

## مقایسه مدل‌های شبکه عصبی مصنوعی، درخت تصمیم و شبکه بیزی در تحلیل منطقه‌ای سیلاب با استفاده از روش‌های گشتاور خطی و حداکثر درستنمایی در حوزه‌های آبخیز کرخه و کارون

کمال قادری<sup>۱</sup>، بهارک معتمدوزیری<sup>۱</sup>، مهدی وفاخواه<sup>۲\*</sup> و امیراحمد دهقانی<sup>۳</sup>

(تاریخ دریافت: ۱۳۹۹/۱۱/۲۰؛ تاریخ پذیرش: ۱۴۰۰/۲/۲۷)

### چکیده

پیش‌بینی صحیح دبی سیل، برای طراحی سازه‌های هیدرولیکی، کاهش خطر شکست و به حداقل رساندن آسیب‌های محیط زیستی پایین‌دست، از اهمیت زیادی برخوردار است. هدف از انجام این مطالعه، بررسی کاربرد روش‌های یادگیری ماشین برای تحلیل منطقه‌ای فراوانی سیلاب است. برای دستیابی به این هدف، ۱۸ پارامتر فیزیوگرافی، اقلیمی، سنگ‌شناسی و کاربری اراضی برای حوضه‌های بالادست ایستگاه‌های هیدرومتری حوزه‌های آبخیز کرخه و کارون (۴۶ ایستگاه با طول آماری ۲۱ سال) در نظر گرفته شد. سپس، بهترین تابع توزیع احتمال با استفاده از آزمون کولموگروف-اسمیرنف در هر ایستگاه برای برآورد دبی سیل با دوره بازگشت ۵۰ سال با استفاده از روش‌های حداکثر درستنمایی و گشتاورهای خطی تعیین شد. در نهایت، تحلیل منطقه‌ای فراوانی سیلاب با استفاده از درخت تصمیم، شبکه بیزی و شبکه عصبی مصنوعی انجام شد. نتایج نشان داد که توزیع لوگ پرسون تیپ ۳ در روش حداکثر درستنمایی و توزیع نرمال تعمیم‌یافته در روش گشتاورهای خطی بهترین توزیع احتمالی منطقه‌ای هستند. بر اساس آزمون گاما، پارامترهای محیط، طول حوضه، عامل شکل و طول جریان اصلی به‌عنوان بهترین ترکیب ورودی انتخاب شدند. نتایج تحلیل منطقه‌ای فراوانی سیلاب نشان داد که مدل بیزی با روش گشتاور خطی ( $R^2=0.7$ ) بهترین برآورد را در مقایسه با روش‌های دیگر دارد. درخت تصمیم و شبکه عصبی مصنوعی در رده‌های بعدی قرار داشتند.

واژه‌های کلیدی: مدل‌سازی، دبی حداکثر لحظه‌ای، مدل‌های یادگیری ماشین، گشتاور خطی، شبکه بیزی

۱. گروه جنگل، مرتع و آبخیزداری، دانشکده منابع طبیعی و محیط زیست، دانشگاه آزاد اسلامی، واحد علوم و تحقیقات

۲. گروه مهندسی آبخیزداری، دانشکده منابع طبیعی و علوم دریایی، دانشگاه تربیت مدرس

۳. گروه مهندسی آب، دانشکده مهندسی آب و خاک، دانشگاه علوم کشاورزی و منابع طبیعی گرگان

\* مسئول مکاتبات: پست الکترونیک: vafakhah@modares.ac.ir

## مقدمه

سیلاب یکی از بلایای طبیعی است که خسارات مالی و جانی را در بردارد. برآورد صحیح دبی حداکثر سیل برای طراحی سازه‌های هیدرولیکی، کاهش خطر شکست آنها و به حداقل رساندن خسارت محیط زیستی در پایین دست اهمیت دارد. دبی حداکثر و فراوانی آن همچنین برای ارزیابی خطر سیل پروژه‌ها و تعیین دبی طراحی سیل مورد نیاز است (۲۵). بیشتر از ۷۵ درصد ایران در مناطق خشک و نیمه‌خشک قرار دارد و با وجود بارندگی سالانه کم، اغلب سیلاب‌های شدیدی رخ می‌دهد. در ایران، تعداد زیادی از سیلاب‌های شدید در ۵۰ سال اخیر اتفاق افتاده است و باعث تلفات اقتصادی فراوانی شده است. امروزه مناطق در معرض سیل پرجمعیت هستند و سریع توسعه می‌یابند. بسیاری از سری‌های سالانه سیل، کوتاه‌مدت هستند و نمی‌توان برآورد دقیق و قابل اعتمادی از رویدادهای سیل داشت یا اینکه هرگز داده ثبت شده‌ای در منطقه مورد نظر وجود ندارد (۲۳ و ۱۹).

یک راه برای برآورد دبی حداکثر سیلاب، استفاده از داده‌های ثبت شده از یک حوزه آبخیز با رفتار سیلابی مشابه به جای استفاده از داده‌های همان حوزه است (۱۹). تحلیل منطقه‌ای فراوانی سیل (Regional Flood Frequency Analysis) نامی عمومی است که برای توصیف روش‌هایی به کار می‌رود که از داده‌های جریان حوزه‌های دارای ایستگاه هیدرومتری در یک منطقه برای برآورد دبی طراحی در حوزه‌های دارای آمار ضعیف یا حوزه‌های بدون آمار و ایستگاه، به کار می‌رود (۸). تحلیل منطقه‌ای دبی‌های حداکثر سیلاب با دوره‌های برگشت مختلف به دو دلیل اهمیت دارد. دلیل اول این است که، یک شبکه به اندازه کافی متراکم از ایستگاه‌های هیدرومتری وجود ندارد و از این رو، کم اتفاق می‌افتد که داده ثبت شده در همان مقطع عرضی از یک رودخانه طبیعی که قرار است سدی ساخته شود، در دسترس باشد. دلیل دوم این است که دوره ثبت داده‌ها فقط چند سال باشد، که برای یک تحلیل فراوانی، قابلیت اعتماد پایینی دارد (۱۳). به همین دلیل تحلیل منطقه‌ای سیلاب با

روش‌های بسیار دقیق مثل روش‌های هوش مصنوعی در دو حوضه حساس کرخه و کارون بسیار ضروری است. سیل فروردین ۱۳۹۸ ضرورت مطالعه سیلاب در این حوضه را بیشتر نشان داد.

اعلمی و همکاران (۲) با توجه به کارایی روش‌های سیستم فازی-عصبی تطبیقی و شبکه بیزی پیش‌بینی سیلاب و شبیه‌سازی فرایندهای با اطلاعات ناقص و غیر دقیق به‌منظور ارزیابی نقش عدم قطعیت در سری دبی رودخانه، این دو مدل را در استان آذربایجان شرقی مقایسه کرده و نتایج نشان‌دهنده کارایی مناسب هر دو مدل عصبی فازی-تطبیقی و بیزی برای مدل‌سازی این‌گونه سیستم‌ها داشته و همچنین شبکه بیزی در نقاط حداقل و حداکثر عملکرد بهتری نسبت به مدل ANFIS داشته است. مقایسه روش‌های ماشین بردار پشتیبان و شبکه‌های بیزی در پیش‌بینی جریان روزانه رودخانه (مطالعه موردی: رودخانه باراندوزچای توسط احمدی و همکاران (۴) نشان داده است دو روش SVM و BNs از نظر تحلیل‌های آماری نزدیک به هم بوده، اما از نقطه نظر گرافیکی و تخمین دبی‌های کمینه و بیشینه، مشخص شده است که هر دو مدل در پیش‌بینی دبی بیشینه خطای بیشتری دارند اما در پیش‌بینی دبی متوسط و کمینه عملکرد مدل SVM به مراتب بهتر از مدل BNs بوده است. در مطالعه مشاری و دانشفراز (۲۲) به بررسی قابلیت عملکرد شبکه بیزی با استفاده از مدل‌های هوشمند از قبیل برنامه‌ریزی ژنتیک و شبکه عصبی مصنوعی به‌منظور تخمین دبی در رودخانه قوره چای با داده‌های مربوط به سال‌های ۱۳۸۶-۱۳۹۰ دبی روزانه جریان رودخانه پرداخته‌اند. با توجه به نتایج بدست آمده از مدل‌ها هر سه روش از دقت قابل قبولی برخوردار بوده‌اند و در نهایت مدل شبکه عصبی مصنوعی دارای بیشترین دقت و مدل‌های برنامه‌ریزی ژنتیک و شبکه بیزی به ترتیب رتبه‌های دوم و سوم را به خود اختصاص داده‌اند. نتایج تحقیق شریفی گرمدره و همکاران (۲۷) عملکرد بهتر ماشین بردار پشتیبان را نسبت به شبکه عصبی مصنوعی در تحلیل منطقه‌ای سیلاب حوزه آبخیز دریاچه نمک نشان داده‌اند. پاپروتی و

طول‌های جغرافیایی ۰۲' ۶۶° و ۰۱' ۵۲° شرقی و عرض‌های ۰۲' ۲۹° تا ۰۲' ۳۴° شمالی در غرب ایران قرار دارد (شکل ۱). ارتفاع بین ۲۲ (در جنوب منطقه و در نزدیکی خلیج فارس) تا ۴۴۱۵ متر از سطح دریا (در جنوب شرق منطقه مورد مطالعه) متغیر است. بارش به‌طور متوسط به‌ترتیب از ۱۵۰ تا ۷۵۰ میلی‌متر به‌ترتیب برای مناطق کم‌ارتفاع و مناطق مرتفع متغیر است. آب و هوای غالب منطقه مورد مطالعه، بیشتر در غرب و جنوب غربی منطقه مورد مطالعه، خشک با تابستان‌های گرم است. شمال منطقه مورد مطالعه دارای آب و هوای نیمه‌خشک است، درحالی‌که مناطق مرتفع در شرق و جنوب شرقی منطقه قرار دارد (به‌عنوان مثال یاسوج و شهرکرد) دارای آب و هوای نیمه‌مرطوب است (۹). منطقه مورد مطالعه جزء حوضه منطقه‌ای شماره ۲ ایران (حوضه خلیج فارس و دریای عمان) است که این حوضه منطقه‌ای بزرگ، تنها حوضه غیر بسته ایران است. حوضه‌های کرخه و کارون مساحت زیادی از کشور را در بر می‌گیرد، به‌طوری‌که قسمت زیادی از استان‌های خوزستان، لرستان، چهارمحال و بختیاری، کرمانشاه، ایلام، کهگیلویه و بویراحمد و قسمتی از استان‌های همدان، مرکزی، اصفهان، فارس و حتی بخشی از کردستان در این حوضه قرار دارند.

#### داده‌های مورد استفاده

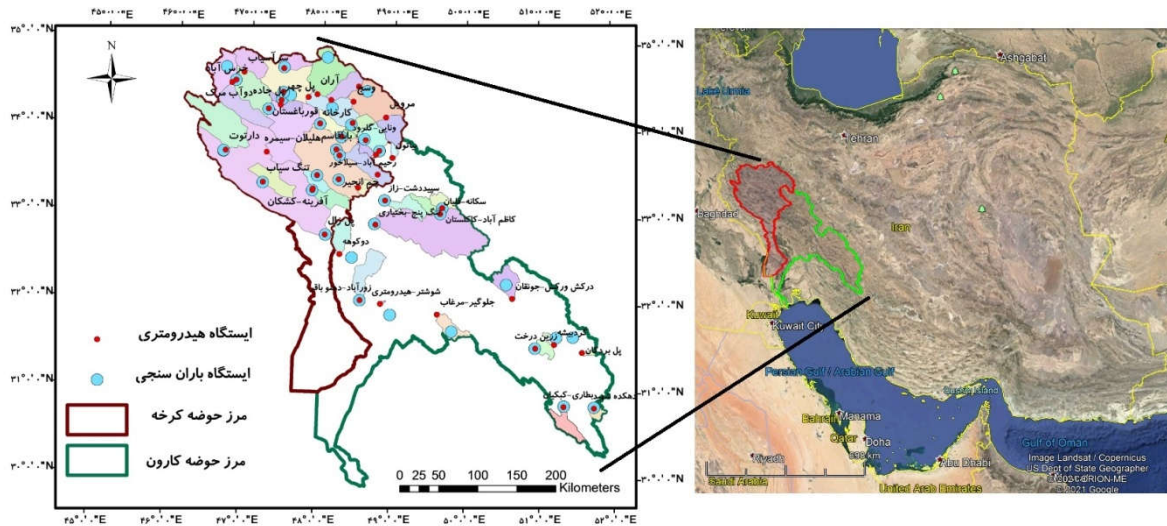
در این پژوهش، ابتدا، ایستگاه‌های هیدرومتری که در بالادست سدها قرار داشته و داده‌های طولانی‌مدت داشتند، انتخاب شدند که بر این اساس ۴۶ ایستگاه انتخاب شد. دبی‌های حداکثر لحظه‌ای ایستگاه‌ها با یک دوره یکسان (۲۱ سال طی سال‌های ۱۹۹۱ تا ۲۰۱۱) مورد استفاده قرار گرفت. سپس ایستگاه‌های باران‌سنجی متناظر تعیین شدند در این مطالعه پارامترهای فیزیوگرافی شامل مساحت، محیط، شیب متوسط، ارتفاع متوسط، طول حوضه، طول آبراهه اصلی، شیب آبراهه اصلی، تراکم زهکشی، حداکثر رتبه آبراهه، عامل شکل، طول حوضه، ضریب کروی حوضه، ضریب فشردگی، متغیرهای

مورالس-ناپولس (۲۴) به برآورد دبی‌های حداکثر در ۱۸۰۰ ایستگاه هیدرومتری اروپا از طریق یک شبکه بیزی پرداخته‌اند و نتیجه گرفته‌اند که این مدل را می‌توان برای برآورد سریع دبی‌های حداکثر در هر نقطه‌ای به کار برد. حسینی و همکاران (۱۷) به ارزیابی خطر سیلاب ناگهانی (Flash Flood) با استفاده از مدل‌های یادگیری ماشین مبتنی بر بیزی پرداخته‌اند. نتایج حاکی از عملکرد بالای مدل‌ها با دقت ۹۰-۹۲ درصد است. تحلیل منطقه‌ای سیلاب را در مناطق نیمه‌خشک با استفاده از گشتاورهای خطی و مدل‌های یادگیری ماشین که توسط الله بخشیان فارسانی و همکاران (۵) انجام شده، نتایج بسیار خوبی در پی داشته داشت.

با توجه پژوهش‌های گذشته، مدل‌های داده کاوی، به‌طور مؤثری در پیش‌بینی متغیرهای هیدرولوژیکی مورد استفاده قرار گرفته‌اند. کارایی این مدل‌ها در شبیه‌سازی و پیش‌بینی مشخص است ولی پژوهش‌های نادری مدل‌های بیزی و درخت تصمیم را در تحلیل منطقه‌ای سیلاب و با روش‌های گشتاور خطی و حداکثر درستی‌نمایی به‌کاربرده‌اند. از این‌رو، در این مطالعه سعی شده است کاربرد سه مدل یادگیری ماشین از جمله درخت تصمیم، شبکه بیزی و شبکه عصبی مصنوعی برای تحلیل منطقه‌ای فراوانی سیل با استفاده از دو روش گشتاور خطی و حداکثر درستی‌نمایی در حوزه‌های آبخیز کرخه و کارون مقایسه شود. مراحل مورد استفاده به صورت زیر است: الف) استفاده از روش‌های یادگیری ماشین برای تحلیل منطقه‌ای سیلاب، (ب) انتخاب توزیع احتمالی مناسب با استفاده از روش‌های حداکثر درستی‌نمایی و گشتاورهای خطی، (ج) تعیین بهترین ترکیب ورودی با استفاده از آزمون گاما، (د) شناسایی بهترین طول آماری داده‌های آموزش و آزمون در فرایند یادگیری ماشین و (ه) مقایسه عملکرد سه مدل انتخابی با دو روش حداکثر درستی‌نمایی و گشتاورهای خطی

#### مواد و روش‌ها

منطقه مورد مطالعه با مساحت ۱۱۸۴۳ کیلومتر مربع بین



شکل ۱. موقعیت منطقه مورد مطالعه (ایستگاه‌های هیدرومتری حوزه‌های آبخیز کرخه و کارون) (رنگی در نسخه الکترونیکی)

آمارای ۲۱ سال) انتخاب شدند. سپس دبی‌های حداکثر لحظه‌ای با دوره بازگشت ۵۰ سال در همان دوره مد نظر قرار گرفت. با توجه به اینکه در روش تحلیل منطقه‌ای، محاسبات در دوره‌های برگشت پایین خطای بیشتری نسبت به دوره‌های برگشت بالا دارد و اینکه برای دوره‌های بازگشت بالاتر (۲۵ و ۵۰ سال) روش گشتاور خطی نسبت به سایر روش‌ها از جمله رگرسیون مناسب‌تر است، دبی سیلاب با دوره بازگشت ۵۰ سال برای مدل‌سازی انتخاب شد. تکمیل و تطویل داده‌های ناقص، به وسیله رگرسیون بین دبی حداکثر لحظه‌ای و حداکثر دبی سالانه به‌عنوان متغیر مستقل با استفاده از نرم افزار SPSS انجام شد. با برآزش دبی اوج سیل به توابع توزیع احتمال مختلف بر اساس آزمون کولموگروف-اسمیرنوف با به کار بردن نرم افزار Easy Fit دبی حداکثر لحظه‌ای با دوره‌های بازگشت مشخص با استفاده از روش‌های حداکثر درستیابی و گشتاورهای خطی پیش‌بینی شد. برای هر ایستگاه هیدرومتری یک ایستگاه باران‌سنجی متناظر تعیین شد و بر اساس آن حداکثر بارش ۲۴ ساعته برای دوره‌های بازگشت مختلف به‌دست آمد. در کل ۱۸ پارامتر فیزیوگرافی، اقلیمی، سنگ‌شناسی و کاربری اراضی برای حوزه‌های بالادست ایستگاه‌های هیدرومتری در نظر گرفته شد. بهترین ترکیب ورودی و بهترین طول داده‌های مجموعه داده‌های

اقلیمی (شامل میانگین بارش سالانه و حداکثر بارش ۲۴ ساعته)؛ مشخصات سنگ‌شناسی (شامل درصد سازندهای نفوذپذیر)؛ و کاربری اراضی (شامل درصد اراضی لخت و درصد اراضی جنگل) برای حوزه آبخیز بالادست ایستگاه‌های انتخاب شده محاسبه شد. این پارامترها به‌عنوان پیش‌بینی‌کننده برای مدل‌سازی سیلاب‌های ایستگاه‌های منتخب مورد استفاده قرار گرفتند (۱۲). میانگین بارندگی سالانه ۴۶ ایستگاه منتخب ۵۳۸ میلی‌متر بوده و بین ۲۲۶ میلی‌متر در ایستگاه زورآباد تا ۱۱۱۴ میلی‌متر در ایستگاه رحیم‌آباد سیلاخور متغیر است. در بین کاربری‌های مختلف، کاربری‌های مرتع، اراضی کشاورزی و دیم و اراضی جنگلی بیشترین درصد را به خود اختصاص می‌دهند. از نظر زمین‌شناسی حدود ۳۰ درصد منطقه را اراضی نفوذپذیر تشکیل می‌دهند (۱۱). داده‌ها و نقشه‌ها از شرکت مدیریت منابع آب ایران، پژوهشکده حفاظت خاک و آبخیزداری، سازمان زمین‌شناسی کشور و سازمان جنگل‌ها، مراتع و آبخیزداری تهیه شدند.

#### روش تحقیق

همان‌طور که اشاره شد ایستگاه‌های هیدرومتری در حوزه‌های آبخیز کرخه و کارون با دوره طولانی‌مدت (۴۶ ایستگاه با طول

وابستگی‌های احتمالاتی بین متغیرها را نشان می‌دهند. این وابستگی‌های شریطی اغلب توسط روش‌های آماری و احتمالاتی مشخص ارزیابی می‌شوند (۳). شبکه‌های بیزی بر اساس تئوری احتمال Bayes به انتشار اطلاعات بین گره‌ها می‌پردازند و به راحتی می‌توانند به یک ابزار حمایتی تصمیم‌گیری در کمک به مدیریت منابع طبیعی تبدیل شوند. این شبکه‌ها قادرند داده‌های ناهمگن را در یک شبکه گرد آورده و همچنین ارتباطات میان آنها را برقرار کنند. مهم‌تر از همه اینکه می‌توانند جمع‌بندی کرده و ما را به یک تصمیم نهایی برسانند (۲۲). نظریه بیز روشی را برای ترکیب اطلاعات جدید با احتمال پیشین به منظور ایجاد یک احتمال شرطی جدید به نام احتمال پسین فراهم می‌کند.

$$P(A \cap B) = P(A)P(B|A) \text{ و } P(A \cap B) = P(B)P(A|B)$$

$$P(B)P(A|B) = P(A)P(B|A) \quad \text{در نتیجه:}$$

$$P(B|A) = \frac{P(B)P(A|B)}{P(A)} \quad \text{بنابراین:}$$

که در آن  $P(B)$  احتمال پیشین  $B$  (هیچ اطلاعاتی پیرامون  $A$  در نظر گرفته نمی‌شوند)،  $P(AB)$  احتمال وقوع  $A$  به شرط وقوع  $B$  را نشان می‌دهد (احتمال شرطی  $B$  و  $A$ )،  $P(A)$  احتمال حاشیه‌ای  $A$  است که به عنوان مقدار ثابت نرمالیزه شده عمل می‌کند و  $P(BA)$  احتمال شرطی  $A$  و  $B$  است و احتمال پسین نیز نامیده می‌شود. نظریه بیز برای به‌روزرسانی منحنی‌های فراوانی سیل و پیش‌بینی آنی سیلاب (نظیر فیلتر کلمن) بسیار مفید است (۲۸).

پس از آماده‌سازی داده‌ها، از Learn New Network برای آموزش ساختار شبکه استفاده شد، برای الگوریتم یادگیری، PC انتخاب شد. در بخش learn parameters ارتباط بین پارامترها و در بخش Update Beliefs، آموزش شبکه انجام شد. شبکه بیزی مورد استفاده در شکل ۳ دیده می‌شود (۳). پارامترهای انتخابی (محیط، طول حوضه، عامل شکل و طول آبراهه اصلی) برای برآورد دبی حداکثر با دوره بازگشت ۵۰ ساله هستند (۱۲).

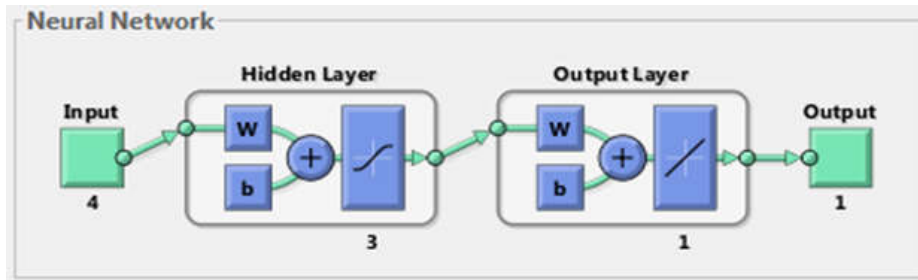
آموزش به ترتیب با استفاده از آزمون‌های گاما در نرم افزار WinGamma و آزمون  $M$  تعیین شد که نتایج آن در مقاله‌ای که توسط نویسندگان این پژوهش تدوین شده، مشخص است (۱۲). در نهایت، تحلیل منطقه‌ای فراوانی سیلاب با استفاده از درخت تصمیم، شبکه بیزی و شبکه عصبی با دو روش گشتاور خطی و حداکثر درستنمایی انجام شد.

### شبکه عصبی مصنوعی

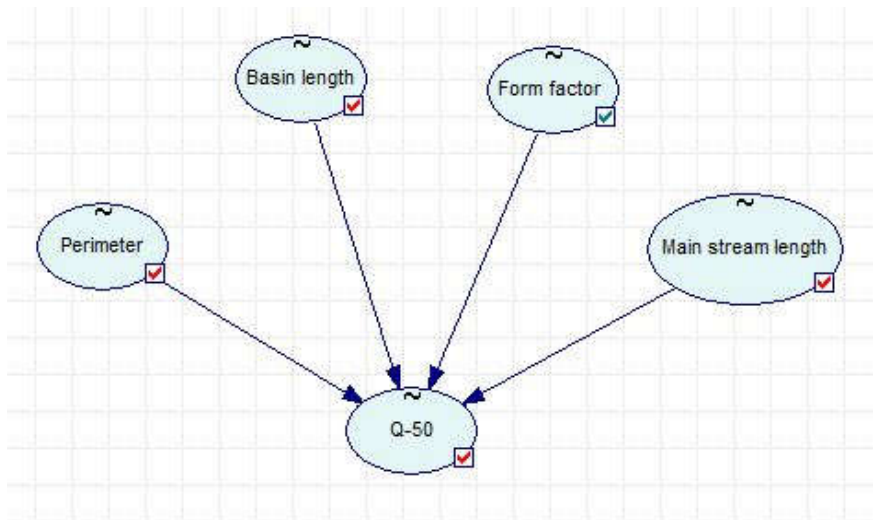
شبکه عصبی مصنوعی که الهام گرفته از شبکه عصبی بیولوژیکی مغز انسان است، شامل شبکه‌ای از نورون‌ها بوده که قادر است رفتار پیچیده بین نورون‌ها را نشان دهد (۳). یکی از مهم‌ترین انواع شبکه‌های عصبی، شبکه‌های پرسپترون چند لایه (Multilayer Perceptron) هستند. در هر لایه تعدادی نورون در نظر گرفته می‌شود که به نورون‌های لایه‌های مجاور وصل می‌شوند. در این شبکه‌ها ورودی مؤثر هر نورون، حاصل ضرب خروجی نورون‌های لایه قبل در وزن‌های میان آن نورون‌ها است (۱).

مدل‌سازی شبکه عصبی مصنوعی در محیط نرم افزار MATLAB انجام شد. از انواع شبکه عصبی، از شبکه پس‌انتشار پیش‌خور و پرسپترون چند لایه (MLP) استفاده شد. برای تابع آموزش، از الگوریتم بهینه‌سازی لوبنبرگ مارکواردت و برای تابع آموزشی تطبیقی نیز از کاهش شیب با مونتوم استفاده شد. برای لایه اول، تابع تانژانت هیپربولیک سیگموئید و برای لایه دوم، تابع خطی به‌عنوان تابع‌های انتقال انتخاب شدند (۱۸). شکل ۲ برای نمونه یکی از مدل‌های شبکه عصبی مصنوعی پیش‌خور با چهار ورودی و سه نورون لایه میانی و یک نورون خروجی را نشان می‌دهد.

شبکه‌های بیزی که با نام شبکه‌های اعتقاد یا باور نیز شناخته می‌شوند، متعلق به خانواده مدل‌های گرافیکی احتمالاتی هستند. این ساختارهای گرافیکی برای نشان دادن اطلاعات در یک حوزه دارای عدم قطعیت به کار می‌روند. به‌طور خاص هر گره در گراف نشان دهنده یک متغیر تصادفی بوده و شاخه‌ها یا کمان‌ها



شکل ۲. یک مدل شبکه عصبی مصنوعی (۱-۳-۴)



شکل ۳. مدل بیزی مورد استفاده

### درخت تصمیم

درخت تصمیم یک ساختار درختی است که فرایند تصمیم‌گیری جهت مشخص کردن دسته یک نمونه ورودی را نمایش می‌دهد. درخت تصمیم با دسته‌بندی نمونه‌ها از ریشه به سمت پایین رشد می‌کند و در نهایت به گره‌های برگ می‌رسد. در هر گره داخلی، سؤالی در رابطه با مقادیر ورودی مطرح می‌شود و به تعداد جواب‌های ممکن شاخه وجود دارد و در نهایت هر برگ این درخت یک کلاس یا دسته را مشخص می‌کند (۳).

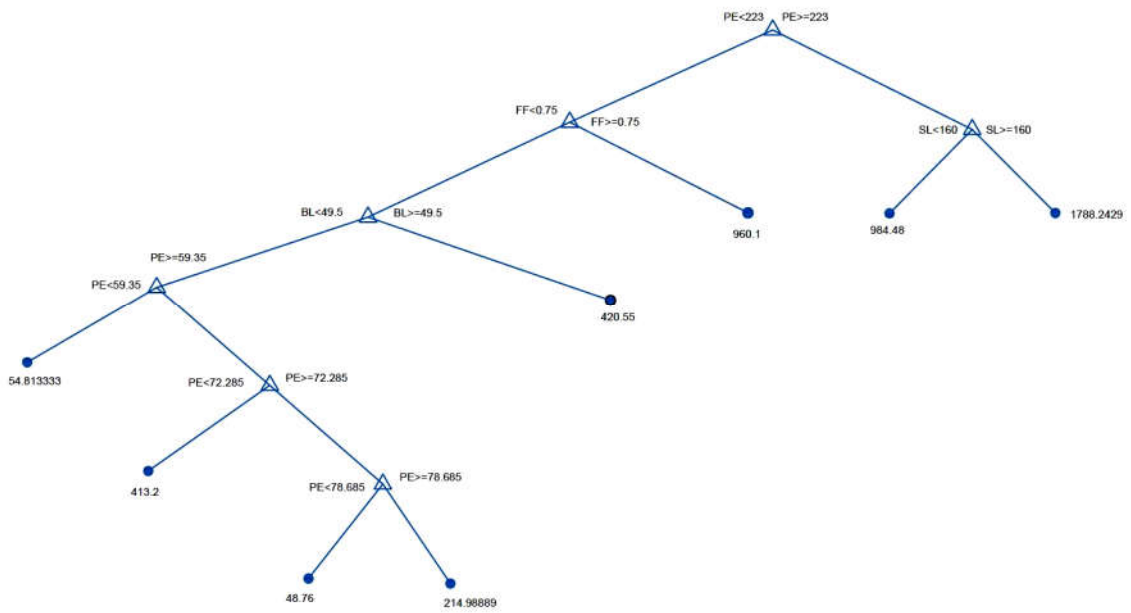
پس از آماده‌سازی داده‌ها، مدل اولیه درخت تصمیم با عنوان Default TREE در نرم‌افزار متلب با استفاده از روش‌های گشتاور خطی و حداکثر درست‌نمایی اجرا شد (شکل ۴ و ۵). نمودار درختی ایجاد شده تا حد ممکن گسترش یافته است. به‌منظور کاهش پیچیدگی مدل و جلوگیری از آموزش اضافی (overfitting)، ۱۰٪ از داده‌های آموزش جهت ارزیابی مدل انتخاب شده و سطحی از

درخت تصمیم با کمترین مقدار خطای ارزیابی مشخص می‌شود. با استفاده از دستور prune، درخت تصمیم اولیه از بهترین سطح، برش زده شد که نمودار درخت تصمیم (Optimal Tree) در شکل ۶ مشاهده می‌شود (۳). هدف از هرس کردن درخت کاهش خطا است. در شکل‌های مربوطه، PE (محیط)، BL (طول حوضه)، SL (طول آبراهه اصلی) و FF (عامل شکل) می‌باشند.

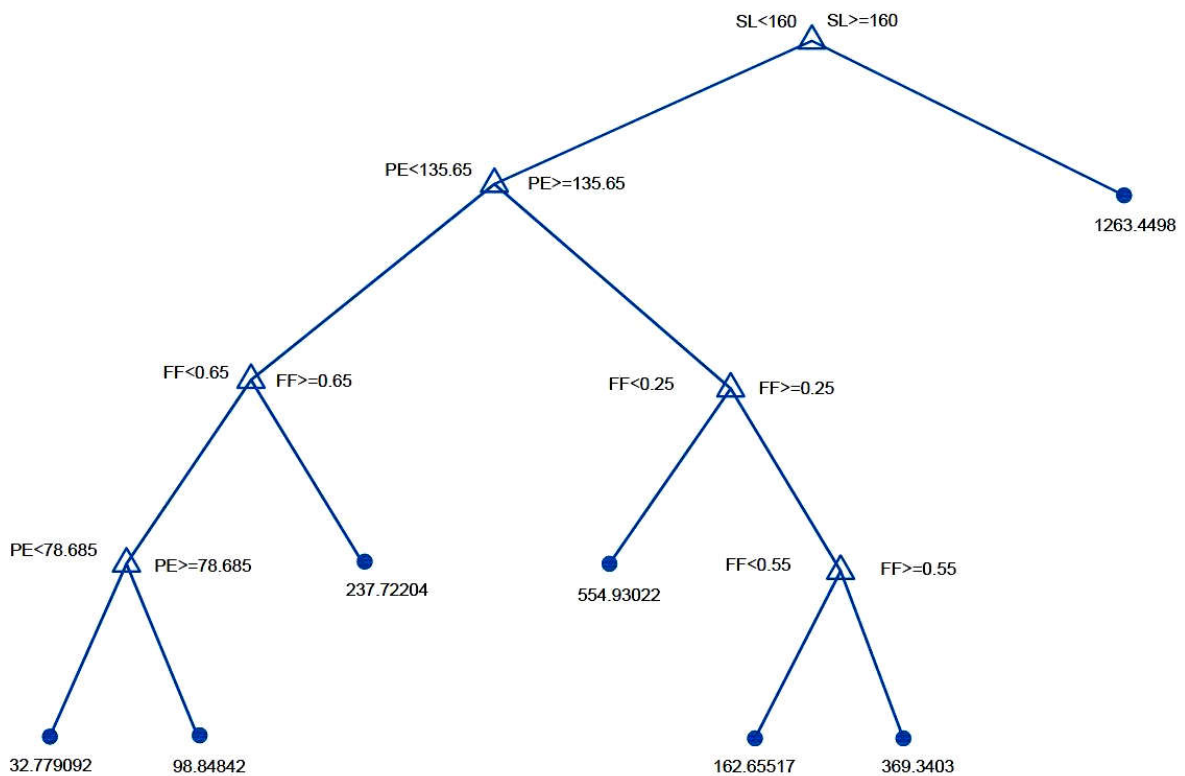
### گشتاورهای خطی

گشتاورهای خطی ترکیب‌های خطی گشتاورهای وزنی احتمالاتی (Probability Weighted Moments) هستند (رابطه ۱) و به داده پرت حساس نیستند و برای نمونه‌های کوچک داده‌های مشاهده‌ای غیر اریب هستند (۱۴):

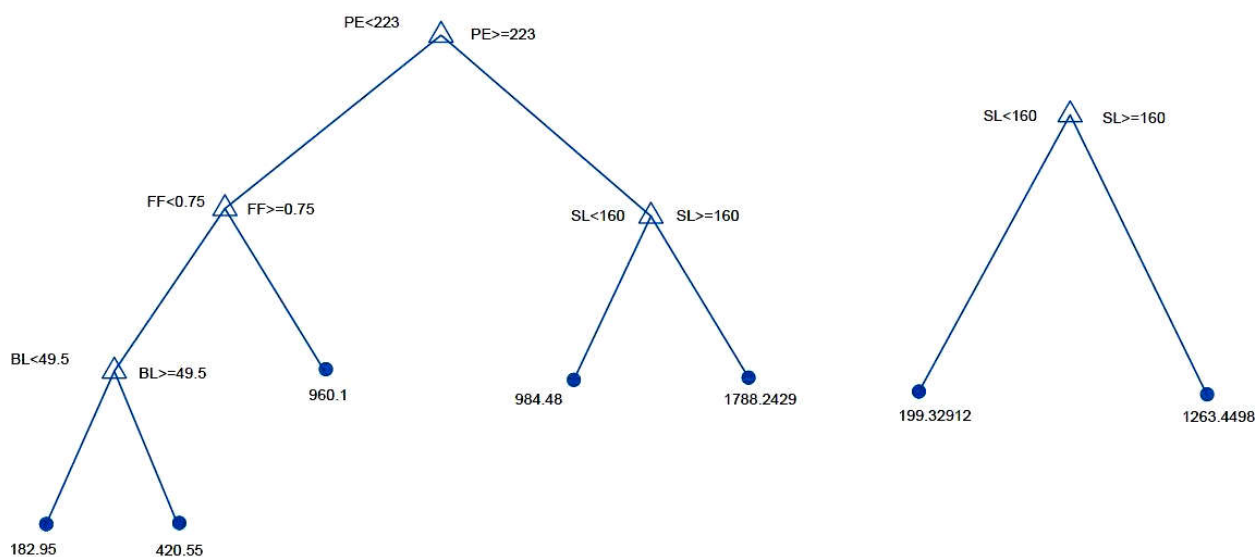
$$\beta_r = E \{ X [F(x)]^r \} \quad (1)$$



شکل ۴. نمودار درخت تصمیم با استفاده از روش حداکثر درستمایی



شکل ۵. نمودار درخت تصمیم با استفاده از روش گشتاور خطی



شکل ۶. نمودار درخت تصمیم پس از هرس

$$\lambda_2 = 2\beta_1 - \beta_0 \quad (۸)$$

$$\lambda_3 = 6\beta_2 - 6\beta_1 + \beta_0 \quad (۹)$$

$$\lambda_4 = 20\beta_3 - 30\beta_2 + 12\beta_1 - \beta_0 \quad (۱۰)$$

در رابطه‌های ۶ تا ۱۰ میانگین گشتاور خطی (L-mean) یا  $\lambda_1$  معیاری از تمایل مرکزی است که مشابه میانگین معمول است و انحراف معیار گشتاور خطی (L-standard deviation) یا  $\lambda_2$  معیاری از پراکندگی است و  $\lambda_3$  و  $\lambda_4$  سومین و چهارمین گشتاورهای خطی هستند. نسبت  $\lambda_2$  به  $\lambda_1$  یا  $\tau_2$  به‌عنوان ضریب تغییرات گشتاور خطی (L-coefficient of variation) نسبت  $\lambda_3$  به  $\lambda_2$  یا  $\tau_3$  به‌عنوان ضریب چولگی گشتاور خطی (L-skewness)، و نسبت  $\lambda_4$  به  $\lambda_2$  یا  $\tau_4$  به‌عنوان ضریب کشیدگی خطی (L-kurtosis) نامیده می‌شود (۱۰). آزمون برازندگی و نیکویی برازش مبتنی بر گشتاورهای خطی انجام می‌گیرد تا یکی از توزیع‌های رایج انتخاب و پارامترهای آن تخمین زده شود. سپس تناوب‌های سیل در داخل یک ناحیه بر اساس توزیع منطقه‌ای منتخب تعیین می‌شود. معیار برازندگی برای هر توزیع بر اساس گشتاورهای خطی تعیین و آماره Z نام می‌گیرد (۱۶). اگر تغییرپذیری ایستگاه‌ها یا

که در آن F(X) تابع توزیع تجمعی x است. تخمین‌های نمونه ناریب از PWM برای هر توزیع از روابط زیر به‌دست می‌آید:

$$\beta_0 = \frac{1}{n} \sum_{j=1}^n X_j \quad (۲)$$

$$\beta_1 = \sum_{j=1}^{n-1} \left[ \frac{(n-j)}{n(n-1)} \right] X_j \quad (۳)$$

$$\beta_2 = \sum_{j=1}^{n-2} \left[ \frac{(n-j)(n-j-1)}{n(n-1)(n-2)} \right] X_j \quad (۴)$$

$$\beta_3 = \sum_{j=1}^{n-3} \left[ \frac{(n-j)(n-j-1)(n-j-2)}{n(n-1)(n-2)(n-3)} \right] X_j \quad (۵)$$

که در آن  $X_j$  داده‌های مرتب‌شده جریان با  $X_1$  به‌عنوان بزرگ‌ترین داده مشاهداتی و  $X_n$  به‌عنوان کوچک‌ترین داده است (۱۶ و ۲۶).

چهار گشتاور خطی نخست که به‌عنوان ترکیبات خطی گشتاور وزنی احتمال بیان می‌شوند عبارت‌اند از:

$$\lambda_1 = \beta_0 \quad (۶)$$

$$\lambda_2 = 2\beta_1 - \beta_0 \quad (۷)$$



تیپ ۳ بهترین توزیع احتمال منطقه است (۱۲). ولی در روش گشتاورهای خطی بر اساس معیار نکویی برازش ( $Z_{dist}$ )، توزیع نرمال تعمیم‌یافته بهترین توزیع برای کل منطقه بود (۱۱) که با پژوهش مصفایی و همکاران (۲۱) و یانگ و همکاران (۳۰) و الله بخشیان فارسانی و همکاران (۶) مطابقت دارد. بهترین ترکیب ورودی از میان پارامترهای فیزیوگرافی، اقلیمی، سنگ‌شناسی و کاربری اراضی با استفاده از آزمون گاما انتخاب شد. نتایج آزمون گاما نشان داد که بهترین ترکیب ورودی با متغیرهای محیط حوضه، طول حوضه، عامل شکل و طول آبراه اصلی در بین سایرین است که در مقاله‌ای که توسط نویسندگان پژوهش حاضر نوشته شده، تحلیل شده است (۱۲). نتایج مدل‌های روش حداکثر درست‌نمایی نشان می‌دهد که در دوره آموزش مدل درخت تصمیم بهترین کارایی را نشان داده است ولی در دوره آزمون، بهترین کارایی را مدل بیزی داشته است که بر اساس مطالعه مورسی و همکاران (۲۰) دارای عملکرد رضایت بخش ( $0.5 < R^2 \leq 0.6$ ) برای پیش‌بینی سیل است (۷). در روش گشتاور خطی، در دوره آموزش، باز هم مدل درخت تصمیم، بالاترین کارایی داشته است و در دوره آزمون، مدل‌های بیزی و درخت تصمیم عملکرد مناسب‌تری از شبکه عصبی مصنوعی نشان داده‌اند که بر اساس مقاله مورسی و همکاران دارای عملکرد بسیار خوب ( $0.7 < R^2 \leq 1$ ) برای پیش‌بینی سیل است (۷). عملکرد خوب مدل بیزی با نتایج حسینی و همکاران (۱۷) و پاپروتنی و مورالس-ناپولس (۲۴) مطابقت دارد که نشان داده‌اند که این مدل قادر به ارائه نتایج قابل اعتماد و با دقت بالا برای برآورد سریع دبی‌های حداکثر است. اشکال ۷ و ۸ نیز عملکرد ضعیف‌تر شبکه عصبی مصنوعی را به وضوح نشان می‌دهد که با نتایج اعلمی و همکاران (۲) همسو است که نشان داده‌اند شبکه بیزی در نقاط حداقل و حداکثر عملکرد بهتری نسبت به مدل عصبی فازی دارد. ولی از لحاظ دقت، مدل شبکه عصبی بیزی با بیشترین ضریب همبستگی (۰/۸۳۲) و کمترین ریشه میانگین مربعات

فضای پراکنش ایستگاه‌ها بزرگ باشد، احتمال تعلق این ایستگاه‌ها به یک مجموعه واحد را می‌توان به وسیله آزمون همگنی گشتاورهای خطی بررسی کرد (۱۵). محاسبه گشتاورهای خطی نمونه‌های غیر اریب و نسبت‌های گشتاور خطی (نوع اول تا چهارم)، آزمون ناهمگنی و ناهماهنگی در نرم افزار R و با به‌کارگیری پکیج گشتاورهای خطی انجام شد (۲۱).

## نتایج

آماره غیر همگنی (Di) که هاسکینگ و والیس آن را برای تشخیص ایستگاه‌های ناهماهنگ پیشنهاد کرده‌اند، برای کلیه ایستگاه‌های مورد بررسی به دست آمد. بنا به تعریف، ایستگاه‌هایی که دارای آماره غیرهمگنی بیش از سه باشند ایستگاه پرت شناخته می‌شوند و از مجموعه ایستگاه‌های مطالعاتی کنار گذاشته می‌شوند. بدین ترتیب، ایستگاه بطاری ناسازگار شناخته شد. آزمون غیرهمگنی گشتاورهای خطی، که مبتنی بر دبی‌های حداکثر سالانه ایستگاه‌هاست، مشخص کرد که مقادیر این آماره، برای دو حوضه بزرگ‌تر از عدد سه است و کل منطقه به صورت یک منطقه غیرهمگن است (جدول ۳). در جدول ۱ معیارهای نکویی برازش  $Z_{dist}$  برای توزیع‌های آماری مختلف و در جدول ۲ مقادیر پارامترها برای بهترین توزیع منتخب ارائه شده است.

نتایج مقایسه کارایی سه مدل (یعنی درخت تصمیم، شبکه عصبی مصنوعی و مدل بیزی) برای تحلیل منطقه‌ای فراوانی سیلاب با استفاده از دو روش حداکثر درست‌نمایی و گشتاورهای خطی برای دو حوزه آبخیز کرخه و کارون در جدول ۴ و شکل‌های ۷ و ۸ نشان داده شده است.

## بحث و نتیجه‌گیری

در روش حداکثر درست‌نمایی بر اساس معیار نکویی برازش (کولموگروف-اسمیرنوف) مشخص شد که توزیع لوگ پیرسون

جدول ۱. معیارهای نکویی برازش  $Z_{dist}$  برای توزیع‌های آماری مختلف

توزیع آماری	لجستیک	مقادیر حدی	نرمال	پاره توی	پیرسون
تعمیم یافته	تعمیم یافته	تعمیم یافته	تعمیم یافته*	تعمیم یافته	نوع سوم
کل منطقه	۲/۹۰	۱/۲۹	-۰/۰۱	-۳/۱۴	-۲/۲۸

\*بهترین توزیع

جدول ۲. مقادیر پارامتر برای بهترین توزیع منتخب

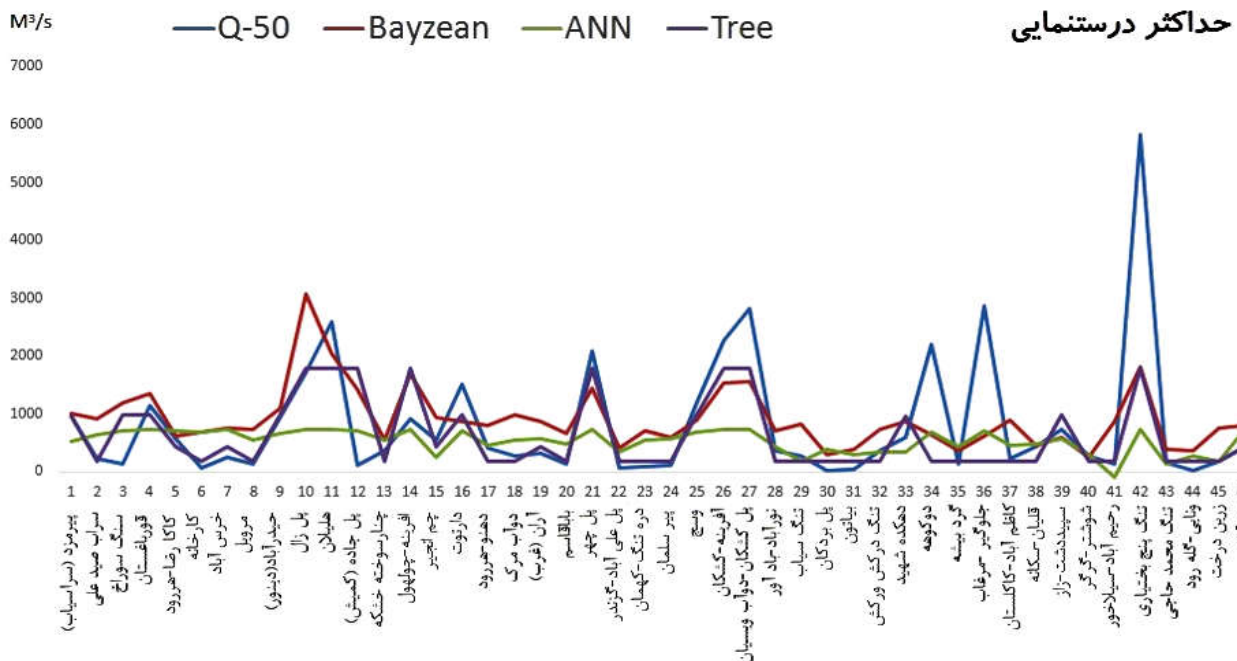
توزیع منتخب		مقادیر پارامتر توزیع	
کل منطقه	نرمال تعمیم یافته	$X_i$	alpha
		۰/۷۷	۰/۶۶
			$k$
			-۰/۶۳

جدول ۳. مقادیر آزمون‌های ناهمگنی ایستگاه‌های مورد بررسی

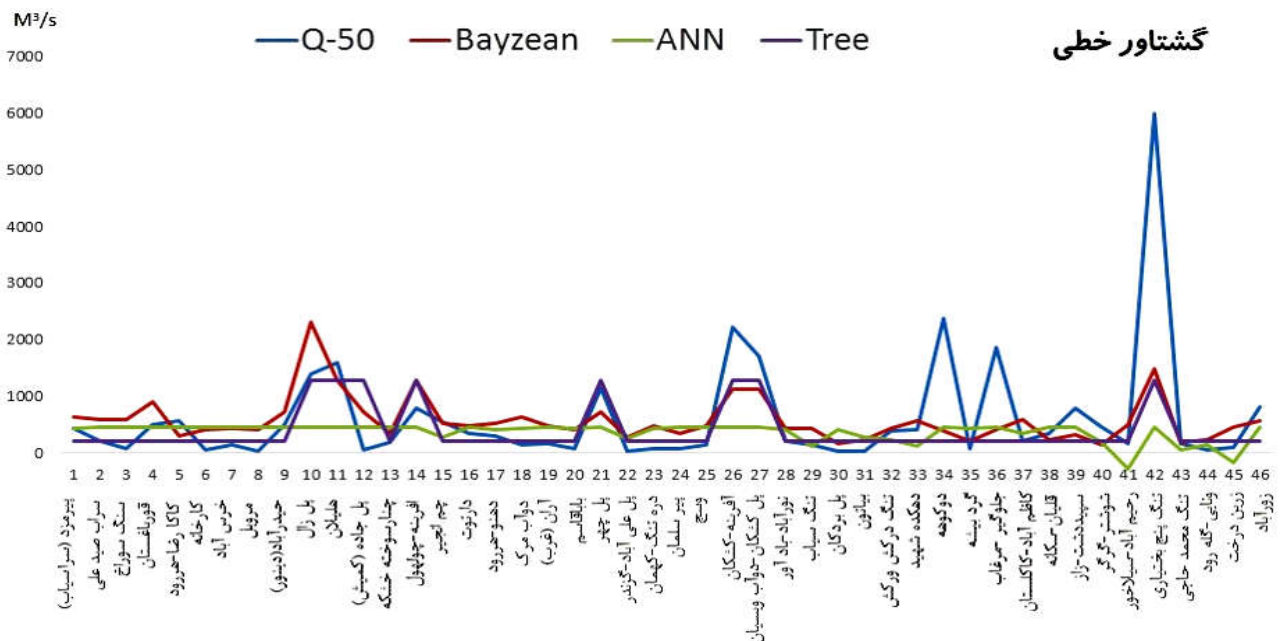
مقادیر آزمون ناهمگنی		
$H_3$	$H_2$	$H_1$
۴/۸۲	۶/۳۴	۶/۹۸

جدول ۴. ارزیابی مدل‌های مورد استفاده در آنالیز منطقه‌ای سیلاب با استفاده از دو روش گشتاور خطی و حداکثر درست‌نمایی

معیار کارایی	RMSE	$R^2$	RMSE	$R^2$
روش	حداکثر درست‌نمایی			
دوره	آموزش		آزمون	
بیزی	۶۴۷/۸۴	۰/۴۷	۱۳۰۳/۵۲	۰/۵۷
مدل شبکه عصبی مصنوعی	۷۴۵/۷۰	۰/۲۷	۱۴۹۰/۹۳	۰/۳۸
درخت تصمیم	۴۷۵/۶۲	۰/۶۵	۱۳۶۳/۰۸	۰/۴۸
روش	گشتاور خطی			
دوره	آموزش		آزمون	
بیزی	۶۱۲/۹۴	۰/۵۵	۱۳۸۷/۲۱	۰/۷
مدل شبکه عصبی مصنوعی	۵۵۱/۴۸	۰/۰۴۶	۱۵۸۰/۶۵	۰/۱۸
درخت تصمیم	۳۴۲/۸۷	۰/۶۳	۱۴۳۱/۹۵	۰/۷



شکل ۷. مقایسه دبی حداکثر با دوره بازگشت ۵۰ ساله (Q-50) با نتایج مدل‌های شبکه عصبی مصنوعی، درخت تصمیم و شبکه بیزی در آنالیز منطقه‌ای سیلاب با استفاده از روش حداکثر درست‌نمایی (رنگی در نسخه الکترونیکی)



شکل ۸. مقایسه دبی حداکثر با دوره بازگشت ۵۰ ساله (Q-50) با نتایج مدل‌های شبکه عصبی مصنوعی، درخت تصمیم و شبکه بیزی در آنالیز منطقه‌ای سیلاب با استفاده از روش گشتاور خطی (رنگی در نسخه الکترونیکی)

تحقیقات مختلف از جمله ملکی نژاد و همکاران (۱۹) همسو است. این پژوهشگران بیان کردند که کاربرد گشتاور خطی در رویکرد منطقه‌سازی برای برآورد دبی سیل با دوره‌های بازگشت مشخص در هیدرولوژی و مهندسی منابع آب مناسب است. نتایج تحلیل منطقه‌ای فراوانی سیلاب نشان داد که مدل بیزی با روش گشتاور خطی بهترین برآورد را در مقایسه با روش‌های دیگر دارد. روش مورد استفاده در این پژوهش می‌تواند در سایر مناطق جهان برای پیش‌بینی سیلاب‌های بزرگ‌تر با دوره بازگشت بیش از ۵۰ سال مورد استفاده قرار گیرد که برای طراحی سازه‌های هیدرولیکی بسیار مهم است.

خطا در مرحله صحت‌سنجی در اولویت قرار گرفته است. عملکرد بهتر درخت تصمیم از شبکه عصبی مصنوعی نیز با نتایج تحقیقات دیگر همسو است که نشان داده‌اند مدل‌های درختی نسبت به روش شبکه عصبی مصنوعی دقت بالاتری دارند و همچنین تفسیر نتایج مدل درختی بسیار راحت‌تر است. در مقایسه عملکرد گشتاور خطی و حداکثر درست‌نمایی، دو مدل بیزی و درخت تصمیم در گشتاور خطی عملکرد بهتری نسبت به حداکثر درست‌نمایی نشان داده‌اند. نتایج ولایتی نژاد و همکاران (۲۹) که نگاهی بر پژوهش‌های صورت گرفته در زمینه تحلیل منطقه‌ای سیلاب انداخته‌اند نشان داده است که روش گشتاورهای خطی کارایی لازم جهت برآورد دبی حداکثر را دارد که نتایج تحقیق کنونی نیز با آن همسو است. کارایی مناسب گشتاور خطی را می‌توان به مزایای آن در منطقه‌ای‌سازی از جمله عدم حساسیت آن‌ها به داده‌های پرت ارتباط داد که با

#### منابع مورد استفاده

1. Aalami, M. and H. Hosseinzadeh. 2014. Modeling rainfall– runoff process in Lighvan Chai basin using conditional threshold temperature neuron. *Water and Soil Science* 20(2): 97-110 (In Farsi).
2. Aalami, M. T., L. Malekani and S. Emami. 2012. Comparison of the efficiency of adaptive neurofuzzy system and Bayesian network methods in runoff simulation. *In: 9th International Congress of Civil Engineering, Isfahan University of Technology, Isfahan, Iran* (In Farsi).
3. Aalami, M.T., S. Sadeghfam, M. Fazelifard and L. Taghipour. 2014. *Data Series Modeling*. Tabriz University Press, Tabriz, Iran (In Farsi).
4. Ahmadi, F., F. Radmanesh and R. Mirabbasi Najafabadi. 2016. Comparing the performance of Support Vector Machines and Bayesian networks in predicting daily river flow (Case study: Baranduz Chai River). *Journal of Water and Soil Conservation* 22(6): 171-186 (In Farsi).
5. Allahbakhshian-Farsani, P., M. Vafakhah and H. Khosravi-Farsani. 2020. Regional flood frequency analysis through some machine learning models in semi-arid regions. *Water Recourses Management* 34: 2887-2909.
6. Allahbakhshian-Farsani, P., M. Vafakhah and H. Rahimiani Iranshahi. 2016. Regional flood analysis using linear moment estimation methods and maximum likelihood in north Karun watershed. *In: 11<sup>th</sup> National Conference on Watershed Management Science and Engineering, Yasuj, Iran* (In Farsi).
7. Ayele, G. T., E. Z. Teshale, B. Yu, I. D. Rutherford and J. Jeong. 2017. Streamflow and sediment yield prediction for watershed prioritization in the Upper Blue Nile River Basin, Ethiopia. *Water* 9(10): 782.
8. Aziz, K., A. Rahman, G. Fang and S. Shrestha. 2014. Application of artificial neural networks in regional flood frequency analysis: a case study for Australia. *Stochastic Environmental Research and Risk Assessment* 28: 541-554.
9. Choubin, B., A. Malekian, S. Samadi, S. Khalighi Sigaroodi and F. Sajedi Hosseini. 2017. An ensemble forecast of semiarid rainfall using large scale climate predictors. *Meteorological Applications* 24(3): 376-386.
10. Eslamian, S. and S. Boroojeni. 2003. Application of L-moment theory in flood frequency analysis of central catchments of Iran. *Journal of Water and Soil Science* 7: 1-17 (In Farsi).
11. Ghaderi, K. 2019. *Regional Flood Analysis Using Artificial Intelligence Methods and L-moment*. Ph.D. Dissertation, Islamic Azad University Science and Research Branch, Iran (In Farsi).

12. Ghaderi, K., B. Motamedvaziri, M. Vafakhah and A. A. Dehghani. 2019. Regional flood frequency modeling: a comparative study among several data-driven models. *Arabian Journal of Geosciences* 12(18): 1-9.
13. Haktanir, T., H. Citakoglu and N. Seckin. 2016. Regional frequency analyses of successive-duration annual maximum rainfalls by L-moments method. *Hydrological Sciences Journal* 61(4): 647-668.
14. Hosking, J. R. M. 2000. FORTRAN Routines for Use with the Method of L-Moments. Version 3.04. IBM Research.
15. Hosking, J. R. M. and J. R. Wallis. 1993. Some statistics useful in regional frequency analysis. *Water Resources Research* 29(2): 271-281.
16. Hosking, J. and J. Wallis. 1997. Regional Frequency Analysis. Cambridge University Press, Cambridge.
17. Hosseini, F. S., B. Choubin, A. Mosavi, N. Nabipour, S. Shamsirband, H. Darabi and A. T. Haghighi. 2020. Flash-flood hazard assessment using ensembles and Bayesian-based machine learning models: application of the simulated annealing feature selection method. *Science of the Total Environment* 7(11): 135-161.
18. Kia, S. M. 2012. Neural Networks in MATLAB, Kian Rayaneh Sabz Publications, Iran (In Farsi).
19. Malekinezhad, H., H. Nachtnebel and A. Klik. 2011. Regionalization approach for extreme flood analysis using L-moments. *Journal of Agricultural Science and Technology* 13: 1183-1196.
20. Moriasi, D. N., J. G. Arnold, M. W. Van Liew, R. L. Bingner, R. D. Harmel and T. L. Veith. 2007. Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. *Transactions of the American Society of Agricultural and Biological Engineers* 50(3): 885-900.
21. Mosaffaie, J., D. Akhzari, S. Rashvand and J. Ataei. 2015. Regional flood frequency analysis using multiple regression method (Case study: hydrometric stations of Qazvin province). *Journal of Range and Watershed Management* 68(4): 821-833 (In Farsi).
22. Moshari, Kh. And R. Daneshfaraz. 2015. Comparison of Bayesian network performance with other intelligent models in estimating Qorechai river flow. In: 10th International Congress of Civil Engineering, University of Tabriz, Tabriz, Iran (In Farsi).
23. Motevalli, A. and M. Vafakhah. 2016. Flood hazard mapping using synthesis hydraulic and geomorphic properties at watershed scale. *Stochastic Environmental Research and Risk Assessment* 30(7): 1889-1900.
24. Paprotny, D. and O. Morales-Nápoles. 2017. Estimating extreme river discharges in Europe through a Bayesian network. *Hydrology and Earth System Sciences* 21(6): 2615-2636.
25. Seckin, N. and A. Guven. 2012. Estimation of peak flood discharges at ungauged sites across Turkey. *Water Resources Management* 26: 2569-2581.
26. Seckin, N., M. Cobaner, R. Yurtal and T. Haktanir. 2013. Comparison of artificial neural network methods with L-moments for estimating flood flow at ungauged sites: the case of East Mediterranean River Basin, Turkey. *Water Resources Management* 27: 2103-2124.
27. Sharifi Garmdareh, E., M. Vafakhah and S. Eslamian. 2019. Assessment the Performance of Support Vector Machine and Artificial Neural Network Systems for Regional Flood Frequency Analysis (A Case Study: Namak Lake Watershed). *Journal of Water and Soil Science* 23 (1): 351-366 (In Farsi).
28. Vafakhah, M. and A. Moghaddamnia. 2019. Flood Control. Tarbiat Modares University Press, Tehran, Iran. (In Farsi).
29. Velayati Nejad, A. S., M. Habibnejad Roshan and J. Mosafaei. 2014. A Look at the Researches on Regional Flood Analysis. In: Iran's First National Congress of Biology and Natural Sciences, Tehran, Iran (In Farsi).
30. Yang, T., C. Y. Xu, Q. X. Shao and X. Chen. 2010. Regional flood frequency and spatial patterns analysis in the Pearl River Delta region using L-moments approach. *Stochastic Environmental Research and Risk Assessment* 24(2): 165-182.

## Comparison of Artificial Neural Network, Decision Tree and Bayesian Network Models in Regional Flood Frequency Analysis using L-moments and Maximum Likelihood Methods in Karkheh and Karun Watersheds

K. Ghaderi<sup>1</sup>, B. Motamedvaziri<sup>1</sup>, M. Vafakhah<sup>2\*</sup> and A. A. Dehghani<sup>3</sup>

(Received: February 8-2021; Accepted: May 17-2021)

### Abstract

Proper flood discharge forecasting is significant for the design of hydraulic structures, reducing the risk of failure, and minimizing downstream environmental damage. The objective of this study was to investigate the application of machine learning methods in Regional Flood Frequency Analysis (RFFA). To achieve this goal, 18 physiographic, climatic, lithological, and land use parameters were considered for the upstream basins of the hydrometric stations located in Karkheh and Karun watersheds (46 stations with a statistical length of 21 years). The best Probability Distribution Function (pdf) was then determined using the Kolmogorov-Smirnov test at each station to estimate the flood discharge with a return period of 50-year using maximum likelihood methods and L-moments. Finally, RFFA was performed using a decision tree, Bayesian network, and artificial neural network. The results showed that the log Pearson type 3 distribution in the maximum likelihood method and the generalized normal distribution in the L moment method are the best possible regional pdfs. Based on the gamma test, the parameters of the perimeter, basin length, shape factor, and mainstream length were selected as the best input structure. The results of regional flood frequency analysis showed that the Bayesian model with the L moment method ( $R^2 = 0.7$ ) has the best estimate compared to other methods. Decision tree and artificial neural network were in the following ranks.

**Keywords:** Modeling, Instantaneous maximum discharge, Machine learning models, L-moments, Bayesian network

1. Department of Forest, Range and Watershed Management, Faculty of Natural Resources and Environment, Science and Research Branch, Islamic Azad University, Tehran, Iran.

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

3. Department of Water Engineering, Gorgan University of Agricultural Sciences and Natural Resources, Gorgan, Iran.

\* Correspondence author: Email: vafakhah@modares.ac.ir