Modeling the Flood Hazard Potential in the Aji Chai basin using Data Mining Algorithms

Tohid Rahimpour [a,*], **Mohammad Hossein Rezaei Moghaddam** [b]
[a] *Postdoctoral researcher in Geomorphology, Faculty of Planning and Environmental Sciences, University of Tabriz, Tabriz, Iran*
[b] *Professor of Geomorphology, Faculty of Planning and Environmental Sciences, University of Tabriz, Tabriz, Iran*

Research Full Paper
Article History (Received: . 2024/10/5 Accepted: 2024/11/15)

Extended abstrac

1- Introduction

Due to the large area and receiving adequate rainfall during the cold and spring seasons, the Aji Chai River and its tributaries become flooded with the beginning of the spring season. In the current study, an attempt was made to prepare a flood hazard potential map in the Aji Chai basin using data mining algorithms. For this purpose, 18 effective parameters in flood occurrence were used. The investigated parameters were Elevation, Slope, Aspect, Topographic wetness index, Sediment transport index, Stream power index, earth curvature, Rainfall, Normalized Difference Vegetation Index, land use, Distance to dam, Distance to bridge, Distance to the river, River density, hydrological soil groups, Drainage texture, Geomorphology and lithology. The information layers of all parameters were prepared in raster format in the ArcGIS software.

2- Research Methodology

The study area of the current study is the Aji Chai basin, which is located in East Azerbaijan province in terms of political divisions. This basin is located in the east of Lake Urmia and its area is about 10985.9 Km2. The elevation changes of the basin are from 1255 meters at the outlet of the basin to 3816 meters in the slopes of Sablan Mountain. The most important river that drains the surface water of this basin is Aji Chai. Four data mining algorithms including Random Forest, Random Subspace, Rotation Forest, and Dagging were used to achieve the purpose of the research. To implement the research models, the location of 274 flood points that happened in the past was used. The map of the location of the flood points in the area was prepared through the information of the regional water company of East Azerbaijan province, field survey, and also the Landsat 8 satellite image of the OLI-TIRS sensor. The implementation steps of all models have been done in WEKA data mining software. WEKA software has been introduced as a machine learning software for the first time in New Zealand and at the University of Waikato. In this research, in order to evaluate the accuracy of flood risk potential maps, receiver operating characteristic curve or system performance characteristic curve (ROC) and area under the curve (AUC) have been used. In the ROC curve, the X-axis shows the detection value or specificity (the percentage of non-flooded pixels that are correctly classified as non-flooded) and the Y-axis shows the sensitivity value (the percentage of flood pixels that are correctly classified as flooded). Variance inflation factor (VIF) and tolerance (T) indexes were used to determine multiple collinearity between independent variables. The presence of collinearity between the selected parameters causes the final maps to be of low accuracy.

3- Results

The results of multiple collinearity analysis showed that except for the drainage texture parameter, other independent variables selected to prepare flood hazard potential maps have low collinearity. Therefore, 17 parameters have been used in flood hazard modeling using data-mining algorithms. The results of examining the importance of each of the parameters in the process of implementing data mining algorithms showed that in the Random subspace model, the parameters of the elevation classes, slope, distance to river and lithology were the most important, respectively. In the Dagging model, the most important effective factors were: Elevation,

* **Corresponding Author: t.rahimpour@tabrizu.ac.ir**

[DOI: 10.61186/jeer.14.4.19]

hydrological soil groups, Topographic wetness index, Distance to the river and Normalized Difference Vegetation Index. In the Rotation Forest model, the parameters of lithology, slope, Rainfall, Elevation and Distance to the river were the most important factors, respectively. The most important effective factors in the Random Forest (RS) model were: Elevation, Distance to the river, slope and Rainfall, respectively.

4- Discussion & Conclusions

Flood hazard potential maps were prepared based on data mining algorithms in the ArcGIS software environment and in five classes with the title of very low, low, moderate, high and very high potential. Examining the final maps shows that the spatial distribution pattern of hazards zones is the same in all maps. So that in all the maps, the heights and steep slopes have very low potential. The distribution of flood points in the classes of the slope map shows that more than 80% of the floods occurred on the slopes of 0-10%, which indicates the effect of this factor on the floods of the region. Examining the area of each of the flood hazard classes in the final maps obtained from data mining algorithms shows that in all maps, more than 30% of the area of the basin are located in high and very high classes. The evaluation of the accuracy of the models using the ROC curve and the area under the curve showed that Random forest model has performed better than other models with the AUC 0.94.

Key Words:. Hazard, Flood, Data Mining, Aji Chai Basin.

Cite this article: Rahimpour, T., & Rezaei Moghaddam, M.H. (2025). Modeling the Flood Hazard Potential in the Aji Chai basin using Data Mining Algorithms. *Journal of* Environmental *Erosion Research. 2025; 14 (4) :19-38.* http//doi.org/

© The Author(s).
DOI: http//doi.org/

Published by Hormozgan University Press.
URL: http://magazine.hormozgan.ac.ir

[DOI: 10.61186/jeer.14.4.19]

مدل‌سازی پتانسیل خطر وقوع سیلاب در حوضه آبریز آجی‌چای با استفاده از الگوریتم‌های داده‌کاوی

توحید رحیم‌پور*: *پژوهشگر پسادکتری ژئومورفولوژی، دانشکدهٔ برنامه‌ریزی و علوم محیطی، دانشگاه تبریز، تبریز، ایران*
محمدحسین رضائی‌مقدم: *استاد گروه ژئومورفولوژی، دانشکدهٔ برنامه‌ریزی و علوم محیطی، دانشگاه تبریز، تبریز، ایران*

نوع مقاله: پژوهشی
تاریخچهٔ مقاله (تاریخ دریافت: ۱۴۰۳/۰۷/۱۴ تاریخ پذیرش: ۱۴۰۳/۰۸/۲۵)
DOI: http//doi.org/

چکیده

حوضه آبریز آجی‌چای واقع در استان آذربایجان شرقی به دلیل گستردگی مساحت و دارا بودن شرایط خاص توپوگرافی همه‌ساله شاهد وقوع سیلاب‌های مخرب هست. هدف اصلی این تحقیق تهیه نقشه پتانسیل خطر وقوع سیل در سطح این حوضه است. با بررسی پژوهش‌های انجام گرفته در ارتباط با موضوع تحقیق و همچنین در دسترس بودن داده‌ها، ۱۸ پارامتر انتخاب شدند که عبارت‌اند از: ارتفاع، شیب، جهت شیب، شاخص رطوبت توپوگرافی، شاخص حمل رسوب، شاخص قدرت آبراهه، انحنای زمین، بارش، شاخص پوشش گیاهی، کاربری اراضی، فاصله از سد، فاصله از پل، فاصله از رودخانه، تراکم زهکشی، گروه‌های هیدرولوژیکی خاک، بافت زهکشی، ژئومورفولوژی و لیتولوژی. از مدل‌های Random forest، Random Subspace، Rotation forest و Dagging به عنوان الگوریتم‌های داده‌کاوی جهت دستیابی به هدف تحقیق استفاده شد. به منظور اجرای مدل‌های تحقیق از موقعیت ۲۷۴ نقطه سیلابی که در گذشته اتفاق افتاده‌اند، استفاده شد. نقشه موقعیت نقاط سیلاب‌های منطقه از طریق اطلاعات شرکت آب منطقه‌ای استان آذربایجان شرقی، بازدیدهای میدانی و همچنین تصویر ماهواره‌ای لندست ۸ سنجنده OLI-TIRS به دلیل در دسترس بودن آن یک روز بعد از وقوع سیل (۲۷ فروردین ۱۳۹۶) تهیه شد. نتایج تحقیق نشان داد که الگوی توزیع فضایی پهنه‌های خطر در تمامی مدل‌ها مشابه هم بوده است. به‌این‌ترتیب که مناطق پایین‌دست حوضه و اطراف آبراهه‌های اصلی منطقه پرخطرترین پهنه‌ها را به خود اختصاص داده‌اند. ارزیابی دقت مدل‌ها با استفاده از منحنی ROC و سطح زیر منحنی نشان داد که مدل Random forest با سطح زیر منحنی ۰/۹۴ نسبت به سایر مدل‌ها عملکرد بهتری داشته است.

واژگان کلیدی: مخاطره، سیلاب، داده‌کاوی، حوضه آبریز آجی‌چای.

* **نویسندهٔ مسئول:** t.rahimpour@tabrizu.ac.ir

[DOI: 10.61186/jeer.14.4.19]

۱ـ مقدمه

سیلاب‌ها از مهم‌ترین و فراوان‌ترین مخاطرات در کشور محسوب می‌شوند که همه ساله خسارت‌های زیادی را برجای می‌گذارند (Rezaei Moghaddam & Rahimpour, 2024a). در طول دو دهه گذشته فراوانی وقوع سیل در سراسر جهان بیش از ۴۰ درصد افزایش داشته است (Hirabayashi et al., 2013). روند افزایشی در وقوع حوادث سیل عمدتاً نتیجه تغییرات آب‌وهوایی و همچنین تغییرات کاربری زمین ناشی از فعالیت‌های انسانی هست (Hirabayashi et al., 2013; Sofia et al, 2017). طبق گزارش‌های سازمان بهداشت جهانی بین سال‌های ۱۹۹۸ تا ۲۰۱۷ حدود ۲ میلیارد نفر در سرتاسر دنیا از سیلاب‌ها متأثر شده‌اند (WHO, 2022). عوامل مختلفی مانند تغییرات آب‌وهوا و فعالیت‌های انسانی می‌توانند باعث وقوع سیل شوند (Kourgialas and Karatzas, 2011). در سال‌های اخیر ارزیابی و برآورد سیلاب‌ها به دلیل افزایش شهرنشینی و تمرکز مناطق مسکونی در اطراف رودخانه‌ها و دشت‌های سیلابی به‌طور قابل‌توجهی افزایش یافته است (Diakakis et al., 2012). به‌منظور ارزیابی پهنه‌بندی و تحلیل حساسیت سیل در هر منطقه، معیارهای متفاوتی برای تعیین رابطه بین هر یک از این عوامل و وقوع سیل مورد نیاز است (Lee et al., 2013; Hudson et al., 2014). روش‌های کمی و کیفی متعددی توسط محققان مختلف به‌منظور مدل‌سازی و تهیه نقشه‌های پتانسیل خطر وقوع سیل ارائه شده است. از جمله این روش‌ها می‌توان به روش‌های تحلیل تصمیم‌گیری چندمعیاره (MCDA)، روش‌های هیدرولوژیکی، مدل‌های آماری و الگوریتم‌های داده‌کاوی و یادگیری ماشین اشاره کرد. عملکرد مدل‌های داده‌کاوی و یادگیری ماشین تا حد زیادی به کیفیت مجموعه داده‌های آموزشی، به‌ویژه دقت پراکنش نقاط سیلابی و بدون سیل بستگی دارد (Khosravi et al., 2020). بررسی مطالعات گذشته نشان می‌دهد که الگوریتم‌های داده‌کاوی در تهیه نقشه‌های پتانسیل خطر وقوع سیل کارایی مناسبی داشته‌اند.

حوضه آبریز آجی‌چای واقع در استان آذربایجان شرقی، به دلیل برخورداری از عوامل ایجادکننده سیل (از قبیل شرایط اقلیمی، توپوگرافی، ژئومورفولوژی، زمین‌شناسی، هیدرولوژی و...) همه ساله در فصل بهار درگیر سیلاب‌های مخرب می‌شود (Rezaei Moghaddam & Rahimpour, 2024b). بنابراین تحقیق حاضر تلاش می‌کند تا یک الگوی فضایی از پراکنش پتانسیل خطر وقوع سیل را در سطح حوضه آبریز آجی‌چای تحت نرم‌افزارهای داده‌کاوی و سیستم اطلاعات جغرافیایی (GIS) ارائه دهد. نتایج تحقیق حاضر برای مدیران و برنامه‌ریزان منطقه جهت مدیریت هر چه بهتر خطر سیلاب کمک شایانی خواهد نمود.

در ارتباط با موضوع تحقیق، پژوهش‌های خوبی در داخل و خارج از کشور توسط محققان مختلف انجام شده است که در ادامه به برخی از تازه‌ترین آن‌ها اشاره می‌شود. Cheraghi Ghalehsari و همکاران (۲۰۲۰)، در تحقیقی اقدام به تهیه نقشه خطر سیلاب با استفاده از مدل ماشین بردار پشتیبان (SVM) و سیستم اطلاعات جغرافیایی نمودند. منطقه مورد مطالعه این تحقیق حوضه آبریز نکارود بود. در این تحقیق از ۱۲ پارامتر ژئومورفولوژیکی، هیدرولوژیکی و فیزیوگرافیکی شامل مقدار شیب، جهت شیب، طبقات ارتفاعی، دما، کاربری اراضی، بارندگی، تراکم و فاصله از گسل، تراکم و فاصله از آبراهه، تراکم و فاصله از جاده استفاده شد. جهت اجرای مدل از ۱۰۰ نقطه سیل گیر که بر اساس بازدیدهای میدانی در طول فصول مختلف سال تهیه شده بود استفاده شد. نتایج تحقیق نشان داد که مناطق حساس به سیل گیری بیشتر در بخش‌های شمالی و شمال غرب حوضه قرار دارند. Luu و همکاران (۲۰۲۱)، با استفاده از مدل‌های

* نویسندهٔ مسئول: t.rahimpour@tabrizu.ac.ir

[DOI: 10.61186/jeer.14.4.19]

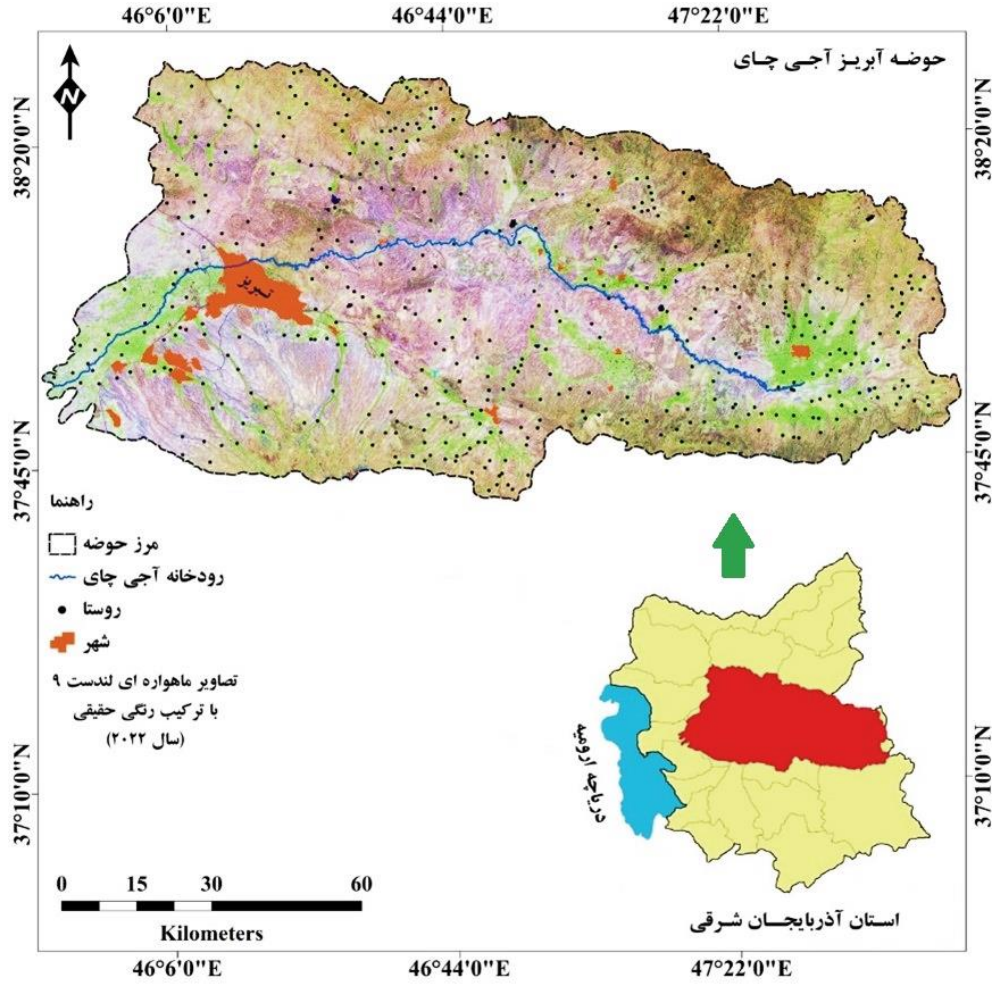
ترکیبی اقدام به تهیه نقشه خطر وقوع سیل در استان کوانگ بین واقع در کشور ویتنام کردند. مدل‌های مورد استفاده در این تحقیق عبارت بودند از: نسبت فراوانی، درخت تصمیم جزئی (PART)، Adaboost، Bagging، Dagging و Random Subspace. جهت تهیه نقشه خطر وقوع سیل از ۱۰ عامل مؤثر در وقوع این پدیده نیز استفاده شد. این عوامل عبارت بودند از: ارتفاع، شیب، انحناء، جهت جریان، جریان تجمعی، تراکم آبراهه، فاصله از رودخانه، بارش، کاربری اراضی و زمین‌شناسی. نتایج تحقیق نشان داد که کاربری اراضی، زمین‌شناسی و شیب بیشترین تأثیر را در وقوع سیلاب‌های منطقه دارند. نتایج ارزیابی مدل‌ها نشان داد که همه مدل‌ها از عملکرد قابل قبولی در تهیه نقشه خطر وقوع سیل برخوردار بودند. Rahimpour و همکاران (۲۰۲۲)، در تحقیقی به تحلیل تغییرات فضایی حساسیت خطر وقوع سیل بر پایه نوعی مدل ترکیبی نوین پرداختند. منطقه مورد مطالعه این تحقیق حوضه آبریز الندچای واقع در شهرستان خوی بود. در این تحقیق از ۱۳ پارامتر استفاده شده که عبارت بودند از: لیتولوژی، گروه‌های هیدرولوژیکی خاک، شاخص پوشش گیاهی، کاربری اراضی، شیب، جهت شیب، ارتفاع، فاصله از آبراهه، تراکم آبراهه، بارش، شاخص رطوبت توپوگرافیک، شاخص قدرت آبراهه و شاخص حمل رسوب. جهت تهیه نقشه خطر وقوع سیل از یک مدل ترکیبی نوین بر اساس طبقه‌بندی فازی (FURIA)، الگوریتم ژنتیک و یک الگوریتم یادگیری ماشین (AdaBoost) استفاده شد. نتایج تحقیق نشان داد مناطقی که حساسیت زیادی از نظر خطر وقوع سیل دارند، اغلب در پایین‌دست حوضه قرار دارند که مناطق مسطح و کم ارتفاع را شامل می‌شوند. Zarei و همکاران (۲۰۲۲)، به ارزیابی پتانسیل وقوع سیل حوضه آبریز فریزی واقع در استان خراسان رضوی با استفاده از مدل‌های داده‌کاوی ماشین بردار پشتیبان، چاید و جنگل تصادفی پرداختند. نتایج تحقیق نشان داد که قسمت‌های شمالی و غربی حوضه که از نظر زمین‌شناسی و لیتولوژی دارای سازندهای سطحی مانند مارن، رس و سیلت بوده و ضریب نفوذپذیری آن‌ها بسیار کم و پوشش گیاهی ناچیز دارند، از پتانسیل بالایی در بروز سیلاب برخوردار هستند. Rahimpour و همکاران (۲۰۲۳)، بر پایه یک رویکرد طبقه‌بندی ترکیبی نوین به مدل‌سازی حساسیت خطر وقوع سیل در حوضه آبریز الندچای پرداختند. در این تحقیق از ۱۳ پارامتر مؤثر در وقوع سیل استفاده گردید. مدل ترکیبی مورد استفاده در این پژوهش جهت تهیه نقشه خطر وقوع سیل FURIA-GA-LogitBoost بوده است. یافته‌های پژوهش نشان داد که مناطق پایین‌دست حوضه حساسیت بالایی را از نظر خطر وقوع سیل دارا هستند. همچنین شهر خوی نیز که در قسمت خروجی حوضه قرار دارد جزو مناطق با خطر سیل‌گیری بالا شناسایی شد. Aldiansyah & Wardani (۲۰۲۳)، در پژوهشی به ارزیابی خطر وقوع سیل بر اساس الگوریتم‌های یادگیری ماشین نمودند. منطقه مورد مطالعه این تحقیق شهر کنداری واقع در کشور اندونزی بود. در این تحقیق ۱۷ پارامتر مؤثر در وقوع سیل مورد استفاده قرار گرفت. این پارامترها عبارت بودند از: جهت شیب، انحناء، ارتفاع، جریان تجمعی، جهت جریان، زمین‌شناسی، شیب، کاربری اراضی، شاخص پوشش گیاهی، بارش، فاصله از رودخانه، خاک، شاخص قدرت آبراهه، شاخص حمل رسوب، شاخص زبری زمین، شاخص رطوبت توپوگرافی و باد. نتایج ارزیابی مدل‌ها نشان داد که مدل جنگل تصادفی عملکرد خوبی در تهیه نقشه خطر وقوع سیل داشته است. همچنین نتایج نشان داد که بیش از ۳۲ درصد از مساحت منطقه از سیل‌خیزی بالایی برخوردار هست. Hitouri و همکاران (۲۰۲۴) با استفاده از الگوریتم‌های یادگیری ماشین نقشه خطر وقوع سیل را برای حوضه آبریز متلیلی واقع در شمال‌شرق مراکش تهیه کردند. در این تحقیق مدل های Random Forest، CART، SVM و

[DOI: 10.61186/jeer.14.4.19]

XGBoost جهت تهیه نقشه‌های نهایی استفاده شدند. نتایج ارزیابی مدل‌ها نشان داد که مدل Random Forest نسبت به سایر مدل‌ها از دقت خوبی برخوردار بوده است.

۲- منطقه مورد مطالعه

حوضه آبریز آجی‌چای از نظر تقسیمات سیاسی در استان آذربایجان شرقی و از نظر موقعیت جغرافیایی بین ″۰۳ ′۴۱ °۳۷ تا ″۰۷ ′۲۹ °۳۸ عرض شمالی و ″۰۷ ′۴۸ °۴۵ تا ″۴۸ ′۵۳ °۴۷ طول شرقی قرار دارد (شکل ۱). این حوضه در شرق دریاچه ارومیه واقع شده و مساحت آن حدود ۱۰۹۸۵/۹ کیلومترمربع است. تغییرات ارتفاعی حوضه از ۱۲۵۵ متر در خروجی حوضه تا ۳۸۱۶ متر در دامنه‌های کوهستان سبلان هست. مهم‌ترین رودخانه‌ای که آب‌های سطحی این حوضه را زهکشی می‌کند آجی چای نام دارد. این حوضه از دامنه‌های جنوب و جنوب غربی کوهستان سبلان با ارتفاع بالغ بر ۳۸۰۰ متر شروع شده و با عبور از قسمت شمالی کلان‌شهر تبریز در قسمت شمال غربی آذرشهر به دریاچه ارومیه منتهی می‌شود. به دلیل گستردگی مساحت و دریافت بارش مناسب در طول فصول سرد و فصل بهار، رودخانه آجی‌چای و آبراهه‌های فرعی آن با شروع فصل بهار حالت طغیانی به خود گرفته و منجر به سیل گرفتگی زمین‌های اطراف خود می‌شوند که خسارت‌های جانی و مالی زیادی را به مردم منطقه وارد می‌سازد.



شکل ۱: موقعیت جغرافیایی منطقه مورد مطالعه

[DOI: 10.61186/jeer.14.4.19]

۳- مواد و روش

در پژوهش حاضر از چهار مدل جنگل تصادفی، Random subspace، Rotation forest و Dagging به‌عنوان شناخته‌شده‌ترین الگوریتم‌های داده‌کاوی جهت تهیه نقشه‌های پتانسیل خطر وقوع سیل استفاده شده است.

مدل جنگل تصادفی

مدل جنگل تصادفی (RF[1]) مجموعه‌ای از درختان تصمیم برای طبقه‌بندی یا رگرسیون به‌منظور پیش‌بینی است که توسط بریمان در سال ۲۰۰۱ معرفی شد. این روش یک الگوریتم یادگیری جمعی هست (Rahman et al., 2019). جنگل تصادفی روش یادگیری نظارت شده‌ای هست که درخت‌های چندگانه را در فرایند طبقه‌بندی به کار می‌گیرد (Breiman et al., 1984). جنگل تصادفی سه ویژگی تعریف شده توسط کاربر را شامل می‌شود: ۱- تعداد متغیرهای مورد استفاده در ساخت هر درخت که قدرت هر درخت مستقل را نشان می‌دهد ۲- تعداد درختان در جنگل تصادفی ۳- حداقل تعداد گره‌های انتهایی (Peters et al., 2008). RF مجموعه‌های مختلفی از نمونه‌ها را توسط نمونه‌برداری با جایگزینی تولید می‌کند و مراحل آموزش درخت رگرسیون منطبق را ایجاد می‌کند و سپس طبقه‌بندی داده‌ها را بر اساس نتایج رأی‌گیری طبقه‌بندی کننده‌های متعدد تخمین می‌زند. در پایان، یک طبقه‌بندی با تعداد آراء قابل توجهی بر روی تمام درختان جنگل پیدا شده و توسط RF انتخاب می‌شود (Towfiqul Islam at al., 2020). قدرت پیش‌بینی مدل جنگل تصادفی با افزایش قدرت درختان مستقل و کاهش همبستگی بین آن‌ها افزایش می‌یابد (Lieb et al., 2012).

مدل Random subspace

هو در سال ۱۹۸۹ مدل Random subspace را به‌عنوان یک مدل جدید الگوریتم یادگیری ماشین جمعی برای حل مسائل محیطی معرفی کرد. این روش به‌عنوان یک الگوریتم یادگیری موازی جهت بهبود عملکرد طبقه‌بندی در مدل‌های داده‌کاوی ارائه شده است. این مدل ابتدا با استفاده از تقسیم داده‌های مکانی اصلی به تعدادی گروه‌های کوچک (زیرگروه‌ها) به p خصوصیات ابعادی مکانی و سپس اجرای طبقه‌بندی کننده پایه برای هر دسته از زیرگروه‌ها و ترکیب نتایج آن‌ها باعث بهبود عملکرد در طبقه‌بندی می‌شود (Piao et al., 2015).

مدل Rotation forest

رودریگز و همکاران در سال ۲۰۰۶ الگوریتم Rotation forest که یکی از روش‌های قدرتمند و رایج کلاسه‌بندی جمعی هست را معرفی کردند. این الگوریتم به عنوان یک روش مهم برای تقویت طبقه‌بندی کننده‌های ضعیف‌تر در نظر گرفته می‌شود (Ozcift, 2012). این الگوریتم مجموعه داده‌های چند متغیره بزرگ را با استفاده از تجزیه‌وتحلیل مؤلفه اصلی (PCA) جهت کاهش ابعاد آن‌ها مورد بررسی قرار می‌دهد (Jolliffe, 2002) و مجموعه داده‌های آموزشی اصلی را به زیرمجموعه‌هایی تقسیم می‌کند که برای آموزش طبقه‌بندی کننده‌ها مورد استفاده قرار می‌گیرند (Roy and Saha, 2022).

مدل Dagging

الگوریتم Dagging توسط تینگ و ویتن در سال ۱۹۹۷ ارائه شد. این الگوریتم یک مدل جمعی نمونه‌گیری مجدد است که بیشترین آراء را برای ادغام طبقه‌بندی کننده‌های مختلف به کار می‌برد تا دقت بهتری در پیش‌بینی طبقه‌بندی

1 . Random forest

[DOI: 10.61186/jeer.14.4.19]

کننده‌های پایه داشته باشد (Zhao et al., 2013; Walia and Kumar, 2019). مدل جمعی Dagging برای حل مشکلات متعددی در ارتباط با طبقه‌بندی در دنیای واقعی استفاده شده است (Towfiqul Islam at al., 2020). یکی از مزیت‌های الگوریتم Dagging این است که شامل تعداد مشخصی از نمونه‌های گسسته است که برای نمونه‌های bootstrap جایگزین می‌شوند تا طبقه‌بندی کننده‌های پایه را به دست آورند (Roy and Saha, 2021). همچنین این الگوریتم دسته‌هایی مجزا و متمایزی از داده‌ها را تولید کرده و مجموعه‌ای از داده‌ها را برای نسخه‌ای تکراری از الگوریتم یادگیری پایه آماده می‌کند (Kotsianti and Kanellopoulos, 2007).

فرایند اجرای تمامی مدل‌ها در نرم‌افزار داده‌کاوی WEKA[1] انجام شده است. نرم‌افزار WEKA به‌عنوان یک نرم‌افزار یادگیری ماشین برای نخستین بار در کشور نیوزیلند و در دانشگاه Waikato معرفی شده است. این نرم‌افزار مجموعه‌ای از به‌روزترین الگوریتم‌های یادگیری ماشین را جهت انجام داده‌کاوی ارائه می‌دهد. WEKA شامل ابزارهایی برای آماده‌سازی داده‌ها، طبقه‌بندی، رگرسیون، خوشه‌بندی، ایجاد قوانین انجمنی و تصویرسازی (کار با نمودارهای مختلف) است (Rahimpour et al., 2023). این نرم‌افزار به زبان برنامه‌نویسی Java نوشته شده است و استفاده از آن بر روی پلت‌فرم‌هایی مثل ویندوز رایگان است.

مدل‌سازی سیلاب با استفاده از الگوریتم‌های داده‌کاوی در محیط نرم‌افزار WEKA طی مراحل زیر انجام شده است. ۱- تعیین داده‌های آموزشی و اعتبارسنجی در محیط نرم‌افزار ArcGIS ۲- تهیه نقشه‌های عوامل مؤثر در وقوع سیل و کلاس‌بندی هر یک از آن‌ها ۳- تبدیل تمامی نقشه‌ها به فرمت ASCII و انتقال آن‌ها به محیط نرم‌افزار SPSS ۴- انتقال لایه‌ها به محیط نرم‌افزار WEKA با فرمت ARFF ۵- انجام پیش‌پردازش‌های مورد نیاز بر روی پارامترها از نوار ابزار Preprocess ۶- اجرای هر یک از الگوریتم‌های داده‌کاوی از نوار ابزار Classify ۷- انتقال نتایج الگوریتم‌های داده‌کاوی به محیط نرم‌افزار ArcGIS و تهیه نقشه‌های پتانسیل خطر وقوع سیل ۸- ارزیابی دقت نقشه‌های نهایی.

تحلیل هم خطی چندگانه

آزمون هم خطی چندگانه وجود یک رابطه خطی بین تمامی متغیرهای توصیفی یک مدل رگرسیونی را تخمین می‌زند (Gujarati, 2004). روش‌های هم خطی چندگانه برای شناسایی مؤثرترین عوامل شرطی‌سازی سیل برای وقوع سیلاب‌ها و افزایش عملکرد مدل‌های پیش‌بینی اتخاذشده، مفید هستند (Bui et al., 2019; Khosravi et al., 2019). در پژوهش حاضر از شاخص‌های عامل تورم واریانس (VIF[2]) و تولرانس (T^3) برای تعیین هم خطی چندگانه بین متغیرهای مستقل استفاده شد.

$$TOL = 1 - R_v^2$$ **رابطه ۱**

$$VIF = 1/TOL$$ **رابطه ۲**

که در آن R_v مقدار مربع R رگرسیون با استفاده از پارامتر رگرسیون v در موارد دیگر هست. مقادیر VIF بزرگ‌تر از ۱۰ و مقادیر TOL کمتر از ۰/۱ نشان‌دهنده هم خطی شدید هست (Hair et al., 2009). وجود هم خطی بین پارامترهای انتخاب شده باعث می‌شود که نقشه‌های نهایی از دقت پایینی برخوردار باشند.

1 . Waikato Environment for Knowledge Analysis
2 . variance inflation factor
3 . tolerance

[DOI: 10.61186/jeer.14.4.19]

ارزیابی و مقایسه عملکرد مدل‌ها

در پژوهش حاضر به‌منظور ارزیابی دقت نقشه‌های پتانسیل خطر وقوع سیل از منحنی مشخصه عملیاتی دریافت‌کننده یا منحنی مشخصه عملکرد سیستم (ROC[1]) و سطح زیر منحنی (AUC[2]) استفاده شده است. در منحنی ROC روی محور X مقدار تشخیص‌پذیری یا ویژگی Specificity (نسبت پیکسل‌های بدون سیل که به درستی به‌عنوان بدون سیل طبقه‌بندی شده‌اند) و روی محور Y مقدار حساسیت Sensitivity (نسبت پیکسل‌های سیل که به درستی به‌عنوان سیل طبقه‌بندی شده‌اند) مشخص است. مقادیر Specificity و Sensitivity با استفاده از روابط زیر محاسبه می‌شوند:

رابطه ۳

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

رابطه ۴

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

که در آن TN[3] تعداد مناطق بدون سیل که درست تشخیص داده شده‌اند؛ FP[4] تعداد نقاط سیل که اشتباه تشخیص داده شده‌اند؛ TP[5] تعداد نقاط سیل که درست تشخیص داده شده‌اند و FN[6] تعداد نقاط بدون سیل که اشتباه تشخیص داده شده‌اند. جهت نشان دادن عملکرد هر یک از مدل‌ها از شاخص آماری صحت (Accuracy) نیز استفاده شد. این شاخص نسبت نقاط سیل گیر و بدون سیل که به درستی طبقه‌بندی شده‌اند را نشان می‌دهد.

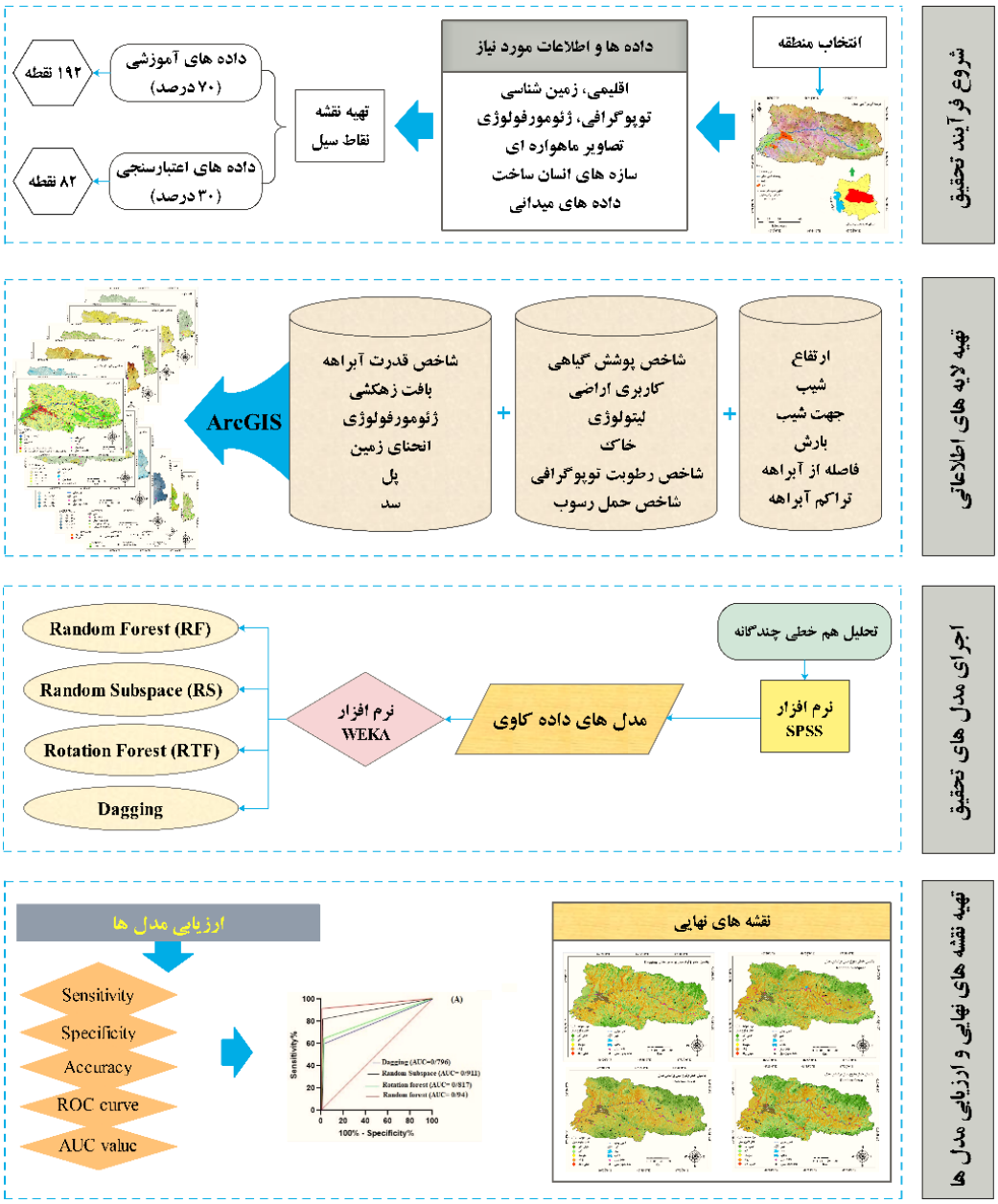
رابطه ۵

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

منحنی ROC در زمینه‌های مختلفی از جمله مهندسی، پزشکی و فیزیک به کار برده شده، اما به طور عمده برای ارزیابی نقشه‌های پیش‌بینی مخاطرات طبیعی استفاده می‌شود (Pradhan et al., 2014; Roy and Saha, 2019). سطح زیر منحنی ROC نشان‌دهنده قابلیت و توانایی یک مدل برای پیش‌بینی پیکسل‌های سیل و بدون سیل هست. مقدار AUC صفر بیانگر این است که مدل غیرواقعی بوده و هر چقدر مقدار AUC بیشتر باشد عملکرد و کارایی مدل در پیش‌بینی بهتر خواهد بود (Tien Bui et al., 2016). در شکل (۲) مراحل انجام این تحقیق به‌صورت فلوچارت نمایش داده شده است.

1 . **R**eceiver **O**perating **C**haracteristics
2 . Area under the ROC curve
3 . True Negative
4 . False Positive
5 . True Positive
6 . False Negative

[DOI: 10.61186/jeer.14.4.19]



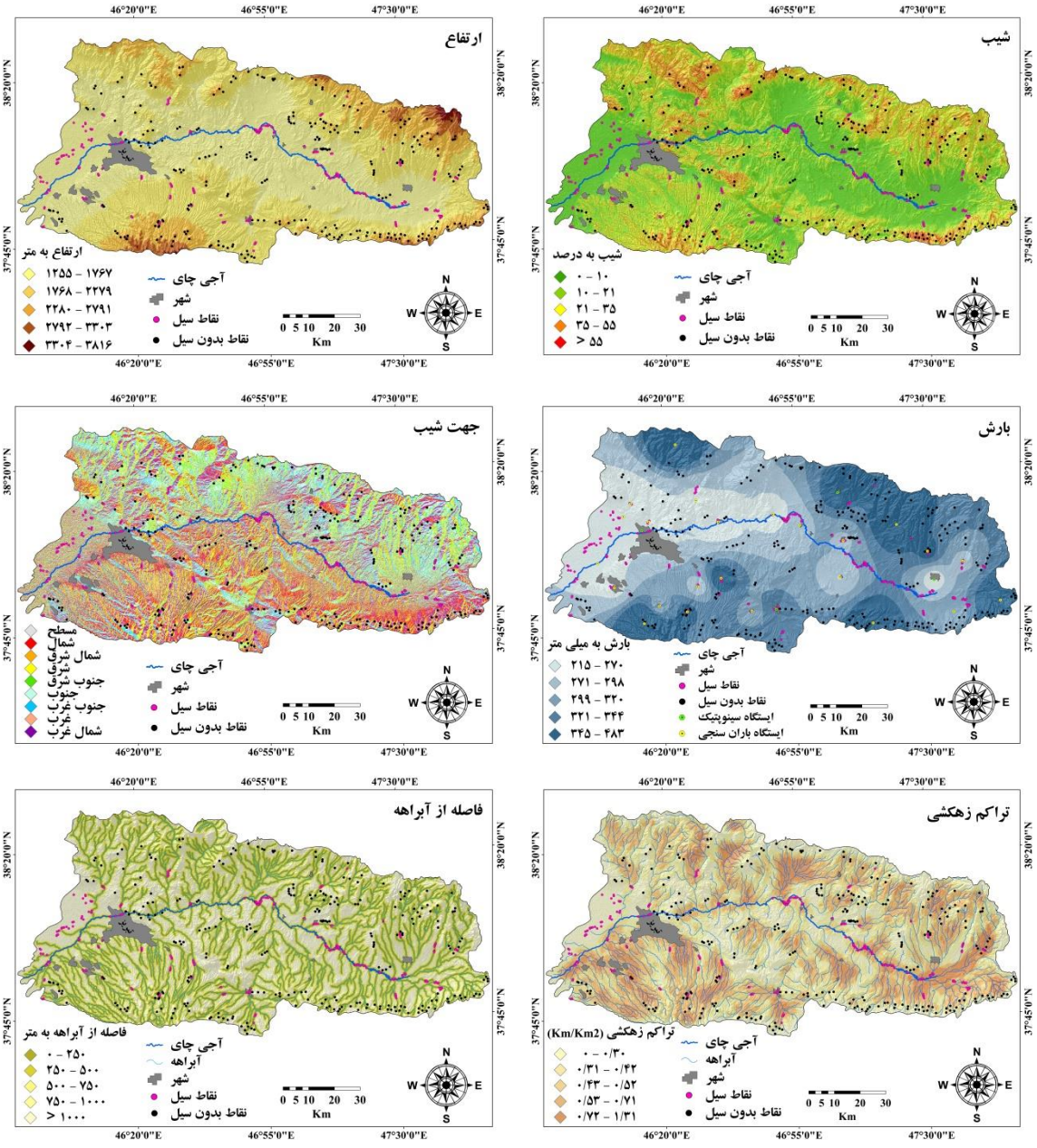
شکل ۲: فلوچارت مراحل تهیه نقشه‌های پتانسیل خطر وقوع سیل

تهیه لایه‌های اطلاعاتی

در پژوهش حاضر با بررسی پژوهش‌های انجام گرفته در ارتباط با موضوع تحقیق و همچنین در دسترس بودن داده‌ها، ۱۸ پارامتر مؤثر در وقوع سیل جهت تهیه نقشه‌های پتانسیل خطر وقوع سیلاب در سطح حوضه آبریز آجی‌چای انتخاب شدند. پارامترهای مورد بررسی عبارتند از: ارتفاع، شیب، جهت شیب، بارش، فاصله از آبراهه، تراکم آبراهه، شاخص پوشش گیاهی، کاربری اراضی، لیتولوژی، گروه‌های هیدرولوژیکی خاک، شاخص رطوبت توپوگرافی، شاخص حمل رسوب، شاخص قدرت آبراهه، بافت زهکشی، ژئومورفولوژی، انحنای زمین، فاصله از پل و فاصله از سد. به‌منظور تهیه نقشه‌های طبقات ارتفاعی، شیب، جهت شیب، شاخص رطوبت توپوگرافی، شاخص قدرت آبراهه، شاخص حمل رسوب و انحنای زمین از مدل رقومی ارتفاعی زمین (DEM) با قدرت تفکیک مکانی ۲۸ متر استفاده شد. جهت تهیه نقشه بارش منطقه از داده‌های بارش ایستگاه‌های سینوپتیک (۴ ایستگاه) و باران‌سنجی (۲۴ ایستگاه) موجود در داخل محدوده

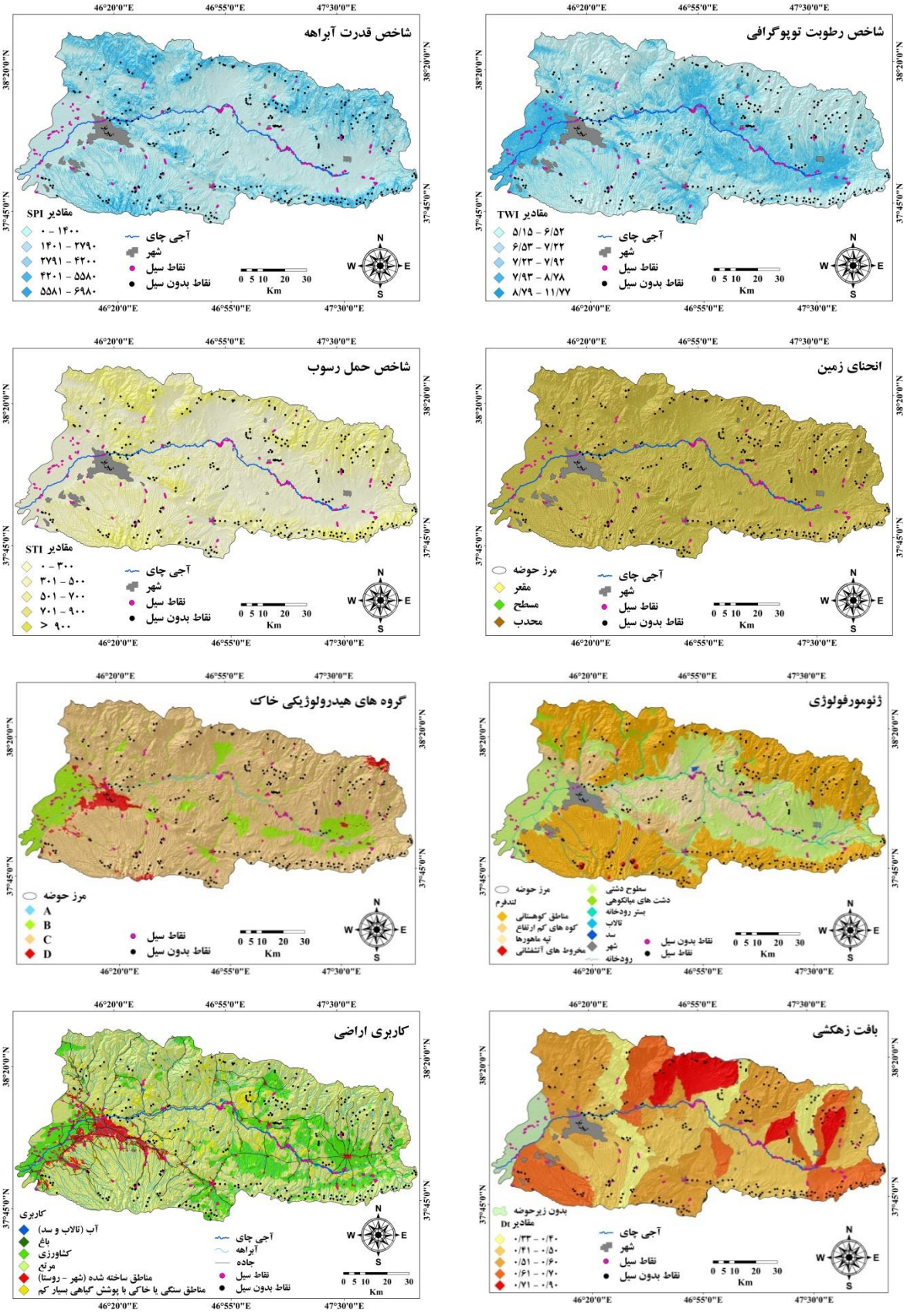
[DOI: 10.61186/jeer.14.4.19]

حوضه آبریز آجی چای استفاده شده است. نقشه نهایی بارش از طریق درون‌یابی و با روش IDW در محیط نرم‌افزار ArcGIS تهیه شد. نقشه شاخص پوشش گیاهی با استفاده از تصاویر ماهواره‌ای لندست ۸ با ردیف و گذرهای ۳۳-۱۶۸، ۳۴-۱۶۸، ۳۴-۱۶۷ و مربوط به سال ۲۰۲۲ تهیه شد. جهت تعیین موقعیت پل‌ها و سدها در منطقه از امکانات سامانه Google Earth استفاده شده و لایه پل‌ها و سدها با فرمت وکتوری تهیه و سپس وارد محیط نرم‌افزار ArcGIS شدند. نقشه نهایی این دو پارامتر با استفاده از ابزار فاصله اقلیدسی و در پنج کلاس تهیه شد. نقشه گروه‌های هیدرولوژیکی خاک منطقه از طریق داده‌های جهانی گروه‌های هیدرولوژیک خاک برگرفته از وب‌سایت ناسا تهیه شد. جهت تهیه نقشه لیتولوژی از نقشه‌های زمین‌شناسی ۱:۱۰۰۰۰۰ و ۱:۲۵۰۰۰۰ (برگه‌های تبریز، سراب، بستان‌آباد، هریس، اسکو و خوجا) برگرفته از سازمان زمین‌شناسی کشور استفاده شده است. (شکل ۳).



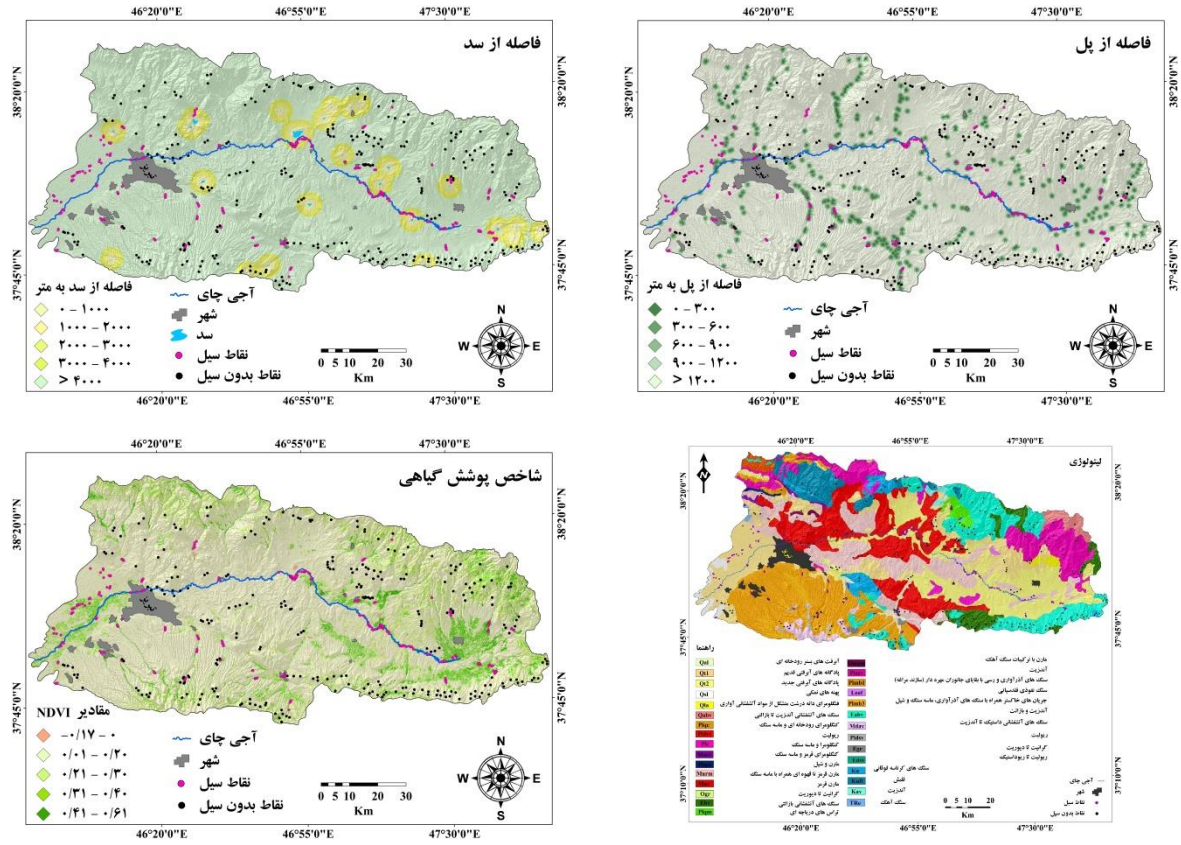
شکل ۳: نقشه لایه‌های اطلاعاتی مورد استفاده در تحقیق

[DOI: 10.61186/jeer.14.4.19]



ادامه شکل ۳: نقشه لایه‌های اطلاعاتی مورد استفاده در تحقیق

[DOI: 10.61186/jeer.14.4.19]



ادامه شکل ۳: نقشه لایه‌های اطلاعاتی مورد استفاده در تحقیق

۴-یافته‌ها

نتایج به‌دست‌آمده از تحلیل هم خطی چندگانه بین پارامترها نشان می‌دهد که به جز پارامتر بافت زهکشی، سایر متغیرهای مستقل انتخاب‌شده جهت تهیه نقشه‌های پتانسیل خطر وقوع سیل دارای هم خطی پایینی هستند (جدول ۱). بنابراین ۱۷ پارامتر در مدل‌سازی خطر وقوع سیل با استفاده از الگوریتم‌های داده‌کاوی مورد استفاده قرار گرفته‌اند. محاسبات مربوط به این بخش در نرم‌افزار SPSS انجام شده است.

جدول ۱: نتایج تحلیل هم خطی چندگانه بین پارامترها

پارامتر	TOL	VIF	پارامتر	TOL	VIF
ارتفاع	۰/۱۴۱	۲/۶	فاصله از سد	۰/۱۲۰	۲/۸
شیب	۰/۷۸۱	۱/۵	لیتولوژی	۰/۷۶۹	۱/۳۰
جهت شیب	۰/۱۴۰	۲/۷	گروه‌های هیدرولوژیکی خاک	۰/۵۴۲	۱/۲۷
بارش	۰/۱۳۳	۲/۹	شاخص رطوبت توپوگرافی	۰/۱۳۰	۳/۵
فاصله از آبراهه	۰/۷۲	۱/۳۷	شاخص حمل رسوب	۰/۱۱۰	۴/۱
تراکم آبراهه	۰/۱۱۴	۳/۹	شاخص قدرت آبراهه	۰/۲۱۱	۲/۳
شاخص پوشش گیاهی	۰/۱۸۰	۲/۵	بافت زهکشی	۰/۰۹۶	۱۰/۱۲
کاربری اراضی	۰/۱۲۶	۳/۶۵	انحنا	۰/۱۱۳	۳
فاصله از پل	۰/۱۳۸	۲/۶۴	ژئومورفولوژی	۰/۸۷۲	۱/۱۴

[DOI: 10.61186/jeer.14.4.19]

در ادامه جهت اجرای هر یک از الگوریتم‌ها، خصوصیات زیر که پس از آزمون‌وخطای گسترده به دست آمده‌اند، تعیین شدند. به‌عنوان‌مثال برای اجرای الگوریتم Dagging از این خصوصیات استفاده شد: اندازه بسته[1]، ۱۰۰؛ تعداد بذر یا دانه[2]، ۱؛ تعداد دسته یا گروه[3]، ۱۰ و decision stump tree به‌عنوان طبقه‌بندی کننده پایه[4]. در جدول (۲) مقادیر بهینه استفاده شده جهت اجرای هر یک از الگوریتم‌های داده‌کاوی ارائه شده است.

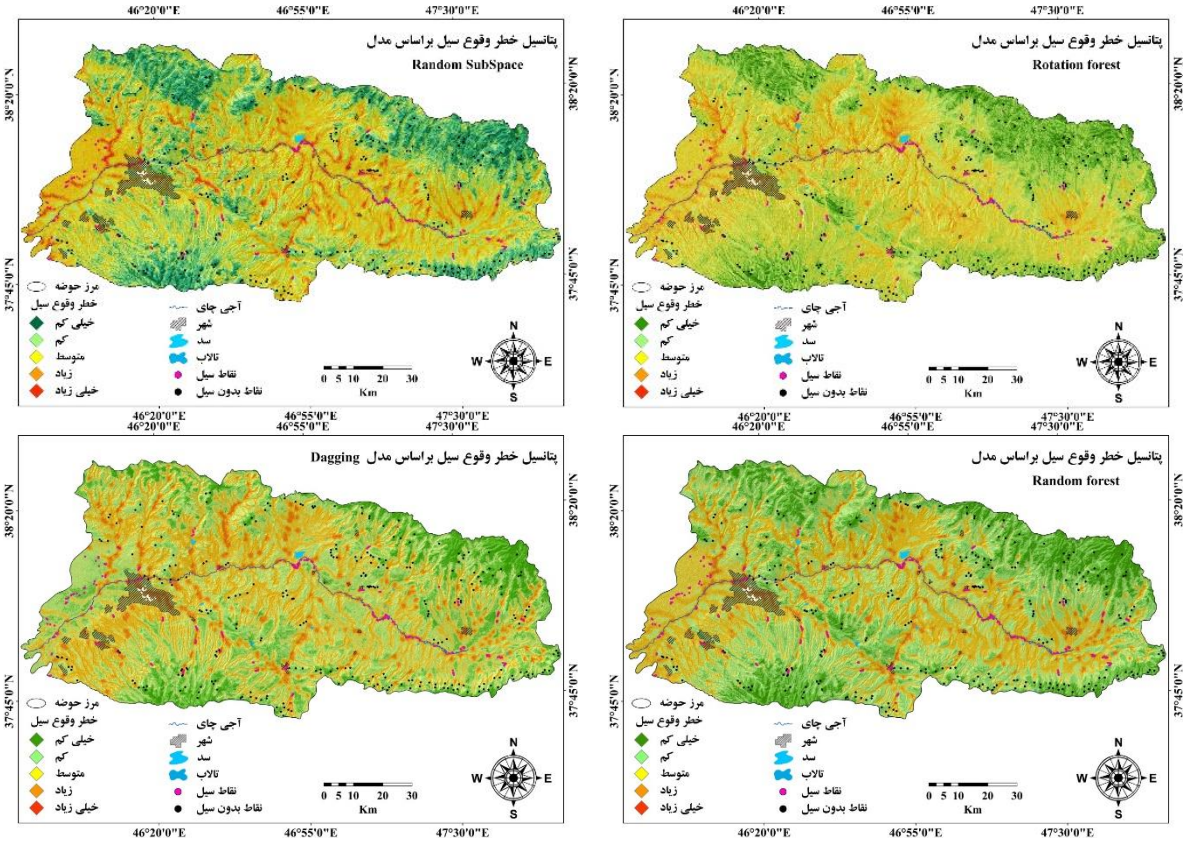
جدول ۲: مقادیر بهینه استفاده شده جهت اجرای الگوریتم‌های داده‌کاوی

الگوریتم	تعداد پارامترهای ورودی	مقادیر استفاده شده
Random forest	۱۷	Batchsize: 100; numExecutionSlots: 1; number of iterations: 21; seed: 1
Random subspace	۱۷	Batchsize: 100; numExecutionSlots: 1; numFolds: 3; number of iterations: 21; seed: 1; base classifier: REPTree
Rotation forest	۱۷	Batchsize: 100; confidenceFactor: 0.25; numExecutionSlots: 1; numFolds: 3; number of iterations: 21; seed: 1; base classifier: J48
Dagging	۱۷	Batchsize: 100; number of folds: 10; seed: 1; base classifier: decision stump tree:

تحلیل اهمیت هر یک از پارامترها بر اساس داده‌های آموزشی در نرم‌افزار WEKA نشان داد که در مدل Random subspace پارامترهای طبقات ارتفاعی، شیب، فاصله از آبراهه و لیتولوژی به ترتیب بیشترین اهمیت را نسبت به سایر پارامترها داشته‌اند. در مقابل عوامل فاصله از پل و شاخص حمل رسوب کمترین اهمیت را در تهیه نقشه پتانسیل وقوع سیل دارند. در مدل Dagging مهم‌ترین عوامل مؤثر عبارت بودند از: ارتفاع، گروه‌های هیدرولوژیکی خاک، شاخص رطوبت توپوگرافی، فاصله از آبراهه و شاخص پوشش گیاهی. پارامترهای شاخص قدرت آبراهه و تراکم آبراهه نیز اهمیت کمتری نسبت به سایر پارامترها داشته‌اند. در مدل Rotation forest نیز پارامترهای لیتولوژی، شیب، بارش، ارتفاع و فاصله از آبراهه به ترتیب مهم‌ترین عوامل بودند. در مقابل انحنای زمین، فاصله از سد و فاصله از پل از اهمیت کمتری برخوردار بوده‌اند. مهم‌ترین عوامل مؤثر در مدل جنگل تصادفی (RS) نیز به ترتیب عبارت بودند از: ارتفاع، فاصله از آبراهه، شیب و بارش. نقشه‌های پتانسیل خطر وقوع سیل بر اساس الگوریتم‌های داده‌کاوی در محیط نرم‌افزار ArcGIS و در پنج کلاس با عنوان پتانسیل خیلی کم، کم، متوسط، زیاد و خیلی زیاد تهیه شدند (شکل ۴). بررسی نقشه‌های نهایی نشان می‌دهد که الگوی پراکنش فضایی پهنه‌های خطر در همه نقشه‌ها یکسان هست. به‌طوری‌که در همه نقشه‌ها ارتفاعات و دامنه‌های شیب‌دار دارای پتانسیل خیلی کم بوده و به سمت مناطق دشتی و ارتفاعات پایین که از شیب بسیار کمی برخوردار هستند، میزان پتانسیل وقوع سیل زیاد می‌شود. توزیع نقاط سیلابی در کلاس‌های نقشه شیب نشان می‌دهد که بیش از ۸۰ درصد از سیلاب‌ها در شیب‌های ۱۰-۰ درصد اتفاق افتاده‌اند، که این امر نشان‌دهنده اثرگذاری این عامل در سیلاب‌های منطقه است.

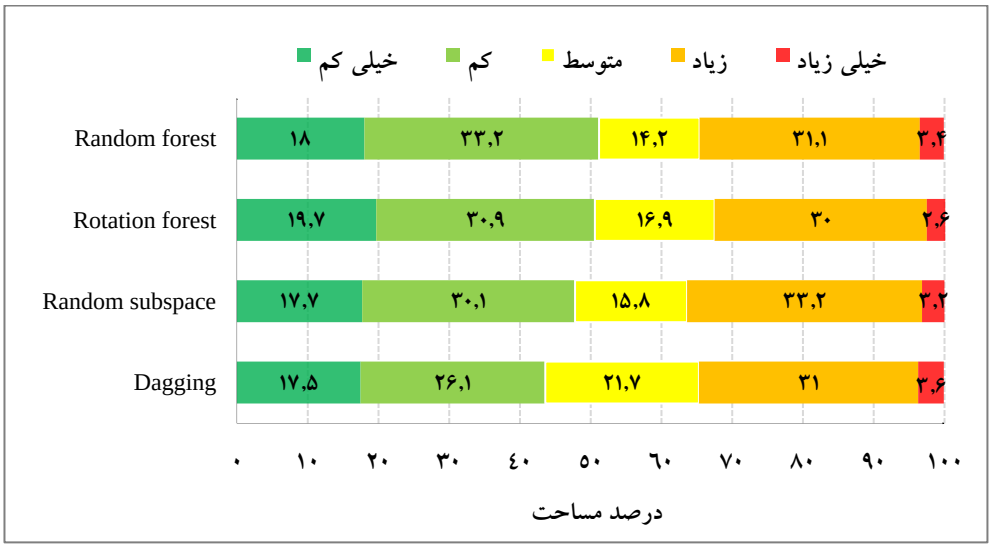
1 . Batchsize
2 . seed
3 . number of folds
4 . base classifier

[DOI: 10.61186/jeer.14.4.19]



شکل ۴: نقشه پتانسیل خطر وقوع سیل بر اساس الگوریتم های داده کاوی

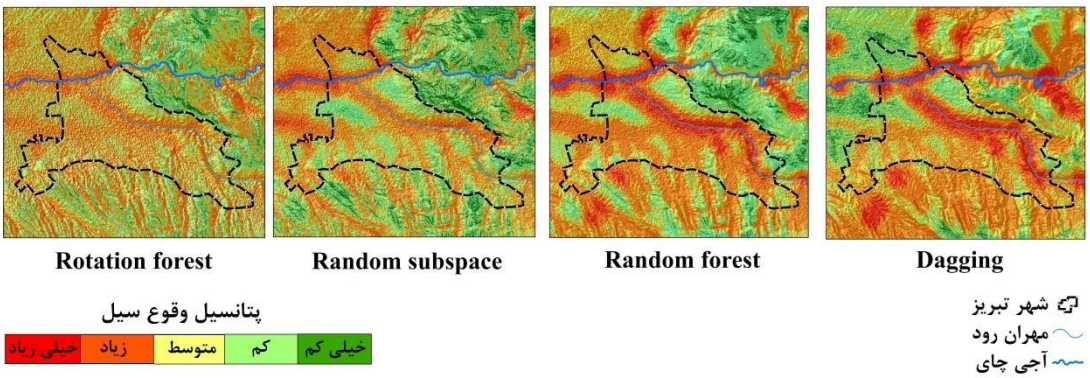
بررسی مساحت هر یک از کلاس‌های خطر وقوع سیل در نقشه‌های نهایی به دست آمده از الگوریتم‌های داده‌کاوی نشان می‌دهد که در تمامی نقشه‌ها بیش از ۳۰ درصد از مساحت منطقه در پهنه‌های زیاد و خیلی زیاد از نظر پتانسیل وقوع سیل قرار دارند. درحالی‌که مساحت کلاس‌های کم و خیلی کم برای مدل‌های Dagging، Random subspace، Rotation forest و Random forest به ترتیب ۴۳/۶، ۴۷/۸، ۵۰/۴ و ۵۲/۲ درصد هست (شکل ۵).



شکل ۵: درصد مساحت هر یک از کلاس‌های خطر وقوع سیل

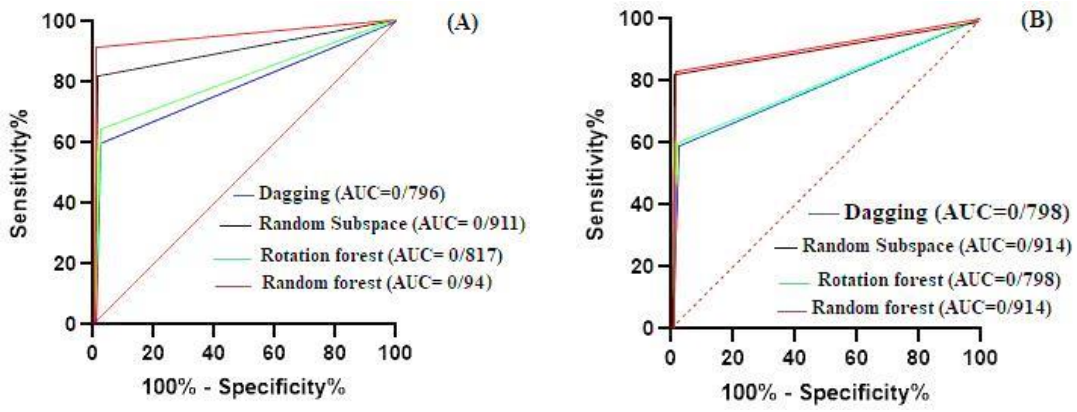
[DOI: 10.61186/jeer.14.4.19]

کلان‌شهر تبریز به‌عنوان مهم‌ترین مرکز جمعیتی و محل تمرکز صنایع سبک و سنگین در داخل حوضه آبریز آجی‌چای هست. وجود دو رودخانه مهران‌رود که از داخل شهر جریان پیدا می‌کند و رودخانه آجی‌چای که از شمال شهر عبور می‌کند، پتانسیل وقوع سیل در این شهر را افزایش داده است. بررسی پهنه‌های خطر وقوع سیل در سطح شهر تبریز در خروجی مدل‌های مورد استفاده در پژوهش حاضر نشان می‌دهد که بخش‌های زیادی از شهر در معرض خطر وقوع سیل قرار دارند (شکل ۶). واحدهای هیدرولوژی کوچک و بزرگ متعددی در قسمت شمال و جنوب شهر تبریز قرار دارند که به هنگام بارندگی، رواناب‌هایی که در این واحدها شکل می‌گیرند به تبعیت از شیب توپوگرافی منطقه به سمت شهر جریان می‌یابند. توسعه و گسترش فیزیکی شهر تبریز به سمت خروجی این واحدها و افزایش ساخت‌وسازها در حریم آبراهه‌های فرعی و اصلی و همچنین عدم توجه به ملاحظات هیدروژئومورفولوژیکی در برنامه‌ریزی شهری، خطر سیل گرفتگی را در این مناطق به شدت افزایش داده است.



شکل ۶: پهنه‌های خطر وقوع سیل در سطح کلان‌شهر تبریز

نتایج ارزیابی دقت نقشه‌های نهایی با استفاده از منحنی ROC نشان داد که مدل Random forest با سطح زیر منحنی ۰/۹۴ نسبت به سایر مدل‌ها عملکرد بهتری داشته است (شکل ۷).



شکل ۷: منحنی ROC برای نتایج الگوریتم‌های داده‌کاوی (A: آموزشی و B: اعتبارسنجی)

[DOI: 10.61186/jeer.14.4.19]

نتایج ارزیابی دقت نقشه‌های نهایی بر اساس شاخص‌های آماری نیز نشان می‌دهد که مدل Random forest با ضریب ۰/۹۰ از دقت بالایی نسبت به سایر الگوریتم‌های داده‌کاوی برخوردار هست. الگوریتم‌های Random subspace، Rotation forest و Dagging نیز به ترتیب در رتبه‌های بعدی قرار دارند (جدول ۴).

جدول ۴: ارزیابی دقت الگوریتم‌های داده‌کاوی بر اساس داده‌های آموزشی و اعتبارسنجی

شاخص آماری	آموزشی			
	Dagging	Random subspace	Rotation forest	Random forest
True positive	۱۱۴	۱۵۸	۱۲۲	۱۷۲
True negative	۱۵۰	۱۷۶	۱۷۷	۱۷۷
False positive	۷۸	۳۴	۷۰	۲۰
False negative	۴۲	۱۶	۱۵	۱۵
Sensitivity (%)	۰/۷۴	۰/۹۰	۰/۸۹	۰/۹۱
Specificity (%)	۰/۶۵	۰/۸۳	۰/۷۱	۰/۸۹
Accuracy (%)	۰/۶۸	۰/۸۶	۰/۷۷	۰/۹۰
	اعتبارسنجی			
True positive	۴۹	۶۸	۴۹	۶۸
True negative	۶۷	۶۲	۷۳	۷۵
False positive	۳۳	۱۴	۳۳	۱۴
False negative	۱۵	۲۰	۹	۷
Sensitivity (%)	۰/۷۶	۰/۷۷	۰/۸۴	۰/۹۰
Specificity (%)	۰/۶۷	۰/۸۱	۰/۶۸	۰/۸۴
Accuracy (%)	۰/۷۰	۰/۷۹	۰/۷۴	۰/۸۷

۵ـ بحث و نتیجه‌گیری

در تحقیق حاضر تلاش شد تا با استفاده از الگوریتم‌های داده‌کاوی، نقشه پتانسیل خطر وقوع سیلاب در حوضه آبریز آجی‌چای تهیه شود. به همین منظور از ۱۸ پارامتر مؤثر در وقوع این پدیده استفاده شد. لایه‌های اطلاعاتی همه پارامترها به فرمت رستری در محیط نرم‌افزار ArcGIS تهیه شدند. بررسی اهمیت پارامترها نشان داد که مناطق نزدیک رودخانه‌ها و پل‌ها، سطوح کم ارتفاع و دارای شیب کم مستعد خطر وقوع سیل می‌باشند. نقشه‌های نهایی در پنج طبقه از پتانسیل خیلی کم تا خیلی زیاد تهیه شدند. نتایج تحقیق نشان داد که الگوی توزیع فضایی پهنه‌های خطر در تمامی مدل‌ها مشابه هم بوده است. به‌این‌ترتیب که مناطق پایین‌دست حوضه و اطراف آبراهه‌های اصلی منطقه پرخطرترین پهنه‌ها را به خود اختصاص داده‌اند. کلان‌شهر تبریز نیز به‌عنوان مهم‌ترین مرکز جمعیتی داخل حوضه به دلیل قرارگیری در مسیر رودخانه‌های آجی‌چای و مهران‌رود در پهنه‌های پرخطر قرار دارد که آسیب‌پذیری آن را در هنگام وقوع سیلاب‌های مخرب نشان می‌دهد. ارزیابی دقت نقشه‌های نهایی با استفاده از شاخص‌های آماری و منحنی ROC نشان داد که مدل Random forest با سطح زیر منحنی ۰/۹۴ از دقت خوبی نسبت به سایر مدل‌ها برخوردار بوده است. بررسی نتایج تحقیق حاضر با نتایج محققانی همچون Kazemi و همکاران (2021) در شهر نوشهر، Karami و همکاران

[DOI: 10.61186/jeer.14.4.19]

(2023) در دشت بیرجند و Chezgi and Poyan (2024) در حوضه آبریز شهرستانک مطابقت دارد. محققین نام برده استفاده از مدل‌های داده‌کاوی را روش مناسبی جهت تهیه نقشه‌های پتانسیل خطر وقوع سیل پیشنهاد کرده بودند که نتایج به‌دست‌آمده از این تحقیق نیز بیانگر تائید نتایج محققین مذکور هست. با توجه به نتایج تحقیق که نشان می‌دهد در تمامی نقشه‌های تهیه شده بیش از ۳۰ درصد از مساحت منطقه در طبقات زیاد و خیلی زیاد از نظر خطر وقوع سیل قرار دارند، لذا جهت مدیریت هر چه بهتر سیلاب در این منطقه پیشنهادات زیر می‌تواند برای مدیران منطقه راهگشا باشد. ۱- اجرای برنامه‌های آبخیزداری مناسب در سطح حوضه آبریز آجی‌چای ۲- احداث سدهای تنظیمی در مسیر آبراهه‌هایی که هر سال حالت طغیانی به خود می‌گیرند ۳- لایروبی منظم بستر آبراهه‌های اصلی منطقه ۴- جلوگیری از دست‌اندازی به حریم و بستر آبراهه‌های اصلی منطقه، به‌ویژه در مناطقی که رودخانه‌ها از داخل روستا یا شهرها عبور می‌کنند.

قدردانی

تحقیق حاضر بخشی از نتایج طرح پژوهشی دوره پسادکتری در دانشگاه تبریز می‌باشد که با حمایت بنیاد ملی نخبگان در قالب طرح شهید دکتر چمران انجام گرفته است. بدین‌وسیله از آن‌ها قدردانی می‌شود.

منابع

1. Aldiansyah, S. & Wardani, F. (2023). Evaluation of flood susceptibility prediction based on a resampling method using machine learning. *Journal of Water and Climate Change*, 14(3), 937-961. https://doi.org/10.2166/wcc.2023.494
2. Breiman, L., HFriedman, J., Olshen, R.A. & Stone, C.J. (1984). *Classification and regression trees*. Chapman & Hall, New York.
3. Bui, D.T., Tsangaratos, P., Ngo P-TT., Thai Pham, T. & Thai Phamet, B. (2019). Flash flood susceptibility modeling using an optimized fuzzy rule based feature selection technique and tree based ensemble methods. *Sci Total Environ*, 668,1038–1054. https://doi.org/10.1016/j.scitotenv.2019.02.422
4. Cheraghi Ghalehsari, A., Habibnejad Roshan, M. & Roshun, S. H. (2020). Flood Susceptibility Mapping Using a Support Vector Machine Models (SVM) and Geographic Information System (GIS). Journal of Natural Environmental Hazards, 9(25), 61-80. (In Persian). doi: 10.22111/jneh.2020.31018.1547
5. Chezgi, J., Poyan, S. (2024). Determining Flood-Prone Areas Using Machine Learning Models in the Shahrestank Watershed Area of Khosef City. *Jwmseir*, 17 (63). 38-50.
6. Costache, R., Popa, M.C., Tien Bui, D., Diaconu, D.C., Ciubotaru, N., Minea, G., Pham, QB. (2020). Spatial predicting of flood potential areas using novel hybridizations of fuzzy decision-making, bivariate statistics, and machine learning. *Journal of Hydrology*, 585:124808. https:// doi. org/ 10. 1016/j. jhydr ol. 2020. 124808.
7. Diakakis, M., Mavroulis, S. & Deligiannakis, G. (2012). Floods in Greece, a statistical and spatial approach. Nat. Hazards, 62, 485–500. https://doi. org/10.1007/s11069-012-0090-z
8. Gujarati, DN. (2004). *Basic econometrics*. 4th ed. New York City (NY), The MacGraw Hill Company, p. 1002.
9. Hirabayashi, Y., Mahendran, R., Koirala, S., Konoshima, L., Yamazaki, D., Watanabe, S., Kim, H. & Kanae, S. (2013). Global flood risk under climate change. *Nature Climate Change*, 3(9), 816.
10. Hitouri, S., Mohajane, M., Lahsaini, M., Ali, S.A., Setargie, T.A., Tripathi, G., D'Antonio, P., Singh, S.K., Varasano, A. (2024). Flood Susceptibility Mapping Using SAR Data and Machine

[DOI: 10.61186/jeer.14.4.19]

Learning Algorithms in a Small Watershed in Northwestern Morocco. *Remote Sens*, 16, 858, 1-21. https://doi.org/10.3390/rs16050858

11. Hudson, P., Botzen, W.J.W., Kreibich, H., Bubeck, P. & Aerts, J.C.J.H. (2014). Evaluating the effectiveness of flood damage mitigation measures by the application of propensity score matching. *Nat. Hazards Earth Syst. Sci*, 14, 1731–1747. https://doi.org/10.5194/nhess-14-1731-2014
12. Jolliffe, I. (2002). *Principal component analysis*. Wiley Online Library.
13. 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. doi: 10.22059/jne.2022.350170.2485
14. Kazemi, H., Mansouri, N., Jozi, S.A. (2021). Flood risk zoning in Nowshahr city using machine learning models. *JHRE*. 40(176), 71-86.
15. Khosravi, K., Panahi, M., Golkarian, A., Keesstra, S.D., Saco, P.M., Bui, D.T., Lee, S. (2020). Convolutional neural network approach for spatial prediction of flood hazard at national scale of Iran. *J. Hydrol*, 591. https://doi.org/10.1016/j.jhydrol.2020.125552
16. Khosravi, K., Shahabi, H., Thai Pham, B., Adamowski, , J., Shirzadi, A., Pradhan, B., Dou, J., Ly, H., Gróf, G., Loc Ho, H., Hong, H., Chapi, K. & Prakash, I. (2019). A comparative assessment of flood susceptibility modeling using Multi-Criteria Decision-Making Analysis and Machine Learning Methods. *Journal of Hydrology*, 573, 311-323. doi: 10.1016/j.jhydrol.2019.03.073
17. Kotsianti, S.B. & Kanellopoulos, D. (2007). *Combining Bagging, Boosting and Dagging for Classification Problems*. In: Apolloni, B., Howlett, R.J., Jain, L. (Eds.), KnowledgeBased Intelligent Information and Engineering Systems. Springer, Berlin Heidelberg, Berlin, Heidelberg, 493–500.
18. Kourgialas, N.N. & Karatzas, G.P. (2011). Flood management and a GIS modelling method to assess flood- hazard areas—a case study. *Hydrological Sciences Journal*, 56(2), 212–225. doi: https://doi.org/10.1080/02626667.2011.555836
19. Lee, G., Jun, K. & Chung, E. (2013). Integrated multi-criteria flood vulnerability approach using fuzzy Atmospheric TOPSIS and Delphi technique. *Nat. Hazards Earth Syst. Sci*, 13, 1293–1312. https://doi.org/10.5194/nhess-13-1293-2013
20. Lieb, M., Glaser, B. & Huwe, B. (2012). Uncertainty in the spatial prediction of soil texture: comparison of regression tree and Random Forest models. *Geoderma*, 170, 70-79. https://doi.org/10.1016/j.geoderma.2011.10.010
21. Luu, C., Thai Pham, B., Van Phong, T., Costache, R., Duy Nguyen, H., Amiri, M., Duy Bui, Q., Thanh Nguyen, L., Van Le, H., Prakash, I. & Trong Trinh, P. (2021). GIS-based ensemble computational models for flood susceptibility prediction in the Quang Binh Province, Vietnam, *Journal of Hydrology*, 599, 126500. https://doi.org/10.1016/j.jhydrol.2021.12650.
22. Ozcift, A. (2012). SVM feature selection based rotation forest ensemble classifiers to improve computer aided diagnosis of Parkinson disease. *J. Med. Syst*, 36(4), 2141–2147.
23. Paryani, S., Bordbar, M., Jun,C., Panahi, M., M. Bateni, S., M. U. Neale, C., Moeini, H. & Lee, S. (2022). Hybrid based approaches for the flood susceptibility prediction of Kermanshah province, Iran. *Natural Hazards*, 116(2), 1-32. https://doi.org/10.1007/s11069-022-05701-4
24. Peters, J., Verhoest, N., Samson, R., Boeckx, P. & De Baets, B. (2008). Wetland vegetation distribution modelling for the identification of constraining environmental variables. *Landscape Ecology*, 23, 1049- 1065.
25. Piao,Y., Piao, M., Hao, Jin, C., Sun, S-H., Chung, J-M., Hwang, B. & Ho, R. (2015). A New Ensemble Method with Feature Space Partitioning for High- Dimensional Data Classification, *Hindawi Publishing Corporation Mathematical Problems in Engineering*, 1-12. http://dx.doi.org/10.1155/2015/590678
26. Pradhan, B., Abokharima, M.H., Jebur, M.N. & Tehrany, M.S. (2014). Land subsidence susceptibility mapping at Kinta Valley (Malaysia) using the evidential belief function model in GIS. *Natural Hazards*, 73(2), 1019–1042.
27. Rahimpour, T., Rezaei Moghaddam, M. H., Hejazi, S. A. & Valizadeh Kamran, K. (2021). Spatial Variations Analysis of Flood hazard Susceptibility based on a new ensemble model

[DOI: 10.61186/jeer.14.4.19]

(Case Study: Aland Chai Basin, Khoy city). *Environmental Management Hazards*, 8(4), 371-393. (In Persian). https://doi.org/10.22059/jhsci.2022.335204.692

28. Rahimpour, T., Rezaei Moghaddam, M. H., Hejazi, S. A. & Vlaizadeh Kamran, K. (2023). Flood Susceptibility Modeling in the Aland Chai Basin using New Ensemble Classification Approach (FURIA-GA-LogitBoost). *Journal of Geography and Environmental Hazards*, 12(1), 1-24. (In Persian). doi: 10.22067/geoeh.2022.74170.1141
29. Rahman, R., Dhruba, S.R., Ghosh, S. & Pal, R. (2019). Functional random forest with applications in dose-response predictions. *Scientific reports*, 9(1), 1-14.
30. Rezaei Moghaddam, M. H. & Rahimpour, T. (2024a). Preparation of flood hazard potential map using two methods: Frequency Ratio and Statistical Index (Case study: Aji Chai Basin). *Environmental Management Hazards*, 10(4), 291-308. (In Persian). doi: 10.22059/jhsci.2024.369163.803
31. Rezaei Moghaddam, M. H., & Rahimpour, T. (2024b). Evaluating of Flood hazard potential using bivariate statistical analysis method (Case study: Aji Chai basin). *Quantitative Geomorphological Research*, *12*(4), 91-107. (In Persian). doi: 10.22034/gmpj.2024.429929.1473
32. Roy, J. & Saha, S. (2019). GIS-based gully erosion susceptibility evaluation using frequency ratio, cosine amplitude and logistic regression ensembled with fuzzy logic in Hinglo River basin, India. *Remote Sensing Applications: Society and Environment*, 15, 100247. https://doi.org/10.1016/j.rsase.2019.100247
33. Roy, J. & Saha, S. (2021). Integration of artificial intelligence with meta classifiers for the gully erosion susceptibility assessment in Hinglo river basin. Eastern India. *Advances in Space Research*, 67(1), 316–333.
34. Roy, J. & Saha, S. (2022). Ensemble hybrid machine learning methods for gully erosion susceptibility mapping: K- fold cross validation approach. *Artificial Intelligence in Geosciences*, 3, 28-45. doi: https://doi.org/10.1016/j.aiig.2022.07.001
35. Sofia, G., Roder, G., Dalla Fontana, G. & Tarolli, P. (2017). Flood dynamics in urbanised landscapes: 100 years of climate and humans' interaction. *Sci. Rep*, 7, 40527. https://doi.org/10.1038/srep40527
36. Tien Bui, D., Pradhan, B., Nampak, H., Bui, Q.T., Tran, Q.A. & Nguyen, Q.P. (2016). Hybrid artificial intelligence approach based on neural fuzzy inference model and metaheuristic optimization for flood susceptibility modeling in a high-frequency tropical cyclone area using GIS. *Journal of Hydrology*, 540, 317–330. https://doi.org/10.1016/j.jhydrol.2016.06.027
37. Towfiqul Islam, A.B., Talukdar, S., Mahato, S., Kundu, S., UddinEibek, K., BaoPham, Q., Kuriqi, A. & ThuyLinh, N.T. (2021). Flood susceptibility modelling using advanced ensemble machine learning models. *Geoscience Frontiers*, 12(3), 101075. https://doi.org/10.1016/j.gsf.2020.09.006
38. Walia, S. & Kumar, K. (2019). Digital image forgery detection: a systematic scrutiny. *Australian Journal of Forensic Sciences*, 51(5), 488-526.
39. WHO (World Health Organization). (2022). *Floods*. 2017. Available online: https://www.who.int/health- topics/floods (accessed on 13 January 2022).
40. Zarei, M., Zandi, R. & Naemitabar, M. (2022). Assessment of Flood Occurrence Potential using Data Mining Models of Support Vector Machine, Chaid and Random Forest (Case study: Frizi watershed). *J Watershed Manage Res*, 13(25), 133-144. (In Persian). doi:10.52547/jwmr.13.25.133
41. Zhao, C., Zhang, X., Zhang, B., Dang, Q. & Lian, J. (2013). Driver's fatigue expressions recognition by combined features from pyramid histogram of oriented gradient and contourlet transform with random subspace ensembles. *IET Intelligent Transport Systems*, 7(1), 36-45.

[DOI: 10.61186/jeer.14.4.19]