



Evaluation of Tree-Based Artificial Intelligence Models to Predict Flood Risk using GIS

S.A. Eslaminezhad¹, M. Eftekhari^{2*}, S. Mahmoodizadeh³, M. Akbari⁴, and A. Haji Elyasi⁵

Abstract

Floods are one of the most devastating types of natural disasters that every year causes the loss of human lives and properties around the world. The purpose of this study is to evaluate and compare the capability of three machine learning models namely Naïve Bayes Tree (NBTree), Alternating Decision Tree (ADTree), and Random Forest (RF) to predict flood risk in Maneh and Samalqan county. The novelty of the present study is the presentation of tree-based hybrid models that have been less used in previous research. To prepare a flood reference map in the study area, 300 flood-prone locations were identified and were divided into training and validation data sets through random selection with a ratio of 70 to 30. The spatial database of the flood was created using 15 hydrogeological and environmental criteria affecting the flood. Finally, three flood risk prediction maps were generated using NBTree, ADTree, and RF models. To validate the flood risk predicting models, the Area Under the Curve (AUC) factor and the statistical criteria of Positive predictive rate, negative predictive rate, sensitivity, specificity, and accuracy were used. The results showed that the RF model had higher accuracy than the NBTree and ADTree models in predicting flood risk in the study area. The results also showed that the risk of flooding in the central areas of the study area is higher than other areas due to lower altitude and slope.

Keywords: Flood Prediction, Naïve Bayes Tree, Alternating Decision Tree, Random Forest.

Received: July 3, 2021

Accepted: August 21, 2021

ارزیابی مدل‌های هوش مصنوعی مبتنی بر درخت به منظور پیش‌بینی خطر سیل در بستر GIS

سید احمد اسلامی نژاد^۱، مبین افتخاری^{۲*}، سعید محمودی زاده^۳، محمد اکبری^۴ و علی حاجی الیاسی^۵

چکیده

سیل یکی از مخرب‌ترین انواع بلایای طبیعی است که هر ساله باعث از دست رفتن جان و مال انسان‌ها در سراسر جهان می‌شود. هدف از تحقیق حاضر ارزیابی و مقایسه قابلیت سه مدل یادگیری ماشین یعنی درخت بیز ساده (NBTree)، درخت تصمیم متناوب (ADTree) و جنگل تصادفی (RF) برای پیش‌بینی خطر وقوع سیل در شهرستان مانه و سملقان می‌باشد. نوآوری تحقیق حاضر ارائه مدل‌های ترکیبی مبتنی بر درخت می‌باشد که کمتر در تحقیقات پیشین مورد استفاده قرار گرفته‌اند. برای تهیه نقشه مرجع سیل منطقه مورد مطالعه، ۳۰۰ موقعیت مستعد سیل شناسایی شدند و از طریق انتخاب تصادفی با نسبت ۷۰ به ۳۰ به مجموعه داده‌های آموزشی و اعتبارسنجی تقسیم شدند. پایگاه داده مکانی سیل با استفاده از ۱۵ معیار هیدروژئولوژیکی و محیطی مؤثر بر سیل ایجاد شد. در نهایت، نقشه‌های پیش‌بینی خطر سیل با استفاده از مدل‌های NBTree، ADTree و RF تهیه شدند. به منظور اعتبارسنجی مدل‌های پیش‌بینی خطر سیل، معیار سطح زیر منحنی (AUC) و معیارهای آماری نرخ پیش‌بینی مثبت، نرخ پیش‌بینی منفی، حساسیت، ویژگی و دقت مورد استفاده قرار گرفتند. نتایج نشان داد که مدل RF دقت بالاتری نسبت به مدل‌های NBTree و ADTree در پیش‌بینی خطر سیل منطقه مورد مطالعه دارد. همچنین، نتایج نشان داد که احتمال وقوع خطر سیل در مناطق مرکزی منطقه مورد مطالعه به دلیل ارتفاع و شیب کمتر، بیشتر از سایر مناطق است.

کلمات کلیدی: پیش‌بینی سیل، درخت بیز ساده، درخت تصمیم متناوب، جنگل تصادفی.

تاریخ دریافت مقاله: ۱۴۰۰/۴/۱۲

تاریخ پذیرش مقاله: ۱۴۰۰/۵/۳۰

1- M.Sc. Graduate, Department of Surveying Engineering, Faculty of Surveying Engineering and Spatial Information, University of Tehran, Tehran, Iran.

2- M.Sc. Graduate, Civil Engineering, Water and Hydraulic Structures, Young Researchers and Elite Club, Mashhad Branch, Islamic Azad University, Mashhad, Iran. Email: mobineftekhari@yahoo.com

3- M.Sc. Graduate, Department of Remote Sensing Engineering, Faculty of Surveying, University of Industrial and Technological Advanced Studies, Kerman, Iran.

4- Associate Professor at Department of Civil Engineering, University of Birjand, Birjand, Iran. Email: Moakbari@birjand.ac.ir

5- M.Sc. Graduate, Department of Water and Hydraulic Structure, K. N. Toosi University of Technology, Tehran, Iran.

*- Corresponding Author

Dor: [20.1001.1.17352347.1400.17.2.12.5](https://doi.org/10.1001.1.17352347.1400.17.2.12.5)

۱- دانش‌آموخته کارشناسی ارشد، گروه مهندسی نقشه برداری، دانشکده مهندسی نقشه‌برداری و اطلاعات مکانی، دانشگاه تهران، تهران، ایران.

۲- دانش‌آموخته کارشناسی ارشد مهندسی عمران آب و سازه‌های هیدرولیکی و عضو باشگاه پژوهشگران جوان و نخبگان، واحد مشهد، دانشگاه آزاد اسلامی، مشهد، ایران.

۳- دانش‌آموخته کارشناسی ارشد مهندسی سنجش از دور، دانشگاه تحصیلات تکمیلی صنعتی فناوری پیشرفته کرمان، ایران.

۴- دانشیار گروه مهندسی عمران، دانشگاه بیرجند، بیرجند، ایران.

۵- دانش‌آموخته کارشناسی ارشد، گروه آب و سازه‌های هیدرولیکی، دانشگاه صنعتی خواجه‌نصیرالدین طوسی، تهران، ایران.

*- نویسنده مسئول

بحث و مناظره (Discussion) در مورد این مقاله تا پایان پائیز ۱۴۰۰ امکانپذیر است.

۱- مقدمه

خطر سیل در شهرستان جهرم استان فارس استفاده کردند. سطح زیر منحنی (AUC) به دست آمده از منحنی مشخصه عملکرد سیستم (ROC) ۰/۹۱، دقت ۰/۹۱/۸٪، ۰/۹۲/۶٪ و ۰/۹۴/۵٪ را به ترتیب برای مدل‌های ترکیبی ANFIS-ACO، ANFIS-GA و ANFIS-PSO نشان داد. بنابراین مدل ترکیبی ANFIS-PSO به عنوان مدل برتر در تهیه نقشه خطر سیل منطقه مورد مطالعه معرفی شد. به دلیل توانایی شناخت دقیق‌تر و قدرت پیش‌بینی بالاتر، پیش‌بینی نواحی مستعد سیل با استفاده از روش‌های ترکیبی جدید برای مطالعات سیل بسیار مهم است. هدف از این پژوهش تهیه نقشه خطر وقوع سیل با استفاده از روش‌های نوین یادگیری ماشین مبتنی بر درخت است. بنابراین، این مطالعه از مطالعات قبلی متمایز است، زیرا در این مطالعه از مدل‌های ترکیبی یادگیری ماشین مبتنی بر درخت یعنی درخت بیز ساده (NBTree)^{۱۸}، درخت تصمیم متناوب (ADTree)^{۱۹} و جنگل تصادفی (RF) جهت تهیه نقشه پیش‌بینی خطر سیل در شهرستان مانه و سملقان استان خراسان شمالی استفاده شده است که نوآوری تحقیق حاضر نیز می‌باشد. در نهایت، معیار سطح زیر منحنی (AUC) و معیارهای آماری شامل نرخ پیش‌بینی مثبت، نرخ پیش‌بینی منفی، حساسیت، ویژگی و دقت جهت اعتبارسنجی مدل‌های پیش‌بینی خطر سیل در منطقه مورد مطالعه مورد استفاده قرار گرفتند.

۲- مواد و روش‌ها

طبق شکل ۱، روند اجرایی این مطالعه شامل سه مرحله اساسی می‌باشد: (۱) تهیه نقشه مرجع سیل (شناسایی ۳۰۰ موقعیت سیل خیز در منطقه مورد مطالعه) و استخراج معیارهای مکانی مؤثر بر سیل، (۲) تولید نقشه‌های پیش‌بینی خطر سیل بر اساس رویکردهای ترکیبی یادگیری ماشین مبتنی بر درخت، (۳) ارزیابی و مقایسه عملکرد مدل‌ها.

۲-۱- منطقه مطالعه

شهرستان مانه و سملقان در شمال غربی استان خراسان شمالی قرار دارد که از شمال با شهرستان راز و جرگلان، از غرب به استان گلستان، از جنوب به شهرستان‌های جاجرم و گرمه و از شرق به جنود محدود گردیده است. این شهرستان دارای ۸ کیلومتر مرز مشترک با کشور ترکمنستان می‌باشد. از لحاظ موقعیت جغرافیایی این شهرستان بین مدار جغرافیایی ۱۷° ۳۷' تا ۱۷° ۳۸' عرض شمالی از خط استوا و ۵۹° ۵۵' تا ۵۷° ۱۷' طول شرقی از نصف‌النهار گرینویچ واقع شده است. شهرستان مانه و سملقان دارای ۶۰۵۳ کیلومتر مربع وسعت است. شکل ۲ محدوده شهرستان مانه و سملقان را نشان می‌دهد.

سیل به عنوان یکی از مخرب‌ترین خطرات طبیعی در سراسر جهان شناخته می‌شود و باعث از دست رفتن جان و مال انسان‌ها و تهدید اکوسیستم‌ها می‌شود (Chapi et al., 2017; Du et al., 2013; Smith and Ward, 1998). بررسی و آمارها نشان می‌دهد سیل در ۵ دهه اخیر بیش از نیمی از کل خسارات ناشی از بلایای طبیعی در دنیا را به خود اختصاص داده است (Imani et al., 2021). رخداد سیل در کشور ایران مختص به منطقه و موقعیت خاصی نبوده و تمام کشور از این پدیده متأثر است، ولی با توجه به خصوصیات هر منطقه، تفاوت در نوع سیل و میزان خسارات وارد شده وجود دارد (Saedi et al., 2020). بنابراین، کاهش آسیب در نواحی مستعد سیل یک موضوع اساسی و ضروری می‌باشد (Chen et al., 2019; Hong et al., 2018). در سال‌های اخیر، با ترکیب فن‌آوری سیستم اطلاعات مکانی (GIS)^۱ و سنجش از دور (RS)^۲، دقت پیش‌بینی خطر وقوع سیل به طور قابل‌توجهی بهبود یافته است (Chen et al., 2015). اخیراً مطالعات متعددی برای شناسایی خطر وقوع سیل انجام شده است که در ادامه به برخی از آن‌ها اشاره شده است:

(Khosravi et al., 2016) نقشه‌های حساسیت به سیل حوزه آبخیز هراز در استان مازندران را با استفاده از چهار مدل مختلف، یعنی نسبت فرکانس (FR)^۳، روش وزن شواهد (WOE)^۴، فرایند تحلیلی سلسله مراتبی (AHP)^۵ و ترکیب نسبت فرکانس با فرایند تحلیلی سلسله‌مراتبی (FR-AHP) تهیه کردند. نتایج نشان داد که مدل FR بیشترین مقدار سطح زیر منحنی (AUC)^۶ را در مقایسه با سایر مدل‌ها دارد. (Liu et al., 2016) روش بیز ساده (NB)^۷ را برای ارزیابی خطر سیل در حوضه بوون^۸ کشور استرالیا پیاده‌سازی کردند. جهت ارزیابی خطر سیل در منطقه مورد مطالعه از ۴ شاخص ارتفاع، شیب، جنس خاک و تراکم زهکشی استفاده کردند. نتایج نشان داد که شاخص‌های ارتفاع و شیب اثرات قابل‌توجهی در ارزیابی خطر سیل نسبت به سایر شاخص‌های ورودی دارد. (Hong et al., 2018) از ترکیب روش‌های رگرسیون لجستیک (LR)^۹، جنگل تصادفی (RF)^{۱۰} و ماشین بردار پشتیبان (SVM)^{۱۱} با روش وزن شواهد (WOE) برای تهیه نقشه حساسیت به سیل در منطقه پویانگ^{۱۲} کشور چین استفاده کردند. نتایج نشان داد که مدل ترکیبی SVM-WOE بالاترین مقدار عملکرد پیش‌بینی (AUC) را نسبت به سایر مدل‌های ترکیبی RF-WOE و LR-WOE دارد. (Razavi Temeh et al., 2018) از ترکیب سیستم استنتاج عصبی فازی (ANFIS)^{۱۳} با الگوریتم‌های فراابتکاری شامل الگوریتم بهینه‌سازی کلونی مورچه‌ها (ACO)^{۱۴}، الگوریتم ژنتیک (GA)^{۱۵} و الگوریتم بهینه‌سازی ازدحام ذرات (PSO)^{۱۶} جهت تهیه نقشه

فرسایش پذیری خاک در شهرستان مانه و سملقان را افزایش داده است. همچنین وجود شیب‌های تند و عدم وجود پوشش گیاهی و درخت از جمله موارد دیگری هستند که موجب به وجود آمدن سیل در چند سال اخیر در این منطقه شده است.

این شهرستان در ناحیه اقلیمی خشک و نیمه‌خشک واقع شده و توزیع مکانی و زمانی بارش باران در این منطقه نیز کاملاً متغیر و نیمه یکنواخت است، به گونه‌ای که اغلب بارش‌ها در استان در فصل سرد و مرطوب رخ می‌دهد و این شرایط، ضرایب سیل‌خیزی و

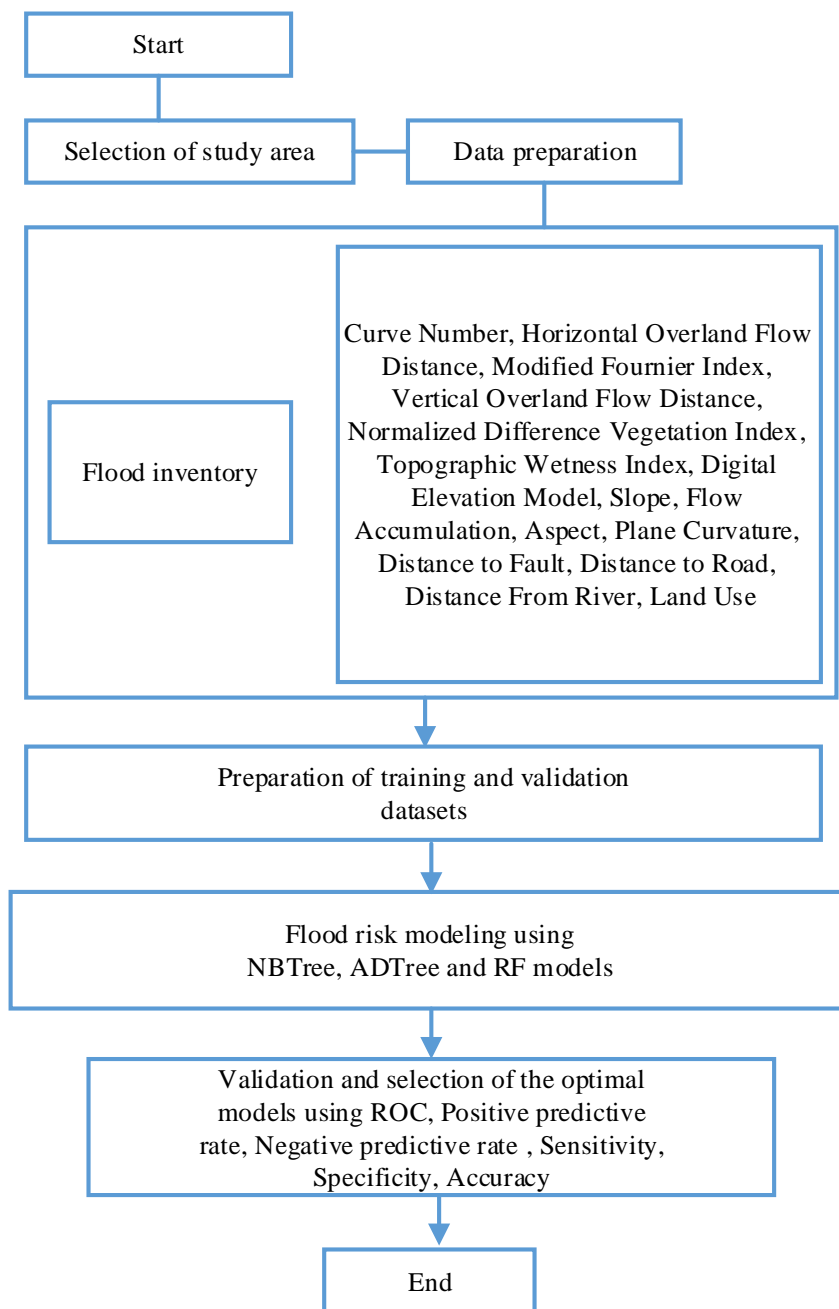


Fig. 1- The flowchart of the research

شکل ۱- روند اجرایی تحقیق

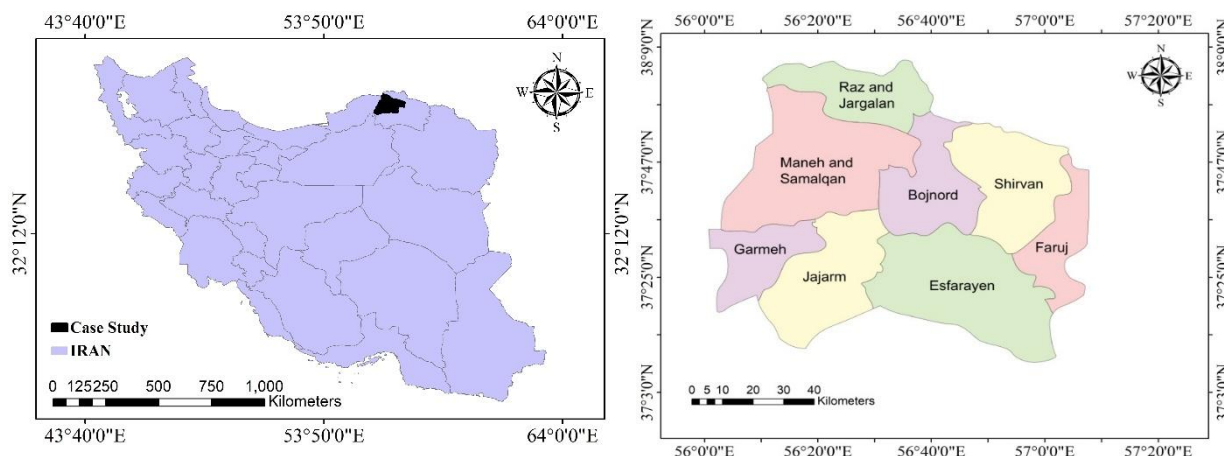


Fig. 2- Geographical location of the study area

شکل ۲- محدوده منطقه مطالعاتی

لایه‌های رودخانه، گسل و جاده نیز طبق بانک اطلاعاتی رودخانه‌ها، گسل‌ها و جاده‌های کشور استخراج شدند و سپس به کمک ابزار فاصله اقلیدسی در نرم‌افزار Arc Map رستری شدند. لایه کاربری اراضی منطقه مورد مطالعه نیز از تصاویر ماهواره‌ای لندست ۸ سنجنده OLI^{۳۶} استخراج گردید. همچنین، از باند ۴ و ۵ تصاویر ماهواره‌ای لندست ۸ جهت تولید لایه NDVI استفاده گردید. به منظور محاسبه شاخص MFI شش ایستگاه سینوپتیک به کار برده شدند و طبق رابطه (۱)، نقشه شاخص مربوطه در نرم‌افزار Arc Map تولید گردید (Kanani-Sadat et al., 2019):

$$MFI = \frac{\sum_{i=1}^n p_i^2}{p} \quad (1)$$

که P_i میانگین بارندگی برای ماه i ام و p میانگین بارندگی سالانه است. همچنین شاخص TWI طبق رابطه (۲) در نرم‌افزار SAGA GIS محاسبه گردید (Jancewicz et al., 2019; Eini et al., 2020):

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

که A_s مساحت حوزه بالادست و β زاویه شیب است. لایه CN نیز از سازمان جنگل‌ها، مراتع و آبخیزداری کشور تهیه شد. سایر لایه‌ها (جریان تجمع، فاصله افقی تا رواناب و فاصله قائم تا رواناب) به کمک تحلیل‌های مناسب در نرم‌افزار Arc Map و SAGA GIS ایجاد شدند.

۲-۳- روش‌ها

۲-۳-۱- درخت بیز ساده (NBTree)

الگوریتم درخت بیز ساده از درخت تصمیم به عنوان یک ساختار

۲-۲- تهیه لایه‌های اطلاعاتی منطقه مورد مطالعه

برای برآورد سیل‌های آینده در منطقه مورد مطالعه، تجزیه و تحلیل رخداد گذشته سیلاب‌ها ضروری می‌باشد. لذا نقشه مرجع سیل یک ابزار حیاتی در مدل‌سازی خطر وقوع سیل است. روش‌های مختلفی مانند سوابق تاریخی (بانک اطلاعاتی از وقوع سیل‌های گذشته) و تفسیر عکس‌های ماهواره‌ای برای به دست آوردن یک نقشه مرجع سیل وجود دارد (Chen et al., 2019). در این مطالعه، با استفاده از بانک اطلاعاتی سازمان جنگل‌ها، مراتع و آبخیزداری کشور در سال ۱۳۹۹، ۳۰۰ موقعیت سیل‌خیز در شهرستان مانه و سملقان شناسایی گردید. بر اساس بررسی تحقیقات پیشین و در دسترس بودن داده‌ها، ۱۵ معیار هیدروژئولوژیکی و محیطی شامل منحنی رواناب (CN)^{۲۰}، فاصله افقی تا رواناب (HOFD)^{۲۱}، شاخص فورنیه اصلاح شده یا شاخص میانگین بارش (MFI)^{۲۲}، فاصله قائم تا رواناب (VOFD)^{۲۳}، شاخص نرمال شده تفاوت پوشش گیاهی (NDVI)^{۲۴}، شاخص رطوبت توپوگرافی (TWI)^{۲۵}، مدل رقمی ارتفاعی (DEM)^{۲۶}، شیب^{۲۷}، جریان تجمع^{۲۸}، جهت شیب^{۲۹}، انحنا^{۳۰}، فاصله از گسل^{۳۱}، فاصله از جاده^{۳۲}، فاصله از رودخانه^{۳۳} و کاربری اراضی^{۳۴} برای مطالعه حاضر انتخاب شدند (Ahmadlou et al., 2018; Chapi et al., 2017; Chen et al., 2019; Khosravi et al., 2019; Razavi Termeh et al., 2016; Tien Bui et al., 2018). هر کدام از این معیارها در قالب نقشه رستری با اندازه پیکسل ۳۰ متر در نرم‌افزار Arc Map تهیه شدند. نقشه توپوگرافی منطقه مورد مطالعه در مقیاس ۱:۲۵۰۰۰ (سازمان نقشه‌برداری کشور) تهیه شد. به منظور تهیه لایه DEM از تصاویر ماهواره‌ای SRTM^{۳۵} با قدرت تفکیک ۳۰ متر استفاده شد و سپس لایه‌های شیب، جهت شیب و انحنا از لایه DEM ایجاد شدند.

است و از طریق اطمینان از اندازه حاشیه طبقه‌بندی ایجاد می‌کند (Pham et al., 2016). همچنین، دارای قابلیت تشخیص و حذف شکاف‌ها است. علاوه بر این، الگوریتم ADTree از تعداد کمتری از تکرارها در عملکرد خود استفاده می‌کند و این قابلیت را دارد که برای تجزیه و تحلیل بهتر مجموعه داده‌های آموزشی و اعتبارسنجی، کلاس‌ها را در مدل باینری کند.

۲-۳-۳- جنگل تصادفی (RF)

الگوریتم جنگل تصادفی یکی از رایج‌ترین الگوریتم‌های به کار رفته برای بررسی مشکلات طبقه‌بندی و پیش‌بینی چندگانه است (de Santana et al., 2018; Quiroz et al., 2018). الگوریتم RF حساسیت کمی به چندخطی بودن دارد و نتایج آن از نظر داده‌های از دست رفته و نامتعادل نسبتاً پایدار است. الگوریتم RF یک نوع مدرن از درخت پایه می‌باشد که شامل انبوهی از درخت‌های کلاس‌بندی و رگرسیونی است. در واقع این الگوریتم یک روش گروهی است که چند الگوریتم درختی را برای تولید یک پیش‌بینی متوالی از هر پدیده ترکیب می‌کند (Rahmati et al., 2016). به‌طور کلی درخت تصمصیم انفرادی مستعد برازش زیادتر بوده و قابلیت تعمیم‌پذیری کمی دارد. در هنگام ساخت یک درخت تصمصیم، تغییرات ناچیز در ساختار آموزش می‌تواند منجر به تغییرات قابل توجه در شکل آن درخت شود. مدل پیش‌بینی‌کننده RF بر اساس میانگین‌گیری از نتایج حاصل از تمامی درخت‌های تصمصیم مربوطه استوار است و برای بسیاری از مجموعه داده‌ها، طبقه‌بندی با صحت بالایی را انجام می‌دهد (de Santana et al., 2018). درختان تصادفی بردار ورودی را گرفته و آن را با هر درخت در جنگل طبقه‌بندی کرده و خروجی برچسب‌های کلاسی هستند که از اکثریت آراء دریافت شده است. روند اصلی الگوریتم RF شامل این مراحل است (Quiroz et al., 2018): (۱) نمونه‌برداری مجدد از مجموعه داده‌های اصلی برای تولید زیرمجموعه‌هایی که اندازه آن‌ها برابر با اندازه مجموعه اصلی است؛ (۲) استفاده از زیرمجموعه‌ها برای ساخت درخت‌های تصمصیم و (۳) به دست آوردن نتایج نهایی با استفاده از پیش‌بینی یا طبقه‌بندی نتایج تمام درخت‌های تصمصیم. با توجه به تحقیقات موجود، عملکرد RF عمدتاً توسط تعداد درختان تصمصیم و ویژگی‌های موجود در زیرمجموعه‌ها تعیین می‌شود (Santana et al., 2018). مقدار زیاد درخت ممکن است باعث افزایش زمان مدل‌سازی شود، در حالی که مقدار کمی از درخت ممکن است منجر به خطا شود.

عمومی استفاده می‌کند. طبقه‌بندی‌کننده‌های بیز هنگامی که مجموعه داده‌های نمونه کوچک باشد، بهتر از درخت‌های تصمصیم کار می‌کنند (Farid et al., 2014). درخت‌های تصمصیم زمانی بهتر کار می‌کنند که مجموعه داده‌های نمونه بزرگ باشد. علت آن این است که پس از چندین مرتبه تقسیم‌بندی هر کدام از ویژگی‌ها، تعداد نمونه‌ها بسیار کم است که براساس آن تصمصیم گرفته شود، در حالی که طبقه‌بندی‌کننده بیز در این مورد بهتر کار می‌کند. بنابراین، به جای ادامه فرآیند تقسیم ویژگی‌ها، دسته‌بندی‌کننده‌های بیز در برگ‌ها مستقر می‌شوند. این مدل ترکیبی به نام درخت بیز ساده نامیده می‌شود (Wang et al., 2006). در واقع الگوریتم NBTree با جایگزین کردن گره‌های برگ درخت تصمصیم با دسته‌بندی‌کننده‌های بیز ساده، مزایای یک درخت تصمصیم و بیز ساده را شامل می‌شود. بنابراین الگوریتم NBTree عملکرد قابل توجهی را در حل مشکلات وابستگی داده دارد. در سال‌های اخیر، الگوریتم NBTree عموماً در زمینه طبقه‌بندی به کار گرفته شده است و کیفیت رضایت‌بخش طبقه‌بندی آن توسط سطح زیر منحنی (AUC) و سایر پارامترهای آماری تایید شده است (Farid et al., 2014; Wang et al., 2006; Wang et al., 2015).

۲-۳-۲- درخت تصمصیم متناوب (ADTree)

الگوریتم ADTree یک الگوریتم درخت تصمصیم تقویت‌شده است که در مقایسه با الگوریتم‌های درخت تصمصیم سنتی، دقت بهتری دارد و سطوح اعتماد بالایی را در هنگام حل مشکلات طبقه‌بندی و پیش‌بینی فراهم می‌کند (Bhowmick et al., 2010; Pham et al., 2016; Pham et al., 2017). الگوریتم ADTree عبارت است از فراخوانی مکرر الگوریتم درخت تصمصیم به کمک روش بوستینگ. به عبارت دیگر الگوریتم ADTree به عنوان یک الگوریتم تطبیقی است که از فراخوانی مکرر درختان تصمصیم مبتنی بر شرایط داده ایجاد می‌گردد (Pham et al., 2016; Pham et al., 2017). الگوریتم ADTree یک ساختار درختی بهینه برای پاسخ‌دهی ایجاد می‌کند. برای ساخت این درخت به هر صفت یک اندیس نسبت داده می‌شود. گره ریشه دارای اندیس صفر بوده و تعداد رکوردهای موجود در مجموعه داده را در خود نگه می‌دارد. از هر گره با اندیس i ، گره‌هایی با اندیس j گسترش داده می‌شوند به طوری که $j > i$ می‌باشد. بنابراین، تمامی صفات از گره ریشه گسترش داده می‌شوند. کاربرد درخت تصمصیم با الگوریتم بوستینگ باعث ایجاد دسته‌بندی‌کننده‌هایی خیلی صحیح می‌شود. ترکیب بوستینگ با درخت‌های تصمصیم قوانین رده‌بندی را به وجود می‌آورد که کوچک‌تر بوده و تفسیر آن‌ها ساده‌تر است (Bhowmick et al., 2010). مزایای الگوریتم ADTree این است که تعداد گره‌های کمتری تولید می‌کند، به راحتی قابل توضیح

$$SPC = \frac{TN}{TN + FP} \quad (6)$$

$$ACC = \frac{TP + TN}{TN + FN + TP + FP} \quad (7)$$

۳- نتایج و بحث

۳-۱- آماده‌سازی داده‌ها

همان‌طور که ذکر شد، در این مطالعه از ۱۵ معیار هیدروژئولوژیکی و محیطی شامل منحنی رواناب، فاصله افقی تا رواناب، شاخص فورنیه اصلاح شده، فاصله قائم تا رواناب، شاخص نرمال شده تفاوت پوشش گیاهی، شاخص رطوبت توپوگرافی، مدل رقمی ارتفاعی، شیب، جریان تجمعی، جهت شیب، انحنای سطح، فاصله از جاده، فاصله از گسل، فاصله از رودخانه و کاربری اراضی جهت پیش‌بینی خطر وقوع سیل در شهرستان مانه و سملقان استفاده گردید. برای تولید نقشه‌های عوامل ذکر شده (شکل ۳)، از درون‌یابی کریجینگ استفاده شده است، زیرا این روش دارای حداقل خطا است (Eftekhari et al., 2021).

۳-۲- اهمیت‌دهی معیارهای مکانی مؤثر بر سیل

عوامل متعددی در یک حوزه آبخیز بر جریان آب در منطقه مورد مطالعه اثرگذار می‌باشند که هرکدام از این عوامل پتانسیل متفاوتی در تغییر جریان و رخداد سیل دارند. تعیین مهم‌ترین و مؤثرترین معیارهای تأثیرگذار بر سیل جهت انجام مدل‌سازی بسیار ضروری می‌باشد. لذا در این تحقیق، میزان اهمیت معیارهای مکانی مؤثر بر خطر وقوع سیل با استفاده از روش اعتبارسنجی متقابل ده‌گانه محاسبه شد (Witten et al., 2011). طبق جدول ۲ نتایج نشان می‌دهد که تمام معیارهای مکانی مؤثر بر خطر وقوع سیل از اهمیت بالایی برخوردار هستند. لذا تمامی معیارها در الگوریتم‌های پیشنهادی تحقیق به کار گرفته شدند. معیار مدل رقمی ارتفاعی، با بالاترین میانگین شایستگی (AM)^{۴۷} برای سه الگوریتم پیشنهادی، مهم‌ترین معیار مکانی در منطقه مورد مطالعه شناسایی شد و پس از آن، به ترتیب نزولی، معیارهای فاصله از رودخانه، شاخص فورنیه اصلاح شده، شاخص پوشش گیاهی، شیب، کاربری اراضی، منحنی رواناب، فاصله افقی تا رواناب، فاصله از جاده، شاخص رطوبت توپوگرافی، فاصله از گسل، فاصله قائم تا رواناب، جریان تجمعی، جهت شیب و انحنای سطح قرار گرفتند.

۴-۲- ارزیابی عملکرد و دقت مدل‌ها

۴-۲-۱- منحنی مشخصه عملکرد سیستم (ROC)

در این مطالعه، روش ROC برای ارزیابی عملکرد مدل‌ها به ترتیب با نرخ‌های مثبت واقعی و مثبت کاذب بر روی محور Y و محور X استفاده شده است (Tehrany et al., 2013). این روش از معیار AUC برای ارزیابی کمی استفاده می‌کند که ویژگی محور X را در مقابل حساسیت محور Y نشان می‌دهد (Chapi et al., 2017; Witten et al., 2011). منحنی ROC کارآمدترین روش در پیش‌بینی سیستم‌ها است که میزان دقت مدل را به صورت کمی برآورد می‌کند. در روش ROC، محدوده مقادیر AUC بین ۰/۵ تا ۱ می‌باشد. هر چه مقدار AUC به ۱ نزدیک‌تر باشد، دقت مدل نیز بیشتر است. همبستگی کمی-کیفی سطح زیر منحنی و ارزیابی تخمین در جدول ۱ ارائه شده است (Chapi et al., 2017).

۴-۲-۲- معیارهای آماری

در این تحقیق مجموعه‌ای از معیارهای آماری شامل نرخ پیش‌بینی مثبت (PPR)^{۳۷}، نرخ پیش‌بینی منفی (NPR)^{۳۸}، حساسیت (SST)^{۳۹}، ویژگی (SPC)^{۴۰} و دقت (ACC)^{۴۱} برای ارزیابی و اعتبارسنجی تمام مدل‌ها استفاده شدند (Chapi et al., 2017; Witten et al., 2011). همه این معیارها براساس ماتریس درهم‌ریختگی^{۴۲} محاسبه می‌شوند که متشکل از چهار نوع پیامد احتمالی شامل مثبت واقعی (TP)^{۴۳}، منفی واقعی (TN)^{۴۴}، مثبت کاذب (FP)^{۴۵} و منفی اشتباه (FN)^{۴۶} برای محاسبه معیارهای آماری است. معیار TP نسبت تعداد مکان‌هایی است که به درستی به عنوان یک مکان با خطر بالای سیل طبقه‌بندی می‌شوند. در حالی که معیار FP نسبت تعداد مکان‌هایی است که به درستی به عنوان یک مکان بدون خطر سیل طبقه‌بندی می‌شوند. علاوه بر این، معیارهای TN و FN تعداد مکان‌هایی هستند که به طور صحیح و نادرست به عنوان مکان با خطر بالای سیل و یک مکان با خطر غیربالقوه طبقه‌بندی می‌شوند. این معیارهای آماری با استفاده از روابط زیر محاسبه می‌شوند (Chapi et al., 2017; Witten et al., 2011):

$$PPR = \frac{TP}{TP + FP} \quad (3)$$

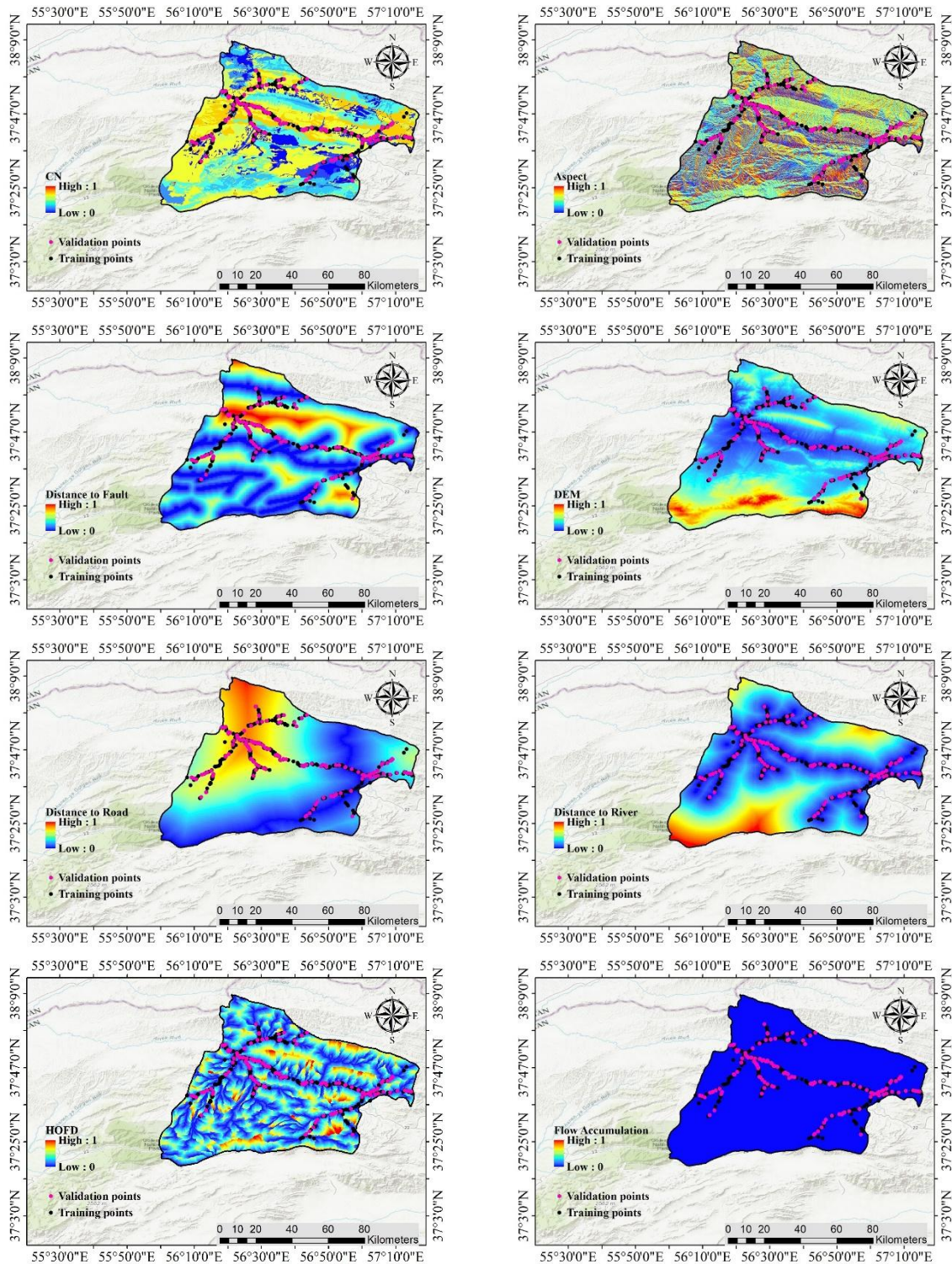
$$NPR = \frac{NP}{NP + FN} \quad (4)$$

$$SST = \frac{TP}{TP + FN} \quad (5)$$

Table 1- Quantitative-qualitative correlation of the area under the curve in ROC method

جدول ۱- همبستگی کمی-کیفی سطح زیر منحنی در روش ROC

Poor	Average	Good	Very Good	Excellent	Qualitative
0.5 - 0.6	0.6 - 0.7	0.7 - 0.8	0.8 - 0.9	0.9 - 1	Quantitative



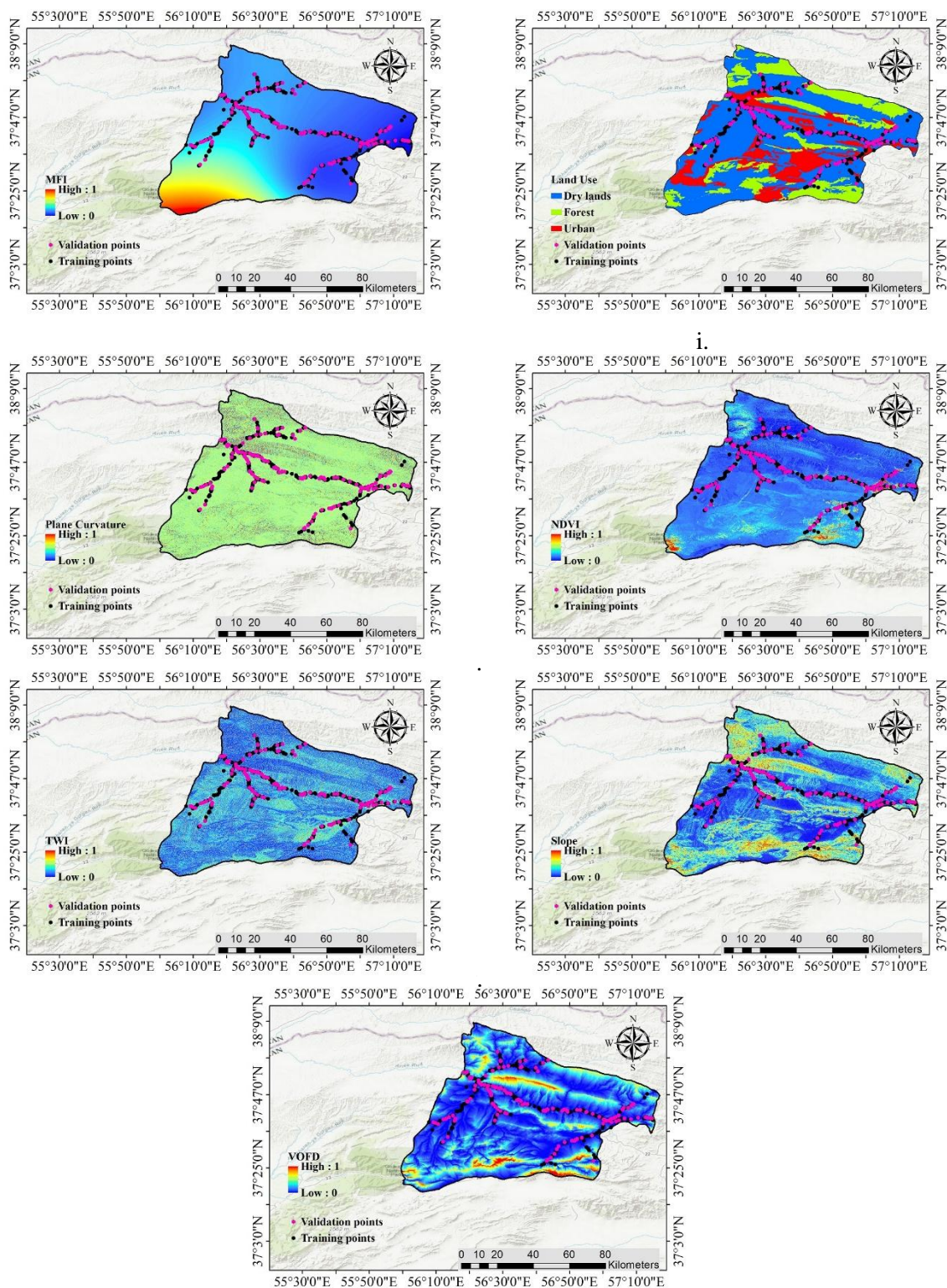


Fig. 3- The maps of spatial criteria affecting the flood

شکل ۳- نقشه معیارهای مکانی مؤثر بر سیل

Table 2- The importance of spatial criteria affecting the flood

جدول ۲- اهمیت معیارهای مکانی مؤثر بر سیل

Order	Criteria	NBTree		ADTree		RF	
		AM	SE	AM	SE	AM	SE
1	DEM	0.391	± 0.006	0.391	± 0.006	0.392	± 0.006
2	Distance to river	0.381	± 0.006	0.381	± 0.006	0.381	± 0.007
3	MFI	0.316	± 0.009	0.315	± 0.009	0.313	± 0.008
4	NDVI	0.255	± 0.013	0.256	± 0.014	0.256	± 0.012
5	Slope	0.204	± 0.005	0.208	± 0.005	0.207	± 0.007
6	Land use	0.190	± 0.009	0.204	± 0.005	0.204	± 0.006
7	CN	0.189	± 0.023	0.200	± 0.011	0.200	± 0.011
8	HOFD	0.187	± 0.006	0.193	± 0.009	0.194	± 0.010
9	Distance to road	0.169	± 0.006	0.172	± 0.007	0.175	± 0.007
10	TWI	0.152	± 0.018	0.160	± 0.013	0.162	± 0.011
11	Distance to fault	0.131	± 0.006	0.142	± 0.007	0.142	± 0.008
12	VOFD	0.115	± 0.006	0.138	± 0.011	0.139	± 0.013
13	Flow Accumulation	0.102	± 0.008	0.124	± 0.007	0.128	± 0.007
14	Aspect	0.098	± 0.009	0.115	± 0.011	0.120	± 0.001
15	Plane curvature	0.089	± 0.009	0.091	± 0.013	0.097	± 0.011

بین کلاس‌ها بیشینه شود (Razavi Temeh et al., 2018). همچنین، در این روش عوارض در کلاس‌هایی قرار می‌گیرند که مرز آنها به طوری تعیین شده است که در این مرزها تغییرات نسبی بزرگی در مقادیر داده‌ها صورت می‌گیرد. شکل ۴ نقشه‌های میزان خطر سیل در منطقه مورد مطالعه را نشان می‌دهد که با استفاده از مدل‌های مورد نظر پیش‌بینی شده‌اند. نقشه حساسیت به وقوع سیل منطقه مورد مطالعه نشان می‌دهد که احتمال وقوع سیل در مناطق مرکزی منطقه مورد مطالعه، به دلیل ارتفاع و شیب کمتر، بیشتر از سایر مناطق است (Kanani-Sadat et al., 2019). همچنین، در مناطق مرکزی منطقه مورد مطالعه به دلیل کاهش سطح اراضی جنگلی و تغییر کاربری‌های شدید صورت گرفته در چند سال اخیر که سبب کاهش اراضی جنگلی و افزایش اراضی دیم و شهری شده، احتمال وقوع خسارات را افزایش داده است. از طرف دیگر کاهش فاصله با حریم اصلی رودخانه و در معرض قرار گرفتن املاک، اراضی و افراد از دیگر عوامل مؤثر در افزایش خسارت در مواقع سیل در این مناطق بوده است.

همچنین، در شکل ۵، مساحت هر یک از کلاس‌های خطر وقوع سیل پیش‌بینی شده توسط مدل‌های مبتنی بر درخت NBTree، ADTree و RF نشان داده شده است. نتایج نشان داد که در مدل RF، ۲ کلاس با خطرپذیری خیلی زیاد و خطرپذیری زیاد، مساحت بیشتری را در منطقه مورد مطالعه نسبت به کلاس‌های مشابه در مدل‌های NBTree و ADTree تحت پوشش قرار می‌دهد.

۳-۳- تهیه نقشه‌های پیش‌بینی خطر وقوع سیل

برای پیاده‌سازی الگوریتم‌های پیشنهادی تحقیق، تنظیم پارامترهای ورودی از مهم‌ترین مراحل می‌باشد. در الگوریتم جنگل تصادفی به منظور تنظیم تعداد درختان تصمیم به طوری که مقدار خطا همگرا شود و تخمین قابل اعتمادتر باشد، مدل از ۱ تا ۵۰۰ درخت ساخته شد. به دلیل این که با افزایش درخت‌ها میزان خطا کاهش می‌یابد، بنابراین تعداد ۵۰۰ درخت برای فراخوانی مدل استفاده شد (Rahmati et al., 2016). پارامتر ویژگی‌های موجود در زیرمجموعه‌ها نیز به روش آزمون و خطا بهینه شد. همچنین پارامترهای لازم جهت پیاده‌سازی الگوریتم‌های NBTree و ADTree از طریق روش آزمون و خطا و بر اساس نتایج تحقیقات پیشین تنظیم شدند (Wang et al., 2006; Pham et al., 2016). در این مطالعه، نقشه‌های پیش‌بینی خطر وقوع سیل در منطقه مورد مطالعه بر اساس مدل‌های ترکیبی مبتنی بر درخت NBTree، ADTree و RF و با استفاده از مجموعه داده آموزشی (۷۰ درصد) و اعتبارسنجی (۳۰ درصد) برآورد شدند. نقشه میزان خطر سیل پیش‌بینی شده با استفاده از روش طبقه‌بندی شکست طبیعی^{۴۸} به کلاس‌های مختلف در نرم‌افزار ARC Map طبقه‌بندی شد. بر این اساس، میزان خطر وقوع سیل پیش‌بینی شده، به پنج کلاس شامل خطرپذیری خیلی کم، کم، متوسط، زیاد و خیلی زیاد طبقه‌بندی گردید. علت استفاده از روش طبقه‌بندی شکست طبیعی در این است که طبقه‌بندی بر اساس گروه‌بندی‌های طبیعی موجود در ذات داده‌ها انجام می‌شود و نقاط شکستگی بین دسته‌ها به گونه‌ای معین می‌شود که به بهترین وجه ممکن مقادیر مشابه در یک گروه جای داده شوند و تفاوت

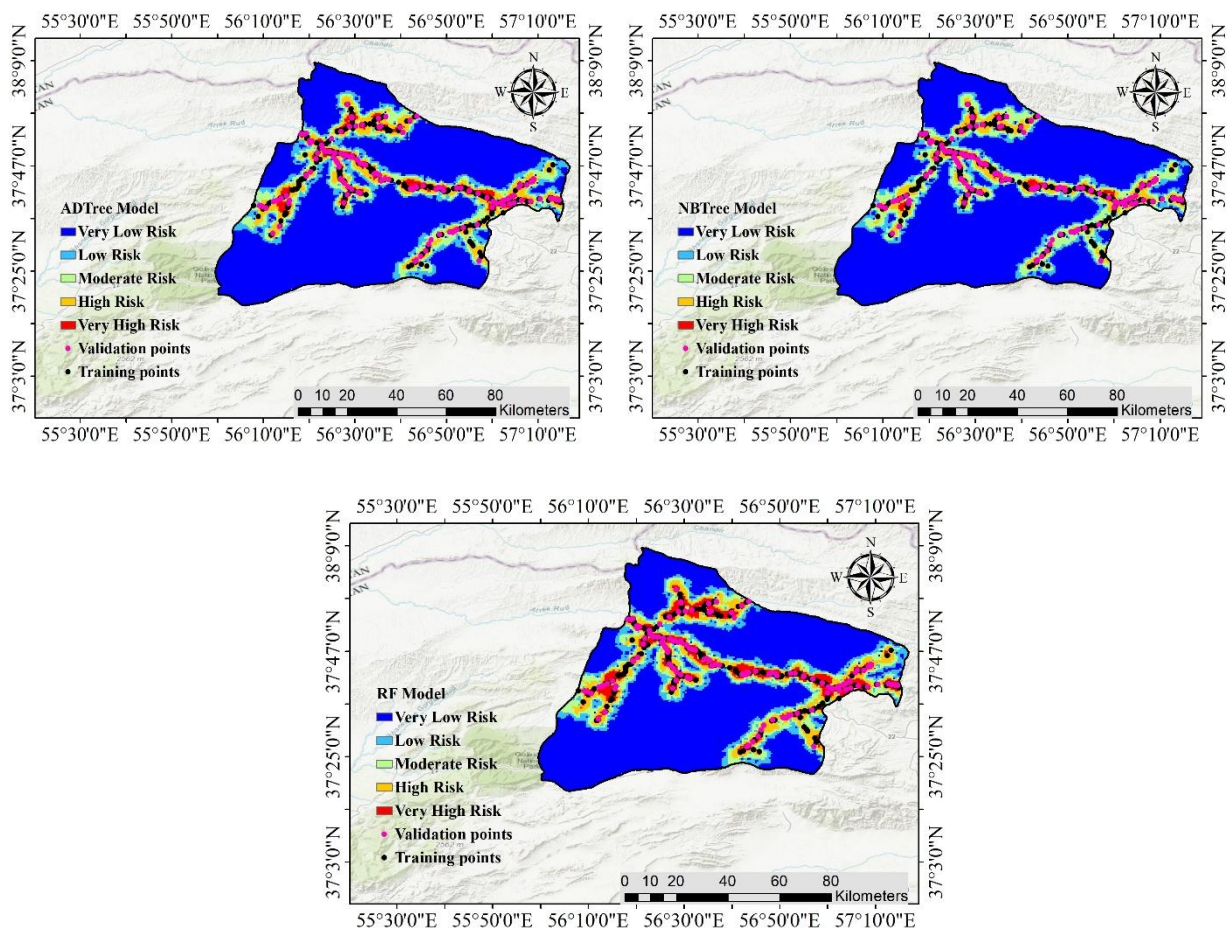


Fig. 4- Flood risk prediction maps: NBTree, ADTree, and RF
 شکل ۴- نقشه‌های پیش‌بینی خطر سیل: NBTree، ADTree، و RF

مجموعه داده‌های اعتبارسنجی، مهارت پیش‌بینی مدل را نشان می‌دهند. طبق شکل ۶ برای مجموعه داده‌های آموزشی، مدل RF دارای بالاترین مقدار AUC (۰/۷۶) می‌باشد و پس از آن مدل‌های NBTree (AUC= ۰/۷۴۰) و ADTree (AUC= ۰/۷۳۰) قرار دارند. همچنین، برای مجموعه داده‌های اعتبارسنجی، مدل RF قوی‌تر بوده و دارای دقت بالاتری (AUC= ۰/۷۴۰) نسبت به مدل‌های ADTree (AUC= ۰/۷۳۰) و NBTree (AUC= ۰/۷۰۰) می‌باشد. بنابراین، می‌توان نتیجه گرفت که اگر چه تمام مدل‌های به کار رفته قدرت پیش‌بینی خوبی دارند، اما عملکرد و دقت مدل RF برای پیش‌بینی خطر سیل در منطقه مورد مطالعه بهتر است که این نتایج با یافته‌های تحقیق Hong et al. (2018) و Khosravi et al. (2016) مطابقت دارد. قابلیت‌های صحت‌سنجی و پیش‌بینی مدل‌های به کار گرفته شده برای پیش‌بینی خطر سیل با استفاده از مجموعه داده‌های آموزشی (جدول ۳) و داده‌های اعتبارسنجی (جدول ۴) ارزیابی شدند.

هم‌چنین در مدل ADTree، کلاس با خطرپذیری متوسط، مساحت بیشتری را در منطقه مورد مطالعه نسبت به کلاس مشابه در مدل‌های NBTree و RF دربر گرفته است. علاوه بر این در مدل NBTree، ۲ کلاس با خطرپذیری خیلی کم و خطرپذیری کم، مساحت بیشتری را در منطقه مورد مطالعه نسبت به کلاس‌های مشابه در مدل‌های RF و ADTree دربر گرفته است.

۳-۴- اعتبارسنجی نقشه خطر وقوع سیل

اعتبارسنجی نقشه‌های تهیه شده، گامی اساسی در توسعه و تعیین مناطق حساس و تعیین کیفیت آن‌ها می‌باشد. لذا در این تحقیق، عملکرد مدل‌های پیاده‌سازی شده با روش منحنی ROC ارزیابی شد. شکل ۶ منحنی ROC را برای مجموعه داده‌های آموزشی و اعتبارسنجی مدل‌های مبتنی بر درخت NBTree، ADTree و RF نشان می‌دهد. بطور کلی، قابلیت مدل‌ها برای پیش‌بینی خطر سیل توسط مجموعه داده‌های آموزشی ارزیابی می‌شود؛ در حالی که

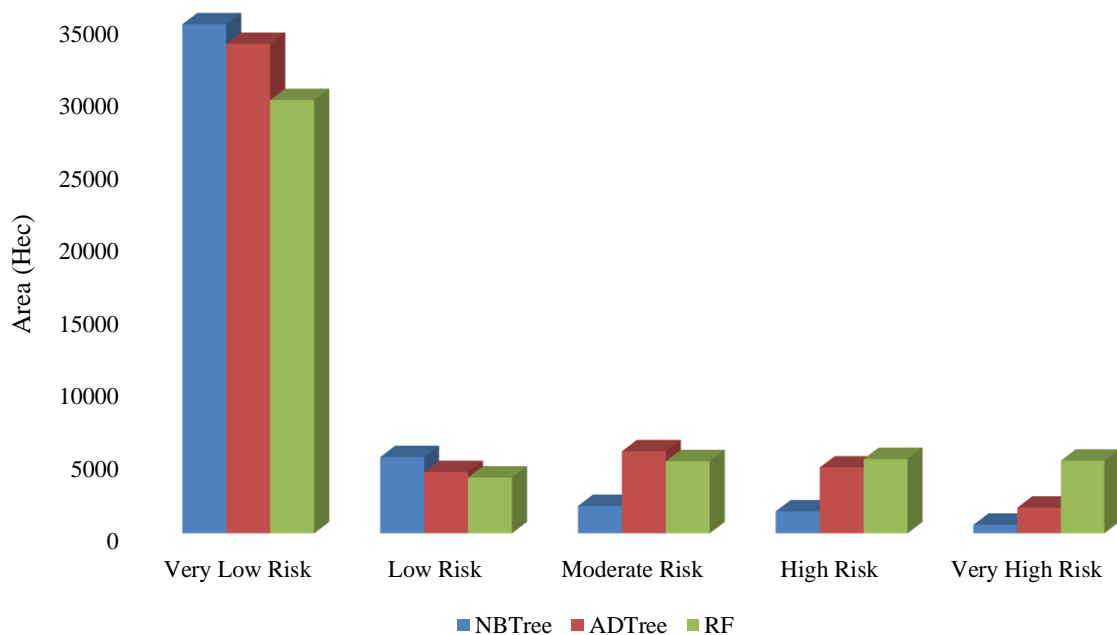


Fig. 5- Area of each flood risk class in the study area
 شکل ۵- مساحت هر یک از کلاس‌های خطر سیل در منطقه مورد مطالعه

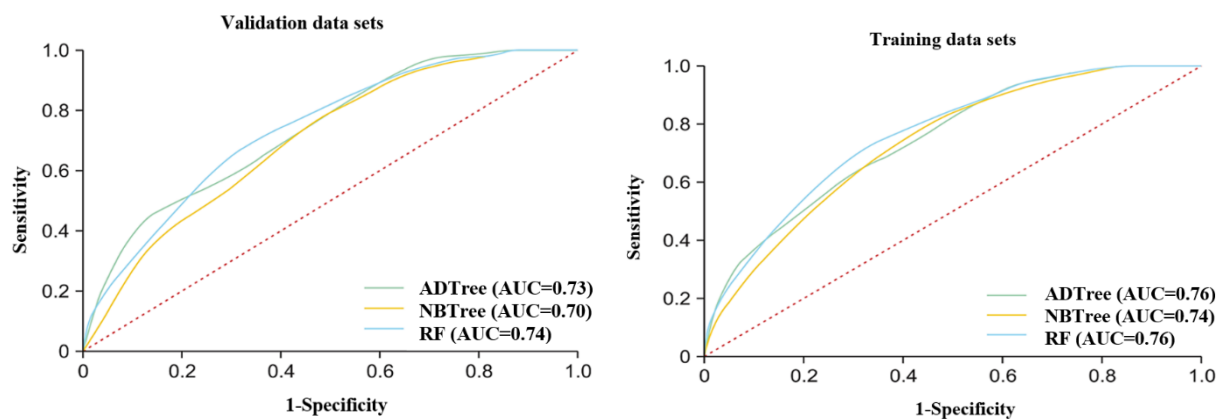


Fig. 6- ROC curve and AUC value for proposed research methods: Training data sets, validation data sets
 شکل ۶- منحنی ROC و مقدار AUC برای روش‌های پیشنهادی تحقیق: مجموعه داده‌های آموزشی، مجموعه داده‌های اعتبارسنجی

دارند. علاوه بر این، مدل RF دارای بالاترین مقدار نرخ پیش‌بینی مثبت، نرخ پیش‌بینی منفی و دقت برای مجموعه داده‌های آموزشی و اعتبارسنجی است که پس از آن مدل‌های ADTree و NBTree قرار گرفته‌اند. نتایج مجموعه داده‌های اعتبارسنجی نشان داد که مدل RF قابلیت بهتری را نسبت به مدل‌های ADTree و NBTree در پیش‌بینی خطر وقوع سیل در منطقه مورد مطالعه دارد. عملکرد بهتر

برای طبقه‌بندی پیکسل‌های سیل، روش RF بیش‌ترین مقدار SST را برای مجموعه داده‌های آموزشی و اعتبارسنجی نشان می‌دهد (۰/۹۴۵ و ۰/۹۴۵) و به دنبال آن مدل‌های ADTree و NBTree قرار دارند. برای طبقه‌بندی پیکسل‌های غیرسیلی، روش RF بهترین مقدار SPC را برای مجموعه داده‌های آموزشی و اعتبارسنجی نشان می‌دهد (۰/۸۸۶ و ۰/۸۳۵) و به دنبال آن مدل‌های ADTree و NBTree قرار

آن‌ها در مدل‌های مورد استفاده برای ارزیابی پتانسیل سیل مفید می‌باشد که با یافته‌های (Rahmati et al. و khosravi et al. (2016) (2016) al. همخوانی دارد. بین شاخص فورنیه اصلاح شده (وابسته به بارش) و خطر وقوع سیل در منطقه مورد مطالعه رابطه مستقیمی وجود دارد به گونه‌ای که با افزایش بارندگی، میزان وقوع سیل و وزن طبقات حساس به سیل نیز افزایش پیدا می‌کند که این نتایج با یافته‌های (Pham et al. (2020) و (Razavi Termeh et al. (2018) مبنی بر وجود رابطه مستقیم بین بارندگی و وقوع سیل در یک منطقه مطابقت دارد. یکی دیگر از عوامل تأثیرگذار شناخته شده در این تحقیق عامل مدل رقومی ارتفاعی می‌باشد؛ بررسی این عامل نشان‌دهنده این امر است که با افزایش ارتفاع حساسیت طبقات نسبت به وقوع سیل کاهش می‌یابد و حداکثر حساسیت سیل در طبقات مربوط به طبقات با دامنه‌های ارتفاعی کم است که دلیل این امر را می‌توان در تجمع آب باران و وقوع سیل در این مکان‌ها دانست که اغلب مطالعات قبلی در این زمینه از قبیل مطالعات (Liu و khosravi et al. (2016) et al. (2016) تأییدکننده این موضوع می‌باشند.

مدل RF می‌تواند به دلیل توانایی آن در مدل‌سازی پایگاه داده‌های بزرگ و توانایی ادغام متغیرهای ورودی زیاد بدون تغییر متغیر باشد (Rahmati et al., 2016). همچنین مطالعات (Zhao et al. (2018)، (Chen et al. (2020) و (Tang et al. (2020) نشان می‌دهند که مدل RF توانایی بسیاری زیادی در تهیه نقشه مناطق حساس به سیل دارد. مدل RF از واریانس بالا در میان درختان فردی استفاده می‌کنند که هر یک از درخت‌ها را برای عضویت در کلاس می‌پذیرد (Chen et al., 2020). سپس، مدل RF بر اساس بیشترین تعداد آراء، طبقه مربوطه را تعیین می‌کند. همچنین در مطالعات مختلف در زمینه‌های آتش‌سوزی، محیط زیست، نقشه پتانسیل آب زیرزمینی و حساسیت به زمین‌لغزش مشخص شده است که مدل RF توانایی خوبی در مدل‌سازی و پیش‌بینی این عوامل دارد (Bui et al., 2018; Levy et al., 2009; Sahoo et al., 2009; Mukerji et al., 2009; et al., 2007). نتایج پیش‌بینی نشان داد که معیارهای مدل رقومی ارتفاعی، فاصله از رودخانه، شاخص فورنیه اصلاح شده (شاخص میانگین بارش) تأثیر بیشتری بر پتانسیل و استعداد سیل‌گرفتگی اراضی داشته و به کارگیری

Table 3- Performance of models using training data sets

جدول ۳- عملکرد مدل‌ها با استفاده از مجموعه داده‌های آموزشی

Number	Criteria	NBTree	ADTree	RF
1	TP	231	234	240
2	TN	223	223	225
3	FP	31	31	29
4	FN	23	10	14
5	PPR (%)	0.882	0.883	0.892
6	NPR (%)	0.907	0.918	0.941
7	SST (%)	0.909	0.921	0.945
8	SPF (%)	0.878	0.878	0.886
9	ACC (%)	0.894	0.900	0.915
10	SE	0.0103	0.0128	0.0102

Table 4- Performance of models using validation data sets

جدول ۴- عملکرد مدل‌ها با استفاده از مجموعه داده‌های اعتبارسنجی

Number	Criteria	NBTree	ADTree	RF
1	TP	92	100	103
2	TN	90	90	91
3	FP	19	19	18
4	FN	17	9	6
5	PPR (%)	0.829	0.840	0.851
6	NPR (%)	0.841	0.909	0.938
7	SST (%)	0.844	0.917	0.945
8	SPF (%)	0.826	0.826	0.835
9	ACC (%)	0.835	0.872	0.890
10	SE	0.0224	0.0214	0.0178

می‌دهد. علاوه بر این، می‌توان نتیجه گرفت که عملکرد آماری کلی مدل ADTree کمی بهتر از مدل NBTtree است. با این حال برای این دو مدل، مقدار AUC بالاتری را برای داده‌های آموزشی نشان داد. این پدیده نشان می‌دهد که عملکرد جامع مدل‌های مختلف باید با استفاده از شاخص‌های چندگانه برای انعکاس قابلیت‌های مدل‌ها در جنبه‌های مختلف ارزیابی و مقایسه شود. هر سه مدل مورد استفاده در این مطالعه می‌توانند نتایج قابل قبولی در پیش‌بینی خطر سیل ایجاد کنند. برای معیار فاصله، در کل با افزایش فاصله از رودخانه احتمال وقوع سیل کمتر می‌گردد. انتخاب ارتفاع از سطح دریا به عنوان یکی از مهم‌ترین معیارهای مؤثر در خطر وقوع سیل است که سیل در ارتفاعات پایین‌تر از سطح دریا اتفاق می‌افتد و در ارتفاعات بالا احتمال وقوع سیل بسیار کم است. با بررسی معیار شاخص فورنیه اصلاح شده (شاخص میانگین بارش) می‌توان به این نتیجه رسید که با افزایش ارتفاع، احتمال بارندگی بیشتر می‌گردد و از طرف دیگر احتمال وقوع سیل در ارتفاعات بالاتر کمتر است. کاربری اراضی نیز از دیگر معیارهای مؤثر بر وقوع سیل می‌باشد؛ زیرا نوع کاربری اراضی بر روی نفوذ رواناب و یا تشدید سرعت آن مؤثر است. به دنبال تغییر معیار کاربری اراضی توسط فعالیت‌های انسانی، توان آبراهه نیز کم شده و در نتیجه به عنوان معیار مهمی در وقوع سیل دخالت دارد. این نتایج با نتایج تحقیقات khosravi et al. (2016)، (2016)، Rahmati et al. (2016)، Pham et al. (2020)، Razavi Termeh et al. (2018) و Liu et al. (2016) مطابقت دارد. بر اساس نقشه پیش‌بینی خطر وقوع سیل ارائه شده می‌توان اقدامات مدیریتی مناسبی جهت کاهش خسارت‌ها و تلفات ناشی از سیل انجام داد. به‌کارگیری تکنیک داده‌کاوی و سیستم اطلاعات مکانی به منظور بررسی پتانسیل خطر وقوع سیل، مخصوصاً در کشورهای در حال توسعه که دسترسی به اطلاعات و داده‌های هیدروژئولوژیکی با مشکل و محدودیت مواجه هستند، می‌تواند مفید باشد. در این تحقیق تلاش بر این شد از مدل‌های یادگیری ماشین مبتنی بر درخت استفاده گردد. با توجه به عملکرد و دقت بالای مدل RF در پیش‌بینی خطر وقوع سیل در منطقه مورد مطالعه پیشنهاد می‌شود در تحقیقات آینده مدل موردنظر در سایر مناطق مستعد وقوع سیل با عوامل هیدروژئولوژیکی و محیطی مختلف مورد استفاده قرار گیرد و از نتایج آن برای جلوگیری و کنترل خسارت‌های ویران‌گر استفاده شود. همچنین برای تحقیقات آتی پیشنهاد می‌شود از الگوریتم‌های فراابتکاری مانند الگوریتم بهینه‌سازی کلونی مورچه، الگوریتم ژنتیک و الگوریتم بهینه‌سازی ازدحام ذرات جهت بهینه‌سازی روش‌های به کار برده شده در این تحقیق استفاده شود و عملکرد و کارایی روش‌های ترکیبی بهینه‌سازی شده بررسی شود.

تجمع سیلاب‌ها عمدتاً در ارتفاعات پایین رخ می‌دهد و سیلاب‌های بزرگ در مناطقی با ارتفاع کم اتفاق می‌افتند. با توجه به نتایج حاصل از خطر وقوع سیل در منطقه مورد مطالعه مشخص گردید که حجم سیلاب و توزیع سیل در نزدیکی رودخانه‌ها و مناطق کم‌ارتفاع قرار دارد و وقوع سیل در این مناطق افزایش می‌یابد. به طور کلی، طبق تحقیقات Pham et al. (2020) و Liu et al. (2016)، مناطقی که بیشترین حساسیت به سیل را دارند، مناطق دارای ارتفاع کم، حداقل شیب، مساحت مسطح و نزدیک به رودخانه‌ها هستند.

۴- جمع‌بندی

وقوع سیل یکی از پدیده‌های مهم مصیبت‌بار و فاجعه‌انگیز در نواحی مختلف محسوب می‌شود. از طریق تحلیل‌های مکانی مناسب می‌توان مناطق حساس به خطر وقوع سیل را شناسایی و از این طریق خسارات ناشی از آن را کاهش داد. هدف اصلی مطالعه حاضر، تهیه نقشه‌های پیش‌بینی خطر وقوع سیل با استفاده از سه رویکرد یادگیری ماشین مبتنی بر درخت برای شهرستان مانه و سملقان استان خراسان شمالی است. در مقایسه با این تحقیق، مطالعات اندکی از مدل‌های مبتنی بر درخت برای پیش‌بینی پتانسیل خطر وقوع سیل در سال‌های اخیر استفاده کرده‌اند (Chen et al., 2019; Hong et al., 2015). بنابراین، عملکرد سه مدل ترکیبی مبتنی بر درخت یعنی NBTtree، ADTree و RF در این تحقیق با استفاده از یک سری شاخص‌های آماری و مقادیر AUC مورد ارزیابی و مقایسه قرار گرفت. نتایج نهایی نشان داد که مدل RF نسبت به دو مدل دیگر برای پیش‌بینی خطر وقوع سیل عملکرد بهتری دارد و قابلیت برتر مدل RF توسط تحقیقات مرتبط دیگر تأیید شده است. (Choubin et al., 2019; Zhao et al., 2018). در مقایسه با مدل‌های ADTree و NBTtree، مدل RF یک الگوریتم ترکیبی است که می‌تواند نتایج رضایت‌بخش‌تری با نرخ‌های دقت بالاتر و واریانس‌ها و سوگیری‌های کم‌تر ایجاد کند (Chen et al., 2019; Hong et al., 2018). در نتیجه، مدل RF بهترین نقشه پیش‌بینی خطر سیل را در شهرستان مانه و سملقان ایجاد می‌کند. مدل‌های ADTree و NBTtree اساساً در دو دسته پایه با قابلیت تعمیم نسبتاً پایین هستند. عملکرد آن‌ها دارای اتصالات محکم با ویژگی‌های مجموعه داده‌های اولیه است. به خصوص برای مدل NBTtree، طبقه‌بندی‌کننده‌های بیز ساده به گره‌های برگ حاصل اعمال می‌شوند. از این‌رو، فرض قبلی استقلال ویژگی باید قبل از مدل‌سازی با استفاده از مدل NBTtree برآورده شود (Chen et al., 2017). برای مدل ADTree، عملکرد مدل ممکن است برای داده‌های آموزشی و اعتبارسنجی در برخی موارد به طور قابل‌توجهی متفاوت باشد که پایداری و تعادل ADTree را کاهش

- (ANFIS) and Biogeography-Based Optimization (BBO) and BAT Algorithms (BA). *Geocarto International* 34(11):1252-1272
- Bui DT, Panahi M, Shahabi H, Singh VP, Shirzadi A, Chapi K, Ahmad AA (2018) Novel hybrid evolutionary algorithms for spatial prediction of floods. *Scientific Reports* 8(1):1-14
- Bhowmick S, Eijkhout V, Freund Y, Fuentes E, Keyes D (2010) Application of alternating decision trees in selecting sparse linear solvers. In: K. Naono, K. Teranishi, J. Cavazos, R. Suda (Eds.), *Software Automatic Tuning: From Concepts to State-of-the-Art Results*. Springer New York, New York, NY, 153-173
- Chapi K, Singh VP, Shirzadi A, Shahabi H, Bui D, Pham BT, Khosravi K (2017) A novel hybrid artificial intelligence approach for flood susceptibility assessment. *Environmental Modelling & Software* 95:229-245
- Chen H, Ito Y, Sawamukai M, Tokunaga T (2015) Flood hazard assessment in the Kujukuri Plain of Chiba Prefecture Japan based on GIS and multicriteria decision analysis. *Natural Hazards* 78(1):105-120
- Chen W, Shirzadi A, Shahabi H, Ahmad BB, Zhang S, Hong H, Zhang N (2017) A novel hybrid artificial intelligence approach based on the rotation forest ensemble and naïve Bayes tree classifiers for a landslide susceptibility assessment in Langao County China. *Geomatics, Natural Hazards and Risk* 8(2):1955-1977
- Chen W, Hong H, Li S, Shahabi H, Wang Y, Wang X, Ahmad BB (2019) Flood susceptibility modelling using novel hybrid approach of reduced-error pruning trees with bagging and random subspace ensembles. *Journal of Hydrology* 575:864-873
- Chen W, Li Y, Xue W, Shahabi H, Li S, Hong H, Ahmad BB (2020) Modeling flood susceptibility using data-driven approaches of naïve bayes tree, alternating decision tree and random forest methods. *Science of the Total Environment* 701:134979
- 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(Pt2):2087-2096
- de Santana FB, de Souza AM, Poppi RJ (2018) Visible and near infrared spectroscopy coupled to random forest to quantify some soil quality parameters. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy* 191:454-462
- 1- Geographic Information System
 - 2- Remote Sensing
 - 3- Frequency Ratio
 - 4- Weight of Evidence
 - 5- Analytical Hierarchy Process
 - 6- Area Under the Curve
 - 7- Naïve Bayes
 - 8- Bowen
 - 9- Logistic Regression
 - 10- Random Forest
 - 11- Support Vector Machine
 - 12- Poyang
 - 13- Adaptive Neuro Fuzzy Inference System
 - 14- Ant Colony Optimization
 - 15- Genetic Algorithm
 - 16- Particle Swarm Optimization
 - 17- Receiver Operating Characteristics Curve
 - 18- Naïve Bayes Tree
 - 19- Alternating Decision Tree
 - 20- Curve Number
 - 21- Horizontal Overland Flow Distance
 - 22- Modified Fournier Index
 - 23- Vertical Overland Flow Distance
 - 24- Normalized Difference Vegetation Index
 - 25- Topographic Wetness Index
 - 26- Digital Elevation Model
 - 27- Slope
 - 28- Flow Accumulation
 - 29- Aspect
 - 30- Plane Curvature
 - 31- Distance to Fault
 - 32- Distance to Road
 - 33- Distance to River
 - 34- Land Use
 - 35- Shuttle Radar Topography Mission
 - 36- Operational Land Imager
 - 37- Positive Predictive Rate
 - 38- Negative Predictive Rate
 - 39- Sensitivity
 - 40- Specificity
 - 41- Accuracy
 - 42- Confusion Matrix
 - 43- True Positive
 - 44- True Negative
 - 45- False Positive
 - 46- False Negative
 - 47- Average Merit
 - 48- Natural Breaks
- ۵- مراجع
- Ahmadlou M, Karimi M, Alizadeh S, Shirzadi A, Parvinnejhad D, Shahabi H, Panahi M (2018) Flood susceptibility assessment using integration of Adaptive Network-Based Fuzzy Inference System

- Levy JK, Hartmann J, Li KW, An Y, Asgary A (2007) Multi-criteria decision support systems for flood hazard mitigation and emergency response in urban watersheds 1. *Journal of the American Water Resources Association* 43:346-358
- Liu R, Chen Y, Wu J, Gao L, Barrett D, Xu T, Li L, Huang C, Yu J (2016) Assessing spatial likelihood of flooding hazard using naïve Bayes and GIS: A case study in Bowen Basin, Australia. *Stochastic Environmental Research and Risk Assessment* 30(6):1575-1590
- Mukerji A, Chatterjee C, Raghuwanshi NS (2009) Flood forecasting using ANN, neuro-fuzzy, and neuro-GA models. *Journal of Hydrologic Engineering* 14(6):647-652
- Pham BT, Tien Bui D, Dholakia MB, Prakash I, Pham HV (2016) A comparative study of least square support vector machines and multiclass alternating decision trees for spatial prediction of rainfall-induced landslides in a tropical cyclones area. *Geotechnical and Geological Engineering* 34(6):1807-1824
- Pham BT, Tien Bui D, Prakash I (2017) Landslide susceptibility assessment using bagging ensemble based alternating decision trees, logistic regression and j48 decision trees methods: A comparative study. *Geotechnical and Geological Engineering* 35(6):2597-2611
- Pham BT, Avand M, Janizadeh S, Phong TV, Al-Ansari N, Ho LS, Jafari F (2020) GIS based hybrid computational approaches for flash flood susceptibility assessment. *Water* 12(683):1-30
- Pourghasemi HR, Razavi-Termeh SV, Kariminejad N, Hong H, Chen W (2020) An assessment of metaheuristic approaches for flood assessment. *Journal of Hydrology* 582:124536
- Quiroz JC, Mariun N, Mehrjou MR, Izadi M, Misron N, Mohd Radzi MA (2018) Fault detection of broken rotor bar in LS-PMSM using random forests. *Measurement* 116:273-280
- Rahmati O, Pourghasemi HR, Zeinivand H (2016) Flood susceptibility mapping using frequency ratio and weights-of-evidence models in the Golestan Province, Iran. *Geocarto International* 31(1):42-70
- Razavi Termeh SV, Kornejady A, Pourghasemi HR, Keesstra S (2018) Flood susceptibility mapping using novel ensembles of adaptive neuro fuzzy inference system and metaheuristic algorithms. *Science of the Total Environment* 615:438-451
- Saedi A, Saghafian B, Moazami S (2020) Uncertainty of flood forecasts via ensemble precipitation forecasts of seven NWP models for Spring 2019 Golestan
- Du J, Fang J, Xu W, Shi P (2013) Analysis of dry/wet conditions using the standardized precipitation index and its potential usefulness for drought/flood monitoring in Hunan Province China. *Stochastic Environmental Research and Risk Assessment* 27(2):377-387
- Eftekhari M, Eslaminezhad SA, Haji Elyasi A, Akbari M (2021) Geostatistical evaluation with Drinking Groundwater Quality Index (DGWQI) in Birjand Plain Aquifer. *Environment and Water Engineering* 7(2):268-279 (In Persian)
- Eini M, Kaboli HS, Rashidian M, Hedayat H (2020) Hazard and vulnerability in urban flood risk mapping: Machine learning techniques and considering the role of urban districts. *International Journal of Disaster Risk Reduction*:101687
- Farid DM, Zhang L, Rahman CM, Hossain MA, Strachan R (2014) Hybrid decision tree and naïve Bayes classifiers for multi-class classification tasks. *Expert Systems with Applications* 41:1937-1946
- Hong H, Tsangaratos P, Ilia I, Liu J, Zhu AX, Chen W (2018) Application of fuzzy weight of evidence and data mining techniques in construction of flood susceptibility map of Poyang County, China. *Science of the Total Environment* 625:575-588
- Imani S, Hassanoli S, Farkhnia A, Javadi F, Najafi M (2021) Evaluating the efficiency of WRF-Hydro Model for development of flood forecasting systems (Case study: Kashkan Watershed). *Iran-Water Resources Research* 16(4):225-240 (In Persian)
- Jancewicz K, Migoń P, Kasprzak M (2019) Connectivity patterns in contrasting types of tablelandsandstone relief revealed by Topographic Wetness Index. *Science of the Total Environment* 656:1046-1062
- Kanani-Sadat Y, Arabsheibani R, Karimipour F, Nasserri M (2019) A new approach to flood susceptibility assessment in data-scarce and ungauged regions based on GIS-based hybrid multi criteria decision-making method. *Journal of Hydrology* 572:17-31
- Khosravi K, Nohani E, Maroufinia E, Pourghasemi HR (2016) A GIS-based flood susceptibility assessment and its mapping in Iran: A comparison between frequency ratio and weights-of-evidence bivariate statistical models with multi-criteria decision-making technique. *Natural Hazards* 83(2):947-987
- Khosravi K, Shahabi H, Pham BT, Adamowski J, Shirzadi A, Pradhan B, Dou J, Ly HB, Gróf G, Ho HL (2019) A comparative assessment of flood susceptibility modeling using multi-criteria decision-making analysis and machine learning methods. *Journal of Hydrology* 573:311-323

- Tien Bui D, Pradhan B, Nampak H, Bui QT, Tran QA, Nguyen QP (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
- Wang LM, Li XL, Cao CH, Yuan SM (2006) Combining decision tree and Naive Bayes for classification. *Knowledge-Based Systems* 19(7):511-515
- Wang S, Jiang L, Li C (2015) Adapting naive Bayes tree for text classification. *Knowledge information system* 44:77-89
- Witten IH, Frank E, Mark AH (2011) *Data mining: Practical machine learning tools and techniques*. Third edition, Morgan Kaufmann, Burlington, USA
- Zhao G, Pang B, Xu Z, Yue J, Tu T (2018) Mapping flood susceptibility in mountainous areas on a national scale in China. *Science of the Total Environment* 615:1133-1142
- Flood. *Iran-Water Resources Research* 16(1):347-359 (In Persian)
- Sahoo GB, Schladow SG, Reuter JE (2009) Forecasting stream water temperature using regression analysis, artificial neural network, and chaotic non-linear dynamic models. *Journal of Hydrology* 378:325-342
- Smith K, Ward R (1998) *Floods: physical processes and human impacts*. John Wiley and Sons Ltd
- Tang X, Li J, Liu M, Liu W, Hong H (2020) Flood susceptibility assessment based on a novel random Naïve Bayes method: A comparison between different factor discretization methods. *Catena* 190:104536
- Tehrany MS, Pradhan B, Jebur MN (2013) Spatial prediction of flood susceptible areas using rule based Decision Tree (DT) and a novel ensemble bivariate and multivariate statistical models in GIS. *Journal of Hydrology* 504:69-79