

## ارزیابی کارایی سامانه‌های ماشین بردار پشتیبان و شبکه عصبی مصنوعی در تحلیل منطقه‌ای سیلاب (مطالعه موردی: حوضه آبخیز دریاچه نمک)

ابراهیم شریفی گرم‌دره<sup>۱</sup>، مهدی وفاخواه<sup>۱\*</sup> و سید سعید اسلامیان<sup>۲</sup>

(تاریخ دریافت: ۱۳۹۴/۳/۲؛ تاریخ پذیرش: ۱۳۹۷/۳/۱۹)

### چکیده

تخمین دبی سیلاب با دوره بازگشت مختلف، یکی از مهم‌ترین عوامل لازم برای طراحی و اجرای سازه‌های آبی است. از طرفی بسیاری از رودخانه‌های موجود در حوضه‌های آبخیز ایران فاقد آمار و اطلاعات آب‌سنجی کامل و دقیق هستند. در چنین مواردی یکی از راه‌حل‌های مناسب برای برآورد دبی‌های سیلابی با دوره بازگشت مختلف، انجام تحلیل منطقه‌ای سیلاب است. در پژوهش حاضر ۵۵ ایستگاه آب‌سنجی مورد استفاده قرار گرفتند. برای این منظور ابتدا دبی‌های حداکثر لحظه‌ای ایستگاه‌های منتخب در دوره بازگشت‌های مختلف با استفاده از نرم‌افزار Easy Fit برآورد شد. سپس متغیرهای مؤثر بر دبی‌های سیلابی جمع‌آوری و متغیرهای ورودی مدل با استفاده از آزمون گاما و به کمک نرم‌افزار WinGamma تعیین شدند. در نهایت مدل‌سازی داده‌ها با استفاده از ماشین بردار پشتیبان، شبکه عصبی مصنوعی و رگرسیون چند متغیره غیرخطی انجام پذیرفت. 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 به‌منظور پیش‌بینی دبی‌های حداکثر لحظه‌ای در حوضه آبخیز دریاچه نمک برخوردار است.

واژه‌های کلیدی: ایستگاه آب‌سنجی، دوره بازگشت، آزمون گاما، ماشین بردار پشتیبان، حوضه آبخیز دریاچه نمک

۱. گروه مهندسی آبخیزداری، دانشکده منابع طبیعی، دانشگاه تربیت مدرس

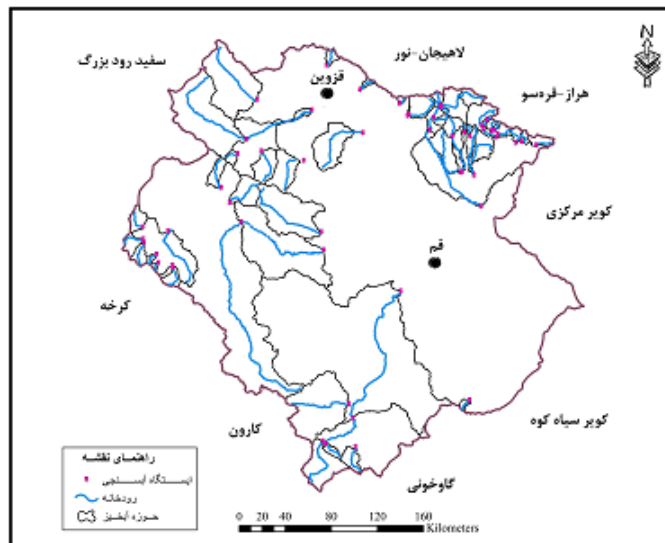
۲. گروه مهندسی آب، دانشکده مهندسی کشاورزی، دانشگاه صنعتی اصفهان، اصفهان

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

## مقدمه

سیلاب، پدیده‌ای است که در آن جریان رودخانه و سطح آب به‌صورت غیرمنتظره افزایش پیدا کرده و باعث خسارات مالی و جانی شود. سیل طرح، سیلابی است که بر اساس آن طراحی یک سازه آبی، انجام می‌شود (۲). برای تعیین حداکثر دبی سیلاب روش‌های مختلفی از جمله آب‌نمود و آب‌نمود واحد، تحلیل فراوانی سیلاب‌های مختلف یک حوضه آبخیز، فرمول‌های تجربی و محاسبه حداکثر سیل محتمل قابل استفاده است (۲۸). جمع‌آوری داده به‌منظور تعیین سیلاب طرح به ایستگاه‌های آب‌سنجی نیاز دارد. از طرفی هزینه احداث ایستگاه‌ها زیاد و زمان زیادی برای جمع‌آوری داده‌ها لازم است. برای حل این مشکل هیدرولوژیست‌ها، از ابزاری با عنوان تحلیل منطقه‌ای استفاده می‌کنند (۴). برآورد دبی حوضه‌های آبخیز بدون آمار با استفاده از انتقال اطلاعات از آبخیزهای دارای آمار، تحلیل منطقه‌ای نامیده می‌شود (۲۳). معمولاً به‌منظور تحلیل منطقه‌ای سیلاب روش‌های استدلالی، سیل شاخص و رگرسیونی مورد استفاده قرار می‌گیرد. بیشتر روش‌های بالا رابطه بین دبی سیلاب و خصوصیات حوضه را به‌صورت خطی در نظر می‌گیرند. از طرفی رابطه بین بیشتر پدیده‌های هیدرولوژیکی به شکل غیرخطی است (۳). از این‌رو، رویکرد به مدل‌های هوشمند به‌عنوان فنون جدید و ابزاری توانمند با انعطاف‌پذیری بالا در فرایندهای آب‌شناختی (Hydrology) نظیر پیش‌بینی، افزایش یافته است. زیرا این روش‌ها، کمتر به ابزار و داده‌های مدل‌های فیزیکی نیاز داشته و قادرند فرایند غیرخطی را مدل‌سازی کنند (۲۶). شبکه‌های عصبی مصنوعی (ANNs)، نروفازی (ANFIS)، برنامه‌ریزی بیان ژن (GEP)، ماشین بردار پشتیبان (SVM) و ... از انواع ماشین‌های یادگیری هستند که به‌منظور مدل‌سازی فرایندهای مختلف هیدرولوژی از جمله تحلیل منطقه‌ای سیلاب مورد استفاده قرار گرفته‌اند. داوسون و همکاران (۶)، در پژوهشی شبکه‌های عصبی مصنوعی را به‌منظور پیش‌بینی وقایع سیل T ساله و تخمین سیل شاخص در ۸۵۰ حوضه آبخیز در سراسر انگلستان استفاده کردند. نتایج به‌دست آمده از این پژوهش نشان

داد که شبکه‌های عصبی مصنوعی نسبت به روش مدل‌سازی رگرسیون چندمتغیره دارای دقت بالاتری هستند. شوو و اوادا (۳۵)، به‌منظور تحلیل منطقه‌ای سیلاب در حوضه‌های آبخیز در ایالت کبک کانادا از روش‌های مدل‌سازی نروفازی، شبکه عصبی مصنوعی و رگرسیون چندمتغیره غیرخطی استفاده کردند. مقایسه نتایج به‌دست آمده از این پژوهش نشان داد که روش مدل‌سازی نروفازی از قابلیت بالاتری نسبت به دو روش مدل‌سازی دیگر به‌منظور برآورد سیلاب حوضه‌های فاقد ایستگاه آب‌سنجی برخوردار است. سکین و گوون (۳۲)، به‌منظور پیش‌بینی دبی‌های حداکثر سیلابی با استفاده از داده‌های ۴۵۳ حوضه در سرتاسر ترکیه، روش برنامه‌ریزی بیان ژن را به‌کار بردند. نتایج حاصل از این پژوهش نشان داد که روش برنامه‌ریزی بیان ژن دقت بالاتری نسبت به برنامه‌ریزی ژنتیک خطی و رگرسیون خطی دارد. عزیز و همکاران (۳)، شبکه عصبی مصنوعی را برای انجام تحلیل منطقه‌ای سیلاب در شرق استرالیا استفاده کردند. برای این منظور تعداد ۴۵۲ حوضه آبخیز و دو متغیر ورودی شامل مساحت حوضه و شدت بارندگی به‌کار گرفته شد. نتایج حاصل از این پژوهش نشان داد که شبکه عصبی مصنوعی از دقت بالاتری نسبت به روش‌های سنتی رگرسیونی برخوردار است. رضایی و همکاران (۳۱)، مدل‌سازی منطقه‌ای دبی‌های اوج را در زیرحوضه‌های آبخیز سد سفیدرود با استفاده از شبکه عصبی مصنوعی انجام دادند. مقایسه نتایج دبی‌های اوج مشاهده‌ای و برآورد شده نشان داد که شبکه عصبی مصنوعی برتری قابل توجهی بر روش‌های سنتی (رگرسیون چندمتغیره) دارد. خسروی و همکاران (۱۹)، پژوهشی را به‌منظور مقایسه کارایی شبکه عصبی مصنوعی و رگرسیون چندمتغیره غیرخطی در پیش‌بینی سیلاب در حوضه آبخیز طالقان انجام دادند. نتایج حاصل از این پژوهش نشان داد که روش مدل‌سازی شبکه عصبی مصنوعی عملکرد بهتری نسبت به روش رگرسیون چندمتغیره غیرخطی دارد. در سال‌های اخیر، مدل‌های ماشین بردار پشتیبان، به‌عنوان یکی از جدیدترین مدل‌های هوش مصنوعی در سطوح وسیعی از علوم مختلف، به‌کار گرفته شده‌اند. از طرفی مدل ماشین بردار



شکل ۱. موقعیت ایستگاه‌های مورد مطالعه در حوضه آبخیز دریاچه نمک

خروجی هدف استفاده کرد، آزمون گاما است (۲۴). بنابراین هدف از پژوهش حاضر، مقایسه عملکرد ماشین بردار پشتیبان با شبکه های عصبی مصنوعی و رگرسیون چندمتغیره غیرخطی به منظور مدل سازی دبی های حداکثر لحظه ای در حوضه آبخیز دریاچه نمک بر اساس توانایی آزمون گاما در انتخاب متغیرهای ورودی مدل ها است.

## مواد و روش ها

### منطقه مورد مطالعه

حوضه آبخیز دریاچه نمک در طول جغرافیایی  $48^{\circ} 00'$  تا  $53^{\circ} 00'$  شرقی و عرض جغرافیایی  $33^{\circ} 00'$  تا  $36^{\circ} 00'$  شمالی با مساحت  $92550$  کیلومتر مربع واقع شده است. شکل (۱) موقعیت این حوضه را در سطح کشور و پراکنش ایستگاه‌های آب سنجی مورد استفاده در پژوهش را نشان می دهد. طیف ارتفاعی از  $800$  متر در اطراف دریاچه نمک تا  $4375$  متر در ارتفاعات جاجرود متغیر است. حوضه آبخیز دریاچه نمک شامل چند حوضه آبخیز اصلی کوچک تر شور، کرج، جاجرود، قره چای، کویر میقان و قم رود است. آبخیز رودخانه های شور، کرج و جاجرود در شمال حوضه آبخیز و در دامنه های جنوبی

پشتیبان برای تحقیقات مرتبط با تحلیل منطقه ای سیلاب، زیاد استفاده نشده است. زکریا و شبری (۴۱)، روش مدل سازی ماشین بردار پشتیبان را به منظور پیش بینی دبی های حداکثر در دوره بازگشت های  $10$  و  $100$  ساله با استفاده آمار  $88$  ایستگاه آب سنجی در مالزی به کار گرفتند و نتایج آن را با روش رگرسیون چندمتغیره مقایسه کردند. نتایج به دست آمده از این پژوهش نشان داد که ماشین بردار پشتیبان عملکرد بهتری نسبت به رگرسیون چندمتغیره دارد. گایزا و گن (۱۳)، به منظور تحلیل منطقه ای سیلاب در کانادا دو روش مدل سازی مبتنی بر هوش مصنوعی شامل رگرسیون بردار پشتیبان و شبکه های عصبی مصنوعی را به کار گرفتند. نتایج حاصله از این پژوهش نشان داد که ماشین بردار پشتیبان، عملکرد بهتری نسبت به روش شبکه عصبی مصنوعی دارد. در نهایت، بررسی منابع مختلف نشان می دهد که پژوهشی به منظور ارزیابی عملکرد ماشین بردار پشتیبان با استفاده از آزمون گاما در ارتباط با تحلیل منطقه ای سیلاب انجام پذیرفته است. از مهم ترین مراحل مدل سازی، انتخاب پارامترهای ورودی بهینه به منظور افزایش دقت و کارایی مدل و کاهش هزینه ها و زمان مصرفی در مدل سازی است. از جمله روش هایی که می توان برای شناسایی پارامترهای ورودی بهینه و مؤثر در مدل سازی یک

### ۳- برآورد متغیرهای ورودی مدل

برای این منظور با استفاده از مدل رقومی ارتفاعی منطقه (DEM) و بسته الحاقی ArcHydro خصوصیات فیزیوگرافی حوضه‌های بالادست ایستگاه‌های آب‌سنجی به‌دست آمد. این خصوصیات شامل مساحت، محیط، ارتفاع حداقل، ارتفاع متوسط وزنی، تراکم زهکشی، طول آبراهه اصلی، شیب متوسط وزنی حوضه آبخیز و شیب آبراهه اصلی هستند. همچنین با استفاده از نقشه‌های کاربری اراضی و زمین‌شناسی با توجه به مرزهای تهیه شده برای هر کدام از ایستگاه‌ها با بسته الحاقی ArcHydro، متغیرهای مربوط به کاربری اراضی شامل درصد اراضی آبی، باغی، مرتع، دیم، شهری و سنگی و پهنه آبی و متغیر مربوط به زمین‌شناسی شامل درصد سازندهای نفوذپذیر تعیین شدند. متغیرهای اقلیمی حوضه‌ها شامل حداکثر بارندگی ۲۴ ساعته، متوسط بارندگی سالانه، متوسط تبخیر سالانه و متوسط درجه حرارت سالانه نیز با استفاده از داده‌های شرکت تحقیقات منابع آب محاسبه شدند (۴۰).

### ۴- آزمون‌های گاما و M

آزمون گاما (Gamma test) یک ابزار مدل‌سازی غیرخطی است که با استفاده از آن می‌توان مؤثرترین متغیرها در مدل‌سازی و بهترین ترکیب از متغیرهای ورودی را تعیین کرد. در مجموعه‌ای از داده‌های ورودی و خروجی که از متغیرهای واقعی ساخته شده‌اند. آزمون گاما برآوردی از بهترین میانگین مربعات خطا روی خروجی انجام می‌دهد که این برآورد آماره گاما ( $\Gamma$ ) نامیده می‌شود.

در یک فرض مجموعه زیر را در نظر بگیرید:

$$\{(x_i, y_i), 1 \leq i \leq m\} \quad (1)$$

که در این جا  $x \in R^m$  به‌عنوان ورودی و مطابق آن عدد  $y \in R$  به‌عنوان خروجی در نظر می‌گیریم و فرض می‌کنیم که بردارهای  $x$  قابلیت پیش‌گویی فاکتورهای مورد استفاده مؤثر در خروجی  $y$  را داشته باشند، از این رو طبق تنها فرض باقیمانده تحت روابط سیستم کنترل رابطه زیر را دنبال می‌کنیم:

$$y = f(x_1, \dots, x_M) + \Gamma \quad (2)$$

که در اینجا  $f$  یک تابع نرم است و  $\Gamma$  یک متغیر تصادفی است که

کوه‌های البرز مرکزی و آبخیز قره‌چای، کویر میغان و قم‌رود در جنوب حوضه آبخیز و در بین ارتفاعات البرز میانی و زاگرس قرار گرفته‌اند (۳۷).  
مراحل انجام تحقیق به شرح ذیل است:

### ۱- انتخاب ایستگاه‌های آب‌سنجی

ابتدا داده‌های مربوط به دبی حداکثر لحظه‌ای کلیه ایستگاه‌های موجود در منطقه جمع‌آوری شدند. سپس ایستگاه‌های مناسب با لحاظ کردن دو معیار حداقل طول دوره آماری ۱۵ سال و رژیم طبیعی جریان رودخانه واقع در حوضه انتخاب شدند (۳۵). بر این اساس تعداد ۵۵ ایستگاه آب‌سنجی با بیشترین طول دوره آماری، در یک دوره آماری مشترک ۲۰ ساله بین سال‌های ۱۳۷۱-۷۰ تا ۱۳۹۰-۸۹ انتخاب شدند. در ادامه نواقص آماری ایستگاه‌های منتخب با استفاده از دبی‌های حداکثر روزانه ایستگاه‌ها با روش‌های همبستگی خطی و غیرخطی برطرف شد (۱۶).

### ۲- تجزیه و تحلیل فراوانی سیلاب

برای انتخاب مناسب‌ترین تابع توزیع منطقه‌ای از نرم‌افزار Easy Fit استفاده شد. به همین منظور، دبی‌های حداکثر لحظه‌ای هر کدام از ایستگاه‌های مورد مطالعه بر اساس آزمون نکوئی برازش کلموگروف-اسمیرنوف (Kolmogorov Smirnov test)، برازش داده شدند و رتبه هفت توزیع متداول در مطالعات هیدرولوژیکی با استفاده از این آزمون مشخص شد. به‌طوری که برای هر کدام از توابع توزیع احتمالی مورد استفاده یک رتبه بین ۱ تا ۷ اختصاص داده شد. مناسب‌ترین توزیع رتبه (۱) و نامناسب‌ترین توزیع رتبه (۷) را نشان می‌دهند. در نهایت بر اساس مجموع نمرات (حداقل خطا) و بیشترین تعداد رتبه اول برای ۵۵ ایستگاه مورد مطالعه، مناسب‌ترین تابع توزیع منطقه‌ای انتخاب شد. در نهایت دبی‌های حداکثر لحظه‌ای ایستگاه‌های منتخب در دوره بازگشت‌های ۲، ۵، ۱۰، ۲۵، ۵۰ و ۱۰۰ ساله با استفاده از توزیع مناسب تعیین شدند (۱).

آزمون‌های گاما و  $M$  از قابلیت‌های نرم‌افزار WinGamma است.

#### ۵- رگرسیون چندمتغیره غیرخطی (MNL)

به‌منظور برآورد دبی‌های حداکثر لحظه‌ای با استفاده از رگرسیون چند متغیره غیرخطی، یک معادله بین دبی‌های حداکثر لحظه‌ای با دوره بازگشت‌های مختلف (متغیرهای وابسته) و خصوصیات فیزیوگرافی، اقلیمی، کاربری اراضی و زمین‌شناسی (متغیرهای مستقل) برقرار می‌شود. شکل عمومی رگرسیون چند متغیره غیرخطی به‌صورت رابطه (۱) است

$$Q_T = mA^a B^b C^c \dots Z^z \quad (6)$$

در این رابطه،  $Q_T$  متغیر پیش‌بینی شده،  $m$  ثابت رگرسیون و  $Z, \dots, C, B, A$  و  $z, \dots, c, b, a$  ضرایب هستند (۳۵).

#### ۶- ماشین بردار پشتیبان (SVM)

ساختار مدل ماشین بردار پشتیبان توسط وپنک (Vapnik) معرفی شد و این مدل، از اصل استقرایی کمینه‌سازی خطای ساختاری بهره می‌جوید. مدل‌های ماشین بردار پشتیبان به دو گروه عمده مدل طبقه‌بندی ماشین بردار پشتیبان (Support Vector Classification) و مدل رگرسیون ماشین بردار پشتیبان (Support Vector Regression) تقسیم‌بندی می‌شوند. از مدل‌های طبقه‌بندی ماشین بردار پشتیبان برای حل مسائل طبقه‌بندی داده‌هایی که در کلاس‌های مختلف قرار می‌گیرند استفاده می‌شود و مدل رگرسیون ماشین بردار پشتیبان در حل مسائل پیش‌بینی کاربرد دارد (۳۸). در صورتی که داده‌ها به‌صورت خطی قابل تفکیک باشند معادله به‌صورت زیر است:

$$y=f(x)=\text{sign}\left[\sum_{i=1}^n y_i a_i \langle x_i, x_j \rangle + b\right] \quad (7)$$

اگر داده‌ها به‌صورت خطی قابل تفکیک نباشند می‌توان با اعمال پیش‌پردازش، نمونه‌ها را با یک تبدیل غیرخطی به یک فضای با بعد زیاد انتقال داد و در آن فضا حاصل ضرب داخلی انجام داد و ثابت کرد که اگر یک هسته متقارن شرایط قضیه مرسر (Mercer) را داشته باشد، اعمال این هسته در فضای ورودی با بعد کم

نشان‌دهنده مقادیر نویز است. به‌طورکلی و بدون کاهش تابع می‌توان توضیح داد که  $\Gamma$  برابر صفر واریانس نویز کران‌دار است. آماره گاما  $\Gamma$  از قسمتی از واریانس خروجی که نمی‌تواند برای یک مدل نرم محاسبه شود، به‌دست می‌آید. فرض کنید  $\{x_{i,k}, y_{i,k}\}$  کمین نزدیک‌ترین همسایه در یک فاصله اقلیدسی برای  $1 \leq k \leq p$  و  $x_i (i \leq M)$  است در این‌صورت معادله اصلی نیازمند برآورد  $\Gamma$  است.

$$\gamma_M(k) = \frac{1}{\sqrt{M}} \sum_{i=1}^M |y_{N[i,k]} - y_i|^2, \quad (1 \leq k \leq p) \quad (3)$$

$$\delta_M(k) = \frac{1}{M} \sum_{i=1}^M |x_{N[i,k]} - x_i|^2, \quad (1 \leq k \leq p) \quad (4)$$

که در اینجا  $\|\cdot\|$  فاصله اقلیدسی را مشخص می‌کند و به‌منظور محاسبه  $\Gamma$  کافی است که یک خط رگرسیونی حداقل مربعات برای  $\gamma_M(k)$  و  $\delta_M(k)$  ایجاد شود.

$$\gamma = A\delta + \Gamma \quad (5)$$

مقدار آماره  $\Gamma$  عرض از مبدأ خط رگرسیون است.  $A$ ، شیب خط رگرسیون است که نشان‌دهنده پیچیدگی مدلی است که از روی داده‌های مشاهداتی ایجاد می‌شود (۱۷). اگر فرض شود که  $N$  متغیر ورودی بر وقوع پدیده‌ای مؤثر باشد، در این صورت تعداد  $2^{N-1}$  حالت وجود دارد که می‌توان از هرکدام برای ساختن مدل استفاده کرد. استفاده از همه ترکیب‌ها، برای یافتن بهترین ترکیب ورودی به مدل بسیار وقت‌گیر است. بنابراین در شرایطی که تعداد متغیرهای ورودی از مدل زیاد باشد، می‌توان از آزمون گاما به‌منظور انتخاب ترکیب بهینه استفاده کرد. به‌منظور تعیین ترکیب بهینه از متغیرها ابتدا با استفاده از تمام ورودی‌ها میزان آماره گاما ( $\Gamma$ ) محاسبه می‌شود. سپس با حذف یکی از متغیرها میزان آماره گاما ( $\Gamma$ ) تعیین می‌شود. در ادامه متغیر حذف شده به مجموعه برگردانده و متغیر دیگری حذف می‌شود و مقدار آماره گاما ( $\Gamma$ ) محاسبه می‌شود. این فرایند برای تمام متغیرهای ورودی مدل انجام می‌شود به‌طوری‌که مهم‌ترین متغیرها بیشترین میزان آماره گاما ( $\Gamma$ ) را به خود اختصاص می‌دهند. همچنین آزمون  $M$  ( $M$ -test) به‌منظور تعیین داده‌های آموزش و آزمون برای رسیدن به مدل هموار مورد استفاده قرار می‌گیرد (۱۲، ۲۹ و ۳۴).

جدول ۱. معادلات توابع مختلف کرنل

تابع کرنل	معادلات	پارامتر
هسته خطی	$K(x_i, x_j) = x_i \cdot x_j$	-
هسته پلی‌نومیال	$K(x_i, x_j) = (\gamma x_i \cdot x_j + \tau)^d$	$\gamma$ و $d$
هسته تابع پایه شعاعی	$K(x_i, x_j) = \exp\left(-\gamma \ x_i - x_j\ ^2\right)$	$\gamma$

می‌تواند به‌عنوان حاصل ضرب داخلی با بعد زیاد تلقی شود و محاسبات را به‌شدت کاهش دهد. در این حالت رابطه (۷) به رابطه (۸) تغییر می‌یابد:

پارامترهای مجهول، مناسب ارزیابی شده است (۲۱). این نوع شبکه دارای یک لایه ورودی (Input Layer)، یک یا چند لایه مخفی (Hidden Layer) و یک لایه خروجی (Output Layer) است. پارامترهای ورودی در لایه اول و پارامترهای خروجی در لایه سوم قرار دارند. در هر لایه تعدادی گره (نرون) وجود دارد که از طریق وزن‌ها به گره‌های بعدی متصل می‌شوند. تعداد نرون‌ها در لایه ورودی وابسته به تعداد پارامترهای ورودی و تعداد نرون در لایه خروجی وابسته به تعداد پارامتر خروجی است. به‌طور کلی تعداد نرون در لایه پنهان تابع ضابطه خاصی نبوده و تعیین تعداد مناسب آن فقط از طریق سعی و خطا امکان پذیر است. به‌عبارت دیگر تعداد بهینه نرون‌های مخفی زمانی حاصل می‌شود که شبکه عصبی بهترین پردازش را با کمترین خطا در هر دو مرحله آموزش و آزمون مدل انجام دهد. در ساختار شبکه عصبی مصنوعی هر نرون عموماً بیش از یک ورودی دارد. بدین ترتیب هر نرون بردار ورودی را در وزن‌های خود ضرب کرده و با بایاس (Bias) جمع می‌کند. حاصل این عبارت ورودی خالص به تابع محرکه خواهد شد که به صورت زیر بیان می‌شود:

$$net_j = \sum_{i=1}^n w_{ij} x_j + b_i \quad (9)$$

که در آن،  $net_j$ : ورودی خالص به تابع محرکه،  $w_{ij}$ : ماتریس وزن‌های شبکه،  $x_j$ : بردار ورودی به شبکه،  $b_i$ : بردار بایاس شبکه هستند (۲۰ و ۲۲). در شبکه‌های عصبی، تابع‌های محرک گوناگونی به‌کار گرفته می‌شود که در پژوهش حاضر سه تابع محرک خطی (Pure line Function)، لجستیکی (Logistic Function) و تانژانت هیپربولیک

تابع  $K(x_i, x_j)$  تابع کرنلی است که برای ایجاد ماشین‌هایی با انواع مختلف از سطوح تصمیم‌گیری غیرخطی در فضای داده‌ها، ضرب‌های داخلی تولید می‌کند. برای مدل رگرسیون ماشین بردار پشتیبان از کرنل‌های مختلفی استفاده می‌شود. دو نمونه از مدل‌های SVM تعریف شده است که عبارتند از: ۱- مدل‌های رگرسیونی نوع اول (ε-SVM) ۲- مدل‌های رگرسیونی نوع دوم (ν-SVM). مسئله مهم در ماشین‌های بردار پشتیبان انتخاب پارامترهای بهینه است. این پارامترها در مدل SVM شامل C (پارامتر ظرفیت تابع)، ε (پارامتر تابع خطا) هستند. در پژوهش حاضر مدل رگرسیونی نوع اول به‌دلیل کاربرد وسیع آن در مسائل رگرسیونی با انواع تابع کرنل خطی، پلی‌نومیال، تابع پایه شعاعی مورد استفاده قرار گرفت. این توابع نیز دارای پارامترهایی هستند. جدول (۱) معادلات هر کدام از توابع موردنظر را نشان می‌دهد. در این معادلات  $\gamma$  (گاما)،  $\tau$  (عدد ثابت جمع شونده) و  $d$  (درجه چندجمله‌ای) هستند (۲۹).

$$y=f(x)=\text{sign}\left[\sum_{i=1}^n y_i a_i K(x_i, x_j) + b\right] \quad (8)$$

۷- شبکه‌های عصبی مصنوعی (ANNs)

شبکه پرسپترون چند لایه پیش‌خور (Multi-Layer Perceptrone) با قانون یادگیری پسانتشار خطا (Back Propagation Error) و تابع آموزشی لونبرگ-مارکوات (Levenberg-Marquardt)، ساده‌ترین و پر کاربردترین نوع شبکه عصبی است که در خصوص تخمین

جدول ۲. انتخاب بهترین تابع توزیع آماری منطقه‌ای

توزیع آماری	لوگ پیرسون نوع سوم	لوگ نرمال سه پارامتری	لوگ نرمال دو پارامتری	پیرسون نوع سوم	گامای دو پارامتری	گمبل	نرمال
جمع نمرات (حداقل خطا)	۱۳۹	۱۶۱	۲۳۶	۲۳۲	۲۱۵	۲۵۸	۳۰۸
تعداد رتبه اول	۱۷	۱۲	۶	۶	۵	۴	۴

که در آن  $q_o(t)$  دبی مشاهده شده،  $q_p(t)$  دبی پیش‌بینی شده،  $n$  تعداد داده‌های مشاهداتی،  $\overline{q_o(t)}$  میانگین دبی مشاهده شده و  $\overline{q_p(t)}$  میانگین دبی پیش‌بینی شده با دوره بازگشت‌های مختلف است (۸ و ۳۵).

### نتایج و بحث

پس از مشخص شدن ایستگاه‌های مورد نظر به منظور محاسبه دبی‌های حداکثر لحظه‌ای در دوره بازگشت‌های مختلف، بهترین تابع توزیع منطقه‌ای انتخاب می‌شود. بر همین اساس توزیع لوگ پیرسون نوع سوم (LP۳) با کمترین میزان خطا و بیشترین تعداد رتبه اول به‌عنوان مناسب‌ترین تابع برای برازش دبی‌های حداکثر لحظه‌ای ایستگاه‌ها انتخاب شد. (۲۵، ۳۶ و ۳۹). معمولاً به منظور انتخاب مناسب‌ترین تابع توزیع، از آزمون‌های برازش کای اسکور و کلموگروف-اسمیرنوف استفاده می‌شود. در انتخاب توزیع مناسب هیچ‌گونه توافقی بین هیدرولوژیست‌ها وجود ندارد. ولیکن می‌توان توزیع مناسب را براساس کمترین میزان خطا در نظر گرفت (۱۸). بر همین اساس با استفاده از آزمون برازش کلموگروف-اسمیرنوف مناسب‌ترین تابع توزیع آماری انتخاب شد. جدول (۲) نتایج انتخاب بهترین تابع توزیع منطقه‌ای نشان می‌دهد.

جدول (۳)، نتایج مربوط به آزمون گاما به منظور انتخاب ترکیب بهینه را نشان می‌دهد. به منظور تعیین ترکیب بهینه از ورودی مدل‌ها با استفاده از آزمون گاما، اثر حذف هر یک از متغیرها از ترکیبی که در آن همه متغیرها وجود داشته باشند انجام پذیرفت و در هر مرحله مقدار آماره گاما محاسبه و با مقدار آماره گاما بدون حذف متغیر مقایسه شد. ترکیباتی که در آنها مقدار

(Hyperbolic Tangent Function) برای لایه‌های مخفی و تابع محرک خطی برای لایه خروجی مورد استفاده قرار گرفت.

### ۸- نرمال‌سازی داده‌ها

قبل از آموزش ماشین بردار پشتیبان و شبکه عصبی مصنوعی می‌بایست داده‌های ورودی استاندارد شوند. به همین منظور، برای استاندارد کردن داده‌ها از رابطه (۱۰) استفاده می‌شود:

$$X_{Normal} = \frac{(X_i - X_{min})}{(X_{max} - X_{min})} + 0.1 \quad (10)$$

که در آن،  $X_{Normal}$ : داده نرمال شده،  $X_i$ : داده مشاهده شده،  $X_{min}$ : داده حداقل و  $X_{max}$ : داده حداکثر هستند (۷).

### ۹- ارزیابی مدل‌ها

به منظور ارزیابی نتایج مدل‌ها از نمایه‌های آماری ضریب تبیین (Determination Coefficient)، ریشه میانگین مربعات خطا (Root Mean Square Error) و ضریب کارایی ناش-ساتکلیف (Nash-Sutcliffe Coefficient) استفاده شد. این شاخص‌ها به صورت زیر محاسبه می‌شوند:

$$R^2 = \frac{\sum_{i=1}^n (q_p(t) - \overline{q_o(t)}) (q_p(t) - \overline{q_p(t)})}{\sum_{i=1}^n \left( (q_p(t) - \overline{q_o(t)})^2 \right)^{0.5} \left( (q_o(t) - \overline{q_p(t)})^2 \right)^{0.5}} \quad (11)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (q_p(t) - q_o(t))^2} \quad (12)$$

$$NSC = 1 - \frac{\sum_{i=1}^n (q_o(t) - q_p(t))^2}{\sum_{i=1}^n (q_o(t) - \overline{q_o(t)})^2} \quad (13)$$

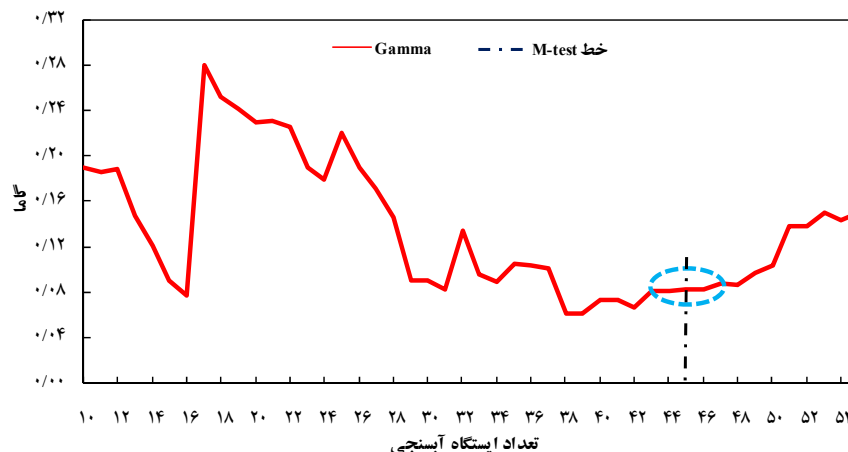
جدول ۳. انتخاب متغیرهای تأثیرگذار بر دبی‌های سیلاب با استفاده از آزمون گاما

شماره ترکیب	متغیر	گاما (Γ)	شماره ترکیب	متغیر	گاما (Γ)
۱	All	۰/۱۸۴۲	۱۲	شیب متوسط وزنی-All	۰/۱۷۸۲
۲	مساحت-All	۰/۲۰۷۷	۱۳	ارتفاع میانگین-All	۰/۱۷۷۳
۳	تراکم زهکشی-All	۰/۲۰۲۰	۱۴	حداکثر بارندگی ۲۴ ساعته-All	۰/۱۷۷۳
۴	ارتفاع حداکثر-All	۰/۱۹۵۰	۱۵	طول آبراهه اصلی-All	۰/۱۷۶۷
۵	محیط-All	۰/۱۸۴۴	۱۶	شیب آبراهه صلی-All	۰/۱۷۵۲
۶	درصد سازندهای نفوذپذیر-All	۰/۱۸۴۰	۱۷	متوسط درجه حرارت سالانه-All	۰/۱۷۰۶
۷	درصد اراضی آبی-All	۰/۱۸۳۲	۱۸	درصد اراضی پهنه آبی-All	۰/۱۶۹۷
۸	درصد اراضی مرتعی-All	۰/۱۸۲۰	۱۹	درصد اراضی مسکونی-All	۰/۱۶۴۵
۹	درصد اراضی باغی-All	۰/۱۸۰۷	۲۰	درصد اراضی دیم-All	۰/۱۶۳۵
۱۰	متوسط بارندگی سالانه-All	۰/۱۷۹۲	۲۱	ارتفاع حداقل-All	۰/۱۵۹۹
۱۱	متوسط تبخیر سالانه-All	۰/۱۷۸۷	-	-	-

لحظه‌ای در حوضه آبخیز دریاچه نمک است (۱۵ و ۳۱). زیرا با حذف این متغیر از ترکیب متغیرهای ورودی، مقدار آماره گاما نسبت به ترکیب شماره ۱ (ترکیبی که در آن همه متغیرهای ورودی حضور دارند) بیشترین مقدار افزایش را داشته است. مساحت حوضه یکی از متغیرهای مؤثر بر خصوصیات هیدرولوژیکی حوضه‌ها، مخصوصاً دبی است. روند افزایش دبی با مساحت حوضه خطی نیست. با افزایش مساحت حوضه از یک آستانه‌ای به بعد، رفته رفته تلفات انتقال افزایش می‌یابد که در این صورت کاهش دبی سیلابی را به دنبال خواهد داشت. در واقع فرایند تلفات بین‌راهی بیشتر برای حوضه‌های آبخیز بزرگ‌تر، دلیل اصلی کاهش ضریب رواناب و به عبارتی کاهش بده جریان ویژه محسوب می‌شود (۹). همچنین بر اساس نتایج آزمون M تعداد ایستگاه‌های آب‌سنجی مورد نیاز برای آموزش و آزمون مدل تعیین می‌شود. برای این منظور با استفاده از آماره گاما (Γ) می‌توان داده‌های مورد نیاز را برای ایجاد خط مجانب پایدار (M) مشخص کرد. همان‌گونه که از شکل (۲) مشاهده می‌شود، هم‌سانی خط مجانب آماره گاما با مقدار ۰/۰۸ در نزدیکی نقاط ۴۳ تا ۴۷ مشخص شده است. به همین دلیل می‌توان داده ۴۵ را به‌عنوان نقطه مرزی برای تعیین تعداد داده‌های آموزش و آزمون

آماره گاما با حذف هر یک از متغیرها از ترکیب همه متغیرهای ورودی (ترکیبات ۲ تا ۲۱)، بیشتر از مقدار آماره گاما بدون حذف متغیر (ترکیب ۱) باشند، به‌عنوان ترکیب بهینه انتخاب می‌شوند (۲۹ و ۳۴). مطابق نتایج جدول (۳)، مقدار آماره گاما در ترکیب شماره ۱، ۰/۱۸۴۲ است. در ترکیب ۲ با حذف متغیر مساحت، مقدار آماره گاما نسبت به ترکیب ۱، به ۰/۲۰۷۷ افزایش می‌یابد. در ترکیب‌های ۳، ۴ و ۵ نیز با حذف متغیرهای تراکم زهکشی، ارتفاع حداکثر و محیط، مقدار آماره گاما نسبت به ترکیب ۱، به ترتیب به ۰/۲۰۲۰، ۰/۱۹۵۰ و ۰/۱۸۴۴ افزایش می‌یابد. از ترکیب ۶ تا ترکیب ۲۱ مقدار آماره گاما نسبت به ترکیب ۱ کاهش یافته است. به عبارتی دیگر متغیرهای غایب ترکیبات ۲، ۳، ۴ و ۵ به‌عنوان، تأثیرگذارترین متغیرها در ایجاد دبی‌های حداکثر لحظه‌ای بوده و متغیرهای غایب هر کدام از ترکیبات ۶ تا ۲۱، از اهمیت کمتری نسبت به متغیرهای غایب چهار ترکیب فوق، به‌منظور مدل‌سازی برخوردار هستند. بنابراین با توجه به نتایج جدول (۳)، می‌توان به ترتیب چهار متغیر مساحت، تراکم زهکشی، ارتفاع حداکثر و محیط را به‌عنوان متغیرهای مؤثر بر دبی‌های حداکثر لحظه‌ای در نظر گرفت. از این بین، مساحت به‌عنوان مهم‌ترین و مؤثرترین متغیر در تولید دبی‌های حداکثر





شکل ۲. نمودار تغییرات مقدار آماره گاما در مقابل تعداد ایستگاه‌های آب‌سنجی (آزمون M)

جدول ۴. نتایج حاصل از ماشین بردار پشتیبان در مرحله آموزش و آزمون

متغیرهای مدل			آزمون			آموزش			تابع کرنل	دوره بازگشت (سال)
$\gamma$	$\epsilon$	C	NSC	RMSE	$R^2$	NSC	RMSE	$R^2$		
۳/۲	۰/۵۰	۶۰	۰/۸۲	۱۱/۳۲	۰/۹۱	۰/۸۳	۸/۸۳	۰/۸۳	RBF	۲
۳	۰/۵۰	۶۰	۰/۸۴	۱۷/۶۰	۰/۹۰	۰/۸۲	۱۷/۳۹	۰/۸۲	RBF	۵
۳/۱	۰/۵۰	۶۰	۰/۸۲	۲۴/۹۷	۰/۹۲	۰/۸۲	۲۶/۰۷	۰/۸۳	RBF	۱۰
۱/۲	۰/۳۵	۵۵	۰/۷۵	۳۷/۳۷	۰/۸۹	۰/۷۷	۵۰/۳۲	۰/۷۷	RBF	۲۵
۱/۲	۰/۳۵	۵۵	۰/۷۵	۴۵/۲۷	۰/۸۳	۰/۷۵	۷۴/۷۴	۰/۷۶	RBF	۵۰
۱/۲	۰/۵۵	۸۰	۰/۷۱	۵۸/۸۴	۰/۷۶	۰/۷۲	۱۱/۰۷	۰/۷۴	RBF	۱۰۰

شبکه درشت ناحیه مناسب پارامترها مشخص می‌شود. سپس این شبکه درشت شکسته شده و مقدار بهینه پارامترها در شبکه‌های ریزتر تعیین می‌شود. جدول (۴) مقادیر بهینه این پارامترها و نتایج مربوط به مدل‌های ماشین بردار پشتیبان را نشان می‌دهد. در فرایند مدل‌سازی، مدلی قابل پذیرش است که ضریب کارایی بالاتری در مرحله آزمون داشته باشد (۲۷ و ۳۵). ارزیابی عملکرد مدل ماشین بردار پشتیبان بر مبنای آماره ضریب کارایی ناش-ساتکلیف نشان می‌دهد که دقت مدل‌سازی تا دوره بازگشت پنج سال رو به افزایش است و از آن به بعد این دقت کاهش می‌یابد. به طوری که آماره ناش-ساتکلیف در بهترین حالت مدل ماشین بردار پشتیبان ۰/۸۴ در مرحله آزمون است. مقادیر بهینه پارامترهای مدل ماشین بردار پشتیبان در بهترین حالت مدل،

مدل‌ها در نظر گرفت. بر این اساس تعداد ۴۵ ایستگاه به‌عنوان داده‌های آموزش و تعداد ۱۰ ایستگاه به‌عنوان داده‌های آزمون مدل‌ها در نظر گرفته شدند.

در نهایت مدل‌سازی دبی‌های حداکثر لحظه‌ای با استفاده از سه روش ماشین بردار پشتیبان، شبکه عصبی مصنوعی و رگرسیون چندمتغیره غیرخطی انجام پذیرفت. در ماشین‌های بردار پشتیبان می‌بایست ابتدا مقادیر بهینه پارامترهای مورد استفاده در توابع، برای دوره بازگشت‌های مختلف مشخص شوند و پس از ایجاد مدل‌ها، مناسب‌ترین تابع انتخاب شود. به‌همین منظور برای انتخاب مقادیر بهینه پارامترها از الگوریتم جستجوی دو مرحله‌ای (Two-Steps Grid Search) با اعتبارسنجی متقاطع (Cross-Validation) استفاده شد. در این روش، ابتدا در یک

جدول ۵. نتایج مدل‌سازی با شبکه عصبی مصنوعی در مرحله آموزش و آزمون

دوره بازگشت (سال)	آرایش شبکه	لایه مخفی	لایه خروجی	تعداد تکرار	آموزش			آزمون		
					NSC	RMSE	R <sup>2</sup>	NSC	RMSE	R <sup>2</sup>
۲	۴-۱۳-۱	Tanh	Pure line	۵۹	۰/۵۶	۱۴/۰۴	۰/۶۳	۰/۹۴	۱۶/۶۴	۰/۶۰
۵	۴-۱۱-۱	Logistic	Pure line	۲۱	۰/۶۵	۲۴/۴۰	۰/۶۶	۰/۹۴	۲۸/۵۲	۰/۵۷
۱۰	۴-۹-۱	Tanh	Pure line	۴۶	۰/۷۷	۲۹/۹۷	۰/۷۸	۰/۹۷	۲۷/۱۰	۰/۷۷
۲۵	۴-۸-۱	Logistic	Pure line	۲۴	۰/۶۳	۶۳/۵۲	۰/۶۶	۰/۸۵	۴۹/۱۳	۰/۵۸
۵۰	۴-۱۲-۱	Tanh	Pure line	۴۰	۰/۶۸	۸۵/۳۰	۰/۶۸	۰/۸۵	۵۴/۲۸	۰/۶۵
۱۰۰	۴-۴-۱	Logistic	Pure line	۳۴	۰/۵۳	۱۵۰/۵۶	۰/۵۴	۰/۸۰	۷۵/۹۰	۰/۵۱

جدول ۶. نتایج مدل‌سازی با رگرسیون چندمتغیره غیرخطی در مرحله آموزش و آزمون

معادلات	آموزش			آموزش			دوره بازگشت (سال)
	NSC	RMSE	R <sup>2</sup>	NSC	RMSE	R <sup>2</sup>	
$\text{Log}(Q_1) = -۰/۴۹۸۰ - ۰/۵۸۵ \text{Log}(A) + ۱/۴۸۳ \text{Log}(P) + ۱/۲۲۷ \text{Log}(H_{\max}) - ۰/۴۰۲ \text{Log}(D_d)$	۰/۲۸	۲۲/۴۰	۰/۹۱	۰/۳۶	۱۶/۸۷	۰/۴۲	۲
$\text{Log}(Q_2) = -۲/۵۸۳ - ۰/۲۶۸ \text{Log}(A) + ۰/۹۹۰ \text{Log}(P) + ۰/۷۰۸ \text{Log}(H_{\max}) - ۰/۴۵۰ \text{Log}(D_d)$	۰/۴۴	۳۲/۴۶	۰/۹۳	۰/۵۲	۲۸/۵۸	۰/۵۷	۵
$\text{Log}(Q_3) = -۱/۲۴۷ - ۰/۱۰۳ \text{Log}(A) + ۰/۷۱۴ \text{Log}(P) + ۰/۴۲۲ \text{Log}(H_{\max}) - ۰/۴۷۰ \text{Log}(D_d)$	۰/۵۰	۳۹/۹۸	۰/۹۱	۰/۵۳	۴۲/۶۰	۰/۵۸	۱۰
$\text{Log}(Q_4) = ۰/۲۰۶ + ۰/۰۷۲ \text{Log}(A) + ۰/۴۱۴ \text{Log}(P) + ۰/۱۱۲ \text{Log}(H_{\max}) - ۰/۴۸۸ \text{Log}(D_d)$	۰/۵۵	۵۰/۶۶	۰/۸۸	۰/۴۷	۷۶/۳۳	۰/۵۳	۲۵
$\text{Log}(Q_5) = ۱/۳۵۱ + ۰/۲۳۲ \text{Log}(A) + ۰/۱۵۹ \text{Log}(P) - ۰/۱۴۱ \text{Log}(H_{\max}) - ۰/۴۲۸ \text{Log}(D_d)$	۰/۵۱	۶۳/۵۴	۰/۸۵	۰/۴۰	۱۱۶/۶۶	۰/۴۷	۵۰
$\text{Log}(Q_{100}) = ۲/۰۰۳ + ۰/۲۸۴ \text{Log}(A) + ۰/۰۴۵ \text{Log}(P) - ۰/۲۷۱ \text{Log}(H_{\max}) - ۰/۵۰۳ \text{Log}(D_d)$	۰/۷۱	۵۸/۲۴	۰/۷۵	۰/۳۸	۱۷۲/۶۰	۰/۳۸	۱۰۰

دقیق‌ترین مدل شبکه عصبی مصنوعی در دوره بازگشت ۱۰ سال با آماره ناش - ساتکلیف ۰/۷۷ در مرحله آزمون است. در بهترین حالت مدل شبکه عصبی مصنوعی، تعداد ۴۶ تکرار برای آموزش، تابع محرکه (Tanh) برای لایه مخفی و (Pure line) برای لایه خروجی و تعداد ۹ نرون در لایه مخفی به دست آمد. همچنین نتایج روش مدل‌سازی رگرسیون چندمتغیره غیرخطی به همراه معادلات مربوط به آن در جدول (۶) ارائه شده است. ارزیابی عملکرد رگرسیون چندمتغیره غیرخطی نشان‌دهنده یک

به ترتیب ( $C=۶۰$ ،  $\epsilon=۰/۵۰$  و  $\gamma=۳$ ) هستند. کالیبراسیون شبکه‌های عصبی مصنوعی، شامل تعداد تکرارهای مناسب برای آموزش، تعیین مناسب‌ترین تابع محرکه در هر یک از لایه‌های پنهان و خروجی شبکه و تعداد بهینه نرون‌ها در لایه پنهان است. در جدول (۵) نتایج مربوط به مدل‌سازی دبی‌های حداکثر لحظه‌ای با استفاده شبکه عصبی مصنوعی ارائه شده است. ارزیابی عملکرد مدل شبکه عصبی مصنوعی بر اساس آماره ناش - ساتکلیف نشان می‌دهد که دقت مدل‌سازی یک روند سینوسی دارد. به طوری که

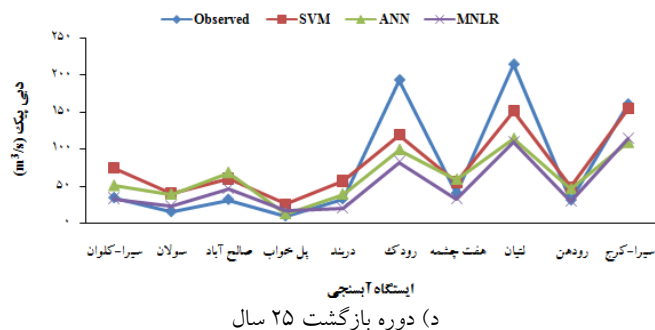
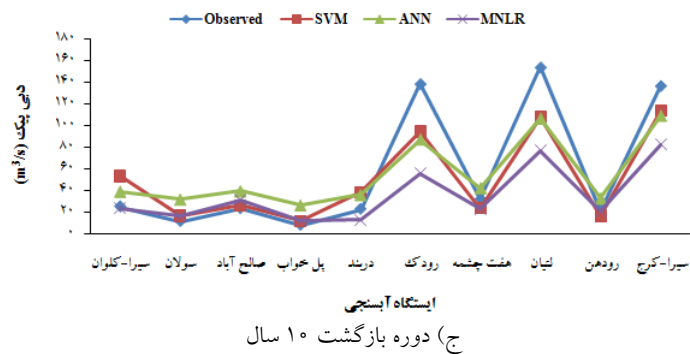
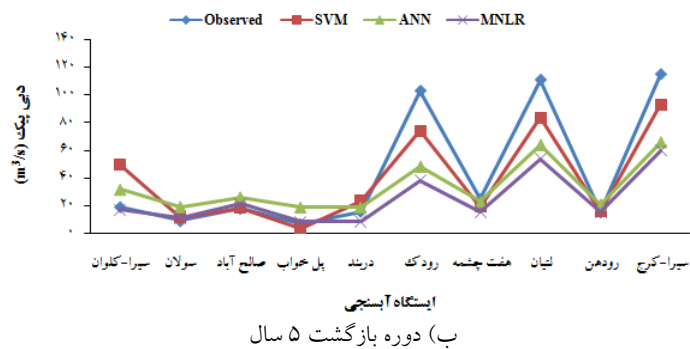
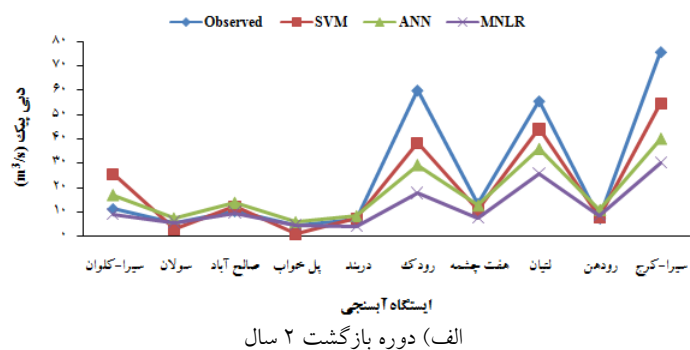
جدول ۷. مقایسه کیفی مدل‌های مورد استفاده در دوره بازگشت‌های مختلف

دوره بازگشت (سال)	SVM	ANNs	MNLR
۲	خیلی خوب	متوسط	ضعیف
۵	خیلی خوب	متوسط	ضعیف
۱۰	خیلی خوب	خوب	ضعیف
۲۵	خیلی خوب	متوسط	متوسط
۵۰	خیلی خوب	متوسط	متوسط
۱۰۰	خوب	متوسط	خوب

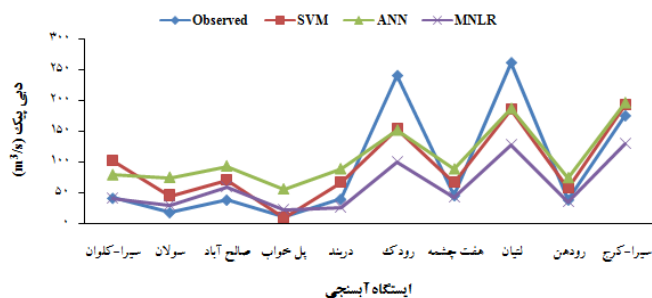
### نتیجه‌گیری

هدف از پژوهش حاضر مقایسه عملکرد ماشین بردار پشتیبان با شبکه عصبی مصنوعی و رگرسیون چندمتغیره غیرخطی بر اساس توانایی آزمون گاما در انتخاب متغیرهای مؤثر بر دبی‌های حداکثر لحظه‌ای در حوضه آبخیز دریاچه نمک است. در این مطالعه تعداد ۵۵ ایستگاه آب‌سنجی با طولانی‌ترین دوره آماری مشترک (۱۳۷۱-۱۳۹۰) مورد استفاده قرار گرفتند. پس از انتخاب توزیع لوگ پیرسون نوع سوم براساس نتایج آزمون برازش کلموگروف-اسمیرنوف به‌عنوان دقیق‌ترین تابع توزیع آماری به‌منظور برآورد دبی‌های حداکثر لحظه‌ای برای دوره بازگشت‌های مختلف، نتایج آزمون گاما نیز نشان داد که به‌ترتیب متغیرهای مساحت، تراکم زهکشی، ارتفاع حداکثر و محیط به‌عنوان متغیرهای مؤثر بر دبی‌های حداکثر لحظه‌ای هستند. براساس نتایج به‌دست آمده برای هرکدام از روش‌های مدل‌سازی ملاحظه می‌شود که دقیق‌ترین مدل ماشین بردار پشتیبان در دوره بازگشت ۵ سال، شبکه عصبی مصنوعی در دوره بازگشت ۱۰ سال و رگرسیون چندمتغیره غیرخطی در دوره بازگشت ۱۰۰ سال به‌دست آمده است. این نتایج نشان می‌دهد که دقیق‌ترین مدل‌های ماشین بردار پشتیبان و شبکه عصبی مصنوعی در دوره بازگشت کوتاه‌مدت و دقیق‌ترین مدل رگرسیون چندمتغیره غیرخطی در دوره بازگشت بلندمدت به‌دست آمده است. درنهایت پس از بررسی کمی و کیفی هرکدام از روش‌های مورد بررسی در دوره بازگشت‌های مختلف ۲، ۵، ۱۰، ۲۵، ۵۰ و ۱۰۰ سال، روش ماشین بردار پشتیبان به‌عنوان یک روش بسیار دقیق و صریح به‌منظور برآورد

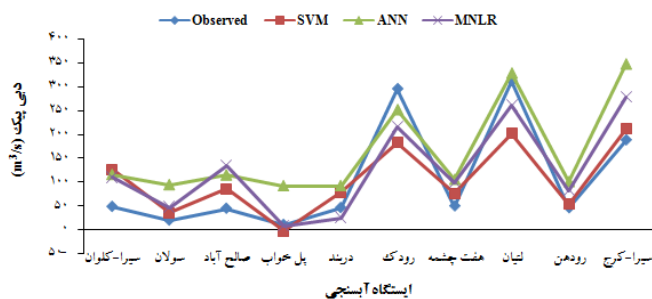
روند افزایشی در دقت مدل‌سازی تا دوره بازگشت ۲۵ سال بر اساس آماره ناش-ساتکلیف است. پس از یک روند کاهشی در میزان این دقت در دوره بازگشت ۵۰ سال، دقیق‌ترین مدل رگرسیون چندمتغیره غیرخطی در دوره بازگشت ۱۰۰ سال با آماره ناش-ساتکلیف ۰/۷۱ در مرحله آزمون به‌دست آمد. درنهایت مقایسه جداول (۴) تا (۶) نشان می‌دهد که روش ماشین بردار پشتیبان از دقت به‌مراتب بیشتری نسبت به دو روش شبکه عصبی مصنوعی و رگرسیون چندمتغیره غیرخطی به‌منظور برآورد دبی‌های حداکثر لحظه‌ای برخوردار است (۸، ۱۰ و ۳۰). در استفاده از روش ماشین بردار پشتیبان، از بین توابع کرنل، تابع کرنل RBF در تمام دوره بازگشت‌های مورد نظر بهترین نتایج را نشان داد (۵، ۱۴ و ۳۳). همچنین به‌منظور مقایسه کیفی مدل‌های مورد استفاده در برآورد دبی‌های حداکثر لحظه‌ای، جدول (۷) شده است. برای این منظور با استفاده از آماره ناش-ساتکلیف،  $0.75 \leq NSC \leq 1$  (خیلی خوب)،  $0.75 \leq NSC \leq 0.95$  (خوب)،  $0.50 \leq NSC \leq 0.75$  (متوسط) و  $0.50 \leq NSC \leq 0.65$  (ضعیف) مدل‌ها در چهار گروه طبقه‌بندی می‌شوند (۱۱). مقایسه کیفی روش‌های مورد استفاده نیز نشان داد که اکثر مدل‌های ماشین بردار پشتیبان در طبقه (خیلی خوب) و مدل‌های شبکه عصبی مصنوعی در طبقه (متوسط) قرار می‌گیرند. از طرفی دیگر نتایج این بررسی نشان داد که بیشتر مدل‌های رگرسیون چندمتغیره غیرخطی در طبقه (ضعیف) قرار دارند. شکل (۳) نیز نمودار برازش دبی‌های حداکثر لحظه‌ای مشاهداتی و تخمینی برای دوره بازگشت‌های مختلف در مرحله آزمون را نشان می‌دهد.



شکل ۳. نمودار برازش داده‌های مشاهداتی و تخمینی در دوره بازگشت‌های مختلف



(و) دوره بازگشت ۵۰ سال



(ه) دوره بازگشت ۱۰۰ سال

ادامه شکل ۳.

روش‌های مدل‌سازی ماشین بردار پشتیبان، شبکه عصبی مصنوعی و رگرسیون چندمتغیره غیرخطی به ترتیب به‌عنوان مناسب‌ترین روش مدل‌سازی داده‌های دبی حداکثر لحظه‌ای در حوضه آبخیز دریاچه نمک هستند.

دبی‌های حداکثر لحظه‌ای در حوضه آبخیز دریاچه نمک نسبت به دو روش شبکه عصبی مصنوعی و رگرسیون چندمتغیره غیرخطی معرفی شد. همچنین این مقایسه دقت مدل‌های شبکه عصبی مصنوعی را بیشتر از دقت مدل‌های رگرسیون چندمتغیره غیرخطی نشان داد. بنابراین می‌توان این‌گونه نتیجه گرفت که

### منابع مورد استفاده

1. Ahn, J., W. Cho, T. Kim, H. Shin and J-H. Heo. 2014. Flood frequency analysis for the annual peak flows simulated by an event-based rainfall-runoff model in an urban drainage basin. *Water* 6(12): 3841-3863.
2. Alizadeh, A. 2006. Principles of Applied Hydrology (18<sup>th</sup> Ed). The Imam Reza University Press. Iran. (In Farsi).
3. Aziz, K., A. Rahman, G. Fang and S. Shrestha. 2013. Application of artificial neural networks in regional flood frequency analysis: a case study for Australia. *Stochastic Environmental Research and Risk Assessment* 28(3): 541-554.
4. Blosclé, G. and M. Sivapalan. 1997. Process controls on regional flood frequency: coefficient of variation and basin scale. *Water Resources Research* 33(12): 2967-2980.
5. Daliri, M., M. Javadi and S. Abedkouhi. 2014. Traffic signs detection and recognition by using a method based on attention mechanism and classical and support vector machine classifier. *Journal of Transportation Engineering* 4(4): 333-344 (In Farsi).
6. Dawson, C. W., R. J. Abrahart, A. Y. Shamseldin and R. L. Wilby. 2006. Flood estimation at ungauged sites using artificial neural networks. *Journal of Hydrology* 319(1-4): 391-409.
7. Dogan, E., M. Gumrukcuoglu, M. Sandalci and M. Opan. 2010. Modelling of evaporation from the reservoir of

- Yuvacik dam using adaptive neuro-fuzzy inference systems. *Engineering Applications of Artificial Intelligence* 23(6): 961-967.
8. Esen, H., M. Inalli, A. Sengur and M. Esen. 2008. Modeling a ground-coupled heat pump system by a support vector machine. *Renewable Energy* 33(8): 1814-1823.
  9. Eslami, A. R. and B. Saghafian. 2008. The role of watershed morphometric and climatic factors in producing of flood flows. *Pajouhesh and Sazandegi* 78: 149-157 (In Farsi).
  10. Eslamian, S. S., J. Abedi-Koupai, M. J. Amiri and S. A. Gohari. 2009. Estimation of daily reference evapotranspiration using support vector machines and artificial neural networks in greenhouse. *Environmental Sciences* 4: 439-447.
  11. Fatehi, I., B. J. Amiri, A. Alizadeh and J. Adamowski. 2015. Modeling the relationship between catchment attributes and in-stream water quality. *Water Resources Management* 29(14): 5055-5072.
  12. Ghabaei Sough, M., A. Mosaedi, M. Hesam and A. Hezarjaribi. 2010. Evaluation effect of input parameters preprocessing in artificial neural networks (ANNs) by using stepwise regression and gamma test techniques for fast estimation of daily evapotranspiration. *Journal of Water and Soil* 24(3): 610-624 (In Farsi).
  13. Gizaw, M. S. and T. Y. Gan. 2016. Regional flood frequency analysis using support vector regression under historical and future climate. *Journal of Hydrology* 538: 387-398.
  14. Goel, A. and M. Pal. 2012. Stage-discharge modeling using support vector machines. *International Journal of Engineering* 25(1): 1-9.
  15. Hall, M. G. and A. V. Minns. 1998. Regional flood frequency analysis using artificial neural networks. *Journal of Hydroinformatics* 98(2): 759-763.
  16. Jaydari, A., A. Fathzadeh, R. Taghizadeh Mehrjardi, M. T. Dastorani and A. Fatahi Ardakani. 2012. Comparison of the efficiency of different methods for reconstruction and prolongation of instantaneous peak flow data. *Journal of Range and Watershed Management (Iranian Journal of Natural Resources)* 64(4): 387-399. (In Farsi).
  17. Jones, A. J. 2004. New tools in non-linear modeling and prediction. *Computational Management Science* DOI: 10.1007/s10287-003-0006-1. 109-149.
  18. Karim, M. A. and J. U. Chowdhury. 1995. A comparison of four distributions used in flood frequency analysis in Bangladesh. *Hydrological Sciences Journal* 40(1): 55-66.
  19. Khosravi, M., A. Salajegheh, M. Mahdavi and M. Mohseni Saravi. 2012. Flood forecasting using artificial neural networks and nonlinear multivariate regression (Case Study: Taleghan Watershed). *Journal of Range and Watershed Management (Iranian Journal of Natural Resources)* 65(3): 341-349 (In Farsi).
  20. Kia, M. 2018. Soft Computing using MATLAB (5<sup>th</sup> Ed). Kian Publication. Iran. (In Farsi).
  21. Kurtulus, B. and M. Razack. 2010. Modeling daily discharge responses of a large karstic aquifer using soft computing methods: artificial neural network and neuro-fuzzy. *Journal of Hydrology* 381(1): 101-111.
  22. Menhaj, M. 2005. Basis of Artificial Neural Network (Calculative Intelligence). Publication Center of Poly Technic University Press. Tehran. Iran. (In Farsi).
  23. Minns, A. W. and M. J. Hall. 1996. Artificial neural network as rainfall-runoff model. *Hydrological Sciences Journal* 41(3): 399-417.
  24. Moghadamnia, A., M. Ghafari, J. Piri and D. Han. 2008. Evaporation estimation using support vector machines technique. *Engineering and Technology* 33: 14-22.
  25. Moghimi, A., S. R. Moussavi, A. Motamed and H. Ahmadi. 2010. Study of morphometry factors regressions of chalus drainage basin using statistical methods and mathematical models. *Journal of Earth and Resources* 2(1): 65-80 (In Farsi).
  26. Nabizadeh, M., A. Mosaedi, M. Hesam and A. A. Dehghani and M. Meftah. 2012. River flow forecasting using fuzzy inference system (FIS) and adaptive neuro-fuzzy inference system (ANFIS). *Iranian Journal of Watershed Management Science and Engineering* 17(5): 7-14 (In Farsi).
  27. Nabizadeh, M., A. Mosaedi, M. Hesam and A. A. Dehghani. 2012. Comparing the performance of fuzzy based models in stream flow forecasting on Lighvan River. *Journal of Water and Soil Conservation* 19(1): 117-134 (In Farsi).
  28. Najmaii, M. 1991. Engineering Hydrology. (2<sup>th</sup> Ed). Science and Technology University Press. Iran. (In Farsi).
  29. Noori, R., A. R. Karbassi, A. Moghaddamnia, D. Han, M. H. Zokaeei-Ashtiani, A. Farokhnia and M. Ghafari Gousheh. 2011. Assessment of input variables determination on the SVM model performance using PCA, Gamma test and forward selection techniques for monthly stream flow prediction. *Journal of Hydrology* 401(3-4): 177-189.
  30. Radhika, Y. and M. Shashi. 2009. Atmospheric temperature prediction using support vector machines. *International Journal of Computer Theory and Engineering* 1(1): 1793-8201.
  31. Rezai, A., M. Mahdavi, K. Lox, S. Feyznaya and M. H. Mehdian. 2007. Regional peak flows modelling of Sefidrood dam's sub basins using artificial neural network. *Journal of Science and Technology of Agriculture and Natural Resources* 11: 25-39 (In Farsi).

32. Seckin, N. and A. Guven. 2012. Estimation of peak flood discharges at ungauged sites across Turkey. *Water Resource Management* 26: 2569-2581.
33. Seefi, A., M. Mirlatifi and H. Reahi. 2013. Introduction and application of least square support vector machine (LSSVM) for simulation of reference evapotranspiration and uncertainty analysis of results (A Case Study of the Kerman City). *Irrigation and Water Engineering* 13(4): 67-79 (In Farsi).
34. Sharifi, A. R., Y. Dinpashoh, A. Fakhri-Fard and A. R. Moghaddamnia. 2014. Optimum combination of variables for runoff simulation in Amameh watershed using gamma test. *Water and Soil Science* 23(4): 59-72 (In Farsi).
35. Shu, C. and T. B. M. J. Ouarda. 2008. Regional flood frequency analysis at ungauged sites using the adaptive neuro-fuzzy inference system. *Journal of Hydrology* 349: 31- 43.
36. Topaloulu, F. 2002. Determining suitable probability distribution models for flow and precipitation series of the Seyhan river basin. *Turkish Journal of Agriculture and Forestry* 26: 187-194.
37. Vafakhah, M. 1999. Estimated regional frequency of low-flow in the season rivers (Case Study the Arid Region of Central Iran). MSc. Thesis, Tarbiat Modares University. (In Farsi).
38. Vapnik, V. 1995. *The Nature of Statistical Learning Theory*. Springer. New York.
39. Vasant, S. A. and S. D. Talegaokar. 2014. Hydrological study of man (Chandrabhaga) river. *International Journal of Advances in Engineering and Technology* 3(7): 807-817.
40. Viola, F., V. Noto, M. Cannarozzo and G. La Loggia. 2011. Regional flow duration curves for ungauged sites in Sicily. *Hydrology Earth System Science* 15: 323-331.
41. Zakaria, A. Z. and A. Shabri. 2012. Streamow forecasting at unengaged sites using support vector machines. *Applied Mathematical Sciences* 60(6): 3003-3014.

## Assessment the Performance of Support Vector Machine and Artificial Neural Network Systems for Regional Flood Frequency Analysis (A Case Study: Namak Lake Watershed)

E. Shrifi Garmdareh<sup>1</sup>, M. Vafakhah<sup>1\*</sup> and S. Eslamian<sup>2</sup>

(Received: May 23-2015 ; Accepted: June 9-2018)

### Abstract

Flood discharge estimation with different return periods is one of important factors for water structures design and installation. On the other hand, a lot of rivers existing in Iran watersheds have no complete and accurate hydrometric data. In these cases, one of the suitable solutions to estimate peak discharges with different return periods is the regional flood analysis. In this research, 55 hydrometric stations were used. For this purpose, at first, peak discharges in different return periods were estimated using the EasyFit software. Then, the effective variables on the peak discharges were collected and the input variables of the models were selected by using gamma test with the help of the WinGamma software. Finally, data modeling was performed using the support vector machine, artificial neural networks and nonlinear multivariate regression techniques. Quantitative and qualitative assessment of the results using various indices including Nash-Sutcliffe Efficiency Coefficient (NSC) showed that SVM modeling method had the most accuracy in comparison to the other two modeling methods to predict the peak discharges in the Namak Lake Watershed.

**Keywords:** Hydrometric Stations, Return Period, Gamma Test, Support Vector Machine, Namak Lake Watershed

- 
1. Department of Watershed Management Engineering, Faculty of Natural Resources, Tarbiat Modares University, Nour, Iran.
  2. Department of Water Engineering, College of Agriculture Engineering, Isfahan University of Technology, Isfahan, Iran.
- \*: Corresponding Author, Email: vafakhah@modares.ac.ir