

ارزیابی مدل‌های هوش مصنوعی در مدل‌سازی جریان رودخانه، مطالعه موردی: رودخانه گاماسیاب

معصومه زینعلی^{۱*}، محمدرضا گلابی^۲، محمدرضا شریفی^۳ و مریم حافظ پرست مودت^۴

^۱ کارشناس ارشد مهندسی منابع آب، دانشکده علوم و مهندسی کشاورزی، دانشگاه رازی، ^۲ دانشجوی دکتری منابع آب، دانشکده مهندسی علوم آب، دانشگاه شهید چمران، ^۳ استادیار، دانشکده مهندسی علوم آب، دانشگاه شهید چمران و ^۴ استادیار، دانشکده علوم و مهندسی کشاورزی، دانشگاه رازی

تاریخ پذیرش: ۱۳۹۷/۰۴/۰۲

تاریخ دریافت: ۱۳۹۶/۰۷/۲۳

چکیده

با پیش‌بینی جریان رودخانه‌ها علاوه بر مدیریت بهره‌برداری از منابع آب، می‌توان حوادث طبیعی نظیر سیل و خشکسالی را نیز پیش‌بینی و مهار کرد. استفاده از مدل‌های جدید در این زمینه می‌تواند به مدیریت و برنامه‌ریزی صحیح کمک کند. در این مطالعه، به ارزیابی سه مدل به نام‌های، برنامه‌ریزی بیان ژن (GEP)، شبکه بیزین (BN) و ماشین بردار پشتیبان (SVM) پرداخته شده است. داده‌های مورد استفاده برای این پژوهش، داده‌های بارش و جریان روزانه رودخانه گاماسیاب نهانوند در یک دوره ۱۰ ساله (۱۳۸۱-۱۳۹۱) می‌باشد. نتایج نشانگر برتری نسبی مدل برنامه‌ریزی بیان ژن نسبت به بقیه مدل‌ها بود و مدل ماشین بردار پشتیبان (SVM) تقریباً عملکرد بهتری نسبت به شبکه بیزین در مدل‌سازی جریان روزانه رودخانه داشت. علاوه بر این، سرعت اجرای مدل برنامه‌ریزی بیان ژن نسبت به بقیه مدل‌ها بیشتر بود و در زمان کوتاهی قادر به ارائه نتایج بود. همچنین، مدل SVM در تخمین مقادیر کمینه پایانی نیز دچار بیش‌برازش شده است. در نهایت مدل برنامه‌ریزی بیان ژن با ضریب تبیین ۰/۹۲۳۰ و جذر میانگین مربعات ۰/۵۸۶۷ در مرحله آموزش و ضریب تبیین ۰/۹۰۲۵ و جذر میانگین مربعات ۰/۴۹۳۶ در مرحله تست، به‌عنوان مدل برتر انتخاب شد.

واژه‌های کلیدی: مدل‌سازی جریان، مدیریت بهره‌برداری از منابع آب، مدل BN، مدل GEP، مدل SVM

مقدمه

مدل‌ها و روش‌های مختلفی برای تخمین و شبیه‌سازی فرایند بارش-رواناب ارائه شده است که با استفاده از آن‌ها شدت جریان رودخانه پیش‌بینی می‌شود. برای پیش‌بینی جریان رودخانه، استفاده از مدل‌های آماری، هیدرولیکی و هیدرولوژیکی دارای سابقه‌ای طولانی هستند. در این میان نبود قطعیت و غیر خطی بودن قوی روابط بین متغیرها، مسأله را پیچیده می‌کند. همچنین، مدل‌های فیزیکی و مفهومی هیدرولوژیکی

بهره‌برداری بهینه و مناسب از منابع آب، یکی از وظایف مهم مدیران منابع آب می‌باشد. از طرف دیگر، عدم تخمین مناسب رواناب حاصل از بارش در حوزه‌های آبخیز، مدیریت بهینه منابع آبی و به‌ویژه مدیریت بهره‌برداری از مخازن سدها را دچار مشکل می‌کند. در این میان، مقوله شبیه‌سازی به‌عنوان راه‌کاری مناسب برای تخمین رواناب، نمود پیدا می‌کند. تا کنون

به دلیل نیاز به اطلاعات فراوان، پارامترهای گوناگون و واسنجی وقت‌گیر، کمتر مورد توجه قرار گرفته‌اند. به همین دلیل، برآورد صحیح و دقیق جریان رودخانه با استفاده از مدل‌های مختلف یکی از موضوعاتی است که در منابع آب مورد بررسی پژوهشگران است. این مسئله به عنوان یکی از چالش‌های مدیریت منابع آب در دهه‌های اخیر مطرح بوده است، در این راستا، استفاده از مدل‌های جدید در این زمینه می‌تواند به مدیریت و برنامه‌ریزی صحیح کمک کند. فرامدل‌های^۱ شبیه‌ساز برنامه‌ریزی بیان ژن (GEP)، شبکه عصبی-فازی (ANFIS)، شبکه بیزین (BN) و ماشین بردار پشتیبان (SVM) از جدیدترین الگوریتم‌های فراکاوشی^۲ هستند که به دلیل دارا بودن دقت کافی، مورد توجه پژوهشگران قرار گرفته‌اند. زیرا با استفاده از آن‌ها، عدم قطعیت‌های موجود را به همراه درک روابط ذاتی میان داده‌ها که ریاضیات کلاسیک قادر به حل آن‌ها نمی‌باشد، می‌توان مدل‌سازی کرد. در زمینه استفاده از مدل‌های فوق، برای پیش‌بینی جریان رودخانه می‌توان به مطالعات زیر اشاره کرد.

Ghorbani و همکاران (۲۰۱۰) عملکرد سه روش برنامه‌ریزی ژنتیک، شبکه عصبی مصنوعی و نروفازی را در روندیابی سیلاب رودخانه قزل ایرماق ترکیه مورد ارزیابی قرار دادند. نتایج این مطالعه نشان داد که از بین سه روش مذکور مدل برنامه‌ریزی ژنتیک با دقت بیشتری هیدروگراف خروجی را شبیه‌سازی کرده است.

Nabzadeh و همکاران (۲۰۱۲) جریان رودخانه را با استفاده از ANFIS و FIS در حوزه آبخیز لیقوان چای در استان آذربایجان شرقی و با استفاده از داده‌های روزانه دبی، به مدت شش سال جریان رودخانه را پیش‌بینی کردند و نتایج پژوهش آن‌ها نشان داد، با توجه به معیارهای ارزیابی و این‌که ANFIS دبی‌های اوج را به خوبی برآورد و پیش‌بینی می‌کند، عملکرد بهتری نسبت به FIS دارد.

Adamowski (۲۰۱۳) مدل‌های ماشین بردار پشتیبان و شبکه عصبی مصنوعی را برای شبیه‌سازی بارش-رواناب حوضه کوهستانی دهرادان در هندوستان به کار

¹ Meta models² Heuristic³ Filyos⁴ Auto regressive moving average

شرایط کمبود داده و اطلاعات، جایگزین مناسبی برای مدل‌های مفهومی باشد.

مواد و روش‌ها

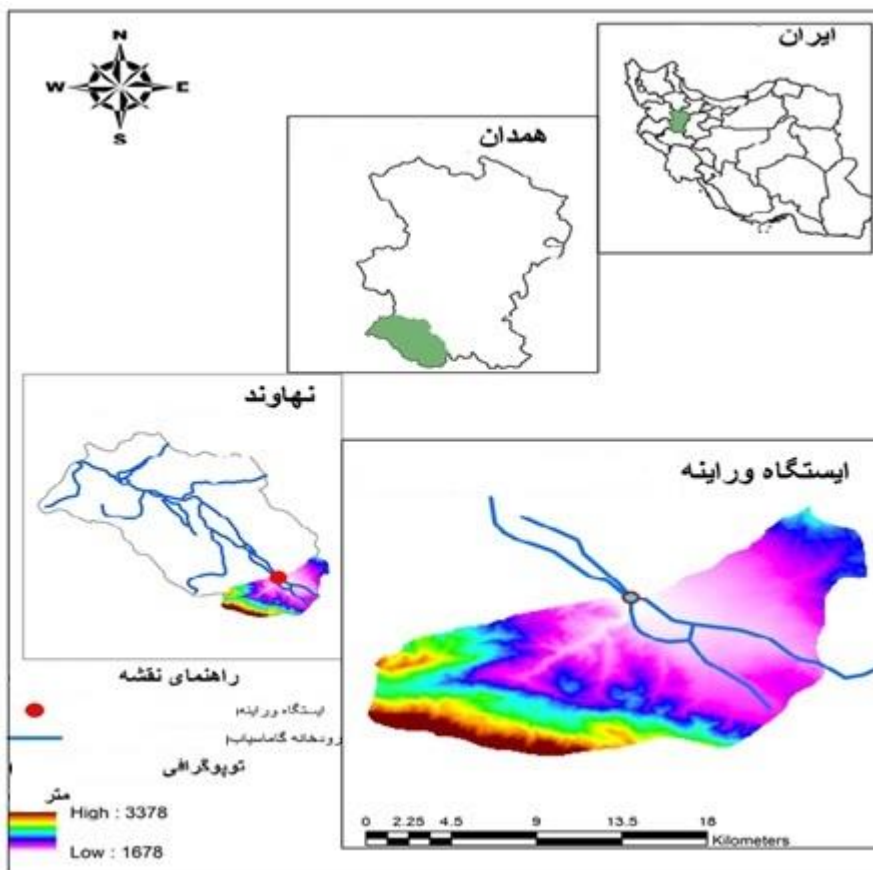
منطقه مورد مطالعه، رودخانه گاماسیاب است که در غرب کشور، در محدوده استان‌های همدان، کرمانشاه و لرستان واقع شده است. رودخانه گاماسیاب از چشمه‌های کارستی گاماسیاب در ۲۰ کیلومتری جنوب غربی شهر نهاوند و در فاصله اندکی از جاده ارتباطی نهاوند به نورآباد لرستان، از ارتفاع ۱۸۶۰ متری از محلی به نام کوه سنگ سوراخ سرچشمه می‌گیرد. در این پژوهش، به بررسی بخشی از حوزه آبخیز رودخانه گاماسیاب از قسمت ابتدا تا نقطه ایستگاه وراینه پرداخته شده است. ایستگاه وراینه در موقعیت جغرافیایی "۱۵' ۲۴" ۴۸' طول شرقی و "۳۲' ۴۰" عرض شمالی قرار دارد. این ایستگاه در سال ۱۳۴۸ تأسیس شده و دارای ارتفاع ۱۷۹۵ متر از سطح دریا با میانگین بارش سالانه دراز مدت ۵۲۱ میلی‌متر می‌باشد. داده‌های مورد استفاده برای این پژوهش، داده‌های بارش و جریان روزانه رودخانه گاماسیاب نهاوند در یک دوره ۱۰ ساله (۱۳۹۱-۱۳۸۱) است. با توجه به این‌که داده‌ها به صورت کامل موجود بودند، بررسی داده‌ها فقط از نظر داده‌های پرت مورد بررسی قرار گرفته، مشکلات موجود برطرف شد. با فرض وابستگی جریان رودخانه به جریان روزهای قبل، اقدام به تخمین جریان روز بعد با استفاده از داده‌های بارش و جریان امروز شد. از بین داده‌های ورودی ۷۵ درصد (۲۷۳۹ داده) از آن‌ها را برای آموزش مدل و ۲۵ درصد (۹۱۳ داده) باقی‌مانده برای آزمون مدل انتخاب شد. برای مرحله پیش‌بینی یا شبیه‌سازی از داده‌های سال آبی (۱۳۹۱-۱۳۹۰) استفاده شد (جدول ۱). شکل ۱ موقعیت ایستگاه وراینه را نشان می‌دهد.

داشته‌اند. Ghorbani و همکاران (۲۰۱۶) کارایی سه مدل شبکه عصبی بیزین، GEP و SVM را در تخمین رواناب ماهانه رودخانه کاکارضا مقایسه کردند. از میان سه مدل ذکر شده، شبکه عصبی بیزین بهترین عملکرد را داشته است. Naeimi Kalourazi و همکاران (۲۰۱۷) سه روش شامل مدل درختی، نزدیک‌ترین همسایگی و رگرسیون خطی چند متغیره را برای برآورد آبدهی در حوزه‌های آبخیز بدون آمار استان گلستان بررسی کردند. Moatamednia و همکاران (۲۰۱۷) با استفاده از داده‌های هواشناسی و هیدرومتری طی دوره زمانی ۱۳۵۰-۱۳۴۹ تا ۱۳۹۱-۱۳۹۰ رواناب در حوزه آبخیز امامه با استفاده از مدل‌های شبکه عصبی پرسپترون چند لایه، تابع پایه شعاعی و سامانه عصبی فازی تطبیقی تخمین زدند. نتایج مطالعه آن‌ها نشان داد، از بین مدل‌های یاد شده سامانه عصبی فازی تطبیقی عملکرد بسیار خوبی داشته است و به خوبی می‌تواند رواناب را پیش‌بینی کند.

لذا، هدف از این مطالعه در ابتدا بیان کردن تفاوت‌ها و شناسایی این سه مدل و سپس مدل‌سازی جریان رودخانه گاماسیاب با استفاده از فرامدل‌های شبیه‌ساز برنامه‌ریزی بیان ژن (GEP)، شبکه بیزین (BN) و ماشین بردار پشتیبان (SVM) است که در بیشتر مطالعات، با توجه به تعداد زیاد مدل‌های موجود برای شبیه‌سازی، برتری خود را ثابت کرده‌اند. سپس، با مقایسه آن‌ها با یکدیگر برای تشخیص این‌که کدام یک از این فرامدل‌های شبیه‌ساز نسبت به بقیه برتری دارد، در نهایت پیش‌بینی جریان رودخانه با استفاده از بهترین مدل انجام می‌شود. سوال اساسی این تحقیق این است که آیا فرامدل شبیه‌ساز برتر در این مطالعه با توجه به دقت در مرحله آموزش و تست می‌تواند در

جدول ۱- متغیرهای اقلیمی ایستگاه وراینه

مقیاس زمانی	متغیر اقلیمی	واحد اندازه‌گیری	میانگین	حداقل	حداکثر
روزانه	جریان	متر مکعب در ثانیه	۳/۷۸	۰/۶۷	۲۲/۸۶
	بارش	میلی‌متر	۱/۴۳	۰/۰	۹۶



شکل ۱- موقعیت ایستگاه وراینه در شهرستان نهاوند، استان همدان و ایران

مشابه با درختان تجزیه در برنامه‌ریزی ژنتیک ترکیب می‌شوند. اولین مرحله در الگوریتم مدل، تولید جمعیت اولیه از راه حل‌هاست. این امر می‌تواند به وسیله فرایند تصادفی و یا در نظر گرفتن اطلاعات ورودی درباره مسأله انجام شود. سپس کروموزوم‌ها به صورت بیان درختی ارائه شده، به وسیله تابع برازش ارزیابی می‌شوند. در صورت دستیابی به راه حل مطلوب و یا رسیدن نسل‌ها به تعداد معین، تکامل متوقف شده و بهترین راه حل ارائه می‌شود. اگر شرایط توقف یافت نشود، نخبه‌گزینی انجام می‌شود و باقی راه‌حل‌ها به فرایندی گزینشی واگذار می‌شوند. این فرایند برای چندین نسل تکرار می‌شود و با پیش رفتن نسل به جلو کیفیت جمعیت نیز به‌طور نسبی بهبود می‌یابد (Ferreira, 2006).

فرامدل شبیه‌ساز بیان ژن نیز همانند الگوریتم ژنتیک و برنامه‌ریزی ژنتیک، یک الگوریتم ژنتیکی است به طوری که از جمعیتی از افراد استفاده می‌کند که آن‌ها را مطابق برازندگی انتخاب می‌کند و تغییرات

برنامه‌ریزی بیان ژن (GEP): برنامه‌ریزی بیان ژن که در ادامه سیر تکاملی مدل‌های هوشمند به وجود آمده است، جزء روش‌های الگوریتم گردشی محسوب می‌شود که مبنای تمامی آن‌ها بر اساس نظریه تکامل داروین استوار است (Danandehmehr و Tabatabaie, 2010). مزیت فرامدل شبیه‌ساز بیان ژن نسبت به مدل‌های دیگر از جمله شبکه عصبی مصنوعی این است که در فرامدل شبیه‌ساز بیان ژن، ابتدا ساختار (متغیرهای ورودی، هدف و مجموع توابع) تعریف شده، سپس ساختار بهینه مدل و ضرایب طی فرایند آموزش تعیین می‌شوند. در حالی که در شبکه‌های عصبی، ابتدا باید ساختار تعیین شده، فقط ضرایب مدل طی فرایند آموزش حاصل می‌شوند. همچنین، این الگوریتم به‌طور خودکار می‌تواند متغیرهای ورودی که در مدل بیشترین تأثیر را دارند، انتخاب کند. در این روش کروموزوم‌های خطی و ساده با طول ثابت، مشابه با آنچه که در الگوریتم ژنتیک استفاده می‌شود و ساختارهای شاخه‌ای با اندازه‌ها و اشکال متفاوت،

استفاده کرده، بر اساس قانون یادگیری شبکه بیزین آموزش دیده، با استفاده از اطلاعات مبهم، ناکافی، متضاد و غیر قطعی، هم در گذشته و هم در حال حاضر، پارامترها را به صورت عصب به یکدیگر ارتباط می‌دهد که فعالیت سلسله اعصاب یک شبکه و چگونگی پیوند بین آن‌ها مبنای هوشمندی شبکه را نشان می‌دهد. شبکه‌های عصبی بیزین در حل مسائلی که روابط بین متغیرها به درستی مشخص نباشد، بسیار مفید هستند.

الگوریتم بیزین: شبکه بیزین یک مدل گرافیکی احتمالاتی است که مجموعه‌ای از متغیرها و احتمالات را نشان می‌دهد. این شبکه، یک گراف مستقیم و غیر چرخه‌ای است که در آن گره‌ها در حکم متغیرهای مسئله هستند. شبکه‌های بیزین محدودیتی در نمایاندن متغیرهای تصادفی ندارند. ساختار یک شبکه بیزین در واقع یک نمایش نموداری از اثرات متقابل متغیرهایی است که باید الگوبندی شوند، علاوه بر این که کیفیت، رابطه بین متغیرهای مسئله را نشان می‌دهد، کمیت، ارتباط بین این متغیرها را نیز به نمایش می‌گذارد.

این روش بر مبنای محاسبات احتمالات شرطی (قانون بیز^۳) است. معادله زیر رابطه بیز را نشان می‌دهد.

$$p(a|b) = \frac{p(b|a) \times p(a)}{p(b)} \quad (1)$$

که در آن، $p(a)$ احتمال وقوع پیشامد a ، $p(b)$ احتمال وقوع پیشامد b ، $p(b/a)$ احتمال شرطی پیشامد b است به شرطی که پیشامد a اتفاق افتاده باشد و $p(a/b)$ احتمال شرطی پیشامد a است. شکل ۲ نمونه‌ای از گراف شبکه بیزین را به صورت چرخه‌ای و مستقیم نشان می‌دهد.

ماشین بردار پشتیبان^۴ (SVM): دسته‌بندی کننده-ای است که جزء شاخه روش‌های کرنلی^۵ در یادگیری ماشین محسوب می‌شود. این نوع سامانه یادگیری هم برای دسته‌بندی و هم برای تخمین و برآورد تابع برازش داده‌ها به کار می‌رود، به طوری که کمترین خطا

ژنتیکی را با استفاده از یک یا چند عملگر ژنتیکی اعمال می‌کند. تفاوت اساسی بین این سه الگوریتم، مربوط به ماهیت افراد آن‌ها است به طوری که در الگوریتم ژنتیک، افراد رشته‌های خطی با طول ثابت (کروموزوم‌ها) و در برنامه‌ریزی ژنتیک، نهاده‌های غیر خطی با اندازه‌ها و اشکال متفاوت (درختان تجزیه) می‌باشند، در حالی که در فرامدل شبیه‌ساز بیان ژن، افراد به صورت رشته‌های خطی با طول ثابت (ژنوم یا کروموزوم‌ها) کد گذاری شده و سپس به شکل نهاده‌های غیر خطی با اندازه‌ها و اشکال متفاوت (یعنی نمایش دیگرام ساده یا بیان درختی) بیان می‌شوند (Ferreira, 2001). برای کسب اطلاعات بیشتر با این مدل به تحقیقات Meshkani و Nazemi (2009) و Singh (2002) مراجعه شود.

شبکه بیزین^۱ (BN): نخستین بار در اوایل دهه ۱۹۹۰ به وسیله بنتین و بعدها MacKay (۱۹۹۲) برای شبکه‌های عصبی مصنوعی به کار برده شد و با سرعت به یک ابزار انتخاب برای کاربردهای هوش مصنوعی تبدیل شد (MacKay, ۱۹۹۲). در شبکه‌های بیزین برخلاف شبکه‌های عصبی متداول هدف یافتن یک بردار وزن بهینه نیست، بلکه ارائه عدم قطعیت صریح با توجه به مقادیر وزن به وسیله توزیع احتمال ثانویه است (Kingston و همکاران، ۲۰۰۵). شبکه بیزین، نمایش با معنی روابط نامشخص مابین پارامترها در یک فرایند می‌باشد و گرافی جهت‌دار غیر حلقوی از گره‌ها، برای نمایش متغیرهای تصادفی و کمان‌ها، برای نمایش روابط احتمالی میان متغیرها به شمار می‌رود. MacKay (۱۹۹۲) دو سطح از استنتاج‌های بیزین را که می‌توان در مدل‌سازی شبکه عصبی بیزین^۲ به کار رود، تشریح کرده است. اولین سطح شامل استخراج شبکه وزن‌ها می‌باشد، البته با این فرض که ساختار شبکه انتخاب شده صحیح است. سطح دوم شامل مقایسه مدل‌ها و انتخاب مدل است (MacKay, ۱۹۹۲). شبکه عصبی بیزین ابزاری مفید برای تجزیه و تحلیل عدم قطعیت در مدل‌سازی شبکه‌های عصبی مصنوعی و یکی از روش‌های سامانه پشتیبان تصمیم‌گیری است که از یک شکل هوش مصنوعی

³ Bayes theorem

⁴ Support Vector Machines

⁵ Kernel methods

¹ Bayesian network

² Bayesian neural network

به وسیله آموزش مدل SVM بر روی یک مجموعه داده به عنوان مجموعه آموزش که شامل فرایندی به منظور بهینه سازی دائمی تابع خطا است، قابل دسترسی است. بر مبنای تعریف این تابع خطا، دو نمونه از مدل های SVM شناخته شده است که عبارتند از الف) مدل های رگرسیون SVM نوع اول که به مدل های SVM-ν مشهورند و ب) مدل های رگرسیون SVM نوع دوم که به مدل های SVM ε مشهورند. در این مطالعه مدل SVM ε به دلیل کاربرد گسترده آن در مسائل رگرسیونی استفاده شد. برای این مدل تابع خطا به صورت زیر تعریف می شود.

$$\frac{1}{2}W^T W + C \sum_{i=1}^N \xi_i + C \sum_{i=1}^N \xi_i^0 \quad (3)$$

تابع خطای مذکور لازم است که با توجه به محدودیت های زیر کمینه شود.

$$W^T \Phi(x_i) + b - y \leq \varepsilon + \xi_i^0 \quad (4)$$

$$y_i - W^T \Phi(x_i) - b \leq \varepsilon + \xi_i^0$$

$$\xi_i, \xi_i^0 \geq 0 \quad i = 1, \dots, N$$

که در آن، C ثابت گنجایش، W بردار ضرایب، W^T ترانهاده بردار ضرایب، 0 ضرایب کمبود، b ضریب ثابت، N الگوی آموزش مدل و Φ تابع کرنل است.

کرنل تابع پایه شعاعی به عنوان بهترین انتخاب از بین دیگر توابع کرنل گزارش شده است (Noori و همکاران، ۲۰۰۸؛ Noori و همکاران، ۲۰۰۹؛ Dibike و همکاران، ۲۰۰۱). بنابراین، در این تحقیق تابع مذکور که به وسیله γ در رابطه زیر تعریف شده است، به کار گرفته شد.

$$K(x_i, x) = \exp(-\gamma |x_i - x|^2) \quad (5)$$

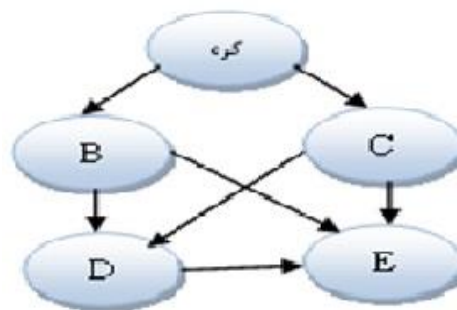
با توجه به روابط (۳) تا (۵)، لازم است برای پیش بینی دبی روزانه، به وسیله مدل SVM سه پارامتر γ ، ε و C بهینه شوند.

آماده سازی داده ها: به منظور هماهنگ تر شدن بهتر داده ها و افزایش سرعت اجرای مدل ها، داده ها با استفاده از روش استاندارد سازی به اعدادی بین صفر تا یک تبدیل شدند. با توجه به پیشنهاد Golabi و همکاران (۲۰۱۳)، از رابطه زیر برای نرمال سازی (استاندارد سازی) استفاده شده است.

$$y = 0.1 + (0.8 \times \left(\frac{x - \bar{x}}{x_{max} - x_{min}} \right)) \quad (6)$$

که در آن، X داده مورد نظر، \bar{X} میانگین داده ها،

در دسته بندی داده ها و یا تابع برازش رخ دهد. ماشین بردار پشتیبان در سال ۱۹۹۲ به وسیله Vapnik که یک ریاضیدان روس بود، بر پایه نظریه یادگیری آماری^۱ بنا شده است (Vapnik و Cortes، ۱۹۹۵). هدف ماشین بردار پشتیبان این است که یک تابع $f(x)$ را برای الگوهای آموزشی x طوری تشخیص دهد که بیشینه حاشیه را از مقادیر آموزشی y داشته باشد. به عبارتی دیگر، ماشین بردار پشتیبان مدلی است که یک منحنی با ضخامت ε را به داده ها برازش می دهد به طوری که کمترین خطا در داده های آزمایشی صورت بگیرد.



شکل ۲- نمایش گراف چرخه ای و مستقیم شبکه بیزین

در مورد تئوری مدل های مطرح شده در این پژوهش، تحقیقات زیادی انجام شده است (Vapnik، ۱۹۹۸؛ Chen و Yu، ۲۰۰۷). لذا، در این تحقیق تنها به شرح خلاصه ای از مدل مورد استفاده SVM به نام SVM ε پرداخته شده است. در یک مدل رگرسیونی SVM لازم است، وابستگی تابعی متغیر وابسته y به مجموعه ای از متغیرهای مستقل x تخمین زده شود. فرض بر این است که مانند دیگر مسائل رگرسیونی، رابطه بین متغیرهای وابسته و مستقل به وسیله یک تابع معین f به علاوه یک مقدار اضافی نویز مشخص می شود.

$$y = f(x) + noise \quad (2)$$

بنابراین، موضوع اصلی پیدا کردن فرم تابع f است که بتواند به صورت صحیح، موارد جدیدی را که SVM تا کنون تجربه نکرده است، پیش بینی کند. این تابع

¹ Statistical learning theory

هر چه مقدار این عدد کمتر باشد، به تبع، آموزش و شبیه‌سازی داده‌ها بهتر صورت گرفته است. همچنین، هرچه مقدار پارامتر نش به یک نزدیک‌تر باشد، نشان‌دهنده تطابق بیشتر مقادیر شبیه‌سازی شده با مقادیر مشاهداتی است.

نتایج و بحث

نتایج برنامه‌ریزی بیان ژن (GEP): با استفاده از انواع ترکیب‌ها از تعداد کروموزوم‌ها و تعداد ژن‌ها و اندازه هد (جدول ۲) و بررسی ترکیب‌های مختلف مدل برنامه‌ریزی بیان ژن با استفاده از معیارهای ارزیابی مدل مانند جذر میانگین مربعات خطا (RMSE) و ضریب تبیین (R^2) بهترین مدل با ضریب تبیین ۰/۹۹۳۰ و جذر میانگین مربعات ۰/۵۸۶۷ در مرحله، آموزش و ضریب تبیین ۰/۹۰۲۵ و جذر میانگین مربعات ۰/۴۹۳۶ در مرحله تست انتخاب شد. شکل ۳، مقایسه مقادیر مشاهداتی با مقادیر برآوردی برای ساختار برتر فرامدل شبیه‌ساز بیان ژن در مرحله تست را نشان می‌دهد.

در شکل ۴، ساختار درختی برای ترکیب برتر فرامدل شبیه‌ساز بیان ژن نشان داده است. این ساختار، در هر مرحله جمعیت اولیه را به صورت ساختار خطی ساده نشان می‌دهد و همه تغییرات فقط بر روی ساختارهای ساده انجام می‌گیرد، در نتیجه، نیازی به ساختارهای نسبتاً پیچیده برای گسترش در هر مرحله نخواهد بود. در این شکل، مقادیر C مربوط به ضرایب فرامدل شبیه‌ساز بیان ژن و مقادیر d مربوط به پارامترهای ورودی می‌باشد که شامل بارش و جریان روز قبل می‌باشد.

X_{max} بیشینه داده‌ها، X_{min} کمینه داده‌ها و y داده استاندارد شده می‌باشد. سپس در این تحقیق برای ایجاد مدل ۷۰ درصد داده‌ها برای آموزش و ۳۰ درصد برای صحت‌سنجی مدل به کار رفت. در این تحقیق، $P(t)$ (بارندگی در زمان t) و $Q(t)$ (دبی در زمان t) به عنوان ورودی و $Q(t+1)$ (دبی در زمان $t+1$) به عنوان خروجی مدل در نظر گرفته شد. به عبارت دیگر، هدف پیدا کردن رابطه‌ای بین پارامتر خروجی (Q_{t+1}) و پارامترهای ورودی (P_t و Q_t) می‌باشد که این را به زبان ریاضی می‌توان به صورت رابطه $Q_{t+1} = f(Q_t, P_t)$ نشان داد.

معیارهای ارزیابی مدل‌ها: برای ارزیابی مدل‌ها از سه معیار، ضریب تبیین (R^2)، جذر میانگین مربعات خطا (RMSE) و پارامتر نش (NASH) استفاده شد. پارامتر آماری نش در سال ۱۹۷۰ به وسیله نش و ساتکلیف معرفی شد (Sutcliffe و Nash، ۱۹۷۰). در زیر روابط این معیارها ارائه شده است.

$$RMSE = \sqrt{\frac{\sum(Q_{obs} - Q_{pre})^2}{n}} \quad (7)$$

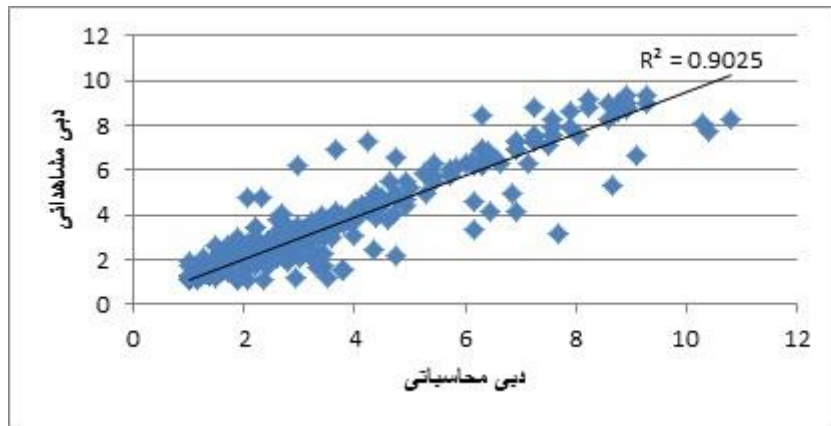
$$R^2 = 1 - \frac{\sum_{i=1}^N (Q_{obs} - Q_{pre})^2}{\sum_{i=1}^N (Q_{obs} - \bar{Q})^2} \quad (8)$$

$$NASH = 1 - \frac{\sum_{i=1}^N (Q_{obs} - Q_{pre})^2}{\sum_{i=1}^N (Q_{obs} - \bar{Q})^2} \quad (9)$$

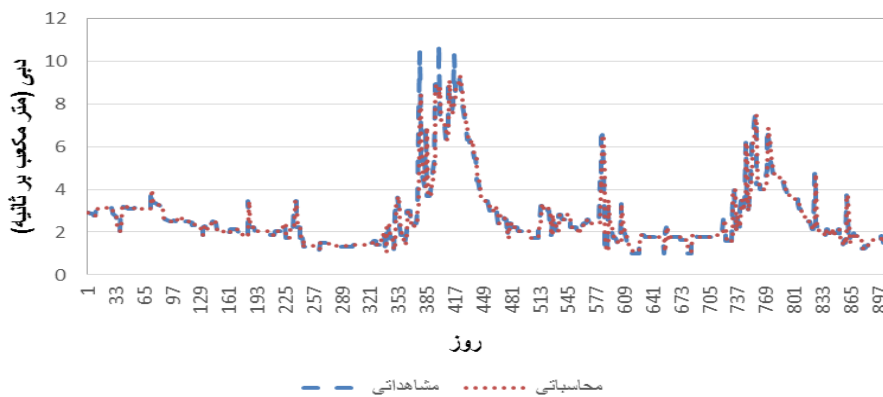
که در آن‌ها، n تعداد داده‌ها، Q_{obs} داده‌های مشاهداتی، \bar{Q} میانگین داده‌های مشاهداتی، Q_{pre} داده‌های محاسباتی، Q_{max} بیشینه داده‌ها، Q_{min} کمینه داده‌ها می‌باشند. ضریب R^2 میزان انطباق داده‌هایی که به وسیله مدل به وجود می‌آید و داده‌های واقعی را نشان می‌دهد. RMSE جذر میانگین مربع خطاهای داده‌های محاسباتی و مشاهداتی را بیان می‌کند. واضح است که

جدول ۲- مقادیر پارامترهای استفاده شده در روش برنامه‌ریزی بیان ژن (یافته‌های پژوهش)

عملکردهای ژنتیکی	تنظیمات کلی
تعداد کروموزوم‌ها	۳۰
اندازه راس	۷
تعداد ژن در هر کروموزوم	۳
تعداد جمعیت تولیدی	۱۰۰۰
تابع پیوند	مجموع (+)
نرخ جهش	۰/۰۴۴
نرخ وارون‌سازی	۰/۱
نرخ ترانهش درج متوالی	۰/۱
نرخ ترانهش ریشه درج متوالی	۰/۱
نرخ ترکیب تک نقطه‌ای	۰/۳

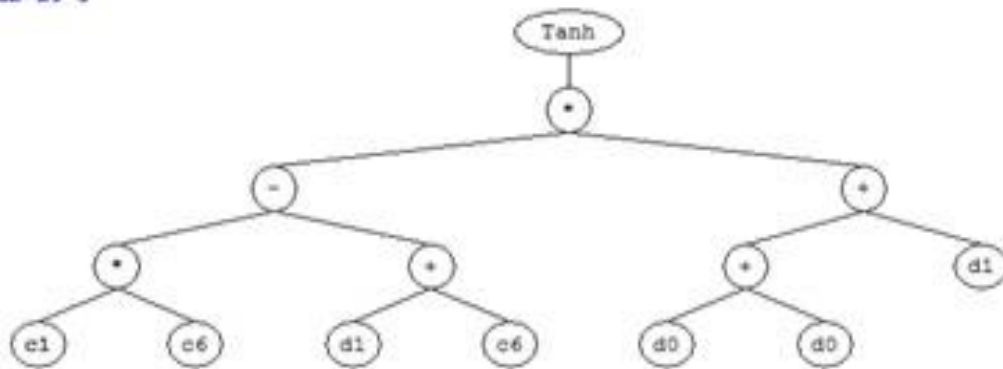


مرحله تست



شکل ۳- مقایسه نتایج مدل برنامه‌ریزی بیان ژن با مقادیر مشاهده‌ای، مرحله تست

Sub-ET 1



Sub-ET 2



شکل ۴- ساختار درختی ترکیب برتر مدل برنامه‌ریزی بیان ژن

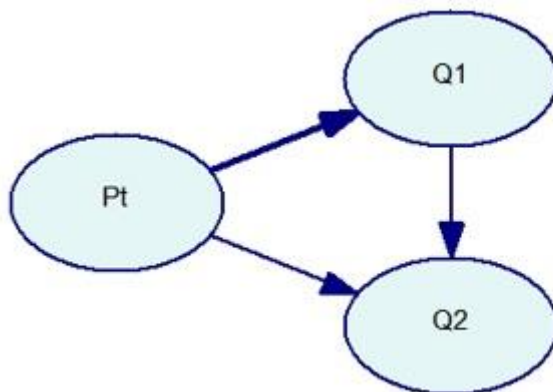
اساس یک مدل مفهومی از سامانه‌ای که قرار است مدل‌سازی و مدیریت شود، شکل می‌گیرد. بنابراین، ساخت مدل مفهومی، اولین قدم در ساختن مدل بیزین است. در این مطالعه، مدل مفهومی بیزین از طریق الگوهای ورودی ساخته شد (شکل ۵).

شبکه بیزین (BN): مراحل ساخت و توسعه مدل شبکه بیزین در مطالعه حاضر برای مدل‌سازی جریان روزانه رودخانه گاماسیاب به صورت زیر انجام شد. ۱- تشکیل چارچوب مدل: شامل تعریف متغیرها و روابط موجود بین آن‌ها در سامانه است و معمولاً بر

پس از تشکیل چارچوب و تعریف مدل، نوبت به آموزش شبکه ساخته شده می‌رسد. در این مرحله، با استفاده از ورودی‌های مختلف (بارش و دبی) در مدل سعی در ایجاد ارتباط بین داده‌ها داشته و شکل ریاضی این ارتباط به دست آمد. حال، با ارتباط ریاضی موجود می‌توان داده مصنوعی تولید کرده و عملکرد مدل را مورد ارزیابی قرار داد.

۲- تعیین نوع ارتباط بین ورودی‌های مختلف مدل در ساختار مدل مفهومی: در مطالعه حاضر از سری روزانه داده‌های بارش و جریان به عنوان ورودی‌های اصلی استفاده شد. حال، در تعریف مفهومی برای مدل بیزین این‌گونه در نظر گرفته شد که دبی فردا متأثر از دبی روز گذشته و بارش می‌باشد (شکل ۵).

۳- آموزش شبکه و به دست آوردن رابطه بین داده‌ها:



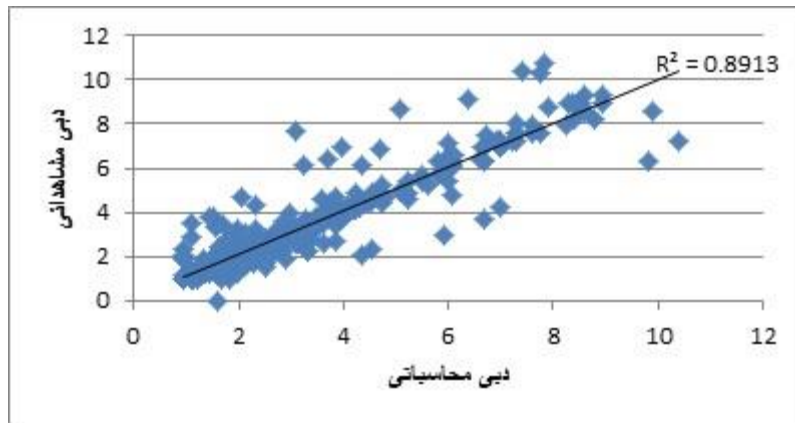
شکل ۵- مدل مفهومی ساخته شده برای شبکه بیزین

مذکور به دست آید که بدین منظور دو مشخصه ε و C به وسیله الگوریتم بهینه‌سازی جستجوی شبکه و متغیر γ نیز به صورت آزمون و خطا محاسبه شد و البته قابل ذکر است که الگوریتم بهینه‌سازی جستجوی شبکه بسیار کند عمل می‌کند و زمان محاسباتی زیادی را به خود اختصاص می‌دهد. برای حل این مشکل، در تحقیق مذکور طبق توصیه Yu و همکاران (۲۰۰۶) از برنامه اصلاح شده الگوریتم جستجوی شبکه که به نام الگوریتم جستجوی شبکه دو مرحله‌ای معروف است، به همراه اعتبارسنجی مقاطع استفاده شد. برای این منظور، ابتدا با انتخاب شبکه‌هایی با ابعاد بزرگ محدوده مشخصه‌های ε و C به ازای مقدار ثابت مشخصه γ تعیین شد. سپس، با مشخص شدن محدوده مذکور و تقسیم آن به شبکه‌هایی با ابعاد ریزتر مقادیر دقیق دو مشخصه ε و C مشخص شدند. روند مذکور برای دیگر مقادیر γ نیز تکرار شد و بدین طریق مدل‌های متفاوتی با تغییر در مقدار γ حاصل شدند. حال، می‌توان از بین مدل‌های توسعه داده شده، مدل با کمترین خطا را تعیین کرده و مشخصه‌های آن را به عنوان مقادیر بهینه ε ، C و γ انتخاب کرد.

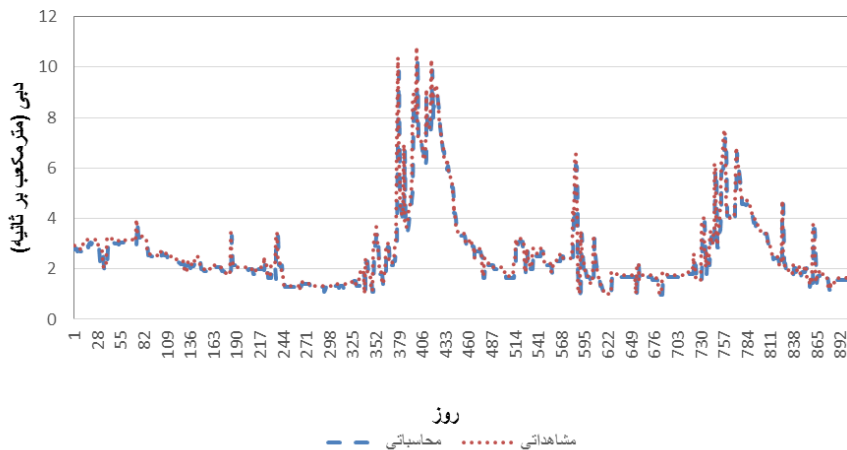
همچنین، رابطه ریاضی حاصل از الگوی ورودی به صورت رابطه (۹) به دست آمد. مدل بیزین با ضریب تبیین ۰/۹۱۲ و جذر میانگین مربعات ۰/۵۷۲۱ در مرحله آموزش و ضریب تبیین ۰/۸۹۱۲ و جذر میانگین مربعات ۰/۵۳۱۷ در مرحله تست انتخاب شد. شکل ۶، مقایسه مقادیر مشاهداتی با مقادیر برآوردی برای ساختار برتر فرامدل شبیه‌ساز بیزین در مرحله تست را نشان می‌دهد. همچنین، رابطه ریاضی حاصل از الگوی ورودی به صورت رابطه (۹) به دست آمد.

$$Q_2 = -0.00861381 \times P_t + 0.961715 \times Q_1 + \text{Normal}(0.129008, 0.600809) \quad (9)$$

ماشین بردار پشتیبان (SVM): در این مرحله، مقادیر بهینه مشخصه‌های مدل SVM شامل ε و C تعیین می‌شود. همچنین، در این مطالعه تابع کرنل مورد استفاده، تابع RBF انتخاب شد، زیرا از دقت بیشتری در برآورد جریان روزانه رودخانه گاماسیاب برخوردار بود. در این تابع باید مشخصه γ تعیین شود. بنابراین، در حالت کلی برای مدل‌سازی و پیش‌بینی جریان روزانه رودخانه گاماسیاب به وسیله مدل SVM رگرسیونی، لازم است که مقادیر بهینه سه مشخصه



مرحله تست



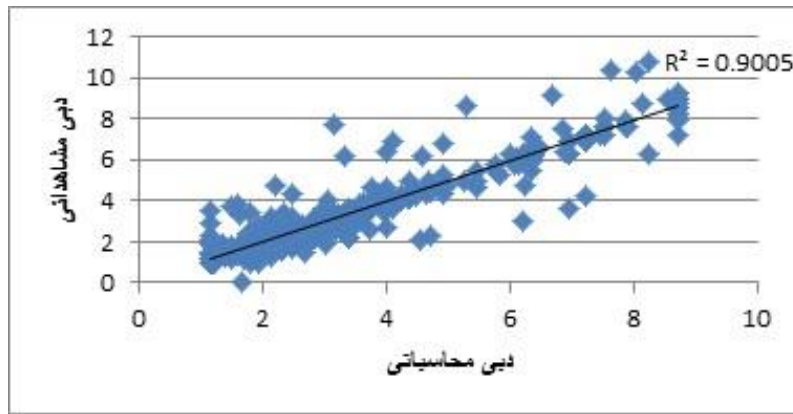
شکل ۶- مقایسه نتایج مدل شبکه بیزین با مقادیر مشاهده‌ای، مرحله تست

است. برای مقایسه کمی بین سه مدل از چهار معیار ارزیابی به نام‌های ضریب تبیین، جذر میانگین مربعات خطا، مقدار آکائیکه و پارامتر نش استفاده شد. نتایج این بررسی در جدول ۳ ارائه شده است.

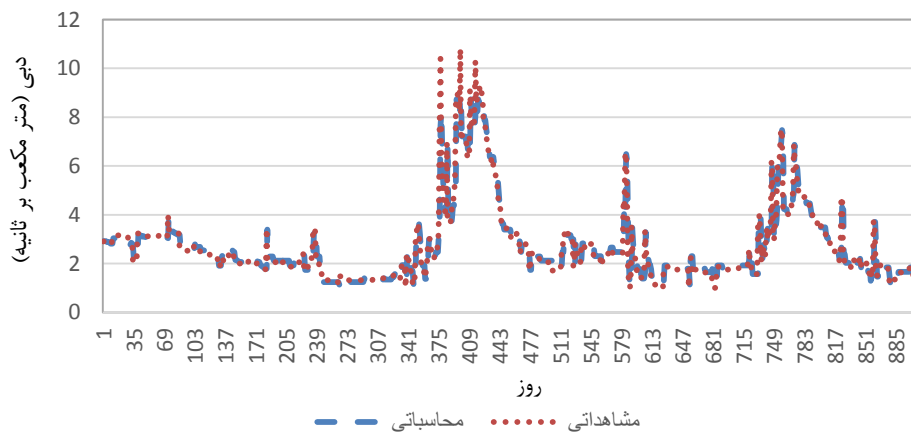
با توجه به جدول ۳ و مقایسه مدل‌ها مشاهده می‌شود، در مرحله تست، ضریب تبیین نتایج هر سه مدل تقریباً با اختلاف بسیار اندک، نزدیک به هم است و تقریباً برتری نسبی مدل GEP را می‌توان مشاهده کرد از طرفی، با توجه به پارامتر جذر میانگین مربعات خطا، اگرچه در اینجا هم اختلاف بسیار اندک در مقدار این پارامتر مشاهده می‌شود، ولی تقریباً برتری نسبی مدل GEP را می‌توان مشاهده کرد. بر اساس معیار AIC که هرچه مقادیر آن کمتر باشد، مدل عملکرد بهتری دارد، مشاهده می‌شود که در هر سه مدل، هم در مرحله آموزش و هم در مرحله تست اختلاف بسیار اندک در مقدار این پارامتر دیده می‌شود و همچنان برتری نسبی مدل GEP قابل مشاهده خواهد بود.

در تحقیق حاضر، از مدل SVM نیز با استفاده از همان الگوی ورودی بقیه مدل‌ها برای مدل‌سازی جریان رودخانه گاماسیاب استفاده شد. برای الگوی ورودی با تغییر ϵ ، C و γ شبکه‌های مختلفی ساخته و آموزش داده شد و در نهایت ساختاری که دارای کمترین مقدار خطا بود، یعنی $\epsilon=0/1$ و $C=10$ و $\gamma=0/210$ به عنوان مناسب‌ترین الگو انتخاب شد. مدل ماشین بردار پشتیبان (SVM) با ضریب تبیین $0/9235$ و جذر میانگین مربعات $0/5841$ در مرحله آموزش و ضریب تبیین $0/9005$ و جذر میانگین مربعات $0/4979$ در مرحله تست انتخاب شد. شکل ۷، مقایسه مقادیر مشاهداتی با مقادیر برآوردی برای ساختار برتر فرامدل شبیه‌ساز ماشین بردار پشتیبان (SVM) در مرحله تست را نشان می‌دهد.

مقایسه نتایج مدل‌های استفاده شده: در این قسمت، به منظور این که مقایسه مناسبی بین عملکرد سه مدل صورت گیرد، مقایسه کمی صورت گرفته



مرحله تست



شکل ۷- مقایسه نتایج مدل ماشین بردار پشتیبان با مقادیر مشاهده‌ای، مرحله تست

جدول ۳- مقایسه مدل‌های استفاده شده در این پژوهش (یافته‌های پژوهش)

مدل	ضریب تبیین R^2		جذر میانگین مربعات خطا RMSE		مقدار آکائیکه AIC		پارامتر نش NASH	
	تست	آموزش	تست	آموزش	تست	آموزش	تست	آموزش
GEP	۰/۹۰۲۵	۰/۹۲۳۰	۰/۵۸۶۷	۰/۴۹۳۶	۵۴۷۶/۹۳	۱۸۲۴/۵۸	۰/۹۱۳۵	۰/۹۰۲۰
BN	۰/۸۹۱۲	۰/۹۱۲	۰/۵۷۲۱	۰/۵۳۱۷	۵۵۴۹/۳۲	۱۸۲۴/۷۳	۰/۸۹۴۲	۰/۸۸۶۲
SVM	۰/۹۰۰۵	۰/۹۲۳۵	۰/۵۸۴۱	۰/۴۹۷۹	۵۴۷۶/۹۲	۱۸۲۴/۶۰	۰/۹۱۲۴	۰/۹۰۰۲

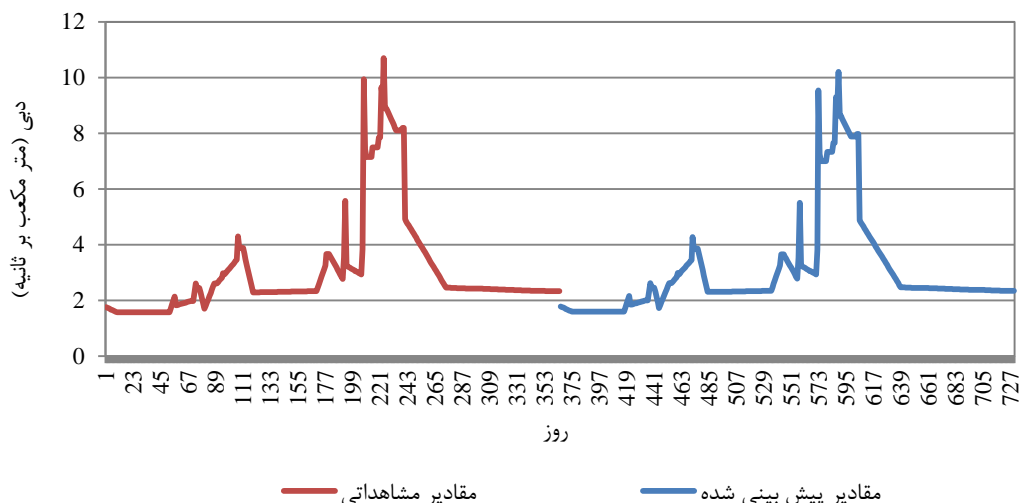
اختلاف بسیار جزئی بهتر از بقیه فرامدل‌های شبیه ساز استفاده شده در این مطالعه است، طبق جدول ۳، برنامه‌ریزی بیان ژن در رتبه اول و ماشین بردار پشتیبان در رتبه دوم و مدل شبکه بیزین در این مطالعه در رتبه سوم قرار می‌گیرد.

در مرحله بعد از مدل‌سازی و انتخاب فرا مدل شبیه‌ساز بیان ژن، هدف پیش‌بینی مقادیر جریان برای روزهای آتی با استفاده از فرا مدل شبیه‌ساز بیان ژن

بر اساس معیار NASH، هرچه مقادیر این پارامتر به یک نزدیک‌تر باشد، مدل عملکرد بهتری دارد، لذا، مشاهده می‌شود که مدل SVM دارای عملکرد بهتری نسبت به مدل BN داشته، ولی همچنان مدل GEP نسبت به بقیه مدل‌ها دارای برتری نسبی است. بنابراین، می‌توان با مشاهده مقادیر جدول به این نتیجه رسید که برای مدل‌سازی جریان روزانه رودخانه گاماسیاب عملکرد فرامدل شبیه‌ساز بیان ژن با

مشخص است، پیش‌بینی جریان رودخانه برای یک سال آبی و به مدت ۳۶۵ روز انجام شده است.

می‌باشد که نمودار آن در شکل ۸ نشان داده شده است. برای انجام این مرحله، از داده‌های سال آبی (۱۳۸۱-۱۳۸۰) استفاده شده است، همان‌طور که



شکل ۸- مقادیر مشاهداتی و پیش‌بینی شده دبی روزانه رودخانه گاماسیاب، با مدل برنامه‌ریزی بیان ژن

باشد. علاوه بر این، سرعت اجرای مدل برنامه‌ریزی بیان ژن نسبت به بقیه مدل‌ها بیشتر بود و در زمان کوتاهی قادر به ارائه نتایج بود. همچنین، مدل SVM در تخمین مقادیر کمینه پایانی نیز دچار بیش برآزش شده است. طبق نتایج Danandehmehr و همکاران (۲۰۱۴)، برنامه‌ریزی ژنتیک خطی عملکرد بهتری نسبت به شبکه عصبی مصنوعی در پیش‌بینی جریان ماهانه رودخانه داشته است. Solgi و همکاران (۲۰۱۷) نشان دادند که می‌توان از ترکیب مدل برنامه‌ریزی بیان ژن با تبدیل موجک به‌عنوان یک ابزار مناسب برای مدل‌سازی و پیش‌بینی جریان رودخانه گاماسیاب بهره برد و همچنین، Baba Ali و Dehghani (۲۰۱۷) نشان دادند که هر دو مدل شبکه عصبی موجک و شبکه عصبی مصنوعی، قابلیت خوبی در تخمین دبی سیلاب دارند. اما، از لحاظ دقت، شبکه عصبی موجک عملکرد بهتری نسبت به شبکه عصبی مصنوعی از خود نشان داده است. نتایج این پژوهش با موارد ذکر شده هم‌خوانی دارد.

نتیجه‌گیری

در این تحقیق، از سه مدل به‌نام‌های برنامه‌ریزی بیان ژن (GEP)، شبکه بیزین (BN) و ماشین بردار پشتیبان (SVM) برای مدل‌سازی جریان روزانه رودخانه گاماسیاب استفاده شد. هدف از این مطالعه، نشان دادن تفاوت‌های بین سه مدل و همچنین، بررسی توانایی آن‌ها برای مدل‌سازی جریان روزانه رودخانه گاماسیاب بوده است. نتایج بیانگر این بود که فرامدل شبیه‌ساز بیان ژن عملکرد بهتری نسبت به دو مدل دیگر دارد. همان‌طور که از نمودار مقادیر پیش‌بینی شده مشخص است، در داده‌های پیش‌بینی، فصل یا تناوب و همچنین سیکل، به درستی شبیه‌سازی و پیش‌بینی شده است و این امر بیانگر آن است که فرامدل شبیه‌ساز بیان ژن توانایی خوبی برای شبیه‌سازی و پیش‌بینی جریان روزانه رودخانه دارد. این فرامدل شبیه‌ساز، با توجه به دقت برآورد در مرحله آموزش و تست می‌تواند در شرایط کمبود داده و اطلاعات، جایگزین مناسبی برای مدل‌های مفهومی

منابع مورد استفاده

- Adamowski, J. 2013. Using support vector regression to predict direct runoff, base flow and total flow in a mountainous watershed with limited data in Uttaranchal, India. *Journal of Land Reclamation*, 45(1): 71-83.

2. Ahmadi, F., F. Radmanesh and R. Mirabbasi Najaf Abadi. 2015. Comparison between genetic programming and support vector machine methods for daily river flow forecasting, case study: Barandoozchay River. *Journal of Water and Soil*, 28(6): 1162-1171 (in Persian).
3. Baba Ali, H.R. and R. Dehghani. 2017. Flood forecast prediction using wavelet neural network. *Journal of Hydro-Geomorphology*. 11: 149- 168 (in Persian).
4. Chen, S.T. and P.S. Yu. 2007. Real-time probabilistic forecasting of flood stages. *Journal of Hydrology*, 340(1-2): 63-77.
5. Cortes, C. and V. Vapnik. 1995. Support-Vector Networks. *Machine Learning*, 20(3): 273-295, <https://doi.org/10.1023/A:1022627411411>
6. Danandehmehr, A. and M.R. Majdzadeh Tabatabai. 2010. Prediction of daily discharge trend of river flow based on genetic programming. *Journal of Water and Soil*, 24(2): 325-33 (in Persian).
7. Danandehmehr, A., E. Kahya and C. Yerdelen. 2014. Linear genetic programming application for successive-station monthly stream flow prediction. *Journal of Computers and Geosciences*, 70: 63-72.
8. Dibike, Y.B., S. Velickov, D.P. Solomatine and M.B. Abbott. 2001. Model induction with support vector machines: Introduction and applications. *Journal of Computing in Civil Engineering*, 15(3): 208-216.
9. Ferreira, C. 2006. Gene expression programming: mathematical modeling by an artificial intelligence (studies in computational intelligence). Springer-Verlag New York, Inc. Secaucus, NJ, USA.
10. Ferreira, C. 2001. Gene expression programming: a new adaptive algorithm for solving problems. *Complex System*, 13:87-129.
11. Ghorbani, M.A., O. Kisi and M.A. Aalinezhad. 2010. A probe into the chaotic nature of daily stream flow time series by correlation dimension and largest Lyapunov methods. *Applied Mathematical Modelling*, 34(12): 4050-4057.
12. Golabi, M.R., A.M. Akhondali and F. Radmanesh. 2013. Comparison of the performance of different artificial neural network algorithms in seasonal rainfall modeling, case study: Selected stations in Khuzestan Province. *Journal of Applied Geosciences Research*, 30: 151-169 (in Persian).
13. Ghorbani, M.A. and R. Dehghani. 2016. Application of bayesian neural networks, support vector machines and gene expression programming analysis of rainfall-runoff monthly, case study: Kakarza River. *Irrigation Science and Engineering*, 39(2): 125-138 (in Persian).
14. Karimi, S., J. Shiri, O. Kisi and A.A. Shiri. 2015. Short-term and long-term streamflow prediction by using 'wavelet-gene expression' programming approach. *ISH Journal of Hydraulic Engineering*, 22(2): 1-15.
15. Kingston, G.B., M.F. Lambert and H.R. Maier. 2005. Bayesian training of artificial neural networks used for water resources modeling. *Water Resources Research*, 41(12), 11 pages, <https://doi.org/10.1029/2005WR004152>.
16. MacKay, D.J.C. 1992. A practical Bayesian framework for backpropagation networks. *Neural Computation*, 4(3): 448-472.
17. Meshkani, A. and A. Nazemi. 2009. Introduction to data mining. Ferdowsi University of Mashhad, 456 pages (in Persian).
18. Motamednia, M., A. Grub, A. Malekian, M. Saberi and K. Karimi. 2017. Routine forecast using smart models. *Journal of Ecohydrology*, 4(4): 968-979 (in Persian).
19. Nabezadeh, M., A. Mosaedi, M. Hessam, A.A. Dehghani and M. Zakerneya. 2012. Investigating efficiency fuzzy logic to predict daily river flow. *Iranian Journal of Watershed Management Science and Engineering*, 5(17): 7-14 (in Persian).
20. Naeimi Kalourazi, Z., Kh. Ghorbani, M. Salarijazi and A.A. Dehghani. 2017. Investigation of effect of basin's physiographic and climatic parameters in seasonal river flow simulation. *Iranian Journal of Ecohydrology*, 3(4): 545-555 (in Persian).
21. Nash, J.E. and J.V. Sutcliffe. 1970. River flow forecasting through conceptual models part I: A discussion of principles. *Journal of Hydrology*, 10(3): 282-290.
22. Noori, R., M.A. Abdoli, A. Ameri and M. Jalili-Ghazizade. 2008. Prediction of municipal solid waste generation with combination of support vector machine and principal component analysis, a case study of Mashhad. *Environmental Progress and Sustainable Energy*, 28(2): 249-258.
23. Noori, R., A. Karbassi, A. Farokhnia and M. Dehghani. 2009. Predicting the longitudinal dispersion coefficient using support vector machine and adaptive neuro-fuzzy inference system techniques. *Environmental Engineering Science*, 26(10): 1503-1510.
24. Noori, N. and L. Kalin. 2016. Coupling SWAT and ANN models for enhanced daily streamflow prediction. *Journal of Hydrology*, 533: 141-151.
25. Singh, V.P. Translation: M.R. Najafi. 2002. Hydrological systems for rainfall modeling. Tehran University Press, First Edition, 578 pages (in Persian).

26. Singh, G., R.K. Panda and M. Lamers. 2015. Modeling of daily runoff from a small agricultural watershed using artificial neural network with resampling techniques. *Journal of Hydroinformatics*, 17(1): 56-74.
27. Solgi, A., H. Zarei and M.R. Golabi. 2017. Performance assessment of gene expression programming model using data preprocessing methods to modeling river flow. *Journal of Water and Soil Conservation*, 24(2): 185-201 (in Persian).
28. Yu, P.S., S.T. Chen and I.F. Chang. 2006. Support vector regression for real-time flood stage forecasting. *Journal of Hydrology*, 328(3-4): 704-716.
29. Zahiri, A. and H.M. Azamathulla. 2014. Comparison between linear genetic programming and M5 tree models to predict flow discharge in compound channels. *Neural Computing and Applications*, 24: 413-420.