



Comparison of the Performance of Machine Learning Algorithms in Flood Susceptibility Zonation of the Doostbiglo Watershed

Sayyad Asghari Saraskanrood ^{1*}, Mahdi Frotan ¹

¹ Department of Physical Geography, Faculty of Social Sciences, University of Mohaghegh Ardabili, Ardabil, Iran

Article Info

Abstract

Article History:

Received: 29 June 2026

Revised: 27 September 2026

Accepted: 28 September 2026

Available online: 01 October 2026

Keywords:

Flood

Random Forest

SVM

GTB

KNN

Northwest Iran

Flood risk management using machine learning methods for identifying susceptible areas plays an effective role in reducing flood damages. This study aimed to compare the performance of machine learning algorithms, including Random Forest (RF), Gradient Tree Boosting (GTB), Support Vector Machine (SVM), and K-Nearest Neighbor (KNN), for flood susceptibility mapping in the Doostbiglo watershed using the Google Earth Engine cloud-based processing environment. A total of 585 flood and 699 non-flood points were extracted from Sentinel-1 imagery and used for model training and validation. The factors affecting flood occurrence, including elevation, slope, runoff, precipitation, Stream Power Index (SPI), Topographic Wetness Index (TWI), drainage and road density, Normalized Difference Vegetation Index (NDVI), soil texture, geological formations, and land use, were considered as model inputs. Flood susceptibility maps were prepared in five classes, ranging from very low to very high, and model performance was evaluated using the ROC curve, AUC, Kappa coefficient, and overall accuracy. The results showed that all four models mainly identified the eastern and northern areas and parts of the western slopes of Sabalan as having high and very high susceptibility. The GTB model, with an AUC of 0.965, showed the highest ability to discriminate flood-susceptible areas and classified 21.94% of the study area as having high and very high susceptibility. The RF, SVM, and KNN models, with AUC values of 0.964, 0.926, and 0.920, respectively, also showed acceptable performance. The map produced by the GTB model can be used to identify flood-susceptible areas and prioritize flood risk management and mitigation measures.

* Corresponding author: Sayyad Asghari Saraskanrood

E-mail address: s.asghari@uma.ac.ir

How to cite this article: Asghari Saraskanrood, S., & Frotan, M. (2026). Comparison of the Performance of Machine Learning Algorithms in Flood Susceptibility Zonation of the Doostbiglo Watershed. *Journal of Geography and Environmental Hazards*, ?(?), ??. <https://doi.org/10.22067/geoeoh.2026.99615.1679>



©2026 The author(s). This is an open access article distributed under Creative Commons Attribution 4.0 International License (CC BY 4.0)

Extended Abstract

Introduction

Flood, one of the most destructive natural hazards, causes enormous damage to infrastructure and communities annually, and its effective management requires a thorough understanding of the factors contributing to its occurrence and the identification of high-risk areas. Despite technological advances, the lack of sufficient data and the complexity of flood patterns continue to pose challenges to the accurate assessment of flood-susceptible areas. Flood susceptibility mapping is considered an essential step toward preventing and reducing the adverse impacts of this phenomenon. Factors such as climate change, improper watershed management, deforestation, and overgrazing play a significant role in intensifying floods, particularly in Iran. Nowadays, advanced methods based on artificial intelligence and machine learning, with their ability to model complex relationships and provide high accuracy, have brought about major improvements in the accuracy of flood prediction and susceptibility mapping. The present study aimed to compare the performance of advanced machine learning algorithms in flood susceptibility mapping of the Doostbiglo watershed. The novelty of this study lies in the simultaneous comparison of the capabilities of several machine learning algorithms, including Random Forest (RF), Gradient Tree Boosting (GTB), Support Vector Machine (SVM), and K-Nearest Neighbor (KNN), for identifying spatial patterns of flood susceptibility and determining areas prone to flood risk. In this regard, the use of the Google Earth Engine cloud-based processing environment, combined with Sentinel-1 radar data for identifying flood locations and a diverse set of topographic, hydrological, climatic, vegetation, soil, geological, and land-use factors, enables the simultaneous investigation of the factors influencing flood susceptibility. Another innovative aspect of the study is the comparative evaluation of model performance using different evaluation metrics and the identification of the model with more suitable performance for detecting high-risk areas. In addition to providing an accurate and comparative representation of the capabilities of different algorithms, this approach can provide a practical basis for identifying priority areas, planning preventive measures, and managing and reducing flood risk in the Doostbiglo watershed and similar watersheds.

Material and Methods

This research aimed to zone flood susceptibility in the Doostbiglo watershed, employing advanced machine learning algorithms within the GEE cloud platform. Initially, flood-influencing environmental factors, including lithology, topography (elevation, slope), hydrological indices (TWI, SPI), distances to rivers and roads, runoff, rainfall, vegetation cover (NDVI), soil texture, and land use, were collected and processed using reliable sources and the GEE environment. Subsequently, all layers were fuzzified to determine the relative importance of each factor in the modeling process. For model training, 585 flood points and 699 non-flood points were collected using Sentinel-1 satellite imagery. The data were split into a 70% training set and a 30% testing set. Four machine learning algorithms, including Random Forest (RF), Gradient Boosting Trees (GTB), Support Vector Machine (SVM), and K-Nearest Neighbors (KNN), were implemented and trained in GEE. Finally, the models' performance was evaluated using statistical metrics ROC and AUC to identify the best model for the final flood susceptibility zoning of the targeted watershed.

Results and Discussion

This research, utilizing Sentinel-1 radar imagery, successfully identified and extracted 585 flood points and 699 non-flood points in the study area, which served as the basis for subsequent analyses. The analysis of factors influencing flood occurrence in the Doostbiglo watershed revealed complex spatial patterns. The northern and eastern parts of the watershed, due to their low elevation and slope, along with high values of Topographic Wetness Index (TWI), were identified as susceptible to runoff accumulation. Conversely, the western, eastern border, and southeastern regions of the watershed are exposed to severe erosion and runoff due to high rainfall intensity and steep slopes; this condition is exacerbated by the Stream Power Index (SPI), which confirms high erosive power. The high density of streams in the north,

east, and southwest of the region plays a key role in guiding and concentrating water flow and increasing flood risk. Furthermore, factors such as road density (as impervious surfaces) in the central and northwestern areas, sparse vegetation cover in the west and north (reducing soil absorption), low soil permeability (including clay loam textures and marl, shale, and conglomerate rocks), and specific land uses (aquatic, urban, rural, and agricultural areas) have led to increased surface runoff and intensified the destructive effects of floods in these areas.

Conclusion

During the modeling phase using machine learning algorithms, the importance of various factors in flood prediction was examined. Among these, slope was identified as the most influential factor in the machine learning process, accounting for 12.24% of the importance. Land use (11.98%) and vegetation cover (11.88%) ranked next in importance. The RF, GTB, SVM, and KNN algorithms generally classified the eastern and northern areas, as well as parts of the western slopes of Sabalan, into high and very high flood risk categories. Based on the models' estimates, the extent of high-risk flood areas constituted 21.42%, 21.94%, 21.31%, and 18.08% of the total study area, respectively. In the performance evaluation, the Gradient Tree Boosting (GTB) algorithm achieved the highest performance in distinguishing between flood-prone and non-flood-prone areas, with an AUC value of 0.965. The Random Forest (RF) algorithm followed closely with an AUC of 0.964. Additionally, the SVM and KNN algorithms achieved AUC values of 0.926 and 0.920, respectively. Overall, all machine learning algorithms demonstrated satisfactory performance in zoning flood susceptibility within the Doost-Bigloo watershed; however, the GTB model outperformed the other models in distinguishing between flood-prone and non-flood-prone areas due to its highest AUC value.



مقایسه عملکرد الگوریتم‌های یادگیری ماشین در پهنه‌بندی حساسیت سیلاب حوضه آبخیز دوست‌بیگلو

صیاد اصغری سراسکانرود^{۱*}، مهدی فروتن^۱

^۱ گروه جغرافیای طبیعی، دانشکده علوم اجتماعی، دانشگاه محقق اردبیلی، اردبیل، ایران

اطلاعات مقاله	چکیده
تاریخچه مقاله: تاریخ دریافت: ۱۴۰۵/۰۴/۰۸ تاریخ بازنگری: ۱۴۰۵/۰۷/۰۵ تاریخ پذیرش: ۱۴۰۵/۰۷/۰۶	مدیریت ریسک سیلاب با استفاده از روش‌های یادگیری ماشین در شناسایی مناطق مستعد، نقش مؤثری در کاهش خسارات سیلاب دارد. این پژوهش با هدف مقایسه عملکرد الگوریتم‌های یادگیری ماشین شامل جنگل تصادفی (RF)، گرادیان تقویت درختی (GTB)، ماشین بردار پشتیبان (SVM) و نزدیک‌ترین همسایه (KNN) در پهنه‌بندی حساسیت سیلاب حوضه آبخیز دوست‌بیگلو، در محیط پردازش ابری Google Earth Engine انجام شد. تهیه نمونه‌های سیلابی و غیرسیلابی، به‌ترتیب ۵۸۵ و ۶۹۹ نقطه بر اساس تصاویر Sentinel-1 استخراج و برای آموزش و اعتبارسنجی مدل‌ها استفاده شد. عوامل مؤثر بر وقوع سیلاب شامل ارتفاع، شیب، رواناب، بارش، شاخص توان آبراهه (SPI)، شاخص رطوبت توپوگرافی (TWI)، تراکم آبراهه‌ها و جاده‌ها، شاخص پوشش گیاهی (NDVI)، بافت خاک، سازند زمین‌شناسی و کاربری اراضی، ورودی مدل‌ها در نظر گرفته شدند. نقشه‌های حساسیت سیلاب در پنج طبقه از خیلی کم تا خیلی زیاد تهیه و عملکرد مدل‌ها با استفاده از منحنی ROC، شاخص AUC، ضریب کاپا و دقت کلی ارزیابی شد. نتایج نشان داد که هر چهار مدل، نواحی شرقی و شمالی و بخش‌هایی از دامنه‌های غربی سبلان را عمدتاً در طبقات حساسیت زیاد و خیلی زیاد شناسایی کردند. مدل GTB با AUC برابر ۰/۹۶۵، بالاترین توان تفکیک سیلاب را داشت و ۲۱/۹۴ درصد از مساحت منطقه را در طبقه حساسیت زیاد و خیلی زیاد قرار داد. مدل‌های RF، SVM و KNN به‌ترتیب با AUC برابر ۰/۹۶۴، ۰/۹۲۶ و ۰/۹۲۰ عملکرد قابل قبولی نشان دادند. نقشه حاصل از مدل GTB می‌تواند برای شناسایی مناطق مستعد سیلاب و اولویت‌بندی اقدامات مدیریت و کاهش خطر سیلاب مورد استفاده قرار گیرد.
کلمات کلیدی: سیلاب Random Forest SVM GTB KNN شمال غرب ایران	

مقدمه

سیلاب یکی از مهم‌ترین بحران‌های طبیعی است که خسارات گسترده‌ای به مناطق متأثر وارد می‌کند. مدیریت مؤثر سیلاب مستلزم شناسایی عوامل مؤثر در وقوع آن و تعیین مناطق مستعد و پرخطر است. با این حال، انجام ارزیابی‌های دقیق مناطق مستعد سیلاب به‌عنوان یکی از بحران‌های طبیعی ویرانگر، به دلیل محدودیت داده‌ها و کمبود اطلاعات کافی درباره الگوهای وقوع سیلاب در مقیاس‌های مختلف، با چالش‌هایی روبرو است (Kazemi, Mansouri & Jozi, 2021). در این راستا، تهیه نقشه‌های حساسیت سیلاب به‌عنوان مرحله‌ای حیاتی برای پیشگیری و مدیریت مؤثر سیلاب‌های آینده شناخته شده است (Kazemi & Jafarpoor, 2024). سیلاب‌ها تقریباً ۹۰ درصد از تلفات ناشی از بلایای طبیعی در آسیا را به خود اختصاص می‌دهند. عوامل انسانی و طبیعی هر دو به طور مستقیم در وقوع سیلاب نقش دارند. علاوه بر اثرات تغییرات اقلیمی، مدیریت ناکافی حوضه‌های آبخیز، از جمله جنگل‌زدایی، چرای بی‌رویه و فقدان اقدامات کنترل سیلاب، نیز در وقوع سیلاب‌های مخرب اخیر در ایران مؤثر بوده است (Shirzadi et al., 2025). پژوهشگران از روش‌های متنوعی برای مطالعه سیلاب استفاده می‌کنند که شامل رویکردهای تصمیم‌گیری چندمعیاره، ادغام داده‌های سنجش از دور با فناوری GIS^۱، تحلیل نسبت فراوانی، رگرسیون لجستیک، منطق فازی، جنگل تصادفی، شبکه‌های عصبی مصنوعی و ماشین‌های بردار پشتیبان می‌شوند. تکنیک‌های پیشرفته قادر به ارائه عملکرد پیش‌بینی سیلاب با وضوح بالا و دقیق هستند (Abdollahzadeh Bina & Fard Moradinia, 2024). در دهه اخیر، استفاده از هوش مصنوعی در نقشه‌برداری سیلاب و پیش‌بینی استعداد سیلاب روندی رو به افزایش داشته است. ادغام هوش مصنوعی و یادگیری ماشین موجب ارتقای چشمگیر دقت و کارایی در تهیه نقشه‌های استعداد سیلاب شده است. روش‌های سنتی در مواجهه با پویایی‌های پیچیده سیلاب با چالش‌هایی روبه‌رو بودند، در حالی که الگوریتم‌های یادگیری ماشین مانند جنگل تصادفی (Random Forest)، ماشین بردار پشتیبان (SVM^۲) و گرادیان بوستینگ قادر به مدل‌سازی روابط غیرخطی و داده‌های با ابعاد بالا هستند. الگوریتم‌هایی نظیر جنگل تصادفی (RF^۳) و گرادیان تقویتی شدید (XGBoost^۴) به دلیل دقت بالا به‌طور گسترده مورد استفاده قرار گرفته‌اند (Arora, Durga, Pandey & Arabameri, 2025). با توجه به اهمیت شناسایی مناطق مستعد سیلاب و نیاز به مدیریت بهینه خطرات ناشی از آن، پژوهش حاضر با هدف مقایسه عملکرد چند الگوریتم پیشرفته یادگیری ماشین در پهنه‌بندی حساسیت سیلاب حوضه آبخیز دوست‌بیگلو انجام شده است. اهمیت این مطالعه در ارائه تصویر دقیق و مقایسه‌ای از توانایی الگوریتم‌ها در پیش‌بینی مناطق پرخطر و کمک به تصمیم‌گیری‌های مدیریتی و کاهش خسارات سیلاب نهفته است. نوآوری این پژوهش در استفاده از چند الگوریتم یادگیری ماشین پیشرفته در یک حوضه آبخیز، شناسایی عوامل مؤثر محیطی، پردازش داده‌ها در محیط ابری Google Earth Engine و ارائه نقشه‌های حساسیت با دقت بالا است که می‌تواند پایه‌ای برای برنامه‌ریزی پایدار و اقدامات پیشگیرانه در حوضه‌های مشابه فراهم کند.

در این زمینه مطالعات فراوانی در دنیا انجام گرفته است از جمله پژوهش گوش و همکاران (Ghosh, Saha & Bera, 2022) با استفاده از مدل‌های یادگیری ماشین شامل SVM، Random Forest و XGBoost حساسیت سیلاب را در پایین‌دست حوضه رودخانه رایداک پهنه‌بندی کردند. ارزیابی مدل‌ها با شاخص AUC^۵ نشان داد که XGBoost با دقت ۹۲ درصد بهترین عملکرد را دارد. همچنین نتایج نشان داد حدود ۲۷ درصد از منطقه، به‌ویژه بخش‌های شرقی و میانی حوضه، به دلیل تغییرات کاربری اراضی و فعالیت‌های انسانی، در معرض خطر بالای سیلاب قرار دارند. موکومبرانوا و مادامومبه (Mukomberanwa & Madamombe, 2025)

1 - Geographic Information System

2 - Support Vector Machine

3 - Random Forest

4 - Extreme Gradient Boosting

5 - Area Under the Curve

با تلفیق روش‌های AHP^۱ و MCDA^۲ و الگوریتم‌های یادگیری ماشین مکانی، حساسیت سیلاب را در چینه‌هوی زیمبابوه پهنه‌بندی کردند. در این پژوهش، ۹ عامل مؤثر به‌عنوان ورودی مدل‌های RF، SVM و CNN^۳ استفاده شد و نتایج نشان داد مدل CNN از نظر دقت و قابلیت تعمیم، عملکرد بهتری نسبت به دو مدل دیگر داشت. فیض‌بهر و همکاران (Feizbahr, Brake, Arbabkhah, Hariri) (Asli & Woods, 2025) با ترکیب داده‌های مکانی و الگوریتم‌های Random Forest، XGBoost و CNN، حساسیت سیلاب را با استفاده از داده‌های تاریخی Sentinel-1 SAR در دوره ۲۰۱۸ تا ۲۰۲۳ پهنه‌بندی کردند. در این پژوهش، ۱۳ عامل مؤثر به‌عنوان ورودی مدل‌ها استفاده شد و نتایج نشان داد مدل XGBoost با مقدار AUC برابر ۹۲ درصد بهترین عملکرد را داشته و ارتفاع، شیب و TWI^۴ مهم‌ترین عوامل پیش‌بینی سیلاب بودند. لیو و همکاران (Liu et al., 2025) به‌منظور تهیه نقشه خطر سیلاب در استان جیانگشی چین، روش MCDA را با الگوریتم‌های Random Forest، XGBoost و LightGBM^۵ تلفیق کردند. مدل‌ها با استفاده از ۱۳ عامل مؤثر و داده‌های تاریخی آموزش داده شدند و نتایج بیانگر برتری مدل RF و کارایی مناسب تلفیق یادگیری ماشین با MCDA در مدیریت ریسک سیلاب بود. در پژوهش آیت ناصر و همکاران (Ait Naceur, Igmoullan & Namous, 2025) مدل آماری FR^۶ با الگوریتم‌های Random Forest، Decision Tree و SVM برای پهنه‌بندی حساسیت سیلاب در حوضه Oued Lakhdar مراکش تلفیق شد. با به‌کارگیری ۱۱ عامل مؤثر، نتایج نشان داد مدل FR-SVM با مقدار AUC برابر ۰/۹۳ درصد بالاترین دقت را داشته است و نواحی شمالی حوضه و مناطق مجاور رودخانه‌های کم‌شیب بیشترین خطر سیلاب را دارند. سیدیکی و نافی (Siddiquie & Nafi, 2025) به‌منظور پهنه‌بندی خطر سیلاب در حوضه‌های شهری، مدل هیدرودینامیکی دوبعدی HEC-RAS^۷ را با الگوریتم Random Forest و روش SHAP^۸ تلفیق کردند. یافته‌ها نشان داد این رویکرد با دقت کلی ۸۹ درصد و ضریب کاپای ۸۴ درصد، عملکرد مطلوبی داشته و شدت بارش و فاصله از رودخانه مهم‌ترین عوامل مؤثر در خطر سیلاب بودند. همچنین، استفاده از XAI^۹ موجب بهبود قابلیت تفسیر نتایج مدل شد. حجی و همکاران (Hajji et al., 2025) با هدف شناسایی خودکار مناطق مستعد سیلاب از شاخص NDFI^{۱۰} استفاده کردند و ۱۷ عامل مؤثر را وارد مدل‌سازی نمودند. آن‌ها از سه الگوریتم AdaBoost، Gradient Boosting و Random Forest برای تهیه نقشه حساسیت سیلاب استفاده کردند و عملکرد مدل‌ها را با معیار AUC ارزیابی کردند. نتایج نشان داد مدل جنگل تصادفی با AUC برابر با ۱، بهترین عملکرد را داشت. همچنین، با استفاده از روش SHAP، سهم عوامل مؤثر در پیش‌بینی مدل‌ها تعیین شد. اسراده و همکاران (Asrade, Abebe, Tadesse, Kerebih & Meshesha, 2026) حساسیت سیلاب در حوضه کوکه را با بهره‌گیری از الگوریتم‌های Random Forest، Gradient Boosting و XGBoost ارزیابی کردند. مدل‌ها بر پایه عوامل محیطی و زمین‌سنجی آموزش داده شدند و نتایج ارزیابی نشان داد الگوریتم‌های GB و XGBoost با دقت ۰/۹۷ عملکرد بهتری نسبت به RF با دقت ۰/۹۶ داشته‌اند. حسن و همکاران (Hassan, Samantaray, Annayat & Satapathy, 2026) نیز با استفاده از الگوریتم‌های Random Forest، ANN، SVM و مدل ترکیبی و تلفیق آن‌ها با GIS، حساسیت سیلاب را در حوضه جهلوم پهنه‌بندی کردند. در این مطالعه، ۱۴ عامل محیطی به‌عنوان متغیرهای مؤثر در مدل‌سازی استفاده شد و عملکرد مدل‌ها با شاخص AUC-ROC

1 - Analytic Hierarchy Process

2 - Multi-Criteria Decision Analysis

3 - Convolutional Neural Network

4 - Topographic Wetness Index

5 - Light Gradient Boosting Machine

6 - Frequency Ratio

7 - Hydrologic Engineering Center – River Analysis System

8 - SHapley Additive exPlanations

9 - Explainable Artificial Intelligence

10 - Normalized Difference Flood Index

ارزیابی شد. نتایج نشان داد مدل ترکیبی با مقدار AUC برابر ۰/۹۲۹ بهترین عملکرد را داشته و پس از آن SVM، ANN و RF قرار گرفتند.

آوند و همکاران (Avand, Janizadeh & Jafari, 2020) به‌منظور تهیه نقشه احتمال خطر سیلاب در حوضه کمیجان استان مرکزی از مدل‌های یادگیری ماشین RF و GBM^۱ استفاده کردند. مدل‌ها با ۲۷۵ نقطه سیل، ۲۷۵ نقطه غیرسیل و ۱۲ عامل محیطی آموزش داده شدند. نتایج نشان داد مدل RF با مقدار AUC برابر با ۰/۸۳ عملکرد بهتری نسبت به GBM داشته و بارندگی، فاصله از رودخانه و ارتفاع مهم‌ترین عوامل مؤثر در وقوع سیلاب هستند. یوسفی و همکاران (Yousefi, Yonesi, Davoudimoghdam, 2022) برای پهنه‌بندی پتانسیل خطر سیلاب در حوضه کشکان از مدل‌های CART^۲، GLM^۳ و GAM^۴ بهره گرفتند. مدل‌ها با ۱۲۳ رخداد سیلاب و ۱۰ عامل محیطی ارزیابی شدند و نتایج نشان داد مدل CART با دقت بالاتر، بخش‌های غربی و مرکزی حوضه را عمدتاً در محدوده خطر زیاد و متوسط قرار می‌دهد. کرمی و همکاران (Karami, Eslamnezhad, Eftekhari, 2023) با تلفیق الگوریتم ژنتیک و مدل‌های SVM و RF، حساسیت سیلاب را در دشت بیرجند بررسی کردند. مدل‌ها بر پایه ۴۲ نقطه سیل خیز و ۱۹ معیار هیدروژئولوژیکی، توپوگرافی، زمین‌شناسی و محیطی توسعه یافتند. نتایج حاکی از برتری مدل RF-GA^۵ نسبت به SVM-GA بود و بیشترین حساسیت سیلاب در نواحی مرکزی با ارتفاع و شیب کمتر مشاهده شد. اشکریز و همکاران (Ashkriz, Falahati & Garakani, 2024) در بررسی سیلاب فروردین ۱۳۹۸ شهرستان آق‌قلا با بهره‌گیری از الگوریتم‌های یادگیری ماشین و تصاویر Sentinel-2، پهنه‌های سیلابی و اراضی آسیب‌دیده را شناسایی کردند. با ترکیب شاخص‌های طیفی NDVI^۶، MNDWI^۷ و NDBI^۸، مدل جنگل تصادفی با دقت کلی ۹۳/۴۰ درصد بهترین عملکرد را داشت و بیشترین خسارت در اراضی زراعی و سپس اراضی ساخته‌شده برآورد شد. در مطالعه ناعمی‌تبار و همکاران (Naemitabar, Zanganeh Asadi & Boroughani, 2025) الگوریتم‌های LT^۹، K-SVM^{۱۰} و KNN^{۱۱} برای شناسایی و پهنه‌بندی مناطق مستعد سیلاب در منطقه نیمه‌خشک طبس به‌کار گرفته شدند. مدل‌ها با ۲۸۸ نقطه سیلابی و ۱۲ عامل محیطی توسعه یافتند و نتایج نشان داد نواحی مجاور آبراهه‌ها و مناطق با پوشش گیاهی کم، بیشترین خطر سیلاب را داشتند. همچنین بخش‌های غربی، مرکزی، شمالی و جنوب‌شرقی منطقه به‌عنوان پهنه‌های پرخطر شناسایی شدند. حنیفی‌نیا و عبقری (Hanifinia & Abghari, 2025) با به‌کارگیری مدل‌های خطی تعمیم‌یافته و بیشینه آنتروپی، حساسیت سیلاب را در حوضه زیوه ارومیه ارزیابی کردند. مدل‌ها با ۹۶ نقطه سیلاب و مجموعه‌ای از عوامل محیطی و انسانی توسعه یافتند. نتایج نشان داد مدل بیشینه آنتروپی با مقدار AUC برابر با ۰/۹۱ عملکرد بالاتری داشت و زمین‌شناسی، فاصله از آبراهه، ارتفاع و شیب مهم‌ترین عوامل مؤثر در وقوع سیلاب هستند. کردی و همکاران (Kordi, Sadoddin, Mohammadi, Sheikh & Asadi Nalivan, 2025) با بهره‌گیری از مدل‌های SVM، Random Forest و بیشینه آنتروپی، احتمال وقوع سیلاب‌های ناگهانی را در حوضه آبخیز سد کرج پهنه‌بندی کردند. ارزیابی مدل‌ها نشان داد SVM با مقدار AUC برابر با ۸۸ درصد و ضریب کاپا ۶۸ درصد، بهترین عملکرد را داشته است. همچنین حدود ۵۳ درصد از منطقه دارای پتانسیل متوسط تا بسیار زیاد سیل‌خیزی بوده و ارتفاع، تراکم زهکشی، شیب و بارندگی از مهم‌ترین عوامل مؤثر در وقوع سیلاب‌های ناگهانی بودند. جوان و

1 - Gradient Boosting Machine

2 - Classification and Regression Tree

3 - Generalized Linear Model

4 - Generalized Additive Model

5 - Genetic Algorithm

6 - Normalized Difference Vegetation Index

7 - Modified Normalized Difference Water Index

8 - Normalized Difference Built-up Index

9 - Logistic Tree

10 - Kernel Support Vector Machine

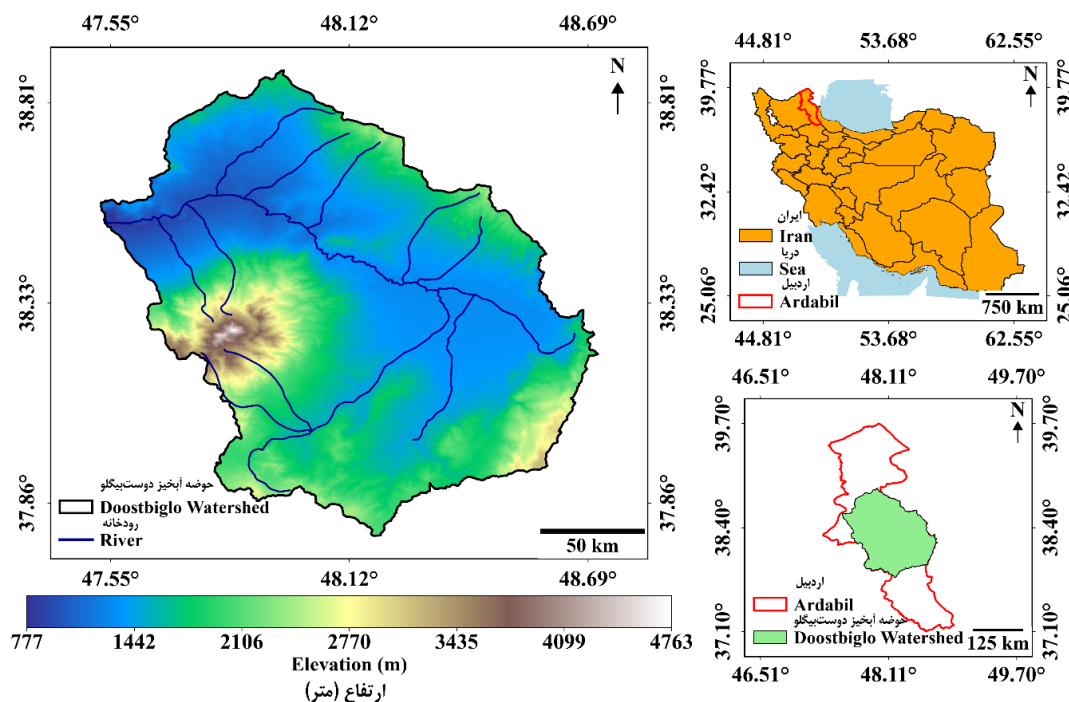
11 - K-Nearest Neighbors

همکاران (Javan, Jamini & Atashbahar, 2026) با استفاده از ۱۳ معیار مؤثر، خطر سیلاب را در منطقه نمونه گردشگری بیستون استان کرمانشاه با الگوریتم‌های CART و Simple RF پهنه‌بندی کردند. یافته‌ها نشان داد اگرچه هر دو مدل از دقت بالایی برخوردار بودند، اما Simple RF عملکرد بهتری داشت. همچنین بر اساس نتایج، حدود ۳۶ تا ۳۹ درصد از منطقه و بیش از ۷۵ درصد روستاها در پهنه‌های با خطر زیاد و بسیار زیاد قرار دارند.

با وجود مطالعات متعدد، عملکرد الگوریتم‌های یادگیری ماشین به شرایط محیطی هر حوضه وابسته است و مقایسه هم‌زمان چند الگوریتم و شناسایی عوامل مؤثر در حوضه دوست‌بیگلو کمتر مورد توجه قرار گرفته است. همچنین، بهره‌گیری از داده‌های چندمنبعی و قابلیت پردازش ابری GEE در این حوضه با محدودیت مطالعاتی مواجه است. از این‌رو، پژوهش حاضر با هدف مقایسه عملکرد الگوریتم‌های یادگیری ماشین، شناسایی مهم‌ترین عوامل مؤثر و تهیه نقشه حساسیت سیلاب در حوضه دوست‌بیگلو انجام شده است. نتایج این پژوهش می‌تواند مبنایی برای شناسایی مناطق پرخطر و بهبود برنامه‌ریزی و مدیریت خطر سیلاب در این حوضه باشد.

منطقه مورد مطالعه

حوضه آبخیز دوست‌بیگلو، واقع در استان اردبیل، دارای اقلیم نیمه‌خشک، یکی از زیرحوضه‌های حوضه آبخیز ارس به‌شمار می‌آید. محدوده مورد مطالعه در طول جغرافیایی ۴۷ درجه و ۳۶ دقیقه تا ۴۸ درجه و ۴۵ دقیقه شرقی و عرض جغرافیایی ۳۷ درجه و ۴۷ دقیقه تا ۳۸ درجه و ۵۲ دقیقه شمالی قرار دارد. شکل ۱، موقعیت جغرافیایی حوضه آبخیز دوست‌بیگلو را در سطح استان اردبیل و ایران نشان می‌دهد (Esmali Ouri, Golshan & Khorrami, 2016).



شکل ۱- نقشه موقعیت جغرافیایی حوضه آبخیز دوست‌بیگلو

Fig. 1. Geographical location map of the Doostbiglo watershed

مواد و روش‌ها

این پژوهش با هدف پهنه‌بندی حساسیت سیلاب در حوضه آبخیز دوست‌بیگلو با بهره‌گیری از الگوریتم‌های یادگیری ماشین در بستر پلتفرم ابری GEE انجام گرفت. در گام نخست، متغیرهای مؤثر بر وقوع سیلاب شامل ارتفاع، شیب، SPI^1 ، TWI ، فاصله از رودخانه، فاصله از جاده، رواناب، بارش، $NDVI$ ، بافت خاک، کاربری اراضی و لیتولوژی گردآوری و آماده‌سازی شدند. داده‌های مورد نیاز از منابع و پایگاه‌های داده مربوطه تهیه و در محیط GEE پردازش شدند. مشخصات کامل داده‌های مورد استفاده، شامل پایگاه داده، باند یا متغیر، تفکیک مکانی و واحد اندازه‌گیری، در جدول ۱ ارائه شده است.

جدول ۱- مشخصات پایگاه داده و باند مورد استفاده پارامترهای مورد مطالعه

Table 1- Database specifications and band used for the parameters studied

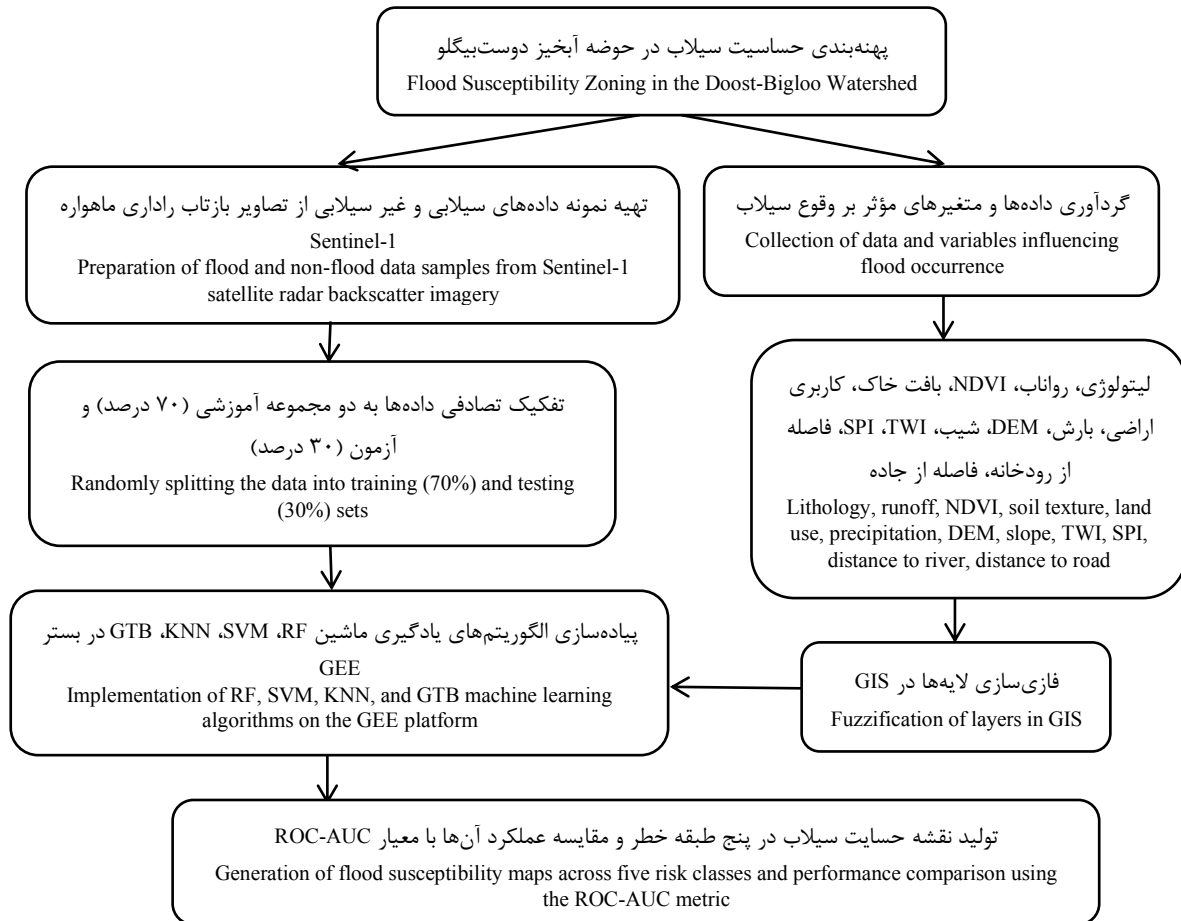
پارامتر Parameter	پایگاه داده Database	باند Band	رزولیشن Resolution	واحد Unit
ارتفاع (Elevation)	SRTM	elevation	30 متر / Meter	متر / Meter
شیب (Slope)	SRTM	DEM	30 متر / Meter	درجه / Degree
شاخص رطوبت توپوگرافی (TWI)	SRTM	DEM	30 متر / Meter	بدون واحد / Dimensionless
شاخص توان آبراهه (SPI)	SRTM	DEM	30 متر / Meter	بدون واحد / Dimensionless
فاصله از رودخانه (Distance from River)	شیپ‌فایل آبراهه Stream Shapefile	لایه آبراهه River Layer	30 متر / Meter	متر / Meter
فاصله از جاده (Distance from Road)	شیپ‌فایل جاده Road Shapefile	لایه جاده Road Layer	30 متر / Meter	متر / Meter
رواناب (Runoff)	TerraClimate	ro	4638 متر / Meter	میلی‌متر / mm
بارش (Rainfall)	CHIRPS	precipitation	5566 متر / Meter	میلی‌متر در روز / (mm/d)
پوشش گیاهی (NDVI)	Landsat 8	B5, B4	$0.67 - 0.64 \mu m$ $0.88 - 0.85 \mu m$	بدون واحد / Dimensionless
بافت خاک (Soil Texture)	OpenLandMap Soil Texture	b0	250 متر / Meter	بدون واحد / Dimensionless
کاربری اراضی (LULC)	GLC-FCS30D	b22	-	بدون واحد / Dimensionless
لیتولوژی (Lithology)	سازمان زمین‌شناسی ایران Geological Survey of Iran	واحدهای لیتولوژی Lithological Units	-	بدون واحد / Dimensionless

به‌منظور استانداردسازی و تبدیل مقادیر متغیرهای مؤثر به مقیاس فازی، فازی‌سازی لایه‌ها با استفاده از ابزارهای Spatial Data Modeller (ArcSDM) در محیط ArcGIS انجام شد. در لایه‌هایی مانند فاصله از رودخانه و جاده، منطق فازی بر اساس رابطه معکوس فاصله اعمال شد؛ یعنی هرچه فاصله کمتر باشد، احتمال وقوع سیلاب بیشتر است. در لایه‌هایی مانند ارتفاع، شیب و $NDVI$ نیز به مقادیر کمتر اهمیت بیشتری داده شد و در سایر لایه‌ها مانند SPI ، TWI ، رواناب، بارش مقادیر زیاد اهمیت بیشتری یافتند و در لایه‌های سازند زمین‌شناسی، بافت خاک و کاربری اراضی با توجه به حساسیت کلاس‌ها به سیلاب، فازی‌سازی شدند. متغیر هدف شامل ۵۸۵ نقطه سیلابی و ۶۹۹ نقطه غیرسیلابی بود که با استفاده از تصاویر بازپراکنش راداری ماهواره Sentinel-1 تهیه شد. بدین منظور، تصاویر Sentinel-1 GRD با قطبش VV و حالت تصویربرداری IW مورد استفاده قرار گرفتند و نقشه میانگین بازپراکنش راداری طی دوره ۲۰۲۵-۲۰۱۶ در محیط GIS تهیه شد. به‌منظور تفکیک مقادیر بازپراکنش و شناسایی مناطق با شرایط متفاوت رطوبتی، نقشه بازپراکنش با استفاده از روش Natural Breaks (Jenks) به پنج طبقه بازطبقه‌بندی شد. بر این اساس، طبقه با کمترین

بازپراکنش، در محدوده ۱۴/۹۱- تا ۲۴/۳۵- دسی بل، به عنوان طبقه پرخطر و نماینده شرایط مرطوب تر و مستعد آبگرفتگی در نظر گرفته شد و ۵۸۵ نقطه نمونه سیلابی از این طبقه استخراج گردید. در مقابل، طبقه با بیشترین بازپراکنش، در محدوده ۷/۵۷ تا ۱۳/۸۵ دسی بل، به عنوان طبقه کم خطر و نماینده شرایط بدون آبگرفتگی یا غیرسیلابی در نظر گرفته شد و ۶۹۹ نقطه غیرسیلابی از آن استخراج گردید. در نهایت، برای نقاط سیلابی برچسب یک و برای نقاط غیرسیلابی برچسب صفر اختصاص یافت. لازم به ذکر است که استفاده از میانگین بازپراکنش طی دوره ۲۰۲۵-۲۰۱۶، به جای یک تصویر منفرد، با هدف کاهش وابستگی نمونه‌ها به شرایط لحظه‌ای و شناسایی الگوهای پایدارتر رطوبتی و مستعد آبگرفتگی انجام شد. پس از استخراج نقاط نمونه از مناطق مستعد سیلاب و غیرسیلابی، داده‌ها با استفاده از روش تقسیم تصادفی ساده و تابع randomColumn در محیط GEE به دو مجموعه شامل ۷۰ درصد داده‌های آموزشی و ۳۰ درصد داده‌های آزمون تقسیم شدند. در این فرآیند، نمونه‌ها به صورت تصادفی بین دو مجموعه تخصیص یافتند و برای تفکیک داده‌های آموزش و آزمون، از نمونه‌گیری طبقه‌بندی شده یا مکانی استفاده نشد. سپس چهار الگوریتم یادگیری ماشین شامل جنگل تصادفی (RF)، گرادیان تقویت درختی (GB¹)، ماشین بردار پشتیبان (SVM) و نزدیک‌ترین همسایه (KNN) برای پهنه‌بندی حساسیت سیلاب در پلتفرم GEE پیاده‌سازی شدند. در الگوریتم جنگل تصادفی، تعداد ۵۰ درخت تصمیم مستقل ایجاد شد که هر یک با انتخاب تصادفی زیرمجموعه‌ای از داده‌های آموزشی و متغیرهای ورودی آموزش دیدند تا پایداری مدل افزایش یابد و از بیش‌برازش جلوگیری شود. خروجی مدل در حالت رگرسیونی احتمال وقوع سیلاب را برآورد نمود. در الگوریتم گرادیان تقویت درختی با پیکربندی ۱۰۰ درخت تصمیم، ضریب کوچک‌سازی برابر با ۰/۱ و حداکثر ۲۰ گره در هر درخت تنظیم گردید تا با بهره‌گیری از تابع زیان حداقل مربعات، روابط غیرخطی پیچیده میان متغیرهای محیطی و پهنه‌های سیل‌گیر را با دقت مدل‌سازی کند. در الگوریتم ماشین بردار پشتیبان (SVM) با هسته تابع پایه شعاعی (RBF) برای دسته‌بندی نقاط سیل‌گیر استفاده شد. در این مدل، پارامتر هزینه (Cost) برابر با ۱۰ و پارامتر گاما برابر با ۰/۵ تنظیم گردید تا مرزهای تصمیم‌گیری بهینه در فضای ویژگی‌های چندبعدی شناسایی شود. با استفاده از پیکربندی C-SVC برای دسته‌بندی داده‌ها، مدل مذکور بر روی مجموعه داده آموزشی مورد یادگیری قرار گرفت تا با بهره‌گیری از ویژگی‌های محیطی استخراج‌شده، توانایی مدل در تفکیک دقیق کلاس‌های سیلاب از مناطق غیرسیلابی مورد ارزیابی قرار گیرد؛ استفاده از هسته RBF² در این مدل، امکان مدل‌سازی روابط پیچیده و غیرخطی میان متغیرهای ورودی را فراهم ساخته و دقت بالایی را در شناسایی الگوهای پهنه‌های سیل‌گیر به همراه دارد. در الگوریتم KNN، هر پیکسل ناشناخته بر اساس شباهت ویژگی‌هایش با نمونه‌های آموزشی طبقه‌بندی گردید. معیار سنجش شباهت، فاصله اقلیدسی بود و مقدار بهینه K انتخاب شد تا تعادل مناسبی میان دقت و پایداری مدل ایجاد شود. در این پژوهش، ابرپارامترهای الگوریتم‌های مورد استفاده به صورت اختصاصی بهینه‌سازی نشدند و مقادیر پیش فرض محیط GEE مورد استفاده قرار گرفتند. بنابراین، مقادیر گزارش شده لزوماً بیانگر مقادیر بهینه برای داده‌ها و شرایط خاص منطقه مورد مطالعه نیستند و این موضوع به عنوان یکی از محدودیت‌های پژوهش در نظر گرفته شده است. در نهایت عملکرد مدل‌ها با مقایسه خروجی پیش‌بینی شده و داده‌های واقعی سیلاب با معیارهای آماری ROC و AUC، همچنین ضریب کاپا و دقت کلی ارزیابی شد و مدلی که بالاترین مقدار معیارها را داشت، به عنوان مدل برتر برای پهنه‌بندی حساسیت سیلاب در حوضه آبخیز دوست‌بیگلو معرفی گردید. در شکل ۲، نمودار جریانی پژوهش حاضر ترسیم شده است.

1 - Gradient Tree Boosting

2 - Radial Basis Function



شکل ۲- نمودار جریانی روش تحقیق
Fig. 2. Research methodology flowchart

شاخص NDVI

برای ارزیابی وضعیت پوشش گیاهی، از شاخص‌های مختلفی استفاده می‌شود که یکی از پرکاربردترین آن‌ها، شاخص تفاضل نرمال شده پوشش گیاهی (NDVI) است. این شاخص که در بازه -۱ تا +۱ قرار می‌گیرد، از طریق تفریق و جمع مقادیر باندهای قرمز (Red) و مادون قرمز نزدیک (NIR) محاسبه می‌شود:

$$NDVI = \frac{NIR - Red}{NIR + Red} \quad (1)$$

در این فرمول، NIR به باند مادون قرمز نزدیک و Red به باند قرمز اشاره دارد (Asghari Saraskanrood & Froitan, 2025).

شاخص رطوبت توپوگرافی (TWI)

شاخص رطوبت توپوگرافی (TWI) به منظور بیان اثر ویژگی‌های توپوگرافی بر پتانسیل تجمع آب و وضعیت رطوبت سطح زمین استفاده می‌شود. این شاخص بر مبنای مفهوم مساحت مشارکت‌کننده بالادست و شیب موضعی در چارچوب مدل TOPMODEL توسعه یافته توسط بون و کیرکبی (Beven & Kirkby, 1979) قرار دارد. در این چارچوب، توزیع مکانی ذخیره آب زیرسطحی با کمیتی

متناسب با لگاریتم نسبت مساحت زهکشی بالادست به تانژانت شیب مرتبط می‌شود. بر این اساس، شاخص رطوبت توپوگرافی به صورت رابطه ۲ محاسبه شد:

$$TWI = \ln\left(\frac{a}{\tan\beta}\right) \quad (2)$$

در این رابطه، a بیانگر مساحت حوضه آبخیز بالادست در واحد طول خط تراز و β زاویه شیب محلی بر حسب رادیان است. مقادیر بالاتر TWI نشان‌دهنده پتانسیل بیشتر تجمع آب و مقادیر پایین‌تر بیانگر زهکشی بهتر و شیب بیشتر سطح هستند (Beven & Kirkby, 1979).

شاخص توان جریان (SPI)

شاخص توان جریان (SPI) یکی از شاخص‌های هیدرولوژیکی استخراج‌شده از مدل رقومی ارتفاع (DEM) است که پتانسیل جریان سطحی و توان فرسایشی جریان آب را نشان می‌دهد. این شاخص بر این فرض استوار است که دبی جریان با مساحت ویژه حوضه آبخیز (A_s) و زاویه شیب محلی (β) ارتباط دارد و به صورت رابطه ۳، محاسبه می‌شود:

$$SPI = A_s \tan\beta \quad (3)$$

که در آن، (A_s) مساحت ویژه حوضه آبخیز و زاویه شیب محلی بر حسب درجه است (Kahrizi, Izadbakhsh, Shabanlou & Yosefvand, 2023).

الگوریتم جنگل تصادفی (RF)

الگوریتم جنگل تصادفی یک روش یادگیری تجمعی است که از مجموعه‌ای از طبقه‌بندی‌کننده‌های درختی تشکیل می‌شود. در این روش، برای ساخت هر درخت، یک بردار تصادفی مستقل از بردارهای تصادفی مربوط به درخت‌های قبلی تولید می‌شود و با استفاده از داده‌های آموزشی و بردار تصادفی مربوطه، یک طبقه‌بندی‌کننده ایجاد می‌شود. پس از تولید تعداد زیادی درخت، طبقه‌بندی‌کننده‌های حاصل برای تعیین کلاس نهایی رأی می‌دهند و کلاسی که بیشترین رأی را به دست آورد، به عنوان کلاس نهایی انتخاب می‌شود. بر این اساس، جنگل تصادفی مجموعه‌ای از طبقه‌بندی‌کننده‌های درختی است که هر یک بر اساس یک بردار تصادفی مستقل ایجاد شده‌اند (Breiman, 2001). برای ارزیابی عملکرد جنگل تصادفی، Breiman در سال ۲۰۰۱، تابع حاشیه را معرفی کرد. این تابع میزان برتری میانگین آرای درخت‌ها برای کلاس صحیح را نسبت به بیشترین میانگین رأی اختصاص یافته به سایر کلاس‌ها نشان می‌دهد و بر اساس رابطه ۴ به دست می‌آید.

$$mg(X, Y) = \text{av} \sum I(h \square(X) = Y) - \max_{j \neq Y} \text{av} \sum I(h \square(X) = j) \quad (4)$$

که در آن I تابع شاخص است. مقدار بزرگ‌تر تابع حاشیه نشان‌دهنده اطمینان بیشتر در طبقه‌بندی است. بر اساس تابع حاشیه، خطای تعمیم جنگل تصادفی به صورت احتمال منفی شدن حاشیه تعریف می‌شود و بر اساس رابطه ۵ به دست می‌آید.

$$PE^* = P_{x, y}(mg(X, Y) < 0) \quad (5)$$

Breiman نشان داد که با افزایش تعداد درخت‌ها، خطای تعمیم جنگل به یک مقدار حدی همگرا می‌شود. همچنین، برای تبیین عملکرد جنگل تصادفی، دو عامل اصلی شامل قدرت طبقه‌بندی‌کننده‌های منفرد و میزان وابستگی یا همبستگی میان آن‌ها معرفی می‌شوند. تابع حاشیه جنگل تصادفی به صورت رابطه ۶، تعریف می‌شود:

$$mr(X, Y) = P_\theta(h(X, \theta) = Y) - \max_{j \neq Y} P_\theta(h(X, \theta) = j) \quad (6)$$

قدرت مجموعه طبقه‌بندی‌کننده‌ها نیز از رابطه ۷، به دست می‌آید:

$$s = E_{x,y} mr(X, Y) \quad (7)$$

با فرض مثبت بودن قدرت، یک کران برای خطای تعمیم طبق رابطه ۸، ارائه می‌شود:

$$PE^* \leq \text{var}(mr) / s^2 \quad (8)$$

در ادامه، با در نظر گرفتن همبستگی میان طبقه‌بندی‌کننده‌ها، کران خطای تعمیم به شکل رابطه ۹، بیان می‌شود:

$$PE^* \leq \bar{\rho}(1 - s^2) / s^2 \quad (9)$$

بر اساس این رابطه، عملکرد جنگل تصادفی به قدرت طبقه‌بندی‌کننده‌های منفرد و میزان همبستگی میان آن‌ها وابسته است. هرچه قدرت طبقه‌بندی‌کننده‌ها بیشتر و میزان وابستگی میان آن‌ها کمتر باشد، عملکرد جنگل از نظر خطای تعمیم مطلوب‌تر خواهد بود. همچنین نسبت همبستگی به مربع قدرت به صورت رابطه ۱۰، تعریف می‌شود:

$$c/s^2 = \bar{\rho}/s^2 \quad (10)$$

ماشین بردار پشتیبان (SVM)

ماشین بردار پشتیبان یک روش یادگیری نظارت‌شده برای طبقه‌بندی است که در آن بردارهای ورودی می‌توانند به وسیله یک نگاشت غیرخطی به یک فضای ویژگی با ابعاد بالاتر منتقل شوند. سپس، در فضای ویژگی، یک سطح تصمیم خطی ایجاد می‌شود. در حالت داده‌های قابل تفکیک خطی، ابرصفحه بهینه به گونه‌ای تعیین می‌شود که حاشیه بین دو کلاس را بیشینه کند. نمونه‌هایی که موقعیت ابرصفحه بهینه و حاشیه را تعیین می‌کنند، بردارهای پشتیبان نامیده می‌شوند (Cortes & Vapnik, 1995).
داده‌های آموزشی برچسب‌دار به صورت رابطه ۱۱، تعریف می‌شوند:

$$(y_1, x_1), \dots, (y_l, x_l), y_i \in \{-1, +1\} \quad (11)$$

یعنی هر نمونه آموزشی شامل یک بردار ورودی x_i و برچسب y_i است که برچسب آن می‌تواند ۱- یا ۱+ باشد.

شرط جدایی خطی نمونه‌های آموزشی به صورت رابطه ۱۲، بیان می‌شود:

$$y_i (w \cdot x_i + b) \geq 1, i = 1, \dots, l \quad (12)$$

ابرفصله بهینه به صورت رابطه ۱۳، تعریف می‌شود:

$$w_0 \cdot x + b_0 = 0 \quad (13)$$

حاشیه بین دو کلاس به صورت رابطه ۱۴، تعریف می‌شود:

$$p(w, b) = \min_{y=+1} [x \cdot w / |w|] - \max_{y=-1} [x \cdot w / |w|] \quad (14)$$

برای ابرصفحه بهینه، مقدار حاشیه برابر است با:

$$p(w_0, b_0) = 2 / |w_0| \quad (15)$$

بنابراین، تعیین ابرصفحه بهینه با بیشینه‌سازی حاشیه، به کمینه‌سازی تحت قیود جداسازی منجر می‌شود. همچنین، بردار وزن ابرصفحه بهینه را می‌توان به صورت ترکیبی از بردارهای آموزشی و ضرایب مربوط به آن‌ها طبق رابطه ۱۶، بیان کرد:

$$w_0 = \sum_{i=1}^l y_i \alpha_i^0 x_i \quad (16)$$

در حالت استفاده از نگاشت غیرخطی، تابع تصمیم در فضای ویژگی به صورت رابطه ۱۷، بیان می‌شود:

$$f(x) = w \cdot \varphi(x) + b \quad (17)$$

با جایگذاری رابطه بردار وزن، تابع تصمیم به صورت رابطه ۱۸، نوشته می شود:

$$f(x) = \sum y_i \alpha_i [\varphi(x_i) \cdot \varphi(x)] + b \quad (18)$$

برای ساده سازی محاسبات در فضای ویژگی، تابع هسته (K) به صورت رابطه ۱۹، تعریف می شود:

$$K(u, v) = \varphi(u) \cdot \varphi(v) \quad (19)$$

در نتیجه، تابع تصمیم را می توان بر اساس تابع هسته به صورت رابطه ۲۰، بیان کرد:

$$f(x) = \sum y_i \alpha_i K(x_i, x) + b \quad (20)$$

تقویت گرادیانی (Gradient Boost)

درخت تقویت گرادیانی یک روش یادگیری تجمعی است که در آن چندین درخت تصمیم به صورت ترتیبی ساخته می شوند و هر درخت جدید برای بهبود خطاهای مدل قبلی آموزش داده می شود. در چارچوب ارائه شده توسط فریدمن در سال ۲۰۰۱، مدل به صورت یک تابع افزایشی از چندین یادگیرنده پایه طبق رابطه ۲۱، تعریف می شود:

$$F(x) = \sum_{j=1}^M \beta_j h(x, a_j) \quad (21)$$

که در آن، $h(x, a_j)$ یادگیرنده پایه یا درخت تصمیم، a_j پارامترهای ساختار درخت، β_j ضریب مربوط به درخت m ام و M تعداد کل درخت های موجود در مدل است. در روش گرادیان تقویتی، درخت ها به صورت مرحله ای به مدل اضافه می شوند، به گونه ای که هر درخت جدید در جهت کاهش تابع زیان مدل قبلی ساخته می شود (Friedman, 2001). در هر مرحله، مدل با استفاده از گرادیان منفی تابع زیان، مقدار خطای مورد نیاز برای بهبود مدل قبلی را تعیین می کند. این مقدار به صورت رابطه ۲۲، محاسبه می شود:

$$-g_j(x_i) = -[\partial L(y_i, F(x_i)) / \partial F(x_i)] \text{ در } F(x) = F_{j-1}(x) \quad (22)$$

که در آن، L تابع زیان، y_i مقدار واقعی متغیر هدف و $F_{j-1}(x_i)$ پیش بینی مدل تا مرحله قبل است. سپس یک درخت تصمیم به عنوان یادگیرنده پایه برای تقریب گرادیان منفی طبق رابطه ۲۳، برازش داده می شود:

$$a_j = \arg \min_{(a, \beta)} \sum_i [-g_j(x_i) - \beta h(x_i, a)]^2 \quad (23)$$

پس از تعیین ساختار درخت، ضریب بهینه آن با کمینه کردن تابع زیان مطابق رابطه ۲۴، تعیین می شود:

$$\rho_j = \arg \min_{\rho} \sum_i L(y_i, F_{j-1}(x_i) + \rho h(x_i, a_j)) \quad (24)$$

در نهایت، مدل با اضافه کردن درخت جدید به مدل قبلی بر اساس رابطه ۲۵، به روزرسانی می شود:

$$F_j(x) = F_{j-1}(x) + \rho_j h(x, a_j) \quad (25)$$

در GBT، یادگیرنده پایه معمولاً یک درخت رگرسیونی است که فضای متغیرهای ورودی را به چند ناحیه یا گره پایانی تقسیم می کند. نمایش این درخت به صورت رابطه ۲۶، ارائه شده است:

$$h(x | b_j, R_j) = \sum_{j=1}^J b_j I(x \in R_j) \quad (26)$$

که در آن، R_j ناحیه مربوط به گره پایانی I_j ، مقدار اختصاص یافته به آن گره و I تابع شاخص است که در صورت قرار گرفتن x در ناحیه R_j مقدار یک و در غیر این صورت صفر می‌گیرد. بنابراین، به‌روزرسانی مدل با استفاده از درخت‌های تصمیم به‌صورت رابطه ۲۷، بیان می‌شود:

$$F(x) = F_{-1}(x) + \rho \sum_{j=1}^J b_j I(x \in R_j) \quad (27)$$

به این ترتیب، مدل نهایی از مجموع متوالی درخت‌هایی تشکیل می‌شود که هر یک با هدف کاهش خطای باقی‌مانده مدل قبلی به آن اضافه شده‌اند. همچنین فریدمن برای کنترل پیچیدگی مدل، استفاده از ضریب کاهش یا نرخ یادگیری (γ) را پیشنهاد می‌کند. در این حالت، رابطه به‌روزرسانی مدل به‌صورت رابطه ۲۸، است:

$$F(x) = F_{-1}(x) + \gamma \rho h(x, a) \quad (28)$$

که در آن، γ نرخ یادگیری است و معمولاً مقداری بین صفر و یک دارد. کاهش نرخ یادگیری باعث می‌شود هر درخت سهم کوچک‌تری در مدل نهایی داشته باشد و در نتیجه معمولاً به تعداد بیشتری درخت برای تشکیل مدل نیاز است.

الگوریتم K-نزدیک‌ترین همسایه (KNN)

Cover و Hart (۱۹۶۷) الگوریتم نزدیک‌ترین همسایه را به‌عنوان یک روش ناپارامتری برای طبقه‌بندی معرفی و ویژگی‌های خطای آن را بررسی کردند. در این روش، مجموعه‌ای شامل n نمونه آموزشی به‌صورت زوج‌های $(x_1, \theta_1), \dots, (x_i, \theta_i), \dots, (x_n, \theta_n)$ در نظر گرفته می‌شود که x_i بیانگر اندازه‌گیری‌های مربوط به نمونه i ام و θ_i نشان‌دهنده طبقه آن نمونه است. داده‌های ورودی در یک فضای متریک X قرار دارند که در آن معیار فاصله d تعریف شده است.

برای یک نمونه جدید x که طبقه آن ناشناخته است، نزدیک‌ترین همسایه نمونه x به‌صورت نمونه‌ای تعیین می‌شود که کمترین فاصله را از x داشته باشد. در مقاله Cover و Hart، نزدیک‌ترین همسایه x ، یعنی x' ، به‌صورت رابطه ۲۹، تعریف شده است:

$$\min d(x_i, x) = d(x', x) \quad (29)$$

سپس قاعده نزدیک‌ترین همسایه، نمونه x را در طبقه θ' مربوط به نزدیک‌ترین همسایه آن قرار می‌دهد. بنابراین، تصمیم طبقه‌بندی تنها بر اساس طبقه نزدیک‌ترین همسایه انجام می‌شود و طبقه سایر نمونه‌های آموزشی در این قاعده مورد استفاده قرار نمی‌گیرد. مبنای این روش آن است که نمونه‌هایی که در یک معیار مناسب به یکدیگر نزدیک هستند، انتظار می‌رود طبقه یکسان یا توزیع احتمال پسین مشابهی داشته باشند. بر همین اساس، در طبقه‌بندی یک نمونه ناشناخته، اطلاعات مربوط به نمونه‌های نزدیک اهمیت بیشتری پیدا می‌کند. از این دیدگاه، قاعده NN یک روش تصمیم‌گیری ناپارامتری است که برای عملکرد خود به دانستن توزیع احتمال مشترک داده‌ها و طبقات نیاز ندارد.

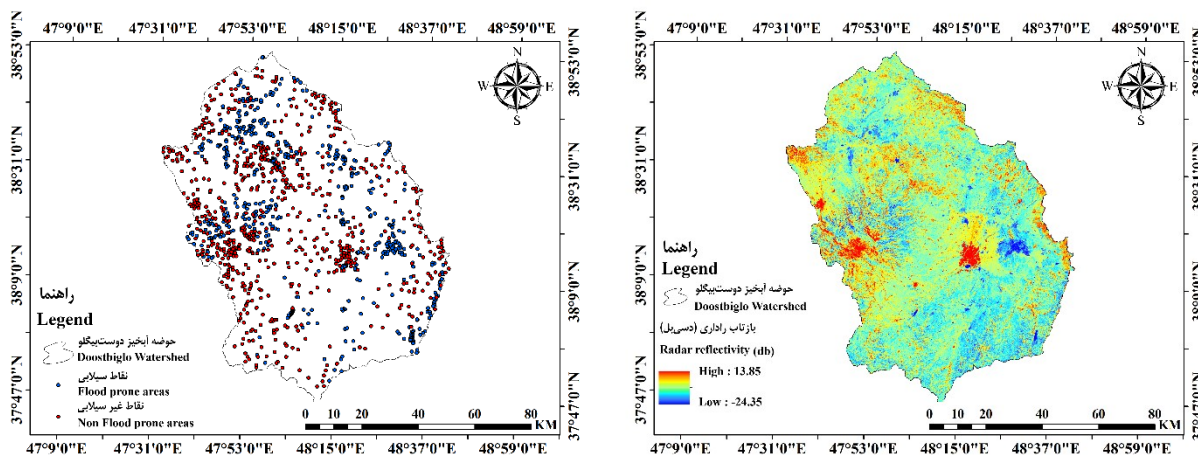
در حالت k نزدیک‌ترین همسایه، به‌جای استفاده از تنها نزدیک‌ترین نمونه، k نمونه‌ای که کمترین فاصله را نسبت به نمونه مورد نظر دارند در فرآیند طبقه‌بندی وارد می‌شوند و تصمیم‌گیری بر اساس طبقه غالب در میان این همسایه‌ها انجام می‌شود. انتخاب k باید به‌گونه‌ای باشد که تعداد همسایه‌ها در مقایسه با تعداد کل نمونه‌های آموزشی کوچک باشد؛ زیرا در این حالت همسایه‌های انتخاب‌شده همچنان در ناحیه نزدیک نمونه مورد طبقه‌بندی قرار می‌گیرند و اطلاعات محلی آن را حفظ می‌کنند. از نظر نظری، اگر R احتمال خطای قاعده نزدیک‌ترین همسایه و خطای بیز باشد، برای مسئله طبقه‌بندی M کلاسه، کران خطای مجانبی قاعده نزدیک‌ترین همسایه به‌صورت رابطه ۳۰، بیان می‌شود:

$$R^* \leq R \leq R^* [2 - MR^*/(M - 1)] \quad (30)$$

این رابطه نشان می‌دهد که خطای قاعده نزدیک‌ترین همسایه، در شرایط مورد بررسی، به خطای بیز وابسته است و کران بالایی آن از دو برابر خطای بیز تجاوز نمی‌کند (Cover & Hart, 1967).

نتایج و بحث

در شکل ۳، میانگین بازپراکنش راداری Sentinel-1 طی دوره ۲۰۱۶-۲۰۲۵ نشان می‌دهد که مقادیر پایین‌تر عمدتاً در نواحی شرقی و شمالی و دامنه‌های شرقی و شمالی سبلان متمرکز شده‌اند. این مقادیر می‌توانند با رطوبت بیشتر سطح، وجود آب آزاد یا سطوح صاف و گل‌آلود مرتبط باشند و از این‌رو بیانگر شرایط مساعدتر برای تجمع آب باشند. در مقابل، مقادیر بالاتر بازپراکنش در ارتفاعات سبلان، شهر اردبیل و مناطق جنگلی مشاهده می‌شوند که می‌تواند با تفاوت در زبری سطح، پوشش گیاهی و شرایط رطوبتی سطح مرتبط باشد. در مجموع، از این لایه ۵۸۵ نمونه سیلابی و ۶۹۹ نمونه غیرسیلابی برای تهیه متغیر هدف استخراج شد. این یافته‌ها از نظر ارتباط بازپراکنش پایین با رطوبت سطحی با مطالعه رحمتی و همکاران (Rahmati et al, 2026) و از نظر لزوم احتیاط در تفسیر بازپراکنش پایین به‌عنوان آب، به دلیل امکان پاسخ مشابه سطوح غیرآبی، با نتایج مارتینز و همکاران (Martinis, Plank, & Ćwik, 2018) همسو است.

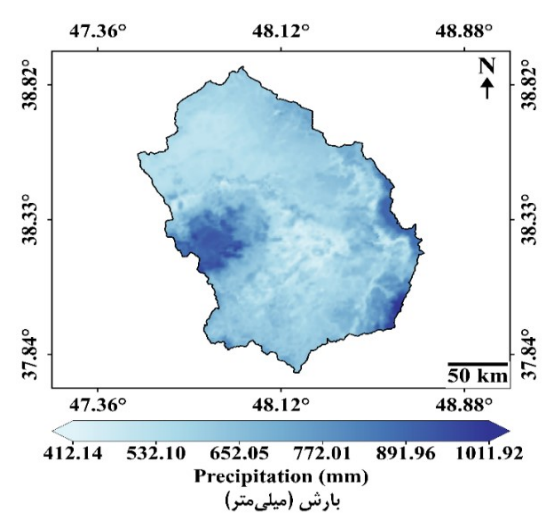
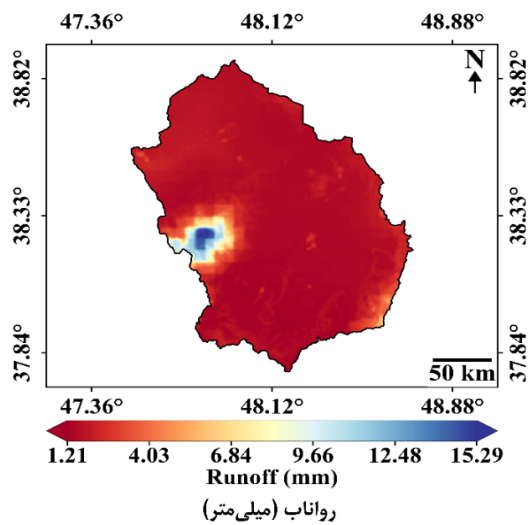
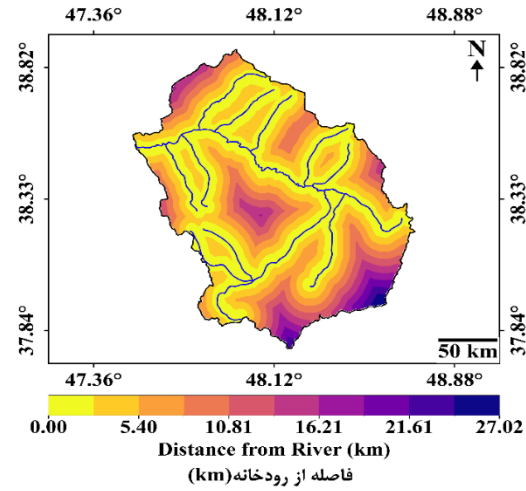
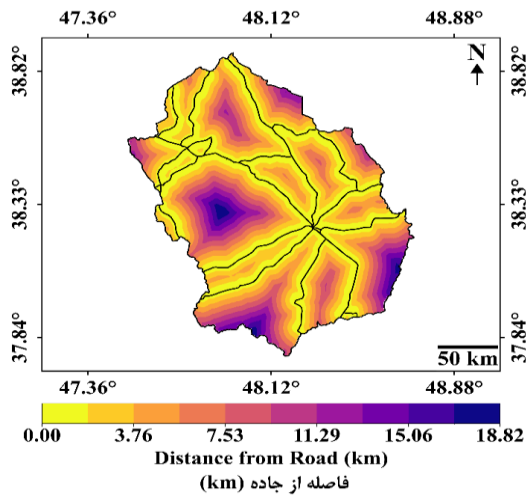
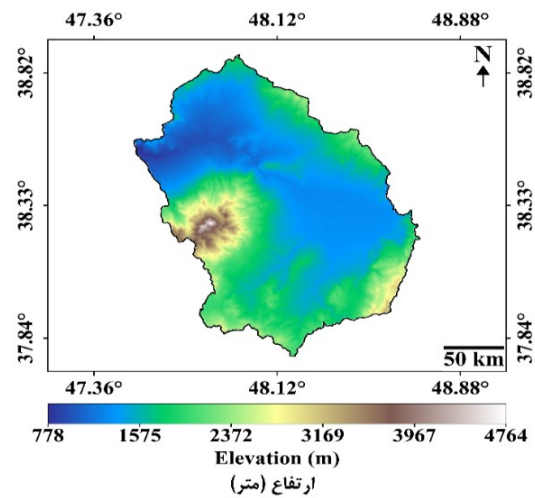
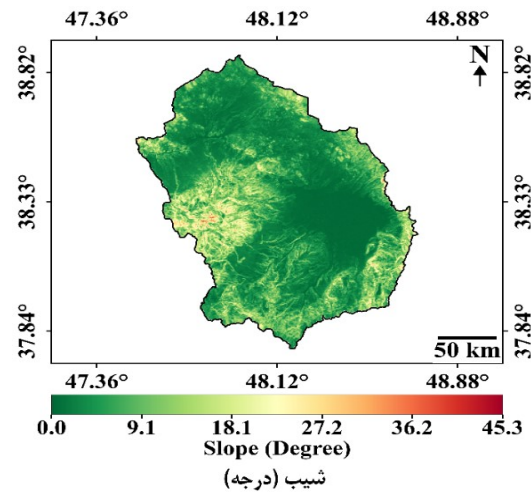


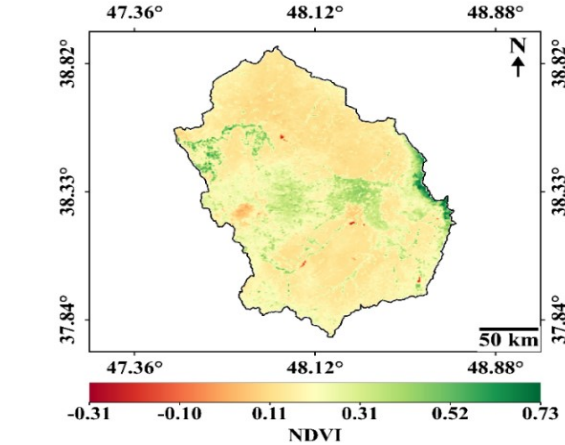
شکل ۳- الف) نقشه بازتاب راداری سنتینل ۱ از سطوح زمین، ب) پراکنش نقاط سیلابی (آبی) و غیر سیلابی (قرمز) در حوضه آبخیز دوست‌بیگللو

Fig. 3. a) Sentinel 1 radar reflectance map of land surfaces, b) Distribution of flood (blue) and non-flood (red) points in the Doostbiglo watershed

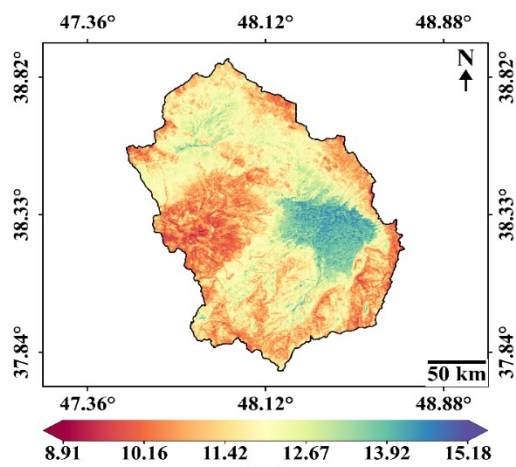
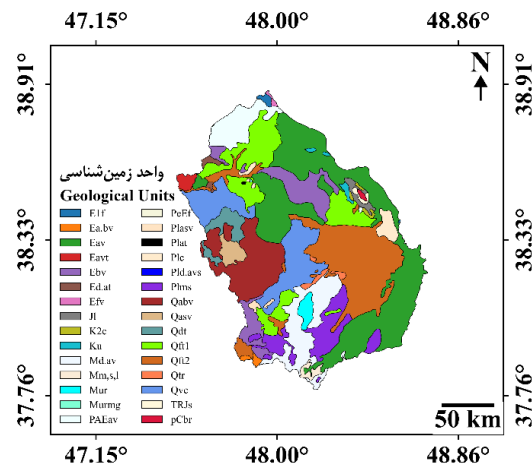
نقشه‌های هر کدام از عوامل مؤثر در وقوع و شدت جریان سیلاب در اشکال ۴ نمایش داده شده است. مطابق شکل (الف)، ارتفاعات غربی حوضه در مجاورت سبلان تا ۴۷۶۳ متر می‌رسد و با هدایت سریع‌تر جریان به سمت دامنه‌های پایکوهی، بر شکل‌گیری رواناب اثر می‌گذارند. در مقابل، نواحی کم‌ارتفاع شمالی و مرکزی، با حداقل ارتفاع ۷۷۷ متر، مستعد تجمع جریان بوده و از پتانسیل بالاتری برای وقوع و تشدید سیلاب برخوردارند. در شکل (ب)، نواحی غربی حوضه با شیب بیشتر، پتانسیل بالاتری برای تولید و انتقال سریع رواناب دارند؛ در مقابل، نواحی کم‌شیب شمالی و مرکزی مستعد تجمع رواناب هستند. همچنین، کاهش فاصله از آبراهه‌ها و رودخانه‌های اصلی با افزایش احتمال تأثیرپذیری از جریان‌های سیلابی همراه است. در شکل (پ)، دو رودخانه اصلی قره‌سو و بالیخلوچای، به ترتیب در جهت‌های شمال‌غربی-جنوب‌شرقی و جنوب‌غربی-شرقی جریان دارند و در شرق حوضه به هم می‌پیوندند. نزدیکی به این رودخانه‌ها، به‌ویژه در بخش مرکزی، از عوامل مؤثر بر الگوی خطر سیلاب است. تراکم سطوح نفوذناپذیر جاده‌ای با کاهش جاده‌های اصلی حوضه، به‌ویژه در بخش مرکزی، از عوامل مؤثر بر الگوی خطر سیلاب است. تراکم سطوح نفوذناپذیر جاده‌ای با کاهش

نفوذ آب و افزایش رواناب سطحی، می‌تواند خطر تجمع و انتقال جریان‌های سیلابی را در مناطق پایین‌دست افزایش دهد. شکل (ث) میانگین بارش سالانه حوضه آبخیز دوست‌بیگلو را طی دوره ۲۰۱۵-۲۰۲۴ نشان می‌دهد. بیشینه بارش، برابر با ۱۰۱۱ میلی‌متر، عمدتاً در مناطق مرتفع غربی و بخش‌هایی از شرق و جنوب‌شرق حوضه مشاهده می‌شود که می‌تواند با افزایش رواناب، حساسیت به وقوع سیلاب را افزایش دهد. شکل (ج) میانگین رواناب را در همین دوره نشان می‌دهد که بیشترین مقدار آن، حدود ۱۵/۲۹ میلی‌متر، در دامنه‌های غربی حوضه و دامنه‌های سبلان ثبت شده و بیانگر پتانسیل بیشتر این نواحی برای شکل‌گیری و تشدید جریان‌های سیلابی است. شکل (ح) الگوی مکانی NDVI را نشان می‌دهد؛ بیشینه مقدار NDVI (۰/۷۳) در بخش شرقی حوضه مشاهده شده است. پوشش گیاهی متراکم‌تر با افزایش نفوذ و کاهش رواناب سطحی، می‌تواند حساسیت سیلاب را کاهش دهد، در حالی که نواحی با پوشش گیاهی کمتر پتانسیل بیشتری برای تولید رواناب و وقوع سیلاب دارند. شکل (خ) هفت کلاس کاربری اراضی حوضه را نشان می‌دهد. پهنه‌های آبی با مساحت ۱۳/۹۵ کیلومتر مربع و مناطق شهری و روستایی با ۱۲۸/۳۱ کیلومتر مربع، از حساس‌ترین کاربری‌ها در برابر سیلاب هستند. اراضی کشاورزی آبی با مساحت ۱۲۰۲/۵۲ کیلومتر مربع و اراضی بایر با ۸۳/۴۴ کیلومتر مربع نیز به دلیل نفوذپذیری کمتر و پوشش گیاهی محدود، پتانسیل بیشتری برای تولید رواناب دارند. در مقابل، کشاورزی دیم (۳۲۴۶/۱۷ کیلومتر مربع)، پوشش گیاهی طبیعی (۳۱۹۹/۶۰ کیلومتر مربع) و جنگل فندقلو (۸۵/۲۳ کیلومتر مربع) به دلیل پوشش گیاهی بیشتر، نقش مؤثرتری در کاهش رواناب و حساسیت سیلاب دارند. در شکل (چ)، پراکنش سازندهای زمین‌شناسی حوضه نشان داده شده است. سازندهای مارن، شیل، کنگلومرا و نهشته‌های تراسی عمدتاً در شمال‌غرب، مرکز و نیمه جنوبی حوضه گسترش یافته‌اند و به دلیل نفوذپذیری کمتر، پتانسیل بیشتری برای تولید رواناب و افزایش حساسیت سیلاب دارند. در مقابل، سازندهای آتشفشانی شامل توف‌های آندزیتی و داسیتیک و سنگ‌های بازالتی در بخش‌هایی از شمال و غرب حوضه قرار دارند و با نفوذپذیری بیشتر، نقش کمتری در تولید رواناب ایفا می‌کنند. شکل (د) وضعیت بافت خاک را در حوضه آبخیز دوست‌بیگلو نشان می‌دهد. بافت شنی لومی عمدتاً در نواحی غربی و جنوبی گسترش یافته و به دلیل نفوذپذیری بیشتر، با کاهش رواناب، حساسیت کمتری نسبت به سیلاب دارد. بافت لومی بخش عمده حوضه را پوشش می‌دهد، در حالی که بافت سیلت لومی بیشتر در نیمه جنوبی مشاهده می‌شود. بافت رسی لومی عمدتاً در بخش‌های مرکزی و جنوبی حوضه قرار دارد و به دلیل نفوذپذیری کمتر، پتانسیل بیشتری برای تولید رواناب و افزایش حساسیت سیلاب نشان می‌دهد. در شکل (ذ)، توزیع شاخص توان فرسایش آبراهه (SPI) نشان داده شده است. بیشینه مقدار SPI در نواحی غربی و پیرامونی حوضه، برابر با ۸/۹۳ واحد، بیانگر انرژی و توان فرسایشی بیشتر جریان آب و در نتیجه پتانسیل بالاتر تغییر و گسترش آبراهه‌ها در این مناطق است. در مقابل، کمینه مقدار SPI در بخش‌های مرکزی و شرقی، برابر با ۲/۶۶ واحد، نشان‌دهنده انرژی و توان فرسایشی کمتر جریان آب است. در شکل (ر)، توزیع شاخص رطوبت توپوگرافی (TWI) نشان داده شده است. بیشینه مقدار TWI در بخش‌های مرکزی تا شرقی و قسمت‌هایی از نیمه شمالی حوضه، برابر با ۱۵/۱۸ واحد، بیانگر تجمع بیشتر رطوبت و پتانسیل بالاتر برای اشباع خاک و تشکیل رواناب است. در مقابل، کمینه مقدار TWI در ارتفاعات و دامنه‌های غربی حوضه، حدود ۸/۹۱ واحد، نشان‌دهنده رطوبت توپوگرافی کمتر و پتانسیل پایین‌تر برای تجمع آب و تولید رواناب است. الگوی مکانی عوامل مؤثر بر حساسیت سیلاب در حوضه دوست‌بیگلو با نتایج مطالعات پیشین در ایران قابل مقایسه است. خسروی و همکاران (Khosravi et al, 2018) در حوضه هراز و جانی‌زاده و همکاران (Janizadeh et al, 2019) در حوضه تفرش، نقش عواملی مانند ارتفاع، شیب، فاصله از رودخانه، بارش، کاربری اراضی، خاک، لیتولوژی، SPI و TWI را در حساسیت سیلاب نشان دادند که با الگوی مشاهده‌شده در حوضه حاضر هم‌راستا است. همچنین، رضایی‌مقدم و همکاران (Rezaei Moghaddam, Karami & Abazari, 2023) در حوضه اوجان‌چای، شیب، کاربری اراضی و بافت خاک را مؤثرترین عوامل شناسایی کردند و پهنه‌های با خطر زیاد را عمدتاً در بستر رودخانه و دشت‌های هموار گزارش نمودند؛ این نتیجه با حساسیت بیشتر نواحی کم‌شیب و مستعد تجمع رواناب در پژوهش حاضر مطابقت دارد.

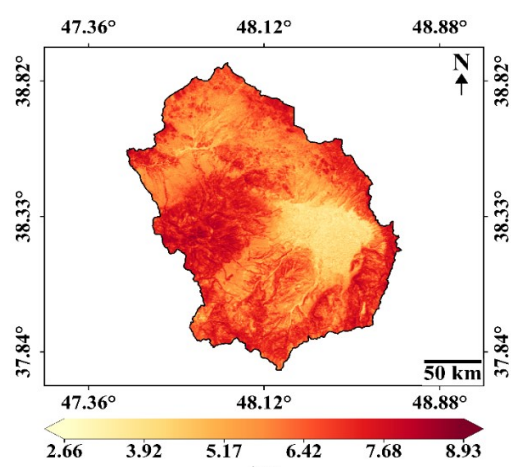




شاخص پوشش گیاهی



شاخص رطوبت توپوگرافی



SFI
شاخص توان جریان

شکل ۴- وضعیت عوامل مؤثر در سیلاب حوضه آبخیز دوست‌بیگلو، الف) ارتفاع، ب) شیب، پ) فاصله از رودخانه، ت) فاصله از جاده، ث) بارش، ج) رواناب، چ) پوشش گیاهی NDVI، ح) کاربری اراضی، خ) سازند زمین، د) بافت خاک، ذ) توان آبراهه (SPI)، ر) رطوبت توپوگرافی (TWI)

Fig. 4. Status of factors affecting flooding in the Doostbiglo watershed, a) elevation, b) slope, c) distance from the river, d) distance from the road, e) precipitation, c) runoff, f) vegetation cover NDVI, h) land use, k) land formation, d) soil texture, i) channel capacity (SPI), r) topographic wetness (TWI)

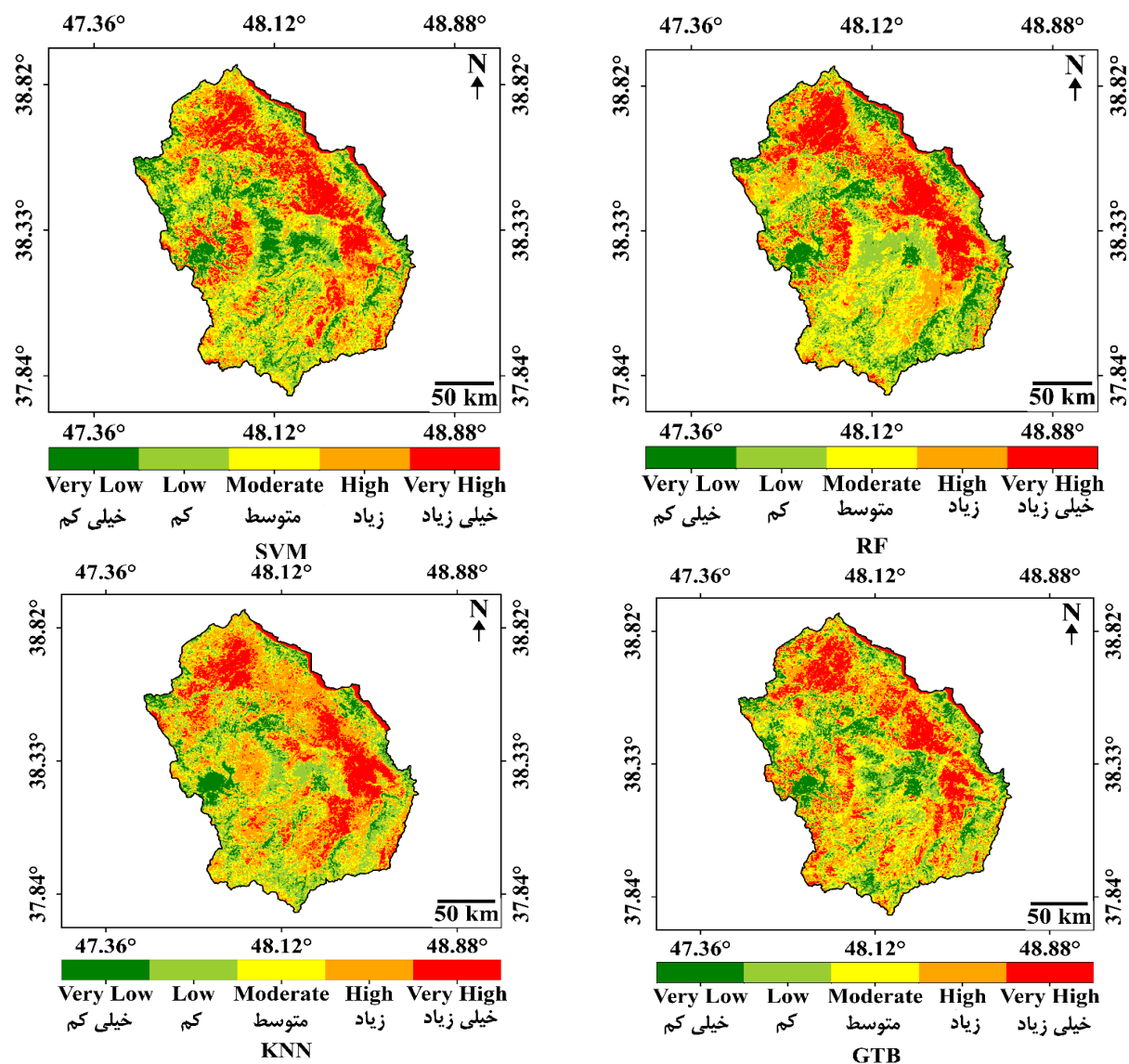
در شکل ۵، نقشه‌ی پهنه‌بندی خطر حساسیت سیلابی حاصل از چهار مدل الگوریتم یادگیری ماشین ارائه شده است. مقادیر خروجی پیوسته مدل‌ها در محیط GIS بازطبقه‌بندی شده و بر این اساس، نقشه‌ها به پنج طبقه حساسیت از خیلی کم تا خیلی زیاد تفکیک شدند. همان‌گونه که مشاهده می‌شود در تمامی مدل‌ها، نواحی شرقی، شمالی و بخش‌هایی از دامنه‌های غربی سبلان در طبقات حساسیت زیاد و خیلی زیاد قرار گرفته‌اند و تفاوت اصلی مدل‌ها بیشتر در وسعت این طبقات است. این الگوی مکانی با ویژگی‌های فیزیوگرافی و کاربری اراضی منطقه مطابقت دارد؛ به‌طوری‌که در نواحی شمالی و شرقی، ترکیب ارتفاع و شیب کمتر، تراکم بالای آبراهه‌ها و جاده‌ها، توسعه اراضی کشاورزی و استقرار مناطق شهری و روستایی، همراه با رطوبت توپوگرافی بیشتر، شرایط مساعدتری برای تجمع آب و افزایش حساسیت سیلابی ایجاد کرده است. در مقابل، نواحی مرتفع غربی و مرکزی سبلان با افزایش شیب و ارتفاع، مقادیر بالاتر SPI و رطوبت توپوگرافی کمتر، بیشتر مستعد تولید و انتقال سریع رواناب هستند؛ باین‌حال، در بخش‌های کوهپایه‌ای و دامنه‌های شمالی که شیب کاهش یافته و انرژی جریان کمتر می‌شود، شرایط برای وقوع سیلاب محلی فراهم‌تر است. این یافته‌ها از نظر نقش هم‌زمان عوامل توپوگرافی و هیدرولوژیکی در تعیین پهنه‌های حساس و اهمیت شیب، رطوبت توپوگرافی، SPI و ویژگی‌های زهکشی با نتایج اسراده و همکاران (Asrade et al, 2026) و از نظر اهمیت شیب، تراکم زهکشی و فاصله از رودخانه در تفکیک مناطق مستعد سیلاب با یافته‌های چوبین و همکاران (Choubin et al, 2019) همسو است.

در جدول ۲، مساحت هر یک از طبقات حساسیت سیلابی توسط مدل‌های مورد مطالعه بیان شده است. در مدل‌های جنگل تصادفی (RF)، ماشین بردار پشتیبان (SVM) و گرادیان تقویتی سه درخت (GTB)، طبقه‌ی خطر خیلی زیاد بیشترین مساحت را در میان طبقات به خود اختصاص داده است و به ترتیب مدل‌های مذکور با ۱۷۰۴/۶۳، ۱۶۹۶/۳۶ و ۱۷۴۶/۱۱ کیلومترمربع وسعت داشته است. این هم‌راستایی نشان می‌دهد که این سه مدل، بخش قابل توجهی از حوضه را به‌عنوان منطقه‌ای با بالاترین حساسیت نسبت به سیلاب شناسایی کرده‌اند. در مقابل، مدل نزدیک‌ترین همسایه (KNN) برای طبقه‌ی خطر زیاد با مساحتی معادل ۲۶۳۹/۹۹ کیلومترمربع، بیشترین اراضی را اختصاص داده است. این در حالی است که طبقه‌ی خیلی زیاد در مدل KNN مساحتی معادل ۱۴۳۸/۷۶ کیلومترمربع دارد. بنابراین مدل KNN نسبت به دو مدل درختی (RF و GTB) و مدل SVM، در تفکیک بین مناطق با خطر زیاد و خیلی زیاد رویکرد محافظه‌کارانه‌تری اتخاذ کرده و یا شدت عوامل مؤثر بر خطر را در مناطق وسیع‌تری در سطح زیاد ارزیابی نموده است.

بر اساس منحنی‌های مشخصه عملکرد گیرنده (ROC) ارائه‌شده در شکل ۶، هر چهار الگوریتم یادگیری ماشین عملکرد قابل قبولی در پیش‌بینی خطر سیلاب نشان دادند. مدل GTB با مقدار AUC برابر با ۰/۹۶۵ بالاترین مقدار را به خود اختصاص داد و پس از آن مدل RF با AUC برابر با ۰/۹۶۴ قرار گرفت. مدل‌های SVM و KNN نیز به‌ترتیب با مقادیر AUC برابر با ۰/۹۲۶ و ۰/۹۲۰ عملکرد مناسبی در تفکیک مناطق سیل‌خیز و غیرسیل‌خیز نشان دادند.

همچنین، مطابق نتایج ارائه‌شده در نمودار شکل ۸، مدل GTB با ضریب کاپای ۰/۷۸۱ و دقت کلی ۰/۸۹۰، بالاترین مقادیر این دو شاخص را به دست آورد. مدل RF نیز با ضریب کاپای ۰/۷۱۸ و دقت کلی ۰/۸۵۷ عملکرد مطلوبی نشان داد. در مدل‌های SVM و KNN نیز به‌ترتیب مقادیر ۰/۶۴۸ و ۰/۵۶۹ برای ضریب کاپا و ۰/۸۲۰ و ۰/۷۷۸ برای دقت کلی به دست آمد. در مجموع، نتایج شکل ۵ و نمودار شکل ۶، نشان می‌دهد که GTB و RF در مقایسه با SVM و KNN عملکرد بالاتری در تفکیک و طبقه‌بندی مناطق از نظر حساسیت به سیلاب داشته‌اند. برتری GTB در پژوهش حاضر با یافته‌های اسراده و همکاران (Asrade et al, 2026) قابل

مقایسه است؛ آنان نیز در ارزیابی سه الگوریتم RF، GB و XGBoost، عملکرد بالاتر GB و XGBoost را نسبت به RF گزارش کردند و مقادیر دقت آزمون ۰/۹۷ را برای هر دو مدل GB و XGBoost و ۰/۹۶ را برای RF به دست آوردند. این شباهت نشان می‌دهد که در هر دو مطالعه، مدل‌های تقویتی مبتنی بر درخت توانایی بالایی در مدل‌سازی روابط غیرخطی میان عوامل مؤثر بر حساسیت سیلاب داشته‌اند.



شکل ۵- الف) نقشه پهنه‌بندی حساسیت سیلاب حوزه آبخیز دوست‌بیگللو با الگوریتم جنگل تصادفی (RF)، ب) مدل بردار پشتیبان (SVM)، ج) گرادیان تقویتی سه‌درخت (GTB)، د) نزدیک‌ترین همسایه (KNN)

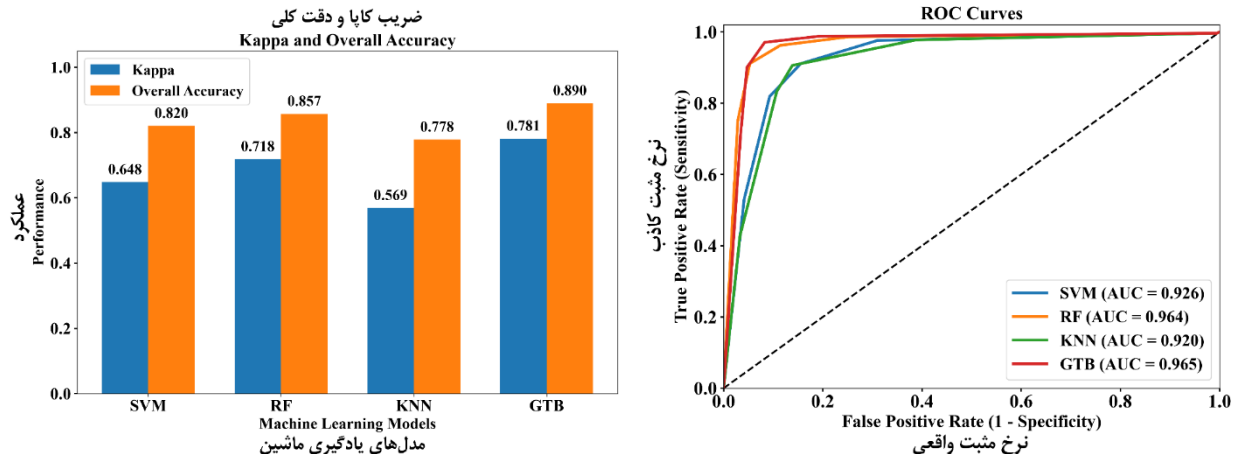
Fig. 5. a) Flood susceptibility zoning map of Doostbiglo watershed with Random Forest (RF) algorithm, b) Support Vector Model (SVM), c) Gradient Boosting Tritree (GTB), d) Nearest Neighbor (KNN)

جدول ۲- مساحت کلاس‌های خطر سیلاب حوزه آبخیز دوست‌بیگلو بر حسب الگوریتم‌های یادگیری ماشین (کیلومتر مربع)

Table 2- Area of flood risk classes of Doostbiglo watershed according to machine learning algorithms (square kilometers)

K نزدیک‌ترین همسایه KNN		تقویت درخت گرادیان GTB		ماشین بردار پشتیبان SVM		جنگل تصادفی RF		طبقات Classes
درصد Percentage	مساحت Area(Km ²)	درصد Percentage	مساحت Area(Km ²)	درصد Percentage	مساحت Area(Km ²)	درصد Percentage	مساحت Area(Km ²)	
14.08	1120.36	19.80	1576.30	19.41	1544.80	19.48	1550.98	خیلی کم (Very Low)
23.36	1859.51	19.36	1540.70	19.78	1574.22	19.89	1583.58	کم (Low)
11.32	900.71	19.70	1568.08	19.62	1561.59	19.73	1570.41	متوسط (Medium)
8.04	639.99	19.20	1528.13	19.88	1582.35	19.47	1549.72	زیاد (High)
18.08	1438.76	21.94	1746.11	21.31	1696.36	21.41	1704.63	خیلی زیاد (Very High)
100	7959.32	100	7959.32	100	7959.32	100	7959.32	جمع مساحت (Total Area)

از سوی دیگر، عملکرد بالای RF در پژوهش حاضر با مطالعات لیو و همکاران (Liu et al., 2025)، حجی و همکاران (Hajji et al., 2025)، کرمی و همکاران (Karami et al., 2023)، اشکرز و همکاران (Ashkriz et al., 2024) و آوند و همکاران (Avand et al., 2020) هم‌راستا است؛ در این مطالعات نیز RF توانایی مناسبی در شناسایی و پهنه‌بندی مناطق مستعد سیلاب نشان داده است. بنابراین، وجه مشترک این مطالعات، قابلیت بالای الگوریتم‌های مبتنی بر درخت در استخراج روابط پیچیده و غیرخطی بین عوامل محیطی و حساسیت سیلاب است، در حالی که تفاوت در عملکرد نسبی مدل‌ها می‌تواند به تفاوت در عوامل ورودی، ویژگی‌های حوزه، نمونه‌های آموزشی و تنظیمات مدل‌ها مربوط باشد.



شکل ۶- الف) نمودار منحنی ROC برای ارزیابی دقت الگوریتم‌های یادگیری ماشین، ب) نمودار عملکرد مدل‌های یادگیری ماشین بر اساس ضریب کاپا و دقت کلی

Fig. 6. a) ROC curve for evaluating the accuracy of machine learning algorithms; b) Performance chart of machine learning models based on the Kappa coefficient and overall accuracy.

نتیجه‌گیری

این مطالعه با هدف مقایسه عملکرد الگوریتم‌های یادگیری ماشین در پهنه‌بندی حساسیت سیلاب در حوزه آبخیز دوست‌بیگلو، با بهره‌گیری از داده‌های مکانی متنوع و تحلیل‌های آماری پیشرفته انجام شد. در این پژوهش، با استفاده از تصاویر راداری Sentinel-1، مجموعاً ۵۸۵ نقطه سیلابی و ۶۹۹ نقطه غیرسیلابی از منطقه مورد مطالعه استخراج شدند و مبنای تحلیل‌های بعدی قرار گرفت.

تحلیل جامع عوامل مؤثر بر وقوع و شدت سیلاب در حوضه دوست‌بیگللو، الگوهای پیچیده‌ای را آشکار ساخت. در نواحی شمال و شرق حوضه، ارتفاع و شیب کم منجر به تجمع رواناب شده و شاخص رطوبت توپوگرافی (TWI) نیز به خوبی میزان تجمع آب در این مناطق را منعکس می‌کند. در مقابل، غرب، بخش شرقی و جنوب شرقی حوضه، به دلیل تمرکز بارش‌های سنگین و شیب تند، در معرض خطر فرسایش و رواناب شدید قرار دارند؛ این وضعیت با شاخص توان آبراهه (SPI) که توان جریان و پتانسیل فرسایش بالاتر را در این نواحی تأیید می‌کند، تشدید می‌شود. تراکم بالای آبراهه‌ها در شمال، شرق و جنوب غرب منطقه، عاملی کلیدی در افزایش ریسک سیلاب محسوب می‌شود، زیرا نزدیکی به این شبکه‌ها، جریان آب را هدایت و متمرکز می‌کند. از سوی دیگر، عواملی چون تراکم جاده‌ها در مرکز و شمال غرب حوضه (به‌عنوان سطوح نسبتاً نفوذناپذیر)، پوشش گیاهی ضعیف در غرب، شمال و جنوب منطقه (کاهنده نفوذ آب در خاک)، نفوذپذیری پایین خاک (شامل سنگ‌های مارن، شیل و کنگلومرا در شمال غرب، مرکز و نیمه جنوبی، و بافت خاک رسی لومی در مرکز و جنوب) و کاربری اراضی خاص (مانند مناطق شهری، روستایی و کشاورزی آبی در مرکز، شرق، جنوب غرب و شمال غرب حوضه)، با افزایش رواناب سطحی و تشدید اثرات سیلاب همراه هستند و پهنه‌های آبی به دلیل قرارگیری در مسیر جریان‌های سیلابی، از نواحی با حساسیت بالاتر در برابر سیلاب محسوب می‌شوند. اما در مرحله مدل‌سازی یادگیری ماشین، مدل‌های RF، GTB، SVM و KNN، نواحی شرقی، شمالی و بخش‌هایی از دامنه‌های غربی سیلان را به‌طور کلی در طبقات خطر زیاد و خیلی زیاد قرار دادند و به ترتیب وسعت اراضی طبقه پرخطر سیلاب را ۲۱/۴۲، ۲۱/۹۴، ۲۱/۳۱ و ۱۸/۰۸ درصد از کل مساحت منطقه برآورد کردند. در این میان، مدل GTB با مقدار AUC برابر با ۰/۹۶۵، بالاترین توانایی را در تفکیک مناطق سیل‌خیز و غیرسیل‌خیز به ثبت رساند و پس از آن، مدل RF با مقدار AUC برابر با ۰/۹۶۴ قرار گرفت. الگوریتم‌های SVM و KNN نیز به ترتیب مقادیر AUC برابر با ۰/۹۲۶ و ۰/۹۲۰ را به‌دست آوردند. عملکرد بالاتر مدل GTB را می‌توان تا حدی به توانایی این الگوریتم در شناسایی روابط غیرخطی و پیچیده میان عوامل مؤثر بر وقوع سیلاب نسبت داد. همچنین، عملکرد مطلوب RF نشان‌دهنده قابلیت این الگوریتم در مدل‌سازی روابط پیچیده میان متغیرهای محیطی و کاهش اثر نویز از طریق تجمیع چندین درخت تصمیم است. در مجموع، هر چهار الگوریتم عملکرد مناسبی در پهنه‌بندی حساسیت سیلاب حوضه آبخیز دوست‌بیگللو نشان دادند؛ با این حال، مدل‌های GTB و RF بر اساس مقادیر AUC، عملکرد بالاتری نسبت به SVM و KNN داشتند.

تعارض منافع

بنا بر اظهار نویسندگان این مقاله تعارض منافع ندارد.

References

- Abdollahzadeh Bina, A., & Fard Moradinia, S. (2024). A novel approach to flood risk zonation: Integrating deep learning models with APG in the Aji Chay catchment. *Aqua - Water Infrastructure, Ecosystems and Society*, 73(7), 1333–1357. <https://doi.org/10.2166/aqua.2024.023>
- Ait Naceur, H., Igoullan, B., & Namous, M. (2025). Machine learning-based optimization of flood susceptibility mapping in semi-arid zone. *DYSONA-Applied Science*, 6(1), 145–159. <http://dx.doi.org/10.30493/das.2024.483211>
- Arora, A., Durga, P. G., Pandey, M., & Arabameri, A. (2025). Machine learning model optimization for flood susceptibility zonation over the Kosi megafan, Himalayan foreland basin, India. *Scientific Reports*, 15, 32757. <https://doi.org/10.1038/s41598-025-07403-w>
- Asghari Saraskanrood, S., & Frotan, M. (2025). Assessing the flood potential of the Khiaochay watershed using the MFFPI model and spectral indices. *Journal of Natural Environmental Hazards*, 15(2), 1–1. [In Persian] <https://doi.org/10.22111/jneh.2025.51600.2109>
- Ashkriz, O., Falahati, F., & Garakani, A. (2024). Detection of flood zones using machine learning algorithms and remote sensing to determine the area of land under flood damage (Case study: March 2019 flood in Aqqla city).

- Journal of Spatial Analysis Environmental Hazards*, 11(3), 21–34. [In Persian]
<http://dx.doi.org/10.61186/jsach.11.3.1>
- Asrade, T. M., Abebe, S. A., Tadesse, K. B., Kerebih, M. S., & Meshesha, T. M. (2026). Flood susceptibility assessment using three machine learning techniques and comparison of their performance. *Scientific Reports*.
<https://doi.org/10.1038/s41598-026-38391-0>
- Avand, M., Janizadeh, S., & Jafari, F. (2020). Evaluating the efficiency of machine learning models in preparing flood probability mapping. *Degradation and Rehabilitation of Natural Land*, 1(1), 19–32. [In Persian]
<https://dor.isc.ac/dor/20.1001.1.27174425.1399.1.1.4.3>
- Beven, K. J., & Kirkby, M. J. (1979). A physically based, variable contributing area model of basin hydrology. *Hydrological Sciences Bulletin*, 24(1), 43–69. <https://doi.org/10.1080/02626667909491834>
- Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32. <https://doi.org/10.1023/A:1010933404324>
- 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(2), 2087–2096.
<https://doi.org/10.1016/j.scitotenv.2018.10.064>
- Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, 20(3), 273–297.
<https://doi.org/10.1007/BF00994018>
- Cover, T. M., & Hart, P. E. (1967). Nearest neighbor pattern classification. *IEEE Transactions on Information Theory*, 13(1), 21–27. <https://doi.org/10.1109/TIT.1967.1053964>
- Esmali Ouri, A., Golshan, M., & Khorrami, K. (2016). Prioritization of suitable axes for construction of underground dam in the Doostbeiglou watershed. *Physical Geography Research*, 48(4), 645–659. [In Persian]
<https://doi.org/10.22059/jphgr.2016.60834>
- Feizbahr, M., Brake, N., Arbakhah, H., Hariri Asli, H., & Woods, K. (2025). Flood susceptibility mapping using machine learning and geospatial-Sentinel-1 SAR integration for enhanced early warning systems. *Remote Sensing*, 17(20), 3471. <https://doi.org/10.3390/rs17203471>
- Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *The Annals of Statistics*, 29(5), 1189–1232. <https://doi.org/10.1214/aos/1013203451>
- Ghosh, S., Saha, S., & Bera, B. (2022). Flood susceptibility zonation using advanced ensemble machine learning models within Himalayan foreland basin. *Natural Hazards Research*.
<https://doi.org/10.1016/j.nhres.2022.06.003>
- Hajji, S., Krimissa, S., Abdelrahman, K., Boudhar, A., Elaloui, A., Ismaili, M., ... & Namous, M. (2025). Enhancing flood prediction through remote sensing, machine learning, and Google Earth Engine. *Frontiers in Water*, 7, 1514047. <https://doi.org/10.3389/frwa.2025.1514047>
- Hanifinia, A., & Abghari, H. (2025). Predicting flood-prone areas using generalized linear and maximum entropy machine learning models. *Journal of Natural Environmental Hazards*, 14(43), 19–34. [In Persian]
<https://doi.org/10.22111/jneh.2024.47730.2021>
- Hassan, S., Samantaray, S., Annayat, W., & Satapathy, D. P. (2026). Flood Susceptibility Assessment in the Jhelum Basin Using Machine Learning and GIS Techniques. *Natural Hazards Review*, 27(3).
<https://doi.org/10.1061/NHREFO.NHENG-2672>
- Janizadeh, S., Avand, M., Jaafari, A., Phong, T. V., Bayat, M., Ahmadisharaf, E., ... & Lee, S. (2019). Prediction success of machine learning methods for flash flood susceptibility mapping in the Tafresh watershed, Iran. *Sustainability*, 11(19), 5426. <https://doi.org/10.3390/su11195426>
- Javan, F., Jamini, D., & Atashbahar, R. (2026). Application of machine learning algorithms in flood risk management with emphasis on tourist-oriented rural settlements (Case study: the Bisotun tourist sample area, Kermanshah

- Province). *Journal of Geography and Environmental Hazards*, 15(1), 226–250. [In Persian] <https://doi.org/10.22067/geoh.2025.95982.1616>
- Kahrizi, A., Izadbakhsh, M. A., Shabanlou, S., & Yosefvand, F. (2023). Investigation the effect of different number of presence points, raster resolution and sample size scenarios on the groundwater potential map prediction using data mining model in Namroud catchment, Tehran. *Iranian Journal of Soil and Water Research*, 54(5), 713–736. [In Persian] <https://doi.org/10.22059/ijswr.2023.325058.668989>
- Karami, P., Eslamnezhad, S. A., Eftekhari, M., Akbari, M., & Rastgoo, M. (2023). Flood susceptibility zoning using machine learning improved by genetic algorithm. *Journal of Natural Environment*, 76(1), 43–60. [In Persian] <https://doi.org/10.22059/jne.2022.350170.2485>
- Kazemi, H., Mansouri, N., & Jozi, S. A. (2021). Flood risk zoning in Nowshahr city using machine learning models. *JHRE*, 40(176), 71–86. [In Persian] <https://doi.org/10.22034/40.176.71>
- Kazemi, M., & Jafarpour, A. (2024). Identifying the threshold of variables affecting flood zones using machine learning technique (Case study: the downstream region of the Karun River). *Water and Soil Management and Modelling*, 4(1), 214–232. [In Persian] <https://doi.org/10.22098/mmws.2023.12285.1220>
- Khosravi, K., Pham, B. T., Chapi, K., Shirzadi, A., Shahabi, H., Revhaug, I., ... & Bui, D. T. (2018). A comparative assessment of decision trees algorithms for flash flood susceptibility modeling at Haraz watershed, northern Iran. *Science of the Total Environment*, 627, 744–755. <https://doi.org/10.1016/j.scitotenv.2018.01.266>
- Kordi, A., Sadoddin, A., Mohammadi, G., Sheikh, V., & Asadi Nalivan, O. (2025). Zoning of flash flood generation probability using machine learning techniques (A case study: Karaj Dam Watershed, Iran). *Desert Ecosystem Engineering*, 14(46), 91–107. [In Persian] <https://doi.org/10.22052/deej.2025.256422.1098>
- Liu, Y., Liu, L., Sun, H., Chen, B., Ma, X., Ning, Y., & Qi, S. (2025). Flood risk assessment combining machine learning with multi-criteria decision analysis in Jiangxi Province, China. *International Journal of Disaster Risk Science*, 16, 858–869. <https://doi.org/10.1007/s13753-025-00423-1>
- Martinis, S., Plank, S., & Ćwik, K. (2018). The use of Sentinel-1 time-series data to improve flood monitoring in arid areas. *Remote Sensing*, 10(4), 583. <https://doi.org/10.3390/rs10040583>
- Mukomberanwa, N. T., & Madamombe, H. K. (2025). Next generation data-driven flood susceptibility modelling with spatial machine learning. *Scientific African*, 30, e03082. <https://doi.org/10.1016/j.sciaf.2025.e03082>
- Naemitabar, M., Zanganeh Asadi, M. A., & Boroughani, M. (2025). Flood susceptibility mapping in the semi-arid region of Tabas using machine learning algorithms. *Journal of Arid Regions Geographic Studies*, 16(61), 46–67. [In Persian] <https://doi.org/10.22034/jargs.2025.509539.1182>
- Rahmati, M., Balenzano, A., Bechtold, M., Brocca, L., Fluhrer, A., Jagdhuber, T., ... & Montzka, C. (2026). Soil moisture retrieval from Sentinel-1: Lessons learned after more than a decade in orbit. *Remote Sensing of Environment*, 333, 115146. <https://doi.org/10.1016/j.rse.2025.115146>
- Rezaei Moghaddam, M. H., Karami, F., & Abazari, K. (2023). Assessment and zoning of flash flood risk in Ojan Chai basin by MFFPI model. *Journal of Hydrogeomorphology*, 10(36), 138–157. [In Persian] <https://doi.org/10.22034/hyd.2023.57462.1701>
- Shirzadi, A., Salvati, A., Hajizadeh Tahan, M., Shahabi, H., Jafari Nodoushan, E., Ramezani, M., ... & Clague, J. J. (2025). Urban flood susceptibility mapping using deep and machine learning algorithms as a management tool: A case study of Sanandaj city, Iran. *Ecological Indicators*, 178, 113886. <https://doi.org/10.1016/j.ecolind.2025.113886>
- Siddiquie, K. T., & Nafi, A. M. (2025). Flood hazard zonation using 2D HEC-RAS and explainable machine learning in urban watersheds. *Research Journal in Civil, Industrial and Mechanical Engineering*, 2(2), 112–130. <https://doi.org/10.61424/rjcime.v2i2.631>
- Yousefi, H., Yonesi, H. A., Davoudimoghdam, D., Azadeh, A., & Shamsi, Z. (2022). Determination of Flood potential Using CART, GLM and GAM Machine learning Models. *Journal of Irrigation and Water Engineering*, 12(48), 84–105. [In Persian] <https://doi.org/10.22125/IWE.2022.150684>