

تهیه نقشه حساسیت سیل با استفاده از مدل های یادگیری ماشین جنگل تصادفی و خطی تعمیم یافته بیزین

محمدتقی آوند، حمیدرضا مرادی و مهدی رمضانزاده لیسوئی

دوره ۶، شماره ۱، بهار ۱۳۹۹، صفحات ۸۳-۹۵

Vol. 6(1), Spring 2020, 83 – 95

DOI: 10.22034/jewe.2020.220593.1351

**Flood Susceptibility Mapping Using Random
Forest Machine Learning and Generalized
Bayesian Linear Model**

Avand M., Moradi H. R. and Ramazanzadeh M.



www.jewe.ir

OPEN ACCESS

ارجاع به این مقاله:

آوند م.، مرادی ح. ر. و رمضانزاده لیسوئی م. (۱۳۹۹). تهیه نقشه حساسیت سیل با استفاده از مدل های یادگیری ماشین جنگل تصادفی و خطی تعمیم یافته بیزین. محیط زیست و مهندسی آب، دوره ۶، شماره ۱، صفحات: ۸۳-۹۵.

Citing this paper: Avand M., Moradi H. R. and Ramazanzadeh M. (2020). Flood susceptibility mapping using random forest machine learning and generalized Bayesian linear model. Environ. Water Eng., 6(1), 83-95. DOI: 10.22034/jewe.2020.220593.1351.

تهیه نقشه حساسیت سیل با استفاده از مدل‌های یادگیری ماشین جنگل تصادفی و خطی تعمیم‌یافته بیزین

محمد تقی آوند^۱، حمیدرضا مرادی^{۲*} و مهدی رمضان‌زاده لسبویی^۳

^۱دانشجوی دکتری علوم و مهندسی آبخیزداری، دانشکده منابع طبیعی و علوم دریایی، دانشگاه تربیت مدرس. تهران. ایران
^۲دانشیار گروه علوم و مهندسی آبخیزداری، دانشکده منابع طبیعی و علوم دریایی، دانشگاه تربیت مدرس. تهران. ایران
^۳استادیار گروه مدیریت جهانگردی، دانشکده علوم انسانی و اجتماعی، دانشگاه مازندران. بابلسر. ایران

*نویسنده مسئول: hmradi@modares.ac.ir

یادداشت فنی

تاریخ پذیرش: [۱۳۹۹/۰۱/۳۱]

تاریخ بازنگری: [۱۳۹۹/۰۱/۲۵]

تاریخ دریافت: [۱۳۹۸/۱۲/۰۳]

چکیده

امروزه پدیده‌ی سیل یکی از پیچیده‌ترین رخدادهای مخاطره‌آمیز است که بیش از سایر بلاهای طبیعی دیگر، همه‌ساله در نقاط مختلف دنیا منجر به ایجاد خسارت‌های جانی و مالی و تخریب اراضی کشاورزی می‌شود؛ بنابراین تهیه نقشه حساسیت به وقوع سیلاب نخستین گام در برنامه مدیریت سیلاب است. هدف از این پژوهش شناسایی مناطق حساس به سیل با استفاده از دو مدل یادگیری ماشین جنگل تصادفی (RF) و خطی تعمیم‌یافته بیزین (GLMbayesian) در حوزه آبخیز تجن در استان مازندران، شهرستان ساری بود. نقشه پراکنش سیلاب‌های گذشته به‌منظور پیش‌بینی سیلاب در آینده تهیه شد. از بین ۲۶۳ رخدادهای سیلاب، ۸۰٪ (۲۱۰ رخداد سیل) به‌منظور مدل‌سازی و ۲۰٪ (۵۳ رخداد سیل) به‌منظور اعتبارسنجی استفاده شد. با بررسی مطالعات قبلی و پیمایش منطقه مورد مطالعه ۱۳ عامل مؤثر به‌منظور پهنه‌بندی سیلاب انتخاب و تهیه شد. نتایج نشان داد که سه فاکتور ارتفاع (۲۱/۵۵)، فاصله از رودخانه (۱۵/۲۸) و شیب (۱۱/۱۸) به ترتیب بیش‌ترین تأثیر را در سیل‌گیری منطقه مورد مطالعه دارند. همچنین نتایج ارزیابی خروجی مدل‌ها نشان داد که مقدار AUC در مدل RF و GLMbayesian به ترتیب ۰/۹۱ و ۰/۸۴۷ بود که نشان‌دهنده برتری مدل RF و دقت بیش‌تر این مدل در تهیه نقشه حساسیت به وقوع سیل در منطقه مورد مطالعه می‌باشد. بیش‌ترین مساحت حساسیت به سیل در مدل RF مربوط به طبقه خیلی کم و در مدل GLMbayesian مربوط به طبقه زیاد است.

واژه‌های کلیدی: حوزه آبخیز تجن؛ خسارت سیلاب؛ مدل GLMbayesian؛ مدل RF؛ منحنی ROC.

۱- مقدمه

همه‌روزه در سراسر جهان وقایع زیادی رخ می‌دهد که تعدادی از آن‌ها تبدیل به بلاهای طبیعی می‌شوند. بلاهای طبیعی در واقع یک فاجعه یا مصیبت ناشی از وقوع یک پدیده طبیعی پرخطر مانند سیلاب، خشک‌سالی، زلزله، رانش زمین، طوفان و یا آتش‌فشان است که باعث وارد آمدن خسارت مالی و جانی فراوان به جوامع انسانی می‌گردد. در مناطقی که ارتباط مستقیم با انسان ندارند و منافع انسانی در میان نیست این وقایع تبدیل به بلاهای طبیعی نمی‌شوند (Fernandez and Lutez 2010). همواره طغیان آب و سرازیر شدن آن به اماکن مسکونی، از جمله حوادث طبیعی شایع در ایران پس از زلزله است که حیات بشر را به مخاطره می‌اندازد، سیل‌ها در مقیاس جهانی ویرانگرترین فاجعه و مؤثرترین عامل در تعداد مرگ‌ومیر و خسارات هستند (Darabi et al. 2016).

پدیده سیل به صورت کنونی خود در کشور ما بیش از آن که ناشی از وقوع بارش‌های شدید باشد، از عوارض به هم خوردن تعادل طبیعی و شرایط جغرافیایی منطقه می‌باشد. به طوری که بروز بارش‌های معمولی نیز موجب جاری شدن سیلاب می‌گردد (Khairizadeh et al. 2014). یکی از روش‌های مهم در کنترل و کاهش رواناب سطحی جهت به تعویق انداختن سیل، عملیات آبخیزداری می‌باشد. یکی دیگر از روش‌های کاهش خطرات سیل، پهنه‌بندی سیل می‌باشد که اطلاعات ارزشمندی را در رابطه با طبیعت، سیلاب‌ها و اثرات آن بر اراضی دشت سیلابی و تعیین حریم رودخانه‌ها ارائه می‌دهند. در نتیجه امکان ارسال هشدارهای مناسب در مواقع خطر سیل و تسهیل عملیات امداد و نجات را فراهم می‌سازد (Khairizadeh et al. 2013).

پژوهش حاضر با هدف پهنه‌بندی حساسیت به سیل در حوزه آبخیز تجن در استان مازندران انجام گرفت. تهیه نقشه حساسیت به سیل از جمله روش‌های جدیدی می‌باشد که در مهر و موم‌های اخیر مورد توجه محققان قرار گرفته است. به دلیل وسعت زیاد منطقه مورد مطالعه، کمبود داده و در دسترس نبودن بسیاری از مناطق این حوزه آبخیز امکان استفاده از مدل‌های هیدرولوژیکی از قبیل HEC-RAS وجود نداشته و یا نیازمند وقت هزینه بسیار زیادی می‌باشد؛ بنابراین در این پژوهش با استفاده از دو روش داده‌کاوی شامل مدل RF و GLMbayes در محیط نرم‌افزار R، نقشه مناطق حساس به سیل در حوزه آبخیز تجن تهیه شده است. در این مطالعه برای اولین بار از مدل GLMbayes به همراه مدل RF به منظور تهیه نقشه حساسیت سیل استفاده شده است.

علیرغم تلاش‌های متخصصان، تصمیم‌گیران، ذینفعان و ادارات دولتی در دهه‌های اخیر برای کاهش اثرات سیلاب، تعداد حوادث و تلفات اقتصادی و انسانی مرتبط با آن، در سراسر جهان در حال افزایش است. این پدیده نه تنها در کشورهای در حال توسعه، بلکه در تمام جهان شایع‌ترین مخاطره طبیعی است (Hazarika et al. 2018). مسئله تهیه نقشه حساسیت به وقوع سیل توسط محققان مختلف بررسی شده است (Motevalli and Vafakhah 2016).

در میان تکنیک‌های پیشرفته‌تری در ارزیابی نقشه‌های حساسیت سیلاب توسط روش‌های محاسبات نرم و داده‌کاوی اجرا شده است. به طور کلی، به دلیل این‌که حوضه‌ها به طور ذاتی پیچیده هستند، مدل‌سازی آن‌ها با روش‌های هیدرولوژی ساده و خطی مطابقت ندارد. به همین دلیل تکنیک‌های مختلفی برای بررسی سیل که رفتار چندبعدی دارد مورد استفاده قرار می‌گیرد. در میان تکنیک‌های مختلف، استفاده از ArcGIS (Ghasemi et al. 2015)، آنالیز سلسله مراتبی (AHP) (Luu and Meding 2018)، نسبت فراوانی (FR) (Rahmati et al. 2016)، رگرسیون لجستیک منطق فازی ((Al-Juaidi et al. 2018)، جنگل تصادفی (RF) (Lee et al. 2017; Janizadeh et al. 2019)، شبکه عصبی مصنوعی (ANN) (Mukerji et al. 2009) و ماشین بردار پشتیبان (SVM) (Tehrany et al. 2014) مورد علاقه محققان می‌باشند.



شکل ۲. بخش‌هایی از مناطق سیل گیر موجود در حوزه آبخیز

تجن: الف- منطقه کردخیل و ب- ورودی پارک ملل

Fig. 2 Parts of the floodplain areas in the Tajan watershed A- Kordkheil area and b- Entrance to Mellal Park

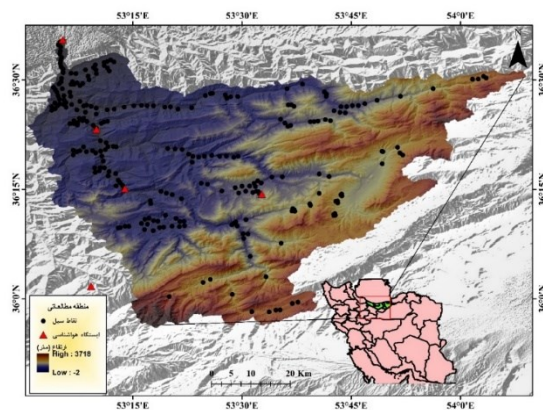
۲-۲- نقشه نقاط سیلاب‌های گذشته

نقاط سیل یک سطح مهم از رابطه بین رخداد سیل و عوامل به وجود آورنده آن می‌باشد. رویدادهای سیل تاریخی به‌عنوان مبنایی برای پیش‌بینی وقوع سیلاب در آینده به‌حساب می‌آیند. به‌طوری‌که مناطق نزدیک به رخدادهای گذشته حساسیت بالایی به سیل‌گیری دارند. حوزه آبخیز تجن از جمله حوضه‌های دارای پتانسیل سیل بالا می‌باشد. بدین ترتیب، ۲۶۳ نقطه سیل در منطقه مورد مطالعه، توسط سازمان آب منطقه‌ای استان مازندران ثبت شده است، از این تعداد ۸۰٪ آن برای آموزش مدل و ۲۰٪ برای اعتبارسنجی مدل استفاده شد. ۲۶۳ نقطه

۲- مواد و روش‌ها

۲-۱- منطقه مورد مطالعه

منطقه مورد مطالعه حوزه آبخیز تجن با وسعت 4000 Km^2 است که در استان مازندران قرار دارد و توسط کوه‌های البرز در جنوب و دریای خزر در شمال احاطه شده است. از لحاظ جغرافیایی منطقه مورد مطالعه بین طول جغرافیایی $18' 53^\circ$ تا $05' 53^\circ$ و عرض جغرافیایی $29' 36^\circ$ تا $09' 36^\circ$ قرار دارد. میانگین سالانه دما حدود 15°C با آب‌وهوای معتدل و مرطوب و بارش متوسط سالانه 700 mm است (Rajaei et al. 2015). توپوگرافی منطقه از ارتفاع 2 m - در خروجی حوزه آبخیز تا 3718 m متغیر می‌باشد (شکل ۱). این منطقه از ۵ دهستان و ۷۷ آبادی تشکیل شده است که این مناطق اکثراً تپ کوهستانی دارند. کاربری‌های غالب موجود در این منطقه شامل: جنگل، مرتع، کشاورزی، باغ و مسکونی می‌باشد. اغلب مساحت این حوزه آبخیز در اراضی جنوبی و چسبیده به مرز استان سمنان مرتعی می‌باشد و در میانه این حوزه آبخیز مساحت اراضی جنگلی بیشتر می‌شود و در مناطق پایین‌دست و نزدیک به خروجی حوزه آبخیز، سطح اراضی مسکونی و کشاورزی افزایش می‌یابد. مهم‌ترین شاخه‌های اصلی رودخانه تجن شاخه‌های چهاردانگه، دودانگه، لاجیم و زارمروید می‌باشد.



شکل ۱- موقعیت حوزه آبخیز تجن در استان مازندران، ایران
Fig. 1 Location of Tajan watershed in Mazandaran Province, Iran

به‌منظور صحت‌سنجی نقاط سیل موجود در منطقه مورد مطالعه و همچنین تعیین کاربری اراضی این منطقه اقدام به مطالعه صحرایی حوزه آبخیز مورد مطالعه شد (شکل ۲).

غیرسیل نیز با استفاده از بازدیدهای میدانی، نقشه توپوگرافی و نرم‌افزار Google Earth با توجه به مناطقی مانند تپه‌ها و کوه‌ها که سیلاب قادر به پیشروی در آنجا

جدول ۱- مشخصات داده‌های ورودی برای مدل‌سازی مناطق حساس به سیل

Table 1- Specifications of input data for modeling flood susceptibility areas

ردیف	داده‌های ورودی اولیه	فرمت اصلی داده‌ها	دقت مکانی	نقشه‌های مشتق شده
۱	مدل رقومی ارتفاع	رستر	12.5 m	شیب، جهت، TRI، TPI، ارتفاع، فاصله از رودخانه
۲	لندست ۸ سنجنده OLI	رستر	30 m	نقشه کاربری اراضی
۳	داده‌های هواشناسی	داده عددی	-	نقشه بارش
۴	نقشه زمین‌شناسی	وکتور	1:100000	نقشه سنگ‌شناسی
۵	نقشه خاک	وکتور	1:100000	نقشه خاک

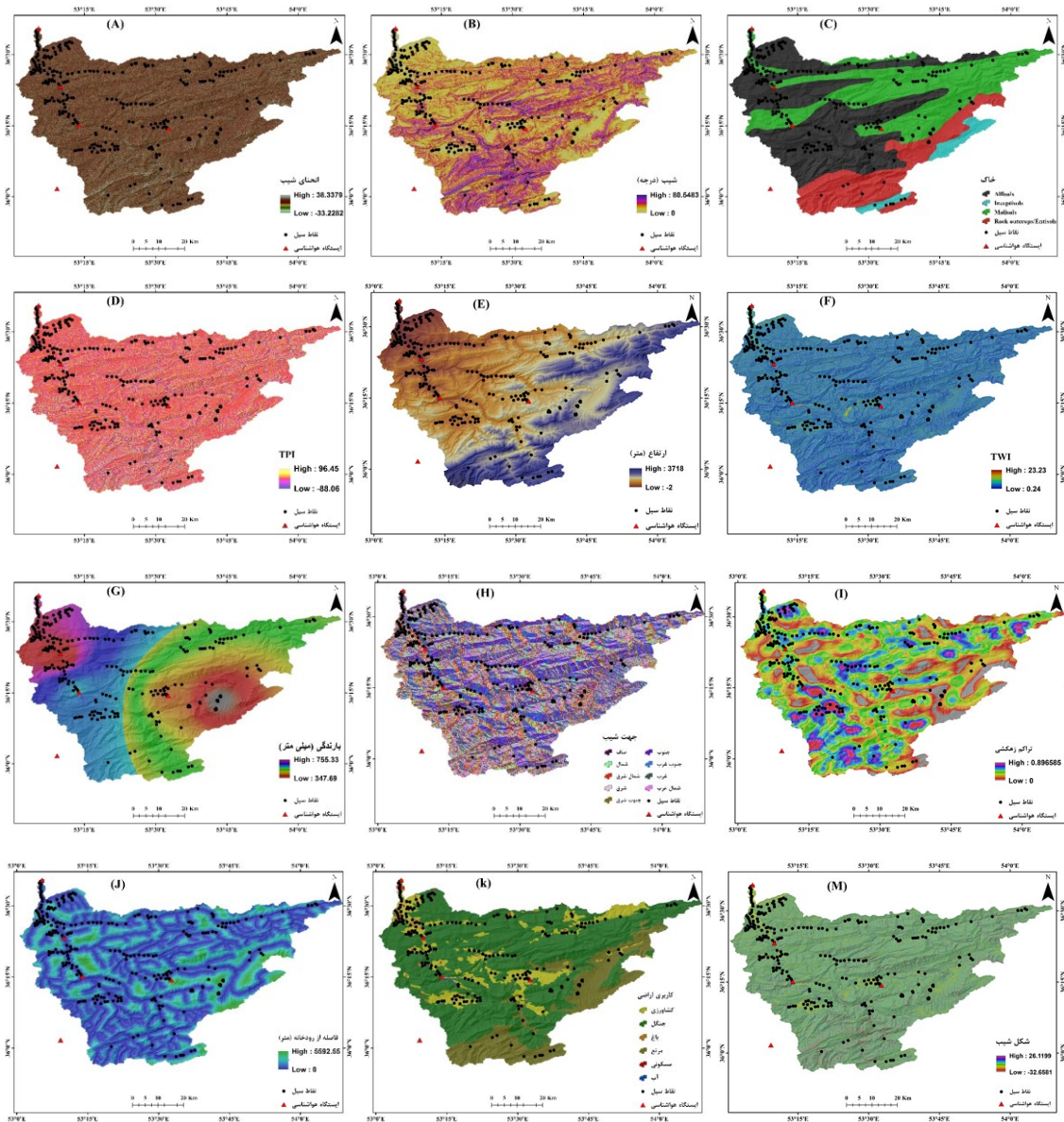
داده و سبب نفوذ بیشتر آب به داخل خاک می‌گردد، در نتیجه تأثیر قابل‌توجهی بر کاهش سیل‌های مخرب خواهد داشت. کاربری اراضی نتیجه روابط متقابل پارامترهای اجتماعی-فرهنگی و توان بالقوه سرزمین است. تغییرات در کاربری اراضی نتایج چشم‌گیری در پتانسیل سیل‌خیزی حوزه‌های آبخیز دارد (شکل ۳K Tehrani et al. 2015). شبکه‌ی رودخانه‌ها به مجموع آبراهه‌هایی گفته می‌شود که در سطح حوضه عمل تخلیه‌ی جریان‌های سطحی را انجام می‌دهند (Alizadeh, 2007) و سنجش درجه تکامل آن‌ها با نمایه‌های تراکم، رده و انشعاب صورت می‌گیرد. تراکم زهکشی به‌طور مؤثری به لیتولوژی، شیب توپوگرافی و شرایط اقلیمی مربوط می‌شود و نقش اساسی در زمان تمرکز و اوج سیلاب دارد (شکل ۳I).

نقشه شیب زمین به دلیل تأثیر مستقیم بر رواناب سطحی، یکی از عوامل مهم در وقوع سیل حوزه‌های آبخیز به شمار می‌رود. این نقشه با استفاده از مدل رقومی ارتفاع با دقت ۱۲/۵ m و با استفاده از نرم‌افزار ArcGIS 10.4 تهیه شد. نقشه شیب زمین نیز بر اساس مدل رقومی ارتفاع و به کمک نرم‌افزار GIS تهیه شد. فاصله از رودخانه یکی از مهم‌ترین عوامل در سیل‌گیری اراضی مجاور است. نقشه فاصله از رودخانه‌های مجاور بر اساس لایه رقومی شبکه جریان حوزه آبخیز تجن در استان مازندران، در نرم‌افزار ArcGIS تهیه گردید (شکل ۳J).

۲-۳- پارامترهای مؤثر در تهیه نقشه پتانسیل سیل

برای تهیه نقشه حساسیت به وقوع سیل و یا به‌طور کلی تولید مدلی برای ارزیابی در معرض آسیب بلایای طبیعی، مجموعه‌ای از فاکتورها مؤثر باید تعریف گردد. مجموعه عوامل مؤثر بر سیلاب توسط ۱۳ فاکتور شامل ارتفاع، شیب، جهت، فاصله از رودخانه، تراکم زهکشی، شاخص رطوبت توپوگرافی، شاخص موقعیت توپوگرافی، انحنای شیب، شکل شیب، زمین‌شناسی، کاربری اراضی ساخته شد. این داده‌ها از منابع مختلف جمع‌آوری شد و سپس به‌اندازه پیکسل ۱۲/۵ m بر اساس نقشه مدل رقومی ارتفاع منطقه تبدیل شدند.

وضعیت زمین‌شناسی علاوه بر ویژگی‌های ساختار زمین‌شناسی، بیانگر نوع لیتولوژی حوضه نیز می‌باشد. لیتولوژی و خاک حاصل از آن تعیین‌کننده شدت و ظرفیت نفوذپذیری خاک بوده و رواناب سطحی را نیز تحت تأثیر قرار می‌دهد. نقشه سنگ‌شناسی منطقه مورد مطالعه با استفاده از نقشه زمین‌شناسی کشور تهیه شد و در نهایت ۲۶ واحد سنگ‌شناسی در منطقه مورد مطالعه شناسایی گردید (شکل ۳L Razavi Termeh et al. 2018). وجود پوشش گیاهی در سطح حوضه به دلیل اثراتی که بر اجزاء سیکل هیدرولوژی حوضه دارد از عوامل کاهش‌دهنده‌ی سیل‌خیزی یک حوضه است. وجود پوشش گیاهی در هر منطقه سرعت جریان‌های سطحی را کاهش



شکل ۳- فاکتورهای مؤثر بر تهیه نقشه حساسیت به سیل: A: انحای شیب؛ B: شیب؛ C: خاک؛ D: TPI؛ E: ارتفاع؛ F: TWI؛ G: بارش؛ H: جهت؛ I: تراکم زهکشی؛ J: فاصله از رودخانه؛ K: کاربری اراضی؛ M: شکل شیب

Fig. 3 Factors affecting the preparation of flood susceptibility map: A: Slope curvature; B: slope; C: soil; D: TPI; E: altitude; F: TWI; G: rainfall; H: aspect; I: drainage density; J: distance from river; K: land use; M: slope shape

یکی از مهم‌ترین فاکتورهای تأثیرگذار بر سیلاب در هر منطقه، بارش می‌باشد. به منظور تهیه لایه بارش در این پژوهش از داده‌های ۲۰ ساله، ۶ ایستگاه سینوئیتیک و هواشناسی (کردخیل، سلیمان تنگه، ریگ چشمه، قراخیل، کیاسر و سنگده) موجود در داخل و خارج از حوضه استفاده گردید. از روش درون‌یابی کریجینگ در محیط نرم‌افزار ArcGIS به منظور تهیه نقشه بارندگی حوزه آبخیز استفاده گردید (شکل ۳). بارش متوسط سالانه منطقه مورد مطالعه از ۳۴۷ تا ۷۵۵ متغیر می‌باشد به طوری که بیش‌ترین

شاخص رطوبت توپوگرافی (TWI) یکی از شاخص‌های تأثیرگذار بر وقوع و پتانسیل سیل در حوزه‌های آبخیز به شمار می‌رود (Vojtek and Vojteková 2019). برای تهیه این نقشه از نرم‌افزار SAGAGIS استفاده شد (شکل ۳F). TWI با توجه به رابطه ۱ تعریف می‌گردد.

$$TWI = A_s / \tan\beta \quad (1)$$

که در آن A_s سطح ویژه حوزه آبخیز و β شیب زمین (برحسب درصد) می‌باشند.

۲-۴-۲- مدل خطی تعمیم‌یافته بیزی

مدل خطی تعمیم‌یافته بیزی (GLMbayesian) تعمیم رگرسیون خطی است برای داده‌هایی که توزیع نرمال ندارند. مدل‌های خطی تعمیم‌یافته بیزی، مدل‌هایی ساده و مناسب برای داده‌های شمارشی هستند که در این مدل‌ها، واریانس تابع معلومی از میانگین است، اما اغلب داده‌های شمارشی ناهمگن‌تر از آن هستند که بتوان واریانس را به صورت تابع معلومی از میانگین در نظر گرفت (Kamata, 2001). اگرچه مدل‌های خطی تعمیم‌یافته بیش پراکندگی، در رابطه بین میانگین و واریانس، سستی بیشتری نشان می‌دهند ولی این مدل‌ها اغلب دارای تفسیر ساده‌ای نیستند و همچنین برازش این مدل‌ها، به آسانی مدل‌های خطی تعمیم‌یافته استاندارد نیست. تجزیه و تحلیل رگرسیون روشی برای تعیین بین متغیر پاسخ و متغیرهای مستقل می‌باشد. برای مواردی که ارتباط بین متغیرها خطی نیست، مدل‌های رگرسیون غیرخطی مورد استفاده قرار می‌گیرد (Hastie and Pregibon, 2009). عمدتاً در این روش‌ها برای نیل به یک پاسخ منحصر به فرد، فرض‌های محدودکننده‌ای در نظر گرفته می‌شود. برای تقلیل این گونه فرض‌ها نظریه مدل‌های خطی تعمیم‌یافته نیز به طور قابل ملاحظه‌ای بسط و گسترش یافته‌اند. در بسیاری از مواقع نوع ارتباط بین متغیرها مشخص نیست و در این صورت به جای مفروض داشتن یک الگوی خاص پارامتری، از روشی استفاده می‌شود که داده‌های ماهیت روند خود را بهتر نشان دهند. مدل‌های خطی تعمیم‌یافته بیزی به دو طریق پارامتری و ناپارامتری برای برآورد منحنی بکار برده می‌شوند (West et al. 1985).

۲-۵- ارزیابی مدل

به منظور ارزیابی مدل تهیه شده، از منحنی تشخیص عملکرد نسبی (ROC) استفاده گردید. منحنی ROC یک نمایش گرافیکی از موازنه بین نرخ خطای منفی و مثبت برای هر مقدار احتمالی از برش‌ها است. سطح زیر منحنی (AUC) ROC بیانگر مقدار پیش‌بینی سیستم از طریق توصیف توانایی آن در تخمین درست وقایع رخ داده (وقوع سیل) و عدم وقوع رخداد (عدم وقوع سیل) آن است. ایده‌آل‌ترین مدل بیشترین سطح زیر منحنی را دارد و

بارندگی مربوط به ایستگاه کردخیل و کم‌ترین بارندگی مربوط به ایستگاه کیاسر است. نقشه انحنای زمین یکی از عوامل مؤثر بر پتانسیل سیل حوزه‌های آبخیز می‌باشد. نقشه انحنای زمین بر اساس مدل رقومی ارتفاع و نرم‌افزار ArcGIS تهیه شد (شکل ۳).

۲-۴- مدل‌های مورد استفاده در تهیه نقشه

حساسیت سیل

۲-۴-۱- مدل جنگل تصادفی

روش‌های درخت پایه روش‌های آماری ناپارامتری (مدل آزاد) برای اجرای آنالیز کلاس‌بندی و آنالیز رگرسیونی با استفاده از الگوریتم افزایشی بازگشتی می‌باشند (Hastie, 1987; Breiman 2001). جنگل‌های تصادفی (RF) یک نوع مدرن از درخت پایه هستند که شامل انبوهی از درخت‌های کلاس‌بندی و رگرسیونی می‌باشند (Breiman 2001). مهم‌ترین ویژگی جنگل‌های تصادفی عملکرد بالای آن‌ها در اندازه‌گیری اهمیت متغیرها است که مشخص شود هر متغیر چه نقشی در پیش‌بینی پاسخ دارد. این الگوریتم در واقع یک الحاقی از درخت رگرسیون و طبقه‌بندی است که اولین بار توسط Breiman (2001) ایجاد و توسعه داده شد. الگوریتم جنگل تصادفی مبتنی بر دسته‌ای از درخت‌های تصمیم است و در حال حاضر یکی از بهترین الگوریتم‌های یادگیری است. مدل پیش‌بینی کننده RF بر اساس میانگین‌گیری از نتایج حاصل از تمامی درخت‌های تصمیم مربوطه استوار است و برای بسیاری از مجموعه داده‌ها، طبقه‌بندی با صحت بالایی را انجام می‌دهد (Lee et al. 2017). 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. جنگل تصادفی درخت تصمیم‌های زیادی تولید می‌کند. برای طبقه‌بندی یک شی جدید بردار ورودی در انتهای هر یک از درختان جنگل تصادفی قرار می‌دهد که هر درخت به یک طبقه‌بندی منجر می‌شود که گفته می‌شود این درخت به آن کلاس رأی می‌دهد. جنگل تصادفی از طبقه‌بندی که بیشترین رأی را داشته باشد (بین همه درختان جنگل) انتخاب می‌شود (Breiman 2001).

³ Receiver Operating Characteristic

⁴ Area Under Curve

¹ Random Forest

² Bayesian Generalization Linear Model

۳- یافته‌ها و بحث

عوامل متعددی در یک حوزه آبخیز بر جریان آب در سطح حوضه اثرگذار می‌باشند که هرکدام از این عوامل پتانسیل متفاوتی در تغییر جریان و رخداد سیلاب دارند. تعیین مهم‌ترین و مؤثرترین فاکتورهای تأثیرگذار بر سیلاب جهت انجام مدل‌سازی بسیار ضروری می‌باشد. در این پژوهش اولویت‌بندی فاکتورهای مؤثر بر سیلاب در حوزه آبخیز تجن در شکل (۴) نشان داده شده است. نتایج نشان داد که فاکتور ارتفاع (۲۱/۵۵)، فاصله از رودخانه (۱۵/۲۸)، شیب (۱۱/۱۸) و کاربری اراضی (۱۱/۱) به ترتیب بیشترین تأثیر را در تهیه نقشه حساسیت به وقوع سیل در منطقه مورد مطالعه دارند که با مطالعات (Darabi et al. 2014; Siahkamari and Zeinivand 2016; Arab ameri et al. 2017) مطابقت دارد. مقدار ارتفاع، شیب و فاصله از رودخانه با وقوع سیل گیری رابطه عکس دارند طوری که هر چه ارتفاع، شیب و فاصله از رودخانه بیش تر باشد احتمال سیل گیری و به طبع آن خسارت کاهش پیدا می‌کند، بنابراین باید بر روی این عوامل و لایه‌های مربوط به آن دقت لازم صورت بگیرد.

مقادیر سطح زیر منحنی از ۰/۵ تا ۱ متغیر است. هر چه سطح زیر منحنی به یک نزدیک تر باشد بیانگر دقت بیشتر نقشه پهنه‌بندی تهیه شده است (Pradhan et al. 2014, avand et al, 2019)

همچنین در این مطالعه از معیارهای حساسیت (Sensitivity)، ویژگی (Specificity)، NPV و PPV نیز به منظور ارزیابی نتایج مدل‌ها استفاده شده است (Pham et al, 2020) که به ترتیب در روابط (۲) تا (۵) ارائه شده است.

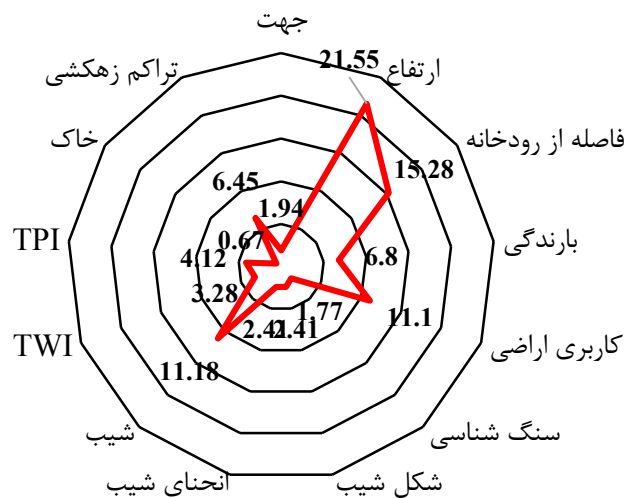
$$\text{حساسیت} = \frac{TP}{TP + FN} \quad (۲)$$

$$\text{ویژگی} = \frac{TN}{TN + FP} \quad (۳)$$

$$NPV = \frac{TN}{TN + FN} \quad (۴)$$

$$PPV = \frac{TP}{TP + FP} \quad (۵)$$

که در این معادلات، TP : تعداد مثبت صحیح، TN : تعداد مثبت کاذب، FN : تعداد منفی کاذب، و FP : تعداد مثبت کاذب می‌باشد.



شکل ۴- اهمیت فاکتورهای مورداستفاده در تهیه نقشه حساسیت سیل

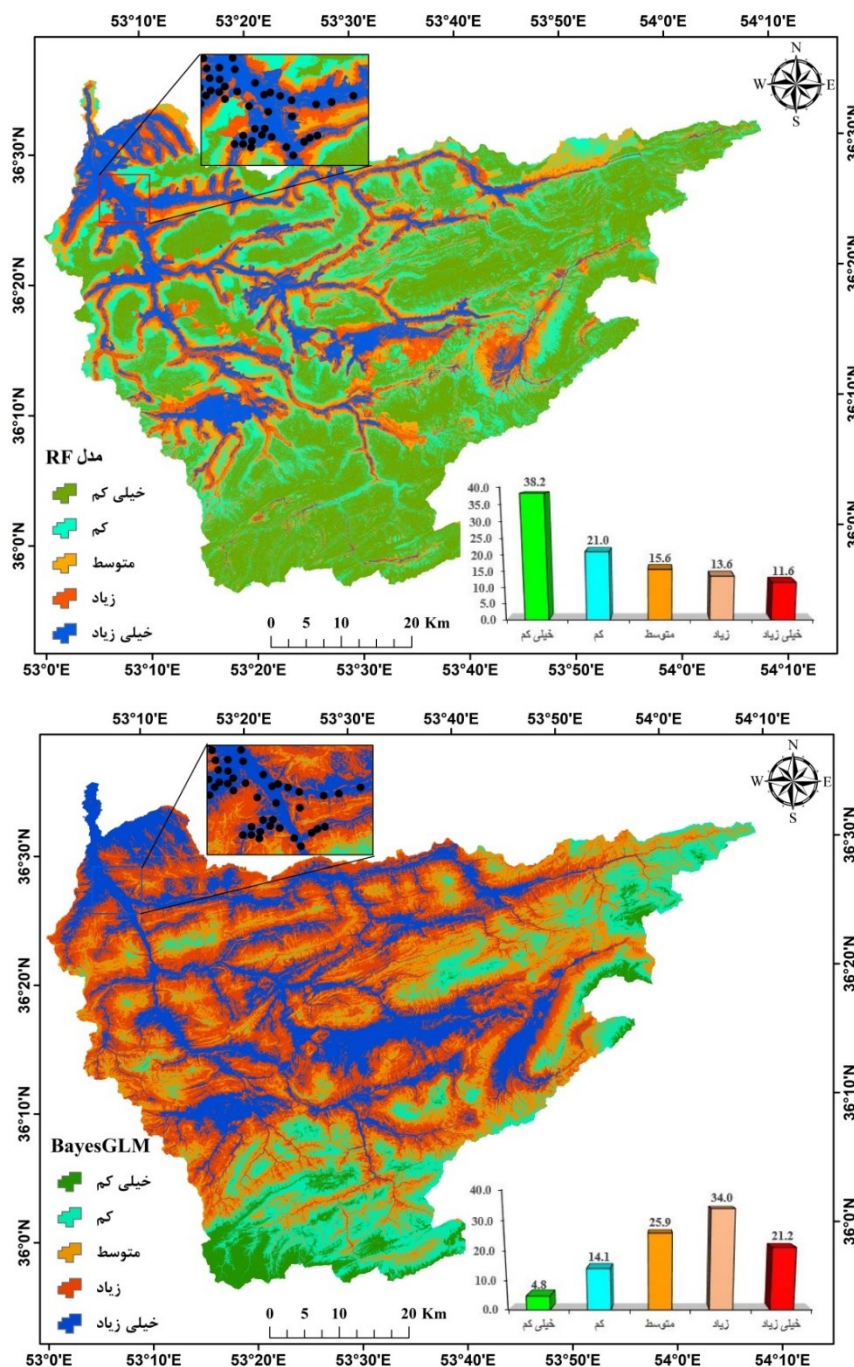
Fig. 4 Importance of factors used in the preparation of flood susceptibility map

نشان داد که احتمال وقوع سیل گیری در مناطق پایین- دست به دلیل ارتفاع و شیب کم‌تر، بیشتر از مناطق بالادست و ارتفاعات است (Costache 2019; Kanani-Sadat et al. 2019). همچنین در مناطق پایین دست به- دلیل کاهش سطح اراضی جنگلی و تغییر کاربری‌های شدید صورت گرفته در چند سال اخیر که سبب کاهش

نقشه‌های احتمال وقوع سیل با استفاده از دو مدل RF و GLMbayes تهیه شد و هرکدام این نقشه بر اساس مطالعات قبلی به ۵ طبقه شامل: طبقه حساسیت خیلی کم، متوسط، زیاد و خیلی زیاد تقسیم گردید. شکل (۴) نقشه احتمال وقوع سیل گرفتگی را نشان می‌دهد. نقشه حساسیت به وقوع سیل این حوضه (شکل ۵) نیز

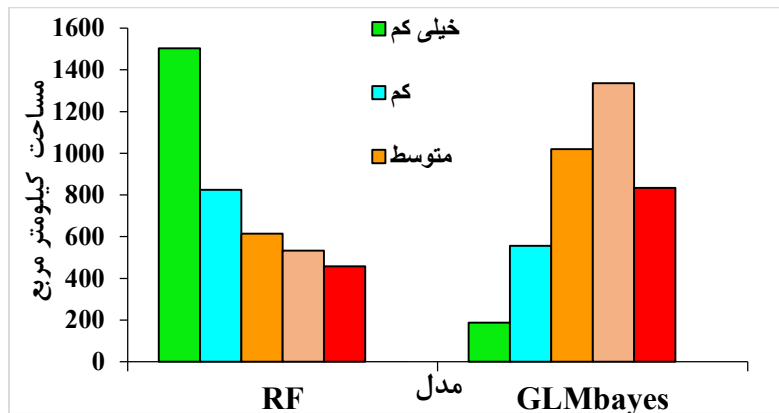
کاهش می‌یابد و در ارتفاعات پایین‌تر و نزدیک به دریا مقدار بارندگی افزایش می‌یابد. از طرفی احتمال وقوع سیل گیری در مناطق پایین‌دست نیز بیشتر می‌باشد، بنابراین یک رابطه مستقیم بین افزایش بارندگی و سیل گیری و خسارت در مناطق پایین‌دست این حوزه آبخیز نیز وجود دارد که باید مورد توجه قرار بگیرد.

اراضی جنگلی و افزایش اراضی کشاورزی و مسکونی 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 سیلابی در این مناطق بوده است. در مناطق شمالی ایران و در حوزه آبخیز مورد مطالعه با افزایش ارتفاع برخلاف مناطق جنوبی رشته کوه البرز مقدار بارندگی



شکل ۵- نقشه حساسیت به سیل تولیدشده توسط مدل‌های مورد استفاده: A: RF و B: GLMbayes

در شکل (۶) مساحت هر یک از کلاس‌های حساسیت دو مدل RF و GLMbayes نشان داده شده است. نتایج نشان داد که در مدل RF بیشترین سطح حساسیت مربوط به طبقه خیلی کم با مساحت $1503/01 \text{ km}^2$ می‌باشد در حالی که در مدل GLMbayes بیشترین سطح مربوط به طبقه حساسیت زیاد با مساحت $1335/54 \text{ km}^2$ است.

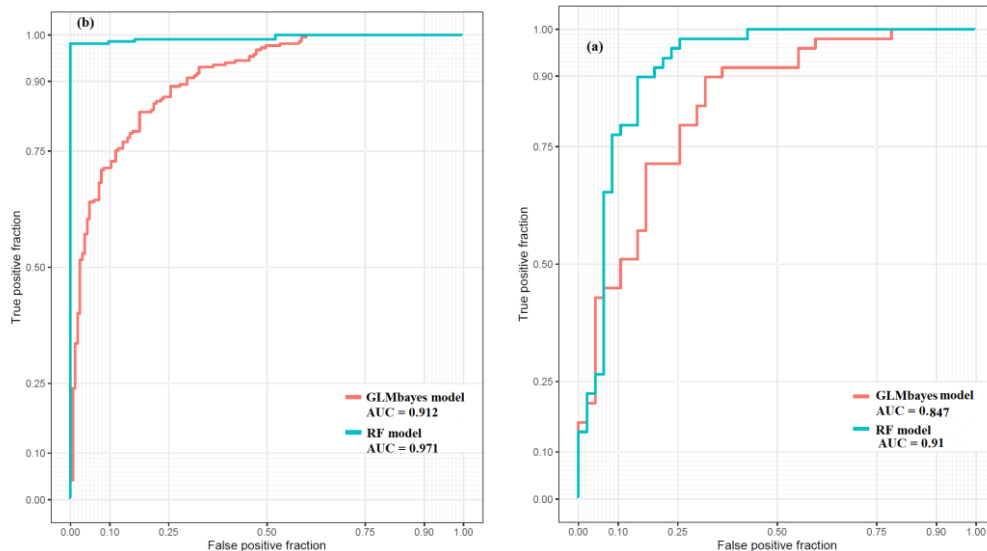


شکل ۶- مساحت هر یک از طبقه‌های حساسیت سیل در منطقه مورد مطالعه

Fig. 6 Area of each flood susceptibility class in the study area

از داده‌های آزمون نشان داد که مدل RF و GLMbayes به ترتیب دارای سطح زیر منحنی $0/847$ و $0/91$ می‌باشند. همچنین نتایج اعتبارسنجی این مدل‌ها با استفاده از داده‌های آموزش نیز نشان‌دهنده سطح زیر منحنی $0/912$ و $0/971$ برای مدل RF و GLMbayes به ترتیب هستند. نتایج نشان داد که مدل RF دارای دقت بالاتری در تعیین مناطق حساس به وقوع سیل در حوزه آبخیز تجن می‌باشد و این نتایج با یافته‌های (Zhao et al. 2018; Chen et al. 2019) مطابقت دارد.

۳-۱- اعتبارسنجی نقشه حساسیت به وقوع سیل اعتبارسنجی نقشه‌های تهیه شده، گامی اساسی در توسعه و تعیین مناطق حساس و تعیین کیفیت آن‌ها می‌باشد. نتایج مربوط به اعتبارسنجی مدل‌های مورد استفاده در این پژوهش در شکل (۷) نمایش داده شده است. به منظور اعتبارسنجی مدل‌های مورد استفاده از منحنی ROC و مساحت زیر منحنی (AUC) جهت ارزیابی استفاده شده است. نتایج اعتبارسنجی مدل‌های استفاده شده با استفاده



شکل ۷- منحنی ROC برای اعتبارسنجی نتایج مدل‌های استفاده شده: (a) آزمون (b) آموزش

Fig. 7 ROC curve for validating the results of the models used: (a) test (b) training

جدول ۲- معیارهای ارزیابی مدل‌های مورد استفاده
Table 2 Evaluation criteria for the models used

Models	Metrics	AUC	Sensitivity	Specificity	PPV	NPV
اعتبارسنجی	RF	0.91	0.918	0.787	0.818	0.902
	GLMbayes	0.847	0.775	0.744	0.760	0.760
آموزش	RF	0.971	0.981	0.993	0.995	0.976
	GLMbayes	0.912	0.792	0.830	0.860	0.752

۴- نتیجه‌گیری

سیل‌گیری یکی از پدیده‌های مهم و مصیبت‌بار در نواحی شمالی ایران محسوب می‌شود. از طریق آنالیز منطقه‌ای سیل‌گیری می‌توان مناطق حساس به سیل را شناسایی و خسارات ناشی از آن را کاهش داد. هدف از این پژوهش تعیین مناطق حساس به سیل با استفاده از روش‌های داده-کاوی RF و GLMbayes می‌باشد. مهم‌ترین نتایج پژوهش حاضر عبارت‌اند از:

۱- عوامل ارتفاع، فاصله از رودخانه و شیب به‌ترتیب مهم‌ترین پارامترهای تأثیرگذار در وقوع سیل در این حوضه می‌باشند.

۲- مدل جنگل تصادفی با مقدار سطح زیر منحنی ۹۱٪ دقت بیش‌تری نسبت به مدل GLMbayes در تعیین مناطق حساس به سیل دارد.

۳- مناطق پایین‌دست و نزدیک خروجی حوزه آبخیز به دلیل ارتفاع و شیب کم‌تر دارای حساسیت بیش‌تری نسبت به سیل می‌باشد.

از نتایج این پژوهش می‌توان به‌منظور شناسایی مناطق مناسب و امن برای توسعه‌های عمرانی استفاده کرد. با توجه به نتایج قابل‌قبول مدل جنگل تصادفی، پیشنهاد می‌شود از این روش به‌منظور تهیه نقشه حساسیت و خطر سیل‌گیری در مناطق مشابه دیگر نیز استفاده گردد.

References

- Alizadeh A. (2007). Principles of applied hydrology. Mashhad, Twenty-third Edition, Astan Qods Razavi Publications, p. 808 [In Persian].
- Al-Juaidi A. E. M., Nassar A. M. and Al-Juaidi O. E. M. (2018) Evaluation of flood susceptibility mapping using logistic regression and GIS conditioning factors. Arab. J. Geosci. 11, 765.
- Arab Ameri A., Pourghasemi H. R. and Shirani K. (2017). Flood susceptibility zoning using a new hybrid Bayesian theory-hierarchical analysis process (Case study: Neka Watershed, Mazandaran Province). Ecohydrol., 4(2), 447-462 [In Persian].
- Avand M., Janizadeh S., Tien Bui D., Pham V. H., Ngo P. T. T. and Nhu V. H. (2020). A tree-based intelligence ensemble approach for spatial prediction of potential groundwater. Int. J. Digit. Earth, 1-22.
- Breiman L. (2001). Random forests. Machine Learn., 45(1), 5-32.
- Costache R. (2019) Flood susceptibility assessment by using bivariate statistics and machine learning models-A useful tool for flood risk management. Water Resour. Manag., 33, 3239-3256.
- Darabi H., Shahedi K. and Mardian M. (2016). Flood susceptibility and probability mapping using frequency ratio method in Pol-Doab Shazand watershed. J. Eng. Watershed Manage., 8(1), 79-68 [In Persian].
- Darabi H., Shahedi K., Solaimani K. and Miryaghoubzadeh M. (2014). Prioritization of subwatersheds based on flooding conditions using hydrological model, multivariate analysis and remote sensing technique.

- Water Environ. J., 28(3), 382-392.
- Fernandez D. S. and Lutz M. A. (2010). Urban flood hazard zoning in Tucuman Province, Argentina, using GIS and multicriteria decision analysis. Eng. Geol., 111, 90-98.
- Hastie T., Tibshirani R. and Friedman J. (2009). The elements of statistical learning: data mining, inference, and prediction. Springer Science & Business Media.
- Hazarika N., Barman D., Das A. K., Sarma A. K. and Borah S. B. (2018). Assessing and mapping flood hazard, vulnerability and risk in the Upper Brahmaputra River valley using stakeholders' knowledge and multicriteria evaluation (MCE). 11(52), 5700-5716.
- Janizadeh S., Avand M., Jaafari A., Phong T., Bayat M., Ahmadisharaf E., Prakash I., Pham B. T. and Lee S. (2019). Prediction success of machine learning methods for flash flood susceptibility mapping in the Tafresh Watershed, Iran. Sustain., 11(19), 5426.
- Kamata A. (2001). Item analysis by the hierarchical generalized linear model. J. Edu. Measure., 38(1), 79-93.
- Kanani-Sadat Y., Arabsheibani R., Karimipour F. and Nasser 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. J. Hydrol., 572, 17-31.
- Khairizadeh M., Maleki J. and Amunia H. (2014). Potential zoning of flood risk in Mardegh Chai catchment using ANP model. J. Quant. Geomorphol. Res., 1(3), 39-65 [In Persian].
- Khairizadeh M., Maleki J. and Amounia H. (2013). Flood hazard zoning using ANP model in watershed, case study: Mardaghchay Watershed. Quant. Geomorphol. Res., 3, 39-56 [In Persian].
- Lee Sunmin Kim J. C., Jung H. S., Lee M. J. and Lee S. (2017). Spatial prediction of flood susceptibility using random-forest and boosted-tree models in Seoul metropolitan city, Korea. Geomat. Nat. Hazard. Risk, 8, 1185-1203.
- Luu C. and Meding J. V. (2018). A flood risk assessment of Quang Nam, Vietnam. Water, 10(4), 461.
- Motevalli A. and Vafakhah M. (2016). Flood hazard mapping using synthesis hydraulic and geomorphic properties at watershed scale. Stoch. Environ. Res. Risk Assess., 30, 1889-1900.
- Mukerji A., Chatterjee C. and Raghuvanshi N.S. (2009). Flood forecasting using ANN, neuro-fuzzy, and neuro-GA models. J. Hydrol. Eng., 647-652.
- Pham B. T., Avand M., Janizadeh S., Phong T. V., Al-Ansari N., Ho L. S., Das S., Le H. V., Amini A., Khosrobeigi Bozchaloei S., Jafari F. and Prakash I. (2020). GIS based hybrid computational approaches for flash flood susceptibility assessment. Water, 12(3), 683.
- Rahmati O., Pourghasemi H. R. and Zeinivand H. (2016). Flood susceptibility mapping using frequency ratio and weights-of-evidence models in the Golastan Province, Iran. Geocarto. Int., 31, 42-70.
- Pradhan B., Hagemann U., Tehrany Shafapour M. and Prechte N. (2014). An easy to use ArcMap based texture analysis program for extraction of flooded areas from TerraSAR-X satellite image. Comput. Geosci., 63, 34-43.
- Razavi Termeh S. V., Kornejady A., Pourghasemi H. R. and Keesstra S. (2018). Flood susceptibility mapping using novel ensembles of adaptive neuro fuzzy inference system and metaheuristic algorithms. Sci Total Environ., 615, 438-451.
- Siahkamri S. and Zainivand H. (2016). Potential of flood prone areas using statistical index model and weight of evidence (Case study: Madarsoo Watershed, Golestan). Remote Sens. GIS Nat. Resour., 7(4), 0-0 [In Persian].
- Souissi D., Zouhri L., Hammami S., Msaddek

- M. H., Zghibi A. and Dlala M. (2019). GIS-based MCDM-AHP modeling for flood susceptibility mapping of arid areas, southeastern Tunisia. *Geocarto. Int.*, 1–27.
- Tehrany M. S., Pradhan. B. and Jebur M. N. (2014). Flood susceptibility mapping using a novel ensemble weights-of-evidence and support vector machine models in GIS. *J. Hydrol.*, 512, 332–343.
- Tehrany Shafapour M., Pradhan B. Mansor S. H. and Noordin A. (2015). Flood susceptibility assessment using GIS-based support vector machine model with different kernel types. *Catena.*, 125, 91-101.
- Vojtek M. and Vojteková J. (2019). Flood Susceptibility Mapping on a National Scale in Slovakia Using the Analytical Hierarchy Process. *Water* 11:364, Multidisciplinary Digital Publishing Institute.
- Wang Y., Hong H., Chen W., Li S., Pamučar D., Gigović L., Drobnjak S., Bui D. T. and Duan H. (2019). A hybrid GIS multi-criteria decision-making method for flood susceptibility mapping at Shangyou, China. *Remote Sens.*, 11(1), 62.
- West M., Harrison P. J. and Migon H. S. (1985). Dynamic generalized linear models and Bayesian forecasting. *J. Am. Stat. Assoc.*, 80(389), 73-83.
- Zhao G., Pang B., Xu Z., Yue J. and Tu T. (2018). Science of the Total Environment Mapping flood susceptibility in mountainous areas on a national scale in China. *Sci. Total Environ.*, 615, 1133–1142,

Flood Susceptibility Mapping Using Random Forest Machine Learning and Generalized Bayesian Linear Model

Mohammadtaghi Avand¹, Hamid Reza Moradi^{2*}, Mehdi Ramezanzadeh³

¹PhD Scholar, Department of Watershed Management Engineering, Faculty of Natural Resources and Marine Sciences, Tarbiat Modares University, Tehran, Iran

²Assoc. Professor, Department of Watershed Management Engineering, Faculty of Natural Resources and Marine Sciences, Tarbiat Modares University, Tehran, Iran

³Assist. Professor, Department of Tourism Management, Faculty of Humanities and Social Sciences, University of Mazandaran, Babolsar, Iran

*Corresponding author: harmoradi@modares.ac.ir

Technical Note

Received: March 14, 2020

Revised: April 07, 2020

Accepted: April 13, 2020

Abstract

Today, the phenomenon of flooding is one of the most complex hazardous events that, more than any other natural disaster, causes deaths and finances every year in different parts of the world. Therefore, flood susceptibility mapping is the first step in a flood management program. The purpose of this study was to identify flood susceptible areas using two methods of random forest (RF) and Bayesian generalized linear model (GLMbayesian) machine learning in the Tajan watershed in Mazandaran province, Sari. Past flood distribution maps were prepared to predict future floods. Of the 263 flood locations, 80% (210 flood locations) was used for modeling and 20% (53 flood locations) was used for validation. Based on previous studies and surveying of the study area, 13 conditional factors were selected for flood zoning. The results showed that three factors of elevation (21.55), distance from the river (15.28) and slope (11.18) had the highest impact on flood occurrence in the study area, respectively. The results also showed that the AUC values for RF and GLMbayesian models were 0.91 and 0.847, respectively, indicating the superiority of the RF model and the accuracy of this model in flood susceptibility mapping in the study area. The highest flood susceptibility area in the RF model is in the very low class and the high class in the GLMbayesian model.

Keywords: Flood Damage; GLMbayesian Model; RF Model; ROC Curve; Tajan Watershed.