



تحلیل خطرپذیری سیلاب حوضه‌های آبریز با بهره‌گیری از الگوریتم‌های یادگیری ماشین و عامل‌های هیدروژئومورفولوژیکی (مطالعه موردی: حوضه شرقی آبریز دریاچه اورمیه)

تاریخ دریافت: ۱۴۰۱/۱۰/۱۲

تاریخ پذیرش: ۱۴۰۲/۰۸/۱۵

سینا فرد مرادی نیا^{۱*}، یوسف زندی^۲

*-استادیار گروه مهندسی عمران، واحد تبریز، دانشگاه آزاد اسلامی، تبریز، ایران (نویسنده مسئول)

fardmoradina@iaut.ac.ir

۲-دانشیار گروه مهندسی عمران دانشکده فنی و مهندسی دانشگاه آزاد اسلامی واحد تبریز، تبریز، ایران

zandi@iaut.ac.ir

چکیده:

نقشه‌های پهنه‌بندی سیل اطلاعات ارزشمندی را در رابطه با طبیعت سیلاب‌ها و اثرات آن بر اراضی دشت سیلابی ارائه می‌دهند. برای تهیه نقشه حساسیت به وقوع سیل، مجموعه‌ای از عوامل مؤثر باید تعریف شود. مجموعه عوامل مؤثر بر سیلاب با استفاده از ۱۳ عامل شامل ارتفاع، شیب، فاصله از رودخانه، شاخص رطوبت توپوگرافی، شاخص موقعیت توپوگرافی، شاخص زبری، سطح انحنای طولی یا عرضی، مقطع انحنای توپوگرافی، انحنای کلی، شاخص نرمال شده تفاوت پوشش گیاهی (NDVI)، کاربری اراضی، سنگ شناسی و بارندگی با توجه به استفاده از تجربیات کارشناسان و پژوهشگران در بررسی‌های صورت گرفته، در حوضه شرقی آبریز دریاچه اورمیه استفاده شد. بعد از آماده سازی لایه‌های مؤثر بر سیلاب و لایه نقطه‌ای نقاط سیل خیز، و نیز انجام آزمون هم‌خطی، در مرحله بعد از پنج روش رگرسیون خطی چندگانه، مدل حداقل مربعات جزئی، رگرسیون چندکی، رگرسیون ستیغی و رگرسیون با ثبات برای مدلسازی و پیش‌بینی استفاده شد. از منحنی مشخصه عملکرد (ROC) برای اعتبارسنجی نتایج استفاده گردید. نتایج این اعتبارسنجی نشان داد که مدل‌های حداقل مربعات جزئی (PLS) و مدل رگرسیون خطی چندگانه (MLR) با دارا بودن حداکثر مساحت زیرمنحنی (AUC)، (به ترتیب ۰,۹۸۳ و ۰,۹۹۷) و کمترین میزان انحراف معیار (به ترتیب ۰,۰۱۵ و ۰,۰۱۸) بهتر عمل کرده‌اند. در بین این دو مدل هم PLS دارای نتایج کمی بهتر نسبت به MLR می‌باشد. در نهایت فاصله از آبراهه و درصد شیب تاثیرگذارترین عوامل روی سیلاب منطقه مورد مطالعه می‌باشند.

کلمات کلیدی: پهنه‌بندی - الگوریتم‌های یادگیری ماشین - عامل‌های هیدروژئومورفولوژیکی - روش رگرسیون - حوضه شرقی آبریز دریاچه اورمیه

۱- مقدمه

سیلاب به عنوان یکی از مخاطرات اصلی طبیعی، سالانه باعث تهدید زندگی، خسارت به دارایی‌ها و زیان به محیط زیست در سراسر جهان می‌شود. شناخت ابعاد و پهنه‌های سیل گیر و بررسی ویژگی‌های مکانی آن از جمله فعالیت‌هایی است که می‌تواند در بحث مدیریت بهینه مناطق در خطر سیل موثر و مفید باشد. کنترل و مدیریت مخاطره به منظور کاهش خطرات جانی و مالی در این موضوع؛ از مواردی است که می‌توان به کمک تحلیل‌های مکانی و بررسی‌های محیط طبیعی انجام داد (گزارش سازمان ملل متحد (UNISDR)، ۲۰۱۵). کشور ما هم یکی از این کشورهایی است که آسیب‌های زیادی از این مخاطره متحمل می‌شود. با شناخت عوامل مؤثر در وقوع سیلاب، از روش‌های مدل‌سازی فرآیندی و تصادفی، می‌توان پهنه‌هایی با خطر سیلاب بالا را شناسایی کرد (بختیاری و همکاران، ۲۰۱۷). پهنه‌بندی را می‌توان تقسیم زمین به واحدهای همگن از نظر استعداد وقوع پدیده‌ای خاص تعریف کرد، که به کمک تحلیل‌های مختلفی قابل انجام است. یکی از این روشها، تکنیک‌های یادگیری ماشین^۱ است که در سالهای اخیر بسیار در بحث‌های پهنه‌بندی و تهیه نقشه‌های خطرپذیری مورد استفاده قرار گرفته است. استفاده از تکنیک‌های یادگیری ماشین در پهنه‌بندی خطر سیلاب، نتایج بهتری نسبت به روش‌های سنتی مانند شبکه عصبی مصنوعی دارد (میرزایی و همکاران، ۲۰۲۱). می‌توان در محیط‌های کم داده، خطر سیلاب را ارزیابی و نقشه‌برداری کرد. در این راستا، با استفاده از تحلیل سلسله مراتبی و شبکه‌های عصبی، شاخص‌های مؤثر بر خطر سیلاب را می‌توان محاسبه کرد (اسلامیان و همکاران، ۲۰۲۰). عوامل محیطی گوناگونی در جاری شدن سیل دخالت دارند که از آن جمله: شدت بارندگی، شیب حوضه، نفوذپذیری زمین، شرایط توپوگرافی، ویژگی‌های پوشش گیاهی و درجه اشباع شدن خاک را میتوان به عنوان عوامل مؤثر در جاری شدن سیلاب نام برد. تغییر کاربری زمین و تغییرات اقلیمی باعث افزایش بارندگی

¹Machine Learning

و جریان سطحی در حوضه شده و میزان سیل در مناطق حوضه رودخانه افزایش می‌یابد (زانگ^۱ و همکاران، ۲۰۲۰). البته امروزه به دلیل دخالت‌های انسانی، ضریب سیل خیزی حوضه‌ها تغییر نموده و عمدتاً موجب کاهش دوره بازگشت آن شده است. فعالیت بشر به شکلهای گوناگونی احتمال وقوع سیل را افزایش داده که از جمله ساختمان سازی در بستر سیلابی رود و تجاوز به حریم رودخانه موجب کاهش ظرفیت طبیعی انتقال جریان رود شده است. نهایتاً کاهش جنگل‌ها و زمین‌های خالی از مراتع، بارش‌های سیلابی را افزایش می‌دهد. در نتیجه، تغییرات اقلیمی و کاربری اراضی هر دو می‌توانند منجر به آسیب‌های جدی اقتصادی و اجتماعی شوند (خانال^۲ و همکاران، ۲۰۲۰).

این تحقیق نوآوری‌هایی در مقایسه با کارهای قبلی در زمینه پهنه‌بندی سیل و ارزیابی مناطق در معرض آسیب بلایای طبیعی دارد که عبارتند از: استفاده از ۱۳ عامل مؤثر بر سیلاب برای تعریف مدل ارزیابی مناطق در معرض آسیب بلایای طبیعی، استفاده از پنج روش رگرسیون خطی چندگانه، مدل حداقل مربعات جزئی، رگرسیون چندکی، رگرسیون ستیغی و رگرسیون با ثبات برای مدلسازی و پیش‌بینی احتمال وقوع سیل در مناطق مختلف، استفاده از منحنی ROC برای اعتبارسنجی نتایج و ارزیابی عملکرد پنج روش رگرسیون و استفاده از مدل جنگل تصادفی برای مشخص کردن اهمیت عوامل مؤثر بر سیلاب و تاثیر آن‌ها در ارزیابی مناطق در معرض آسیب بلایای طبیعی. در این زمینه در تحقیق انجام شده توسط رجبی و همکاران (۱۳۴۱)، برای ارزیابی پتانسیل سیلاب در حوضه آبخیز زاب، از پارامترهای مورفومتریکی مانند ارتفاع، شیب، فاصله از رودخانه، شاخص رطوبت توپوگرافی، شاخص موقعیت توپوگرافی، شاخص زبری، سطح انحنا طولی یا عرضی، مقطع انحنا توپوگرافی، انحنا کلی، شاخص NDVI، کاربری اراضی، لیتولوژی و بارندگی استفاده شده است. همچنین از روش آماری تحلیل همبستگی برای بررسی رابطه بین پارامترهای مختلف و پتانسیل سیلاب استفاده شده است. همچنین فیضی فر رضوانی و همکاران (۱۳۴۱)، برای پیش‌بینی وقوع سیلاب در حوضه رودخانه کارون از داده‌های گرانشی ماهواره GRACE و روش‌های سنجش از دور مانند تصاویر ماهواره‌ای و نقشه‌های ساختار زمین‌شناسی استفاده شده است. همچنین از روش‌های پردازش سیگنال موجک، رگرسیون خطی چند متغیره و شبکه عصبی مصنوعی برای پیش‌بینی وقوع سیلاب استفاده شده است. در نهایت، نتایج به‌دست‌آمده توسط هر سه روش مورد بررسی قرار گرفته و دقت هر یک از روش‌ها با یکدیگر مقایسه شده است.

¹ Zhang, Q.

² Khanal, N. R.

تاکنون تحقیقات متعددی در زمینه پهنه‌بندی خطر سیل با روشهای مختلف صورت گرفته است که در ادامه به برخی از آنها اشاره می‌گردد:

Ganvati و همکاران (۲۰۱۳)، نواحی با بیشترین خطر رخداد سیلاب در حوضه فرحزاد را با استفاده از مدل فازی تعیین نمودند. در این زمینه لایه‌های اطلاعاتی مربوط به شیب، انحنای پروفیل و پلانیمتریک، ارتفاع، تراکم زهکشی، فاصله از آبراهه، شکل زمینی^۱ و کاربری اراضی مورد استفاده قرار گرفته است. در نقشه نهایی پهنه‌بندی به دست آمده مشخص گردید که خطر سیل خیزی نواحی با خطر بسیار بالا در پایین دست حوضه و منطبق بر دره اصلی فرحزاد هستند. **Salehi** و همکاران (۲۰۱۴)، پهنه‌بندی خطر سیل شهری بر پایه آنالیز تصمیم‌گیری چندمعیاره و سیستم اطلاعات جغرافیایی (GIS) را توسعه دادند. نقشه خطر سیل به دست آمده شهر تهران نشان داد که آبراهه‌ها و حریم آنها، همچنین نواحی مرکزی تا جنوب شهر بیشترین پتانسیل را برای سیل خیزی دارند. **Sattari** و همکاران (۲۰۱۷)، قابلیت‌های روشهای رگرسیون ماشین بردار پشتیبان، مدل درختی **M5** و مدل رگرسیون خطی در برآورد دبی سیلاب یک و دو ساعت آینده ایستگاه تازه کند در رودخانه اهرچای را مورد مقایسه قرار دادند. نتایج نشان داد که روش رگرسیون ماشین بردار پشتیبان با ضریب تبیین ۰٫۹۶ و جذر میانگین مربعات خطا ۰٫۰۴۷۲، برای سیلاب یک ساعت بعد و $R^2 = 0.90$ و $RMSE = 0.1596$ برای سیلاب دو ساعت بعد بهترین نتیجه را ارائه نمود. **Abedini** و همکاران (۲۰۱۸)، پهنه‌بندی قسمتی از شهرستان مشکین شهر به مساحت ۳۱۸ کیلومتر مربع، از لحاظ پتانسیل وقوع سیلاب را انجام دادند. نتایج مطالعه نشان داد، عوامل ارتفاع، لیتولوژی، بارش و شیب به ترتیب بیشترین تأثیر را بر ایجاد سیل در منطقه مطالعاتی دارند. همچنین، با توجه به نتایج به دست آمده به ترتیب ۲۳/۳۷ و ۵۲/۵۰ کیلومتر مربع از مساحت محدوده، در طبقه بسیار پرخطر و پرخطر قرار دارند. **Ebrahimipour** و **Ziari** (۲۰۱۹)، با کمک سامانه اطلاعات جغرافیایی و تبادل اطلاعات با مدل‌های شبیه‌سازی ریاضی، مقاطع عرضی از رودخانه چشمه کیله تهیه و ظرفیت ایمن هر یک از مقاطع با توجه به دوره‌های بازگشت مختلف تعیین نمودند. براساس نتایج حاصل شده، زمینهای پیرامون این رودخانه در فواصل مختلف به رنگهای متفاوت بر اساس دوره بازگشت سیل پهنه‌بندی شده اند که این پهنه‌ها شامل بستر و حریم رودخانه (پهنه پرخطر و با خطر متوسط) می‌باشد. **Nouri** و همکاران (۲۰۱۹)، سه روش اصلی در حوزه هوش مصنوعی (مدل شبکه عصبی پرسپترون چند لایه، مدل الگوریتم ترکیبی شبکه عصبی با ژنتیک و مدل ترکیب خوشه‌بندی کاهشی و روش نرو فازی (ANFIS)) را با یکدیگر مقایسه کردند. به منظور پیش‌بینی دبی حداکثر رودخانه یلفان در محل ایستگاه هیدرومتری و رسوب‌سنجی یلفان (یکی از زیرحوضه‌های سد اکباتان همدان)، بهترین روش پیش‌بینی انتخاب شد.

¹ landform

می باشد مورد استفاده قرار دادند. ارزیابی کمی و کیفی نتایج با استفاده از آماره‌های گوناگون از جمله آماره ناش - ساتکلیف نشان داد که روش مدلسازی ماشین بردار پشتیبان، از بیشترین دقت نسبت به دو روش مدلسازی دیگر به منظور پیش‌بینی دبی‌های حداکثر لحظه‌ای در حوضه آبخیز دریاچه نمک برخوردار است. Hejazi و همکاران (۲۰۲۰)، سطوح سیل‌گیر در دوره‌های بازگشت ۲۵ و ۵۰ ساله، در طول رودخانه اصلی ورکش - چای را تعیین نمودند. در همین راستا، به منظور شبیه‌سازی هندسه سه بعدی زمین، سیستم اطلاعات جغرافیایی (GIS) و مدل HEC-RAS و الحاقیه HEC-GeoRAS مورد استفاده قرار داده و نقشه‌های پلان رودخانه، سواحل چپ و راست رودخانه و سرعت جریان آن به دست آمد و روستاهایی که در معرض وقوع سیلاب با دوره‌های بازگشت مذکور قرار داشتند، شناسایی و سپس رفتار هیدرولیکی رودخانه را شبیه‌سازی نمودند. Faramarzi و همکاران (۲۰۲۰)، متغیرهای مهم در ایجاد مناطق سیل‌گیر و سیل خیز و ارائه پتانسیل مخاطره سیل پارک ملی گلستان با استفاده از تکنیک‌های یادگیری ماشین شامل مدل جنگل تصادفی، درخت رگرسیون تقویت‌شده و آنتروپی بیشینه را شناسایی کردند. نتایج به دست آمده از ارزیابی صحت مدل‌ها با استفاده از ۳۰ درصد از داده‌های وقوع سیل که در مدل‌سازی وارد نشده بود نیز دقت زیاد مدل درخت رگرسیون تقویت‌شده و جنگل تصادفی را با مقدار ROC، ۰/۹۹ و دقت مناسب آنتروپی بیشینه را با مقدار ROC، ۰/۸۹ نشان داد، به طوری که نقشه‌های به دست آمده از این مدل‌ها به طور مشترک ۴۵۰۰ هکتار از مساحت پارک را دارای احتمال زیاد خطر سیل برآورد کردند.

Goodarzi و Fatehifar (۲۰۲۰)، خطرپذیری سیلاب حوضه آذرشهر چای را در اثر تغییرات اقلیمی مورد بررسی قرار دادند. نتایج ارزیابی مدل SDSM نشان می‌دهد که با داشتن ضریب تعیین و نش-ساتکلیف به میانگین ۰,۹۵، این مدل به خوبی در ریزمقیاس‌نمایی داده‌های بزرگ مقیاس عمل کرده است همچنین افزایش ۰,۲۳ درجه سانتی‌گراد دما و ۴,۵۳ درصدی بارش و ۸,۴ درصدی افزایش دبی سیلاب در دوره آتی ۲۰۳۰-۲۰۵۹ پیش‌بینی شده است. Rostami و همکاران (۲۰۲۰)، برای تعیین پتانسیل خطر سیلاب در شهرک امام علی (ع) شهر مشهد از روشهای تصمیم‌گیری چندمعیاره استفاده کردند. برای این منظور از شش لایه اطلاعاتی شامل لایه فاصله از کانال، شیب، کاربری اراضی، تراکم زهکشی، شیب کانال اصلی و ارتفاع استفاده گردید. نتایج و بررسی‌های میدانی نشان داد روش تصمیم‌گیری چندمتغیره روشی دقیق و مناسب برای پهنه‌بندی خطر سیلاب شهری می باشد و با استفاده از این روش می‌توان مناطقی را که دارای خطر آبگرفتگی

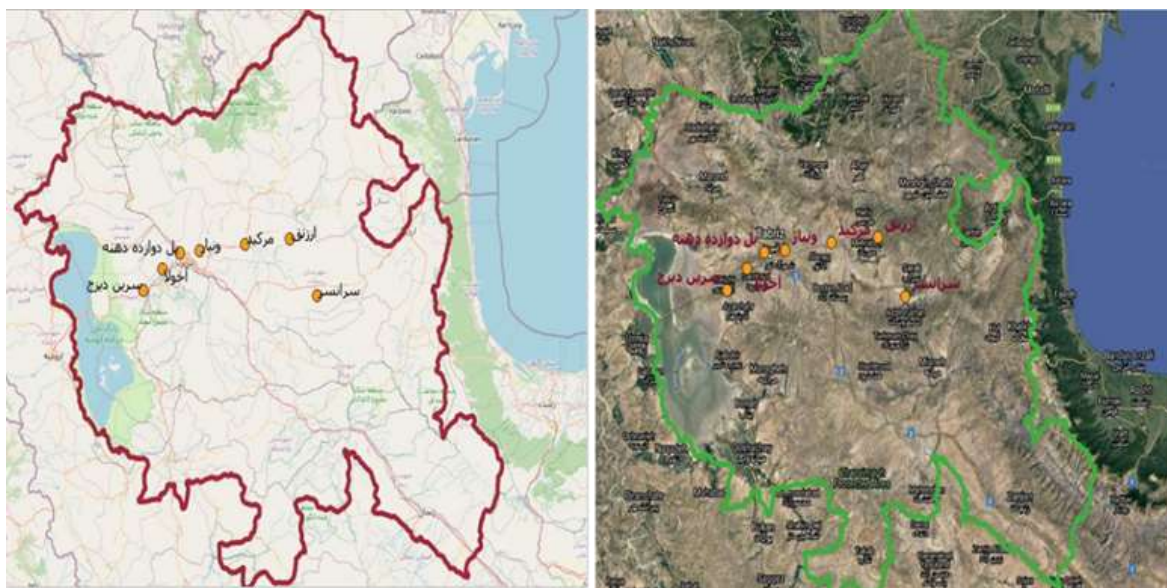
بیشتری هستند مشخص کرد. Rahimi و همکاران (۲۰۲۰)، ه منظور کاهش خسارات ناشی از سیلاب، میزان ارتباط بین هر یک از متغیرهای مکانی، ادراک خطر و رفتارهای پیشگیرانه را مورد بررسی قرار دادند. نتایج آزمون رگرسیون چند متغیره نیز نشان می دهد که اتصال به محل بر ادراک خطر با توان ۱، ارتباط بیشتری نسبت به اتصال به محل بر رفتار پیشگیرانه با توان ۰.۹۷ دارد. نتایج آزمون همبستگی پیرسون، نشان داد که متغیرهای دل بستگی به مکان و ادراک خطر با ضریب ۰/۴۷۲ بیشترین همبستگی را دارند. Malazehi و همکاران (۲۰۲۰)، پهنه بندی و پتانسیل سنجی خطر سیلاب در حوضه آبریز سرباز را انجام دادند. نتایج حاصل از این تحقیق نشان داد که از کل مساحت حوضه سرباز فقط ۵/۰ درصد پتانسیل سیل خیزی زیاد تا بسیار زیاد را داراست و نواحی مرکزی حوضه به سبب توپوگرافی و شرایط هیدروگرافی ویژه، بالاترین پتانسیل سیل خیزی را دارند. Avand و همکاران (۲۰۲۰)، مناطق حساس به سیل با استفاده از دو مدل یادگیری ماشین جنگل تصادفی (RF) و خطی تعمیم یافته بیزین ادر حوضه آبخیز تجن در استان مازندران، شهرستان ساری را مورد شناسایی قرار دادند. نتایج نشان داد که سه عامل ارتفاع (۲۱/۵۵)، فاصله از رودخانه (۱۵/۲۸) و شیب (۱۱/۱۸) به ترتیب بیشترین تأثیر را در سیل گیری منطقه مورد مطالعه دارند. در این تحقیق چارچوب جامع جهت ارزیابی خطر سیلاب ارائه گردید. سیزده عامل مؤثر بر سیلاب شامل ارتفاع، شیب، فاصله از رودخانه، شاخص رطوبت توپوگرافی، شاخص موقعیت توپوگرافی، شاخص زبری، سطح انحنای طولی یا عرضی، مقطع انحنای توپوگرافی، انحنای کلی، شاخص NDVI، کاربری اراضی، لیتولوژی و بارندگی که از تجربیات کارشناسان و پژوهشگران در بررسی های صورت گرفته، به دست آمده در حوضه شرقی آبخیز دریاچه اورمیه مورد استفاده قرار گرفت. نهایتاً بعد از آماده سازی لایه های مؤثر بر سیلاب و لایه نقطه ای نقاط سیل خیز، و نیز انجام آزمون هم خطی از پنج روش رگرسیون خطی چندگانه، مدل حداقل مربعات جزئی، رگرسیون چندکی، رگرسیون ستیغی و رگرسیون با ثبات برای مدلسازی و پیش بینی مورد استفاده قرار گرفت. در انتها از منحنی ROC برای اعتبارسنجی نتایج استفاده گردید.

۲- مواد و روشها

۲-۱- منطقه مورد مطالعه

تلخه‌رود (آجی‌چای، شورسو) یکی از مهم‌ترین و اصلی‌ترین رودخانه‌های حوضه شرقی دریاچه اورمیه واقع در استان آذربایجان شرقی می‌باشد. مساحت حوزه آبریز این رودخانه، ۹۲۲۰۰ کیلومتر مربع و حداکثر دبی آن ۴۰/۶ متر مکعب در ثانیه است. طول این رودخانه ۲۶۵ کیلومتر است و در مناطق مرکزی استان

آذربایجان شرقی در جریان می‌باشد (شکل ۱). تلخه‌رود (آجی‌چای) از دامنه‌های سبلان سرچشمه گرفته و پس از طی حدود ۲۲۰ کیلومتر و عبور از دره و نیار در شمال شهر تبریز و پس از پیوستن آب‌های سرچشمه گرفته از کوه سهند، به دشت وسیع تبریز وارد می‌شود. تلخه‌رود پس از مشروب کردن دشت تبریز به دریاچه ارومیه می‌ریزد. در جدول ۱ موقعیت ایستگاه‌های هیدرومتری منطقه آمده است.



شکل (۱): نقشه DEM حوضه شرقی آبریز دریاچه اورمیه

Fig.(1): DEM map of the eastern catchment area of Lake Urmia

جدول (۱): موقعیت های ایستگاههای هیدرومتری منطقه مورد مطالعه

Table (1): Locations of hydrometric stations in the study area

StCode	St_Name _1	Lon_dms	Lon_dd	Lat_dms	Lat_dd	Riv_Name_1	Elv	Year	Degree
۳۱۰۰۵/۰۰	سرانسر	۴۷-۲۹	۴۷/۴۸۰۰	۳۷-۵۳	۳۷/۸۸۰۰	آجی چای	۱۶۶۰	NA	۴/۰۰۰
۳۱۰۱۴/۰۰	ارزئق	۴۷-۱۴-۰۰	۴۷/۲۳۳۳	۳۸-۱۲-۰۰	۳۸/۲۰۰۰	آجی چای	۱۶۶۰	۱۳۸۰	۲/۰۰۰
۳۱۰۱۵/۰۰	ونیار	۴۶-۲۴-۲۲	۴۶/۴۰۶۱	۳۸-۰۷-۵۵	۳۸/۱۳۱۹	آجی چای	۱۴۵۰	۱۳۲۸	۱/۰۰۰
۳۱۰۲۵/۰۰	پل دوازده دهنه	۴۶-۱۳	۴۶/۲۲۰۰	۳۸-۰۷	۳۸/۱۲۰۰	آجی چای	۱۳۴۵	NA	۴/۰۰۰
۳۱۰۴۵/۰۰	آخولا	۴۶-۰۳-۵۳	۴۶/۰۶۴۷	۳۸-۰۱-۴۶	۳۸/۰۲۴۹	آجی چای	۱۳۱۰	۱۳۶۲	۱/۰۰۰
۳۱۰۸۵/۰۰	سرين ديزج	۴۳-۵۳-۱۵	۴۵/۸۸۷۵	۳۷-۵۴-۵۴	۳۷/۹۱۵۰	آجی چای	۱۵۰۰	۱۳۸۰	۳/۰۰۰
۳۱۱۱۷/۰۰	مرکید	۴۶-۴۹-۲۲	۴۶/۸۲۲۷	۳۸-۱۰-۰۶	۳۸/۱۶۸۳	آجی چای	۱۵۲۰	۱۳۷۰	۱/۰۰۰

۲-۲- مراحل انجام این تحقیق

می توان الگوریتم کار برای تحلیل خطرپذیری سیلاب حوضه های آبریز با بهره گیری از الگوریتم های یادگیری ماشین و عامل های هیدروژئومورفولوژیکی به صورت زیر رسم کرد:

- ۱- جمع آوری داده ها و شناسایی متغیرهای مربوط به خطرپذیری سیلاب حوضه ی آبریز
- ۲- پیش پردازش داده ها شامل پاک سازی، تبدیل، نرمال سازی و استانداردسازی داده ها
- ۳- اعمال روش های آماری و تحلیلی برای تحلیل داده ها، از جمله رگرسیون خطی چندگانه، رگرسیون با ثبات، رگرسیون ستیغی، رگرسیون چندکی و مدل حداقل مربعات جزئی
- ۴- تعیین متغیرهای مهم و ارتباط آن ها با خطر سیلاب
- ۵- طراحی و پیاده سازی مدل یادگیری ماشین با استفاده از الگوریتم جنگل تصادفی
- ۶- آموزش مدل با داده های آموزشی و ارزیابی کیفیت مدل با داده های آزمایشی
- ۷- استفاده از عامل های هیدروژئومورفولوژیکی، برای تحلیل خطرپذیری سیلاب در حوضه های آبریز
- ۸- ترکیب نتایج حاصل از روش های آماری، تحلیلی و یادگیری ماشین به منظور ارائه نقشه ی پهنه بندی خطرپذیری سیلاب در حوضه های آبریز

۹- بررسی دقت و کیفیت نقشه‌ی پهنه‌بندی خطرپذیری سیلاب با استفاده از معیارهای مختلف از جمله ROC curve.

۱۰- ارائه پیشنهادات و توصیه‌ها برای کاهش خطر سیلاب در حوضه‌های آبریز با استفاده از نتایج حاصل از تحلیل خطرپذیری سیلاب

۳-۲- پارامترهای مؤثر در تهیه نقشه پتانسیل سیل

برای تهیه نقشه حساسیت به وقوع سیل و یا به طور کلی تولید مدلی برای ارزیابی نواحی در معرض آسیب بلایای طبیعی، مجموعه‌ای از عوامل مؤثر باید تعریف گردد. مجموعه عوامل مؤثر بر سیلاب با استفاده از ۱۳ عامل شامل ارتفاع، شیب، فاصله از رودخانه، شاخص رطوبت توپوگرافی، شاخص موقعیت توپوگرافی^۱، شاخص زبری سطح (Roughness)، انحنای مقطع (Profile curvature)، انحنای توپوگرافی (Plan curvature)، انحنای کلی (Curvature)، شاخص NDVI، کاربری اراضی، لیتولوژی و بارندگی با توجه به استفاده از تجربیات کارشناسان و پژوهشگران در بررسی‌های صورت گرفته، در منطقه مطالعه شده استفاده شد. داده‌های ماهواره‌ای SRTM با قدرت تفکیک ۹۰ متر به منظور تهیه مدل رقومی ارتفاعی در منطقه مطالعاتی مورد استفاده قرار گرفت. طبقات ارتفاعی مختلف باعث ایجاد شرایط اقلیمی متفاوت شده و در نتیجه باعث ایجاد خاک و نوع پوشش گیاهی متفاوت می‌گردد (Jothibasu و Anbazhagan، ۲۰۱۶). شیب عمدتاً فرآیند تغذیه آبهای زیرزمینی، نفوذ و رواناب را کنترل میکند. بنابراین یک عامل مؤثر جهت پیش بینی فضایی پتانسیل آبهای زیرزمینی می‌باشد. نقشه شیب بر اساس مدل رقومی ارتفاعی با استفاده از ابزار Spatial Analysis در محیط نرم افزار ArcGIS 10.3 تهیه گردید. انحنای سطح بر روی همگرایی و واگرایی جریان تأثیر می‌گذارد و انحنای مقطع مطابق با جهت حداکثر شیب می‌باشد و عمدتاً بر روی سرعت جریان در سطح تأثیرگذار است (Al-Abadi و همکاران، ۲۰۱۶). نقشه انحنای توپوگرافیک، انحنای مقطع و انحنای کلی با استفاده از مدل رقومی و در محیط نرم افزار SAGA-GIS تهیه گردید. انحنای توپوگرافیک به ۳ کلاس مقعر، هموار و محدب تقسیم بندی می‌گردد. مقادیر منفی انحنای توپوگرافی بیانگر تفرق و مقادیر مثبت آن بیانگر تحذب سطح می

¹ Topographic position index

باشد. مقادیر صفر نشان دهنده همواربودن سطح می‌باشد (Naghibi و همکاران، ۲۰۱۵). انحنای مقطع نیز به ۳ کلاس تقسیم بندی می‌گردد. شاخص رطوبت توپوگرافی (TWI) تعیین کننده تاثیر توپوگرافی بر روی میزان سطوح اشباع جهت تولید رواناب است که با استفاده از معادله ۱ محاسبه می‌شود (Moore و همکاران، ۱۹۹۱).

$$TWI = Ln(A_s / \tan \theta) \quad (1)$$

که در آن A_s مساحت ویژه حوضه (مترمربع) و θ درجه شیب می‌باشد. برای تهیه لایه فاصله از رودخانه از الحاقیه Arc hydro استفاده شد و ابتدا آبراهه‌های موجود در حوضه شبیه‌سازی شد و سپس با استفاده از تابع Distance لایه فاصله از رودخانه‌ها محاسبه گردید. لایه موقعیت توپوگرافیک و شاخص زبری با استفاده از مدل رقومی ارتفاع و در نرم افزار SAGA تهیه شد. جهت تهیه لایه NDVI ابتدا تصویر لندست مربوط به حوضه مورد مطالعه دانلود گردید، سپس در محیط نرم افزار ENVI ابتدا تصحیحات رادیومتریک و اتمسفری انجام شد و سپس این شاخص محاسبه گردید. برای تهیه لایه کاربری و لیتولوژی و بارش نیز از نقشه‌های سازمان جنگلها، مراتع و آبخیزداری استفاده شد. برای تهیه نقشه موقعیت نقاط سیل‌خیز از نقشه‌ای که سازمان جنگلها در آن پهنه‌های سیل‌گیر را مشخص کرده استفاده شد و نقاط سیل‌خیز استخراج گردید. شایان ذکر است از ۷۰ درصد داده‌ها جهت آموزش و از ۳۰ درصد داده‌ها جهت اعتبارسنجی استفاده گردید.

۴-۲-آزمون هم‌خطی عوامل موثر بر سیلاب

بعد از آماده سازی لایه‌های موثر بر آب زیرزمینی و نیز لایه نقطه ای چاهها، ابتدا برای تعیین هم‌خطی از پارامترهای Tolerance یا رواداری و VIF یا عامل تورم واریانس استفاده شد. اگر در جریان شبیه سازی بین پارامترها هم‌خطی وجود داشته باشد، دقت پیش بینی مدل کاهش می‌یابد، در نتیجه بررسی آزمون هم‌خطی بین پارامترهای مؤثر ضروری می‌باشد. در این تحقیق برای تعیین هم‌خطی از پارامترهای Tolerance یا رواداری و VIF یا عامل تورم واریانس استفاده شده است. اگر مقدار ضریب تحمل کمتر از ۰/۱ و مقدار عامل تورم واریانس بزرگتر از ۱۰ باشد، بین پارامترها هم‌خطی وجود دارد (Al-Abadi و همکاران، ۲۰۱۶). برای محاسبه آزمون هم‌خطی، مقادیر عددی هر یک از پارامترها در محل نقاط سیل‌خیز، استخراج گردید و در محیط SPSS آزمون هم‌خطی محاسبه گردید.

۵-۲- معرفی مدل‌های استفاده شده در تحقیق

بعد از آماده سازی لایه‌های موثر بر سیلاب و لایه نقطه ای نقاط سیل خیز، و نیز انجام آزمون هم‌خطی در مرحله بعد، از پنج روش رگرسیون خطی چندگانه^۸، مدل حداقل مربعات جزئی^۹، رگرسیون چندکی^{۱۰}، رگرسیون ستیغی^{۱۱} و رگرسیون با ثبات^{۱۲} جهت مدلسازی و پیش‌بینی بهره گرفته شد، که در ادامه به معرفی مختصر این روشها پرداخته می‌شود.

۱-۵-۲- رگرسیون خطی چندگانه (MLR)

روش رگرسیون خطی چندگانه یکی از موثرترین و پرکاربردترین روش‌های تحلیل چندمتغیره محسوب می‌شود. در روش رگرسیون خطی چندگانه، یک رابطه خطی بین متغیر وابسته^{۱۳} با یک یا چند متغیر مستقل^{۱۴} برقرار می‌شود. البته گاهی به متغیر وابسته، متغیر پاسخ^{۱۵} و به متغیرهای مستقل متغیرهای پیش‌گو^{۱۶} نیز می‌گویند. روش رگرسیونی را چندگانه^{۱۷} می‌گویند اگر چندین متغیر مستقل قابلیت پیشگویی متغیر وابسته را داشته باشند و بتوان رابطه را به صورت خطی در نظر گرفت. ارتباط خطی بین متغیرهای مستقل X_1, X_2, \dots, X_n و متغیر Y در این مدل به صورت رابطه ۲ است:

$$Y = a_0 + a_1X_1 + a_2X_2 + \dots + a_nX_n \quad (2)$$

که در آن، a_1, a_2, \dots, a_n ضرایب رگرسیونی و a_0 عرض از مبدا می‌باشند (Balan و همکاران، ۱۹۹۵). در روش رگرسیون خطی چندگانه، پارامترهای مدل شامل ضرایب عددی برای هر یک از متغیرهای وابسته است. برای تنظیم این پارامترها، از یک الگوریتم بهینه‌سازی مانند الگوریتم کمترین مربعات (OLS) استفاده می‌شود. برای اجرای الگوریتم OLS، معمولاً از ماتریس داده‌های آموزشی و ماتریس پاسخ‌های مشاهده شده

⁸ Multiple Linear Regression

⁹ Partial Least Squares

¹ Quantile Regression 0

¹ Ridge Regression 1

¹ Robust Regression 2

¹ Dependent Variable 3

¹ Independent Variable 4

¹ Respond Variable 5

¹ Predictor Variables 6

¹ Multiple Linear 7

¹ Ordinary Least Squares 8

در آن‌ها استفاده می‌شود. در صورت وجود متغیرهای فرعی یا برهم‌کنش‌هایی بین متغیرها، این موارد نیز در ماتریس داده‌ها اضافه می‌شوند تا با توجه به آن‌ها نیز پارامترهای مدل تنظیم شوند.

۲-۵-۲- مدل حداقل مربعات جزئی (PLS)

هنگامی که تعداد متغیرهای توضیحی بسیار زیاد است و بین متغیرها هم‌خطی شدید وجود دارد رگرسیون PLS در ساختن مدل‌هایی برای پیش‌بینی متغیر پاسخ بسیار مفید است. شایان ذکر است که این روش یک روش رگرسیونی اریب است که بیش‌تر در پیش‌بینی متغیر پاسخ استفاده می‌شود و معمولاً کاربردی در تفسیر ارتباط بین متغیر پاسخ و متغیرهای توضیحی ندارد. همچنین هنگامی که بیش‌از یک متغیر پاسخ در مسئله وجود دارد نیز میتوان از این روش استفاده نمود. این روش بر اساس الگوریتم‌های نسبتاً پیچیده‌ای تعریف شده که این موضوع فهم آن را کمی مشکل میکند (Garthwaite و همکاران، ۱۹۹۵). در روش رگرسیون مدل حداقل مربعات جزئی، برای یافتن پارامترهای بهینه، از روش حل معادلات نرمال یا روش گرادیان نزولی استفاده می‌شود. در این روش، همانند روش رگرسیون خطی ساده و رگرسیون خطی چندگانه، پارامترهای مورد بررسی ضرایب عددی برای هر یک از متغیرهای مستقل هستند.

۲-۵-۳- رگرسیون چندکی (QR)

شکل و شیوه محاسبات رگرسیون چندکی با روش‌های معمول رگرسیونی کاملاً متفاوت است. در روش رگرسیون معمولی با استفاده از ضرب داخلی، به یک زیر فضا رسیده و در عمل به کمک جبر خطی قادر به برآورد پارامترهای مدل رگرسیون خطی خواهیم شد. در حالیکه در رگرسیون چندکی چنین ساختاری وجود ندارد و روش حل مانند روش‌های برنامه ریزی خطی (روش سیمپلکس) است (Koenker، ۲۰۰۵). رگرسیون چندکی یک روش آماری با قابلیت محاسبه و رسم منحنیهای رگرسیونی متفاوت و منطبق با نقاط صدکی مختلف میباشد، که ضمن بیان تصویری کاملتر و جامع‌تر از داده‌ها، امکان سنجش ارتباط متغیرهای مستقل با چندکهای مورد نظر متغیر وابسته را بدون نیاز به نرمال بودن داده‌ها و حتی در حضور نقاط دور افتاده فراهم میکند یعنی این رگرسیون نسبت به داده‌های دور افتاده نیرومند می‌باشد. از سوی دیگر برخلاف رگرسیون حداقل مربعات که روی میانگین شرطی یعنی پارامتر مکان متمرکز است، رگرسیون چندکی استراتژی منظمی را برای تعیین چگونگی تاثیر متغیرهای مستقل روی مکان و مقیاس و شکل توزیع پیشنهاد می‌کند (Assaf و همکاران، ۲۰۱۹). از جمله پارامترهایی که می‌توان در رگرسیون چندکی تنظیم کرد، تعداد متغیرهای ورودی^۱، نوع تابع فعال‌سازی، تعداد لایه‌ها و اندازه هر لایه، نرخ یادگیری و تعداد ایپوک‌ها^۲ است.

1 features	9
2 learning rate	0
2 epochs	1

تعیین بهترین مقادیر این پارامترها با استفاده از تجربه و آزمایشات مکرر با داده‌های مختلف انجام می‌شود. برای تنظیم پارامترهای رگرسیون چندکی، ابتدا باید مجموعه داده‌های مورد نظر را به دو بخش آموزشی^{۲۲} و آزمایشی^{۲۳} تقسیم کرد. سپس با استفاده از الگوریتم‌های یادگیری ماشین، مدل رگرسیون چندکی را برای داده‌های آموزشی ساخت. در این مرحله، پارامترهای مدل به صورت خودکار توسط الگوریتم بهینه سازی تعیین می‌شوند.

۴-۵-۲- رگرسیون ستیغی (RR)

در صورتیکه مدل رگرسیونی دچار بیش‌برازش شود، خطای آن برای برآورد مقادیر جدید متغیر وابسته زیاد خواهد بود؛ در حالیکه وجود متغیرهای کمتر از حد لازم در مدل (کم‌برازش) واریانس مدل را افزایش می‌دهد. بنابراین با افزایش تعداد متغیرها مشکل هم‌خطی و بیش‌برازش^{۲۴} ظاهر شده و با کاهش آن‌ها، واریانس مدل افزایش خواهد یافت. یکی از روش‌های غلبه بر این مسائل در رگرسیون چندگانه، استفاده از مدل رگرسیون ستیغی است (Buitendag و همکاران، ۲۰۱۹). از آن جایی که در زمانی که متغیرهای مدل، زیاد و یا هم‌خطی چندگانه وجود داشته باشد، واریانس برآوردگرها متورم شده و به شکل قله (ستیغ) در می‌آید، از همین روی، این روش رگرسیونی که بر این مشکل غلبه می‌کند، رگرسیون ستیغی نام‌گذاری شده است. در رگرسیون ستیغی به کمک ترکیب تابع مجموع مربعات خطا و مقدار جریمه مرتبط با تعداد پارامترها، تابع جدیدی ساخته می‌شود که برای برآورد پارامترهای مدل رگرسیونی به کار می‌رود. حالت کلی‌تر این روش را به نام قاعده‌سازی تیکانوف^{۲۵} می‌شناسند (Lin و همکاران، ۲۰۱۹). تنظیم پارامترها در این روش از طریق یک پارامتر به نام α انجام می‌شود. این پارامتر به صورت مستقیم تاثیرگذار بر تعداد ضرایبی است که به صفر بر می‌گردد. برای مقادیر بزرگتر از صفر، تعداد ضرایب صفر بیشتر می‌شود و مدل ساده‌تر و کمتر انعطاف‌پذیر می‌شود. در صورتی که مقدار α بسیار کوچک باشد، رگرسیون ستیغی به رگرسیون خطی سنتی نزدیک می‌شود.

2 train	2
2 test	3
2 Overfitting	4
2 Tikhonov Regularizatio	5

۵-۵-۲- رگرسیون باثبات (RLR)

رگرسیون باثبات یا رگرسیون مقاوم به روش‌های رگرسیون گفته می‌شود که رفتار باثبات و مقاومی در برابر وجود داده غیرمعمول دارند (Liu و همکاران، ۲۰۱۹). بعضی از روش‌های معمول رگرسیون مانند کمترین مربعات در صورت صدق فرض‌های آنان به خوبی کار می‌کنند اما در مورد داده‌هایی که از فرض‌های آنان تخلف می‌کنند شاید به خوبی عمل نکنند. به ویژه، روش کمترین مربعات نسبت به داده پرت حساس است. مسئله دیگر وجود ناهم‌وابستگی در داده است که رگرسیون با ثبات به خوبی با این مسائل کار می‌کند (Yesilnacar، ۲۰۰۵). در روش رگرسیون با ثبات، پارامترهای مدل با استفاده از یک تابع هزینه‌ی مناسب تخمین زده می‌شوند که در آن تابع هزینه، تاثیر داده‌های پرت و نویزی کاهش یافته است. یکی از توابع هزینه‌ای که معمولاً در این روش به کار می‌رود تابع هابِر (Huber loss) است.

۶-۲- اعتبارسنجی شبیه سازی

به منظور ارزیابی مدل با استفاده از مساحت زیرمنحنی^{۲۶} (AUC) در منحنی ویژگی عملگر نسبی^{۲۷} (ROC) قدرت پیش بینی مدل با استفاده از شاخص سطح سلول هسته^{۲۸} (SCAI) دقت تفکیک بین طبقات مورد بررسی و تأیید قرار گرفت. در مساحت زیرمنحنی، درصد نقاط سیلابی که به درستی به وسیله مدل پیش بینی شده است در مقابل درصد نقاطی که به درستی پیش بینی نشده است ترسیم می‌گردد. رابطه کمی- کیفی بین AUC و دقت پیش بینی که دامنه بین ۰/۵ تا ۱ را شامل می‌شود به شرح ذیل می‌باشد: ۱- ۰/۹ دقت عالی، ۰/۸ - ۰/۹ دقت خیلی خوب، ۰/۷-۰/۸ دقت خوب، ۰/۶-۰/۷ دقت متوسط و ۰/۶ - ۰/۵ دقت ضعیف (Yesilnacar، ۲۰۰۵). با توجه به این که برای ارزیابی مدل نمی‌توان از موقعیت نقاط شبیه سازی استفاده نمود، از بین موقعیت نقاط، ۷۰ درصد برای اجرای مدل و ۳۰ درصد برای ارزیابی مدل مورد استفاده قرار گرفت (Constantin و همکاران، ۲۰۱۱).

² Area under the curve 6

² receiver operating characteristic curve

² Seed Cell Area Index 8

عنوان خلاصه شده مقاله: تحلیل خطرپذیری سیلاب حوضه‌های آبریز با بهره‌گیری از الگوریتم‌های یادگیری ماشین و عامل‌های هیدروژئومورفولوژیکی (مطالعه موردی: حوضه شرقی آبریز دریاچه اورمیه)

سینا فرد مرادی نیا، یوسف زندی

۳۸

۳- نتایج و بحث

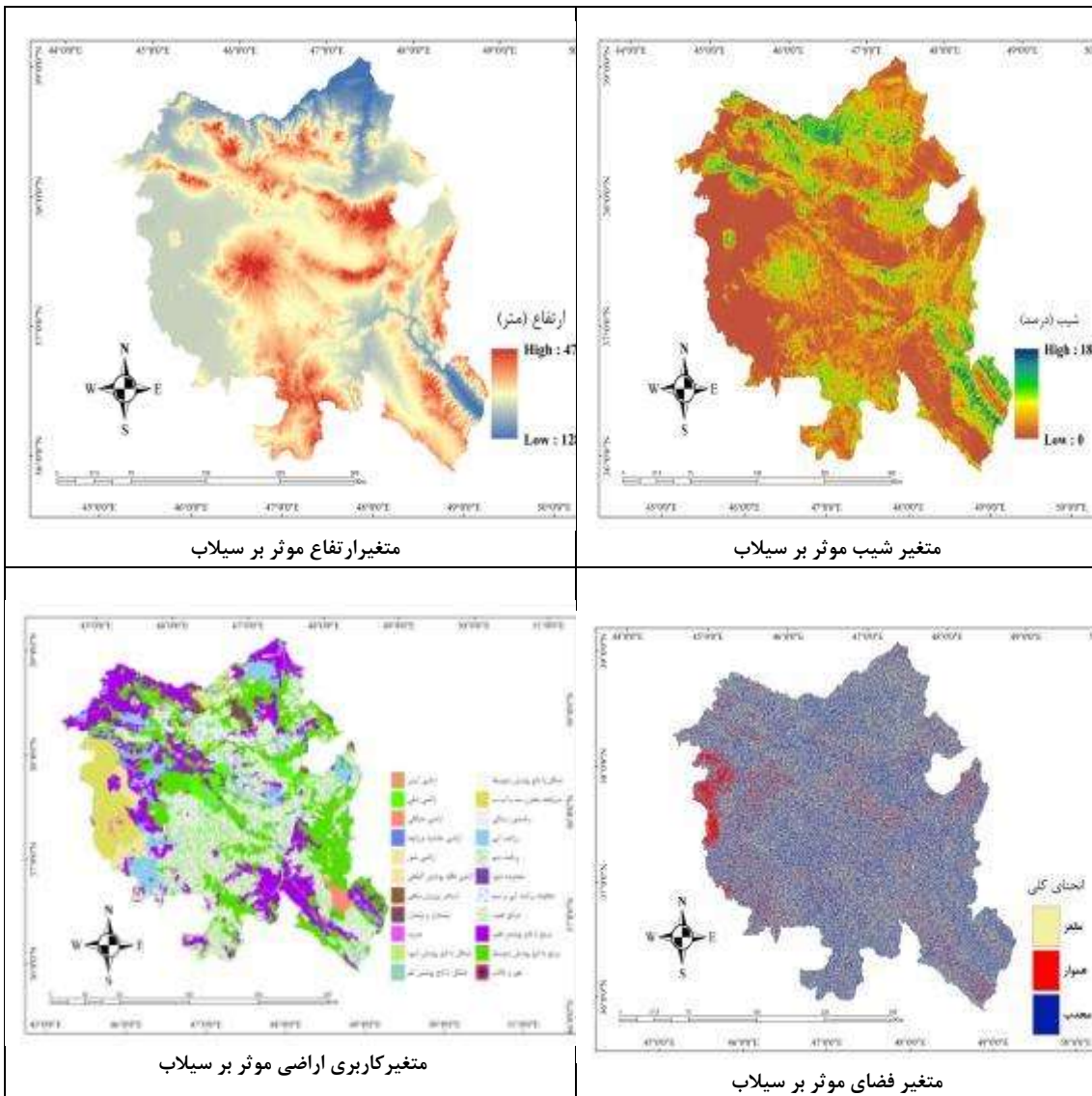
۳-۱- تهیه لایه‌های موثر بر سیل خیزی

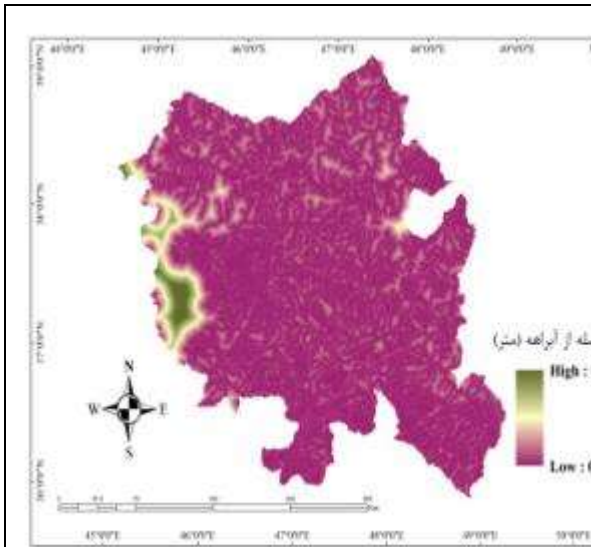
همان طور که در مراحل قبل نیز گفته شد در این تحقیق ۱۳ متغیر از بین مهم ترین عوامل موثر بر سیلاب شامل ارتفاع، شیب، فاصله از رودخانه، شاخص رطوبت توپوگرافی، شاخص موقعیت توپوگرافی، شاخص زبری سطح، انحناى مقطع^۱، انحناى توپوگرافیک^۲، انحناى کلی^۳، شاخص NDVI، کاربری اراضی، لیتولوژی و بارندگی در نظر گرفته شد و لایه GIS مربوط به این عوامل به وسیله روش‌های گفته شده در قسمت‌های قبل تهیه گردید که نتایج در اشکال ۲ نشان داده شده است.

۳-۲- آزمون هم خطی متغیرهای موثر

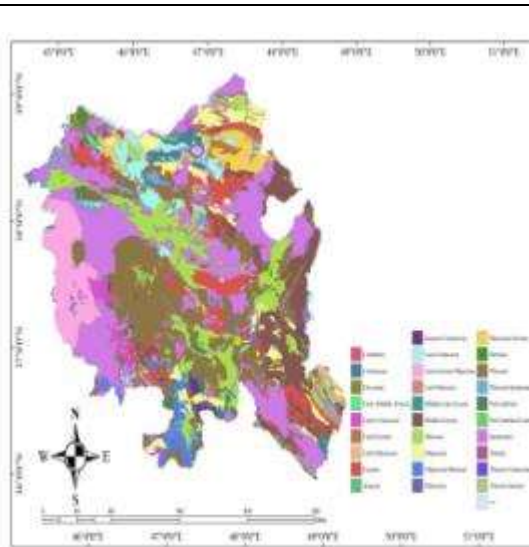
در این تحقیق برای تعیین هم خطی از پارامترهای Tolerance یا ضریب تحمل و VIF^۴ یا عامل تورم واریانس استفاده شده است. اگر مقدار ضریب تحمل کمتر از ۰/۱ و مقدار عامل تورم واریانس بزرگتر از ۱۰ باشد، بین پارامترها هم خطی وجود دارد. برای محاسبه آزمون هم خطی مقادیر عددی هر یک از پارامترها در نقاط سیل خیز استخراج گردید و در محیط SPSS آزمون هم خطی محاسبه گردید. نقاط سیل خیز مشخص شده در منطقه مورد مطالعه در شکل شماره ۳ آورده شده است. همچنین نتایج حاصل از آزمون هم خطی در جدول شماره (۲) ارائه شده است. بیشترین مقدار عامل تورم واریانس ۲/۲۴۸ و کمترین مقدار ضریب تحمل ۰/۴۴۵ میباشد، بنابراین هیچ هم خطی بین متغیرهای مستقل مشاهده نشد.

2	Topographic position index	9
3	Roughness	0
3	(Profile curvature	1
3	Plan curvature	2
3	Curvature	3
3	Variance Inflation Factor	4

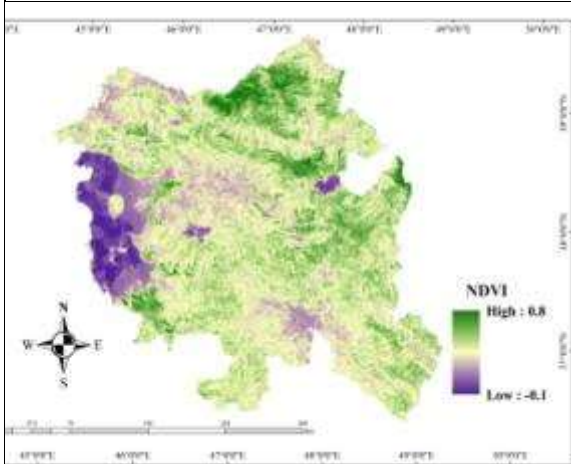




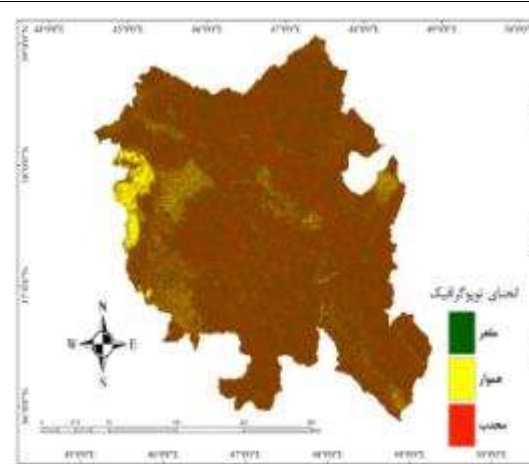
متغیر فاصله از آبراهه موثر بر سیلاب



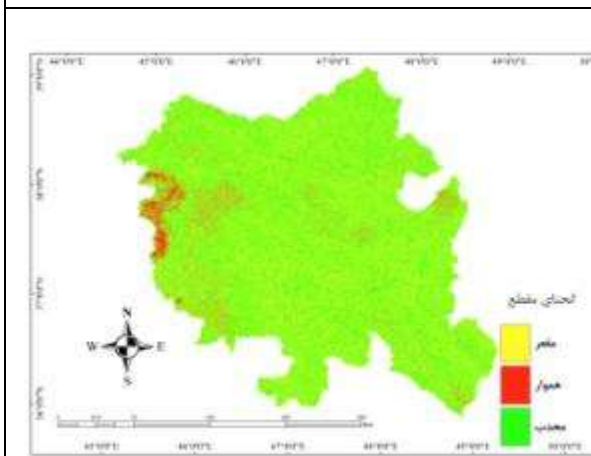
متغیر لیتولوژی موثر بر سیلاب



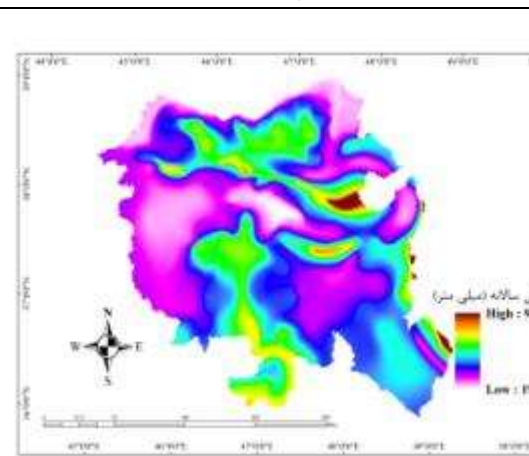
متغیر NDVI موثر بر سیلاب



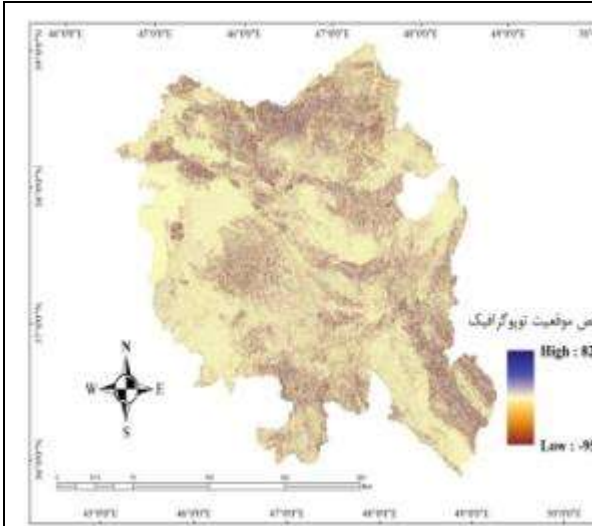
متغیر انحنای توپوگرافیک موثر بر سیلاب



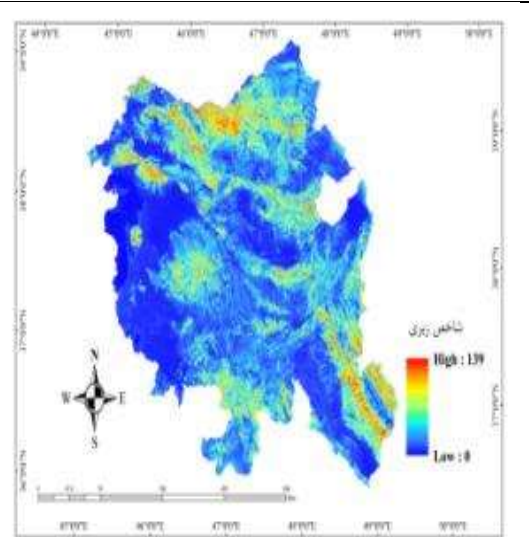
متغیر انحنای مقطع موثر بر سیلاب



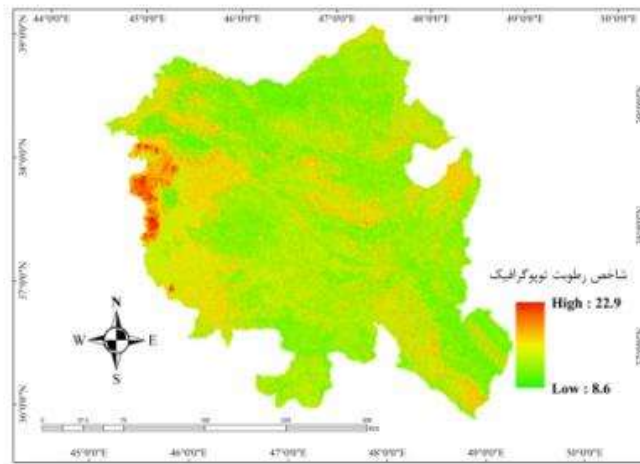
متغیر بارندگی موثر بر سیلاب



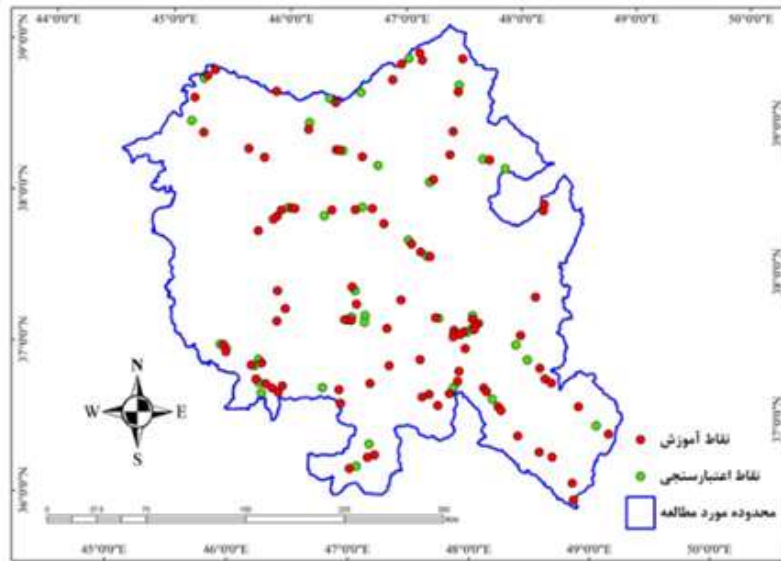
متغیر شاخص موقعیت توپوگرافیک موثر بر سیلاب



متغیر شاخص زبری موثر بر سیلاب



متغیر شاخص رطوبت توپوگرافیک موثر بر سیلاب



شکل (۳): نقاط سیل خیز منطقه مورد مطالعه

Fig. (3): Flood-prone points of the studied area

جدول (۲): نتایج آزمون هم خطی متغیرهای تاثیرگذار بر سیلاب

Table (2): The results of the collinearity test of variables affecting floods

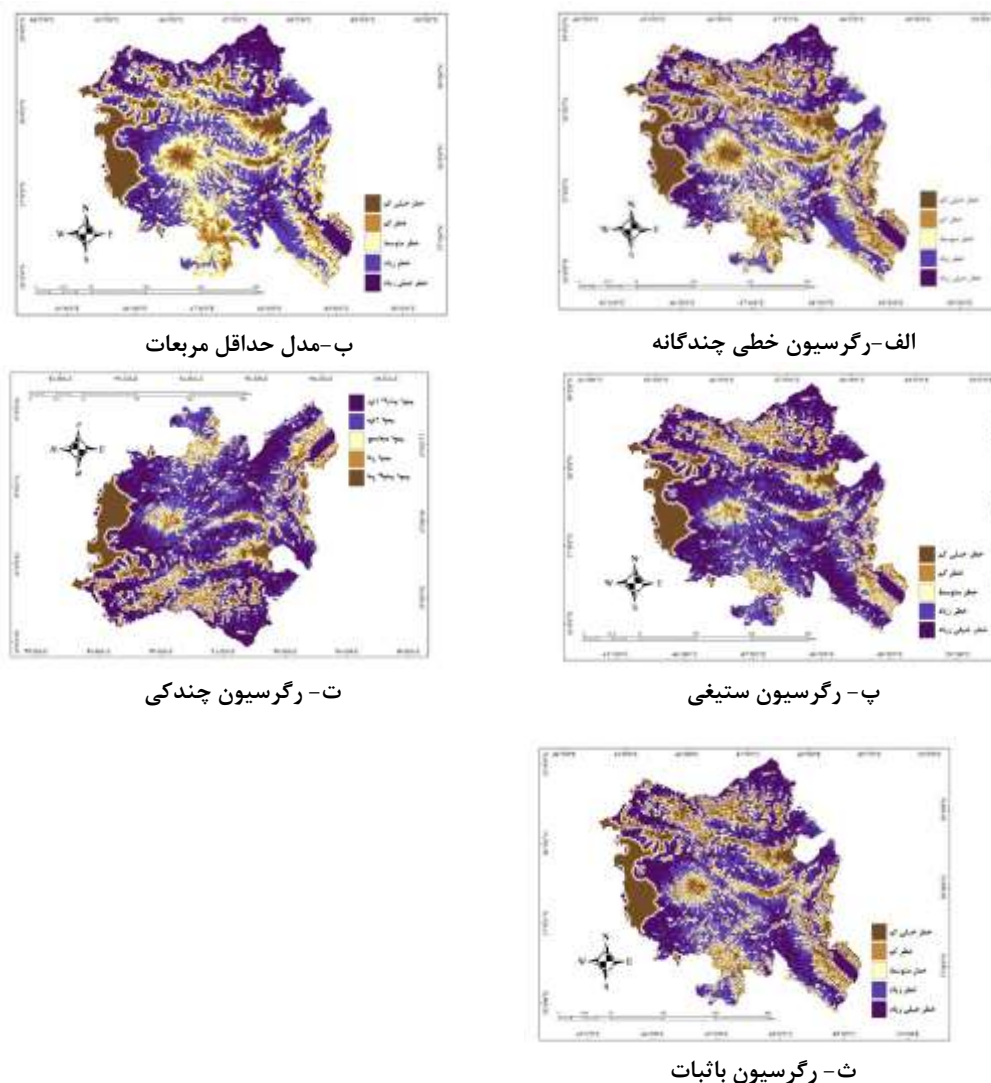
	Tolerance	VIF
انحنای کلی	۰/۹۹۹	۱/۰۰۱
زمین شناسی	۱/۰۰۰	۱/۰۰۰
کاربری اراضی	۰/۹۹۹۶	۱/۰۰۴
NDVI	۰/۹۹۹	۱/۰۰۱
انحنای توپوگرافیک	۰/۹۹۹	۱/۰۰۱
انحنای مقطع	۰/۹۹۳	۱/۰۰۷
بارندگی	۰/۴۴۵	۲/۲۴۸
فاصله از رودخانه	۰/۹۲۲	۱/۰۸۵
شیب	۰/۹۲۵	۱/۰۸۱
شاخص موقعیت توپوگرافیک	۰/۹۹۵	۱/۰۰۵
شاخص زبری سطح	۰/۹۲۶	۱/۰۸۰
شاخص رطوبت توپوگرافیک	۰/۹۶۱	۱/۰۴۰
ارتفاع	۰/۹۲۳	۱/۰۸۳

۳-۳- نقشه‌های خطر سیل خیزی

بعد از آماده سازی لایه‌ها و اطلاعات اولیه از ۷۰ درصد نقاط جهت آموزش مدلها استفاده شد و نقشه‌های خطر سیلاب تهیه گردید. شایان ذکر است که تمامی مراحل مدلسازی در نرم افزار R و با استفاده از بسته Caret انجام گرفت. پس از مدلسازی نقشه‌های رستری تولید شده وارد نرم افزار ArcMap شده و با استفاده از دستور

(Natural Breaks Jenks)

یا شکستهای طبیعی به ۵ کلاس خطر خیلی کم (۰ - ۰/۲۰)، خطر کم (۰/۲ - ۰/۴)، خطر متوسط (۰/۴ - ۰/۶)، خطر زیاد (۰/۶ - ۰/۸) و خطر خیلی بالا (۰/۸ - ۱) تقسیم‌بندی گردید که نتایج آن در شکل شماره ۴ ارائه شده است.



شکل (۴): نقشه‌های خطر سیل خیزی

Fig. (4): Flood risk maps

همچنین جدول (۳) میزان مساحت هر کلاس در مدل‌های استفاده شده را نشان می‌دهد.

عنوان خلاصه شده مقاله: تحلیل خطرپذیری سیلاب حوضه‌های آبریز با بهره‌گیری از الگوریتم‌های یادگیری ماشین و عامل‌های هیدروژئومورفولوژیکی (مطالعه موردی: حوضه شرقی آبریز دریاچه اورمیه)

سینا فرد مرادی نیا، یوسف زندی

۴۴

جدول ۳. مساحت (هکتار) و درصد مساحت طبقات مختلف در مدل‌های استفاده شده

Table (3): The area (hectares) and percentage of the area of different floors in the used models

خطر خیلی زیاد		خطر زیاد		خطر متوسط		خطر کم		خطر خیلی کم		
درصد مساحت	مساحت	درصد مساحت	مساحت	درصد مساحت	مساحت	درصد مساحت	مساحت	درصد مساحت	مساحت	
۱۵	۱۱۸۱۳۴۲	۳۰	۲۳۸۶۳۳۰	۲۸	۲۲۰۹۰۶۷	۱۷	۱۳۷۵۷۱۲	۸/۲	۶۴۱۷۸۳	MLR
۱۵	۱۱۹۴۱۷۰	۳۲	۲۵۳۸۸۲۹	۳۰	۲۳۵۵۶۳۰	۱۲	۹۹۲۱۸۹	۹	۷۱۳۷۶۲	PLS
۳۰	۲۳۷۹۱۴۵	۳۳	۲۵۸۸۴۳۸	۱۹	۱۵۳۵۸۶۹	۸	۶۸۹۹۰۱	۷	۶۰۰۹۷۵	QR
۳۰	۲۳۷۹۱۴۵	۳۳	۲۵۸۸۴۳۸	۱۹	۱۵۳۵۸۶۹	۸	۶۸۹۹۰۱	۷	۶۰۰۹۷۵	RR
۲۳	۱۸۲۶۳۳۵	۳۲	۲۵۰۲۰۳۶	۲۲	۱۷۶۴۷۷۳	۱۱	۹۱۱۵۳۳	۱۰	۷۸۹۳۷۶	RLR

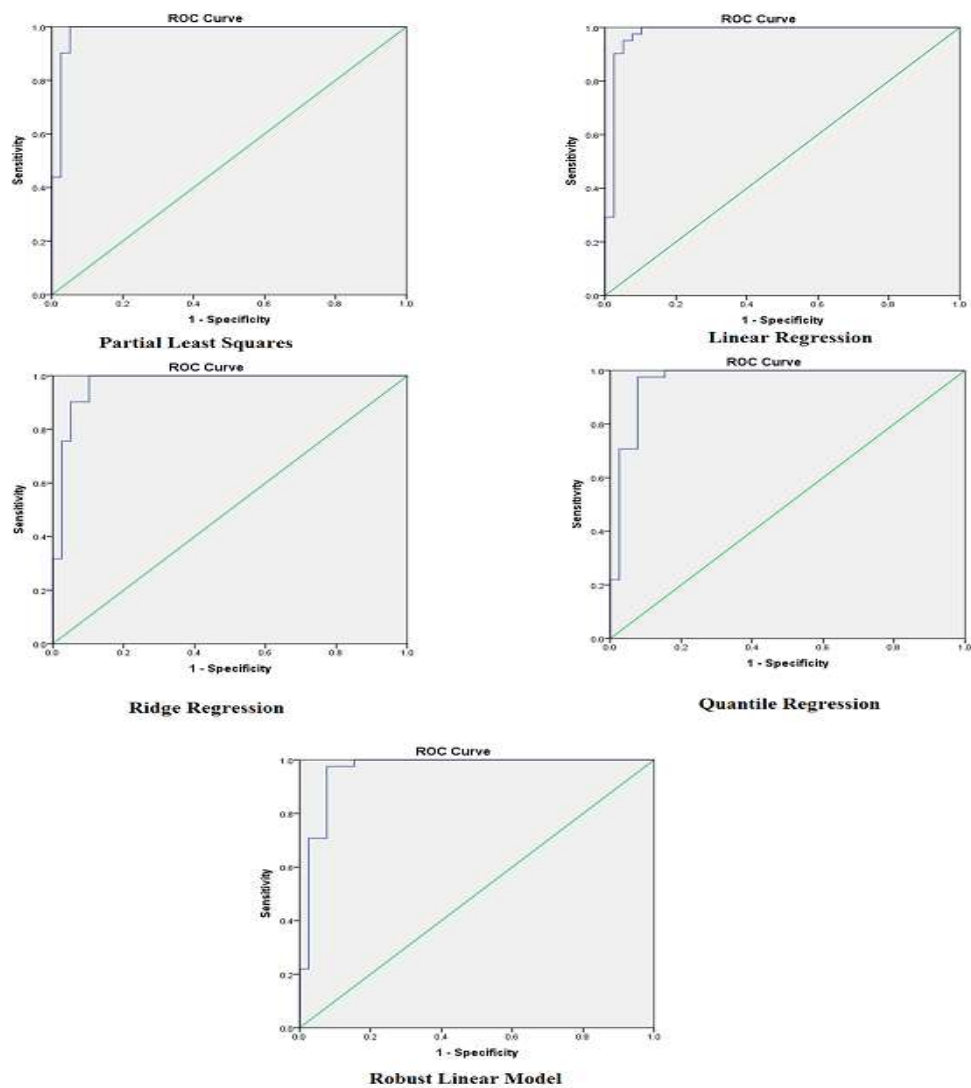
در جدول ۳، مدل رگرسیون خطی چندگانه (MLR)، مدل حداقل مربعات جزئی (PLS)، مدل رگرسیون چندکی (QR)، مدل رگرسیون ستیغی (RR) و مدل رگرسیون باثبات (RLR) می باشند. انطباق و هم‌پوشانی نقشه‌های خطر سیلاب با نقشه‌های شیب و مدل رقومی ارتفاع نشان می دهد که پهنه‌های دارای خطر زیاد و خیلی زیاد بیشتر منطبق بر اراضی پر شیب می باشد و پهنه‌های کم خطر منطبق بر حداقل ارتفاعات به علت نفوذپذیری خیلی زیاد میباشد. روی هم‌گذاری نقشه‌های خطر سیلاب و نقشه‌های پوشش گیاهی نشان می دهد که مناطق دارای خطر کم منطبق بر جاهایی است که دارای تراکم پوشش گیاهی بالا می باشد که این امر به دلیل نگهداری رواناب و سیلاب با استفاده از پوشش گیاهی می باشد که باعث افزایش میزان نفوذپذیری و تغذیه آبخوان‌ها می‌شود. از دیگر نتایج هم‌پوشانی نقشه‌های خطر با عوامل موثر می توان به تاثیر نزدیکی یا دوری از آبراهه‌ها اشاره کرد، که پهنه‌های دارای خطر بالا در فواصل نزدیک آبراهه‌ها قرار دارند. انطباق نقشه‌های خطر با نقشه انحنا توپوگرافیک و انحنا مقطع نشان می دهد که پهنه‌های دارای خطر کم و خیلی کم در پهنه‌های مقعر می باشد که به دلیل شرایط مناسب توپوگرافیک محدوده‌های مقعر می باشد. لازم به ذکر است که انحنا توپوگرافیک یا خم شدگی مسطح، یک شاخص ژئومورفیک است که نشان دهنده نوع خم شدگی یا توپوگرافی سطح زمین است. این شاخص نشان می دهد که یک نقطه روی سطح زمین به سمت چپ یا راست خم می شود. با استفاده از این شاخص، می توان برای تحلیل های ژئومورفیک، مانند

شناسایی محل های کندلک و سرریز، تحلیل پایداری توده های خاک و سنگ، و بررسی ارتباط بین خم شدگی سطح زمین و فرسایش، استفاده کرد.

۳-۴- ارزیابی مدل های استفاده شده

پس از تهیه لایه های خطر سیلاب حاصل از مدل های استفاده شده از ۳۰ درصد نقاط جهت اعتبارسنجی و ارزیابی استفاده شد. جهت این امر نقاط اعتبارسنجی را روی لایه های خطر انداخته و مقادیر ارزش هر لایه استخراج گردید و جهت تهیه منحنی ROC به نرم افزار SPSS منتقل شد که منحنی های ROC در شکل شماره ۵ ارائه شده است. همچنین مساحت زیرمنحنی (AUC) و سایر اطلاعات منحنی های ROC در جدول شماره ۴ آورده شده است.

همانطور که از جدول مشخص است هر ۵ مدل استفاده شده دارای عملکرد عالی در بررسی خطر سیلاب می باشند و انتخاب مدل بهتر قدری مشکل است، با این وجود مدل های PLS و MLR با دارا بودن حداکثر AUC و کمترین میزان انحراف معیار بهتر عمل کرده اند. در بین این دو مدل هم PLS دارای نتایج کمی بهتر نسبت به MLR می باشد.



شکل (۵): نمودار ROC مربوط به مدل‌های استفاده شده
Fig. (5): ROC diagram related to the used models

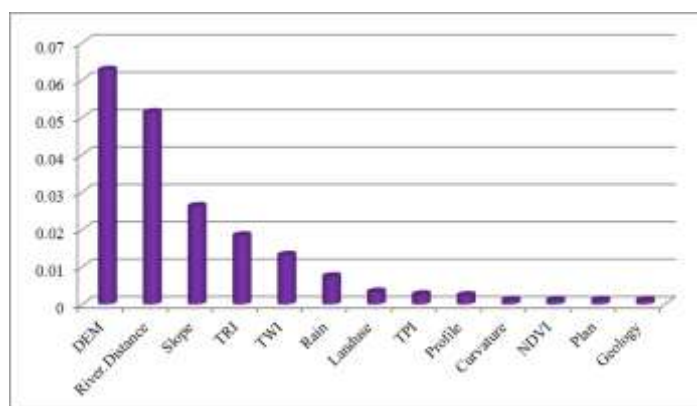
جدول (۴): سطح زیر منحنی ROC مدل‌های استفاده شده

Table (4): The area under the ROC curve of the used models

مدل	مساحت زیر منحنی (AUC)	انحراف معیار	سطح معنی داری	فاصله اطمینان ۹۵ درصد	
				حد پایین	حد بالا
MLR	۰,۹۹۷	۰,۰۱۸	۰,۰۰۰	۰/۹۴۱	۱,۰۰۰
PLS	۰,۹۸۳	۰,۰۱۵	۰,۰۰۰	۰/۹۵۴	۱,۰۰۰
QR	۰,۹۶۳	۰,۰۲۳	۰,۰۰۰	۰/۹۱۹	۱,۰۰۰
RR	۰,۹۷۱	۰,۰۱۹	۰,۰۰۰	۰/۹۳۴	۱,۰۰۰
RLR	۰,۹۶۳	۰,۰۲۳	۰,۰۰۰	۰/۹۱۹	۱,۰۰۰

۵-۳- تعیین میزان اهمیت متغیرهای استفاده شده

در ادامه میزان اهمیت هر کدام از عوامل تاثیرگذار استفاده شده با استفاده از مدل جنگل تصادفی مشخص گردید که نتایج در شکل ۶ ارائه گردیده است. همان طور که از شکل مشخص است معیارهای ارتفاع، فاصله از آبراهه و درصد شیب بالاترین تاثیر را روی سیل خیزی منطقه مورد مطالعه داشته اند. همچنین معیارهای عامل ژئولوژی، انحنا و توپوگرافیک و شاخص NDVI کمترین تاثیر را روی سیل خیزی حوضه مورد بررسی داشته اند.



شکل (۶): میزان اهمیت هر کدام از عوامل تاثیرگذار با استفاده از مدل جنگل تصادفی

Fig. (6): The importance of each influencing factor using the random forest model

پدیده سیل به صورت کنونی خود در کشور ما بیش از آنکه ناشی از وقوع بارشهای شدید باشد، از عوارض به هم خوردن تعادل طبیعی و شرایط جغرافیایی منطقه می باشد. به طوری که بروز بارشهای معمولی نیز موجب جاری شدن سیلاب می گردد. در این تحقیق تلاش گردید یک چارچوب جامع جهت ارزیابی خطر سیلاب ارائه گردد. جهت این امر ابتدا مهم ترین عامل‌های تاثیرگذار بر سیل خیزی از منابع مختلف گردآوری شد و لایه‌های اطلاعاتی مربوط به آنها با استفاده از روشهای مختلف در محیط ARCGIS تهیه گردید. سپس نقاط سیل خیز موجود در منطقه مورد مطالعه از نقشه پهنه‌های سیل خیز که با استفاده از سازمان جنگل‌ها، مراتع و آبخیزداری کشور تهیه شده بود، استخراج گردید و این نقاط با استفاده از تصاویر ماهواره ای گوگل ارث و بازدیدهای میدانی اصلاح گردید. در مرحله بعد جهت آماده‌سازی اطلاعات اولیه، آزمون هم‌خطی بر روی عامل‌های تاثیرگذار اعمال گردید و مشاهده گردید هیچ نوع هم‌خطی بین عامل‌های ورودی وجود ندارد.

۴- نتیجه گیری کلی

در این تحقیق از پنج روش مهم یادگیری ماشین شامل رگرسیون خطی چندگانه، مدل حداقل مربعات جزئی، رگرسیون چندکی، رگرسیون ستیغی و رگرسیون با ثبات جهت پیش بینی مکانی خطر سیلاب استفاده گردید. در ادامه از منحنی ROC جهت اعتبارسنجی نتایج استفاده گردید. نتایج نشان دهنده عملکرد عالی هر ۵ روش استفاده شده می باشد. در ادامه از مدل جنگل تصادفی جهت مشخص کردن اهمیت عامل‌های ورودی استفاده گردید، و مشخص گردید عامل‌های ارتفاع، فاصله از آبراهه و درصد شیب تاثیرگذارترین عوامل روی سیلاب منطقه مورد مطالعه می باشد. در نهایت به‌کارگیری تکنیک‌های یادگیری ماشین و سیستم اطلاعات جغرافیایی به منظور بررسی پتانسیل و استعداد سیل، مخصوصاً در کشورهای در حال توسعه که دسترسی به اطلاعات و داده‌های هیدروژئولوژیکی و اداکیکی (اطلاعات مربوط به ویژگی‌های خاک و ارتباط آن با عواملی مانند آب و هوا، دما، رطوبت و ... است) با مشکل و محدودیت مواجه است، پیشنهاد می شود. از آن جایی که سیل به عنوان یک پدیده فاجعه‌انگیز در کل جهان شناخته می شود، بنابراین تهیه نقشه حساسیت سیل به منظور مدیریت بهتر حوضه‌های آبخیز امری حیاتی و اجتناب ناپذیر می باشد.

ارزیابی حساسیت سیل توسط بسیاری از محققین انجام شده است. در بین این تحقیقات نیاز به یک روش جامع و قابل اعتماد جهت شناسایی مناطق مستعد سیل بیش از پیش احساس می شود. نتایج به دست آمده از این تحقیق با نتایج Darabi و همکاران (۲۰۱۴)، Siahkamari و همکاران (۲۰۱۸)، و نیز با نتایج Khosravi و همکاران (۲۰۱۸)، هم‌خوانی دارد. نتایج مدل PLS که دارای بالاترین سطح زیرنمودار ROC

بود، نشان داد که ۳۲ و ۱۵ درصد از مساحت منطقه مورد مطالعه دارای خطر سیلاب زیاد و خیلی زیاد می باشد. آگاهی از خصوصیات مورفومتریک یک حوضه، همراه با داشتن اطلاعات از شرایط آب و هوایی منطقه می تواند تصویر نسبتاً دقیقی از کارکرد کمی و کیفی سیستم هیدرولوژیک حوضه و هم چنین، وضعیت حاکم بر آن را از نقطه نظرهای مختلف به دست آورد. یافته‌های این تحقیق حاکی از آن است که دقت تکنیکهای یادگیری ماشین در شناسایی مناطق دارای استعداد سیل مناسب می باشد. بنابراین استفاده از این

ارزیابی خطر سیل، مفید و قابل اعتماد است که با نتایج مطالعات Tehrany و همکاران (۲۰۱۴) مطابقت دارد. هم چنین این نتیجه با نتایج Zhao و همکاران (۲۰۱۸)، و نیز Kanani-Sadat و همکاران (۲۰۱۹) مطابقت دارد. در این تحقیق، رویکردی ارائه شده است که با به کارگیری برخی متغیرهای موثر بر وقوع سیل، که این پارامترها برآیند عوامل مختلف محیطی و انسانی هستند، می توان به کاهش خسارت‌ها و تلفات ناشی از سیل کمک کرد. با توجه به نقشه‌های پیش‌بینی خطر سیل، اقدامات مدیریتی مناسبی می توان برای کاهش خسارت‌ها و تلفات ناشی از سیل انجام داد.

با توجه به نتایج حاصل از تحلیل خطرپذیری سیلاب حوضه‌های آبریز با استفاده از الگوریتم‌های یادگیری ماشین و عامل‌های هیدروژئومورفولوژیکی، پیشنهاد می‌شود که در ادامه کار، به موارد زیر توجه شود:

۱. تحلیل دقیق‌تر موانع و چالش‌های موجود در حوضه‌ی آبریز شرقی دریاچه ارومیه و تاثیر آن‌ها بر خطر سیلاب، به منظور بهبود دقت مدل‌سازی.
۲. استفاده از داده‌های بیشتر و مشاهدات واقعی مربوط به شرایط جغرافیایی و آب و هوایی حوضه‌ی آبریز برای بهبود دقت و قابلیت اعتماد مدل‌های یادگیری ماشین.
۳. مطالعه تاثیر تغییرات اقلیمی بر روی خطر سیلاب در حوضه‌ی آبریز شرقی دریاچه ارومیه و ارائه راهکارهایی برای مدیریت و کاهش خطر سیلاب در آینده.
۴. بررسی تاثیر تغییرات کاربری اراضی بر روی خطر سیلاب و ارائه راهکارهایی برای کنترل و کاهش اثرات آن.
۵. استفاده از روش‌های دقیق‌تر و پیشرفته‌تر برای تحلیل داده‌ها و اعمال مدل‌های پیش‌بینی جهت تعیین خطر سیلاب در حوضه‌های آبریز.

۶. بررسی تأثیر تراکم جمعیت و میزان توسعه شهری بر خطر سیلاب در حوضه‌های آبریز و ارائه راهکارهایی برای کاهش اثرات آن.

۷. مقایسه‌ی نتایج حاصل از مدل‌های یادگیری ماشین با نتایج حاصل از روش‌های سنتی مانند مدل‌های آماری و تحلیلی

۸. مقایسه‌ی نتایج حاصل از مدل‌های یادگیری ماشین با نتایج حاصل از روش‌های سنتی مانند مدل‌های آماری و تحلیلی

۶- محدودیت‌های تحقیق

در مورد محدودیت‌هایی که در تحقیق مورد نظر می‌تواند وجود داشته باشد، می‌توان به موارد زیر اشاره کرد:
- داده‌های کم: تعداد داده‌های موجود در مورد خطرپذیری سیلاب در حوضه‌ی مورد مطالعه، ممکن است کم باشد و به همین دلیل نمی‌تواند تمامی جنبه‌های خطر را پوشش دهد.

- نویز در داده‌ها: داده‌هایی که برای تحلیل استفاده می‌شوند، ممکن است دارای نویز باشند و این موضوع می‌تواند به دقت نتایج تأثیر بگذارد.

- کیفیت داده‌ها: کیفیت داده‌های جمع‌آوری شده، به طور مستقیم بر دقت نتایج و پایداری مدل‌های پیشنهادی تأثیر می‌گذارد.

- تحلیل پویایی: تحلیل خطرپذیری سیلاب، به طور کلی یک پروسه پویایی است و ممکن است تحلیل یکباره و ثابت در زمان، نتایج دقیقی ارائه ندهد.

- انتقال پذیری: ممکن است مدل‌های پیشنهادی برای تحلیل خطرپذیری سیلاب حوضه‌ی مورد مطالعه، قابلیت انتقال پذیری به حوضه‌های دیگر را نداشته باشند.

- پوشش جغرافیایی: ممکن است بخشی از حوضه‌ی آبریز، برای جمع‌آوری داده‌ها یا انجام تحلیل، در دسترس نباشد و این موضوع ممکن است باعث شود که داده‌های تحلیلی کامل نباشد.

مراجع:

- Al-Abadi, A. M., Al-Temmeme, A. A., & Al-Ghanimy, M. A. 2016. A GIS-based combining of frequency ratio and index of entropy approaches for mapping groundwater availability zones at Badra–Al Al-Gharbi–Teeb areas, Iraq. *Sustainable Water Resources Management*, 2, 265-283.
- Amiri, F., & Tabatabaie, F. 2020. Integration of GIS, remote sensing and Multi-Criteria Evaluation tools in the search for healthy walking paths. *Journal of Civil Engineering*, 22(1), 279-291. doi:<https://doi.org/10.1007/s11104-009-0053-7>. (In Persian).
- Assaf, A. G., M. Tsionas, & Tasiopoulos, A. 2019. Diagnosing and correcting the effects of multicollinearity: Bayesian implications of ridge regression. *Tourism Management*, 71, 1-8.
- Avand, M.T., Moradi, H.R., & Ramazanzadeh, M. 2020. Preparation of flood sensitivity map using two random forest machine learning models and Bayesian generalized linear model. *Environment and Water Engineering*, 6(1), 83-95. DOI: 10.22094/joeee.2020.108812.1791.
- Bakhtiari, B., Pourghasemi, H., R., Khorshiddoost, N., Fakheri Fard, A., Mosavi, A. 2017. Flood hazard zoning using process-based and stochastic modeling approaches. *Natural Hazards*, 2, 1159-1179.
- Balan, B., Mohaghegh, S., & Ameri, S. 1995. State-of-the-art in permeability determination from well log data: Part 1-A comparative study, model development. In *SPE Eastern Regional Meeting*. OnePetro.
- Bedini, M., Piroozi, A., Aghayari, L., & Ostadi, A. 2018. Flood risk zoning in Meshkinshahr city using Vikor model. *Geography of the Land*, 14. OPA: 0.22059/JGEL.2018.250727.1007473.
- Buitendag, S., Beirlant, J. & de Wet, T. 2019. Ridge regression estimators for the extreme value index. *Extremes*, 22(2), 271-292. doi:10.1007/s10687-018-0320-8.
- Constantin, M., Bednarik, M., Jurchescu, M. C., & Vlaicu, M. 2011. Landslide susceptibility assessment using the bivariate statistical analysis and the index of entropy in the Sibiciu Basin (Romania). *Environmental earth sciences*, 63, 397-406.

- Darabi, H., Shahedi, K., Solaimani, K., & Miryaghoubzadeh, M. 2014. Prioritization of subwatersheds based on flooding conditions using hydrological model, multivariate analysis and remote sensing technique. *Water and environment journal*, 28(3), 382-392.
- Ebrahimipour, M., & Ziari, K. 2019. Zoning of urban lands against flood risk with physical resilience approach (Case study: Cheshmeh Kileh river). *New Attitudes in Human Geography (Human Geography)*, 11(1), 83-104.
- Esmaili, H., Akhond Ali, A. M., Zarei, H., & Taghian, M. 2018. Regional Flood Analysis Via Comparison of The M5 Decision Tree Algorithm and Regression Models. *Irrigation Sciences and Engineering*, 40(4), 183-195.
- Faramarzi, H., Hosseini, S. M., Pourghasemi, H. R., & Farneghi, M. 2019. Assessment and Zoning of Flood Risk in Golestan National Park. *Iranian journal of Ecohydrology*, 6(4), 1055-1068.
- Garthwaite, P.H., Jolliffe, I.T. & Jones, B. 1995. *Statistical Inference*. Prentice-Hall.
- Ghanavati, E., Karam, A., & Aghaalikhani, M. 2013. Flood risk zonation in the farahzad basin (Tehran) using Fuzzy model. *Geography and Environmental Planning*, 23(4), 121-138.
- Goodarzi, M., & Fatehifar, A. 2019. Flood risk zoning due to climate change under RCP 8.5 scenario using hydrologic model SWAT in Gis (Azarshahr basin). *Journal of Applied Researches in Geographical Sciences*, 19(53), 99-117.
- Guo, L., Liu, R., Men, C., Wang, Q., Miao, Y., & Zhang, Y. 2019. Quantifying and simulating landscape composition and pattern impacts on land surface temperature: A decadal study of the rapidly urbanizing city of Beijing, China. *Science of The Total Environment*, 654, 430-440. doi:<https://doi.org/10.1007/s11104-009-0053-7>.
- Hejazi, A., & Khodayee Geshlag, F. 2020. Flood risk zoning in the Varkash Chay catchment using the HEC-RAS model and the HEC-GEO-RAS supplement. *Applied Research in Geographical Sciences (Geographical Sciences)*, 19(53), 75-89.
- Jothibasu, A., & Anbazhagan, S. 2016. Modeling groundwater probability index in Ponnaiyar River basin of South India using analytic hierarchy process. *Modeling Earth Systems and Environment*, 2, 1-14.
- Kanani-Sadat, Y., Arabsheibani, R., Karimipour, F., & Nasserli, 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., 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.

Koenker, R. 2005. *Quantile Regression*. Economic Society Monographs No 38, ed. by Chesher A and Jackson M.

Lin, M., Song, X., Qian, Q., Li, H., Sun, L., Zhu, S., & Jin, R. 2019. Robust gaussian process regression for real-time high precision GPS signal enhancement. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining* (pp. 2838-2847)

Liu, A., G. Shi, S. J. Chung, A. Anandkumar, & Y. Yue. 2019. Robust regression for safe exploration in control. *arXiv preprint arXiv:1906.05819*.

Malazehi, A., M. Pudineh, M. Khosravi, M. Armesh, & A. Dehvari. 2020. Flood risk potential assessment in Sarbaz catchment. *Applied Research in Geographical Sciences (Geographical Sciences)*, 20 (58).

Moore I.D., R.B. Grayson, A.R. Ladson. 1991. Digital terrain modelling: a review of hydrological, geomorphological, and biological applications. *Hydrological processes*, 5(1), 3-30.

Naghbi, S. A., Pourghasemi, H. R., Pourtaghi, Z. S., & Rezaei, A. 2015. Groundwater qanat potential mapping using frequency ratio and Shannon's entropy models in the Moghan watershed, Iran. *Earth Science Informatics*, 8, 171-186.

Nouri, H., Ildoromi, A., Sepehri, M., & Artimani, M. 2019. Comparing Three Main Methods of Artificial Intelligence in Flood Estimation in Yalphan Catchment. *Geography and Environmental Planning*, 29(4), 35-50.

Rahimi, L., Saboori, S., & Bordbar, H. 2019. Investigating the effect of flood risk perception on preventive behaviors through the attachment to place component (Case Study: Babol city). *Geography and Development*, 17(57), 49-68.

Rajabi, M., Roostaei, S., Barzkar, M. 2022. Evaluation of flood potential under basins based on morphometric parameters and correlation test(Case: Zab catchment to Mirabad), *Journal of Geography and Planning*, 26(79), 127-139. magiran.com/p2431487

Rezvani Faezifar, R., Safari, A., Bahroudi, A., & Ramouz, S. 2022. Feasibility study of forecasting the flood occurrence using GRACE satellite gravity data in the Karun river basin, Iran. *Iranian Journal of Geophysics*, 16(1), 103-117.

Rostami Khalaj, M., Hesami, D., Salmani, H., & Tymoriyan, T. 2020. Urban Flood Hazard Zoning Using Multicriteria Decision Analysis (Emam Ali town, Mashhad city). *Journal of Environmental Science and Technology*, 21(11), 173-185.

Salehi, A., Rafiei, Y., Farzad Behtash, M. & Aghababaei, M. 2014. Urban flood risk zoning using GIS and fuzzy hierarchical analysis process (Case study: Tehran). *Environmental Science*, 39(3).

Sattari, M., Abdollah Pourazad, M. & Mir Abbasi Najafabadi, R. 2017. Technical report: Predicting hourly floods of Ahrachai river using machine learning methods. *Watershed Engineering and Management*, 8(1).

Sharifi Garmadreh, E., Vafakhah, M., Eslamian, S. S. 2020. Evaluation of the efficiency of support vector machine systems and artificial neural network in flood zone analysis (Case study: Salt Lake watershed). *Journal of Soil and Water Sciences*, 23(1), 351-366.

Siahkamari, S., Haghizadeh, A., Zeinivand, H., Tahmasebipour, N., & Rahmati, O. 2018. Spatial prediction of flood-susceptible areas using frequency ratio and maximum entropy models. *Geocarto international*, 33(9), 927-941.

Tehrany, M. S., B. Pradhan, S. Mansor, & N. Ahmad. 2015. Flood susceptibility assessment using GIS-based support vector machine model with different kernel types. *Catena*, 125, 91-101.

Yesilnacar, E.K. 2005. The application of computational intelligence to landslide susceptibility mapping



Flood risk Analysis of Watersheds Using Machine Learning Algorithms and Hydrogeomorphological Factors (Case study: Eastern Catchment Area of Lake Urmia)

Abstract

Flood zoning maps provide valuable information about the nature of flood and its effect on floodplain lands. A set of effective factors must be defined to map flood susceptibility or, in general, to develop a model for assessing natural disaster risk. The factors affecting flood were used in eastern catchment area of lake Urmia which include altitude, slope, distance from the river, topographic moisture index, topographic position index, roughness index, curvature level, topographic curvature section, total curvature, NDVI index, land use, lithology and rainfall according to the experiences of experts and researchers reported in previous studies. After preparing the effective layers on the flood and the point layer of the flood points, as well as performing the linear test, five methods including Multiple Linear Regression, Partial Least Squares Model, Quantile Regression, Ridge Regression and Robust Regression were used for modeling and predictions. Then the ROC curve was used to validate the results. The results of this validation showed that the partial least squares (PLS) and multiple linear regression (MLR) models with the maximum area under the curve (AUC) (0.983 and 0.997, respectively) and the lowest standard deviation (0.015 and 0.018, respectively)) have performed better. Among these two models, PLS has slightly better results than MLR. Finally, a random forest model was used to determine the importance of the input factors, and it was found that the factors of height, distance from the waterway and slope percentage are the most influential factors on floods in the study area.

Key words: Flood Zoning - Machine Learning Algorithms - Hydrogeomorphological Factors - Regression Method- Easten Catchment Urmia Lake