



## مقایسه عملکرد مدل‌های ماشین بردار پشتیبان، برنامه‌ریزی بیان ژن و شبکه بیزین در پیش‌بینی جریان رودخانه‌ها (مطالعه موردی: رودخانه کشکان)

\* رضا دهقانی<sup>۱</sup>، حجت‌اله یونسی<sup>۲</sup> و حسن ترابی‌پوده<sup>۳</sup>

<sup>۱</sup> دانشجوی دکتری سازه‌های آبی، دانشکده کشاورزی، دانشگاه لرستان، استادیار گروه مهندسی آب، دانشکده کشاورزی، دانشگاه لرستان،

<sup>۲</sup> دانشیار گروه مهندسی آب، دانشکده کشاورزی، دانشگاه لرستان

تاریخ دریافت: ۹۵/۱۰/۳؛ تاریخ پذیرش: ۹۶/۴/۲۷

### چکیده

**سابقه و هدف:** پیش‌بینی جریان رودخانه‌ها یکی از مهم‌ترین موارد کلیدی در مدیریت و برنامه‌ریزی منابع آب به‌ویژه اتخاذ تصمیمات صحیح در مواقع سیلاب و بروز خشک‌سالی‌ها است. برای پیش‌بینی میزان جریان رودخانه‌ها رویکردهای متنوعی در هیدرولوژی معرفی شده است که مدل‌های هوشمند از مهم‌ترین آن‌ها می‌باشند. در این پژوهش جهت ارزیابی دقت مدل‌ها در پیش‌بینی جریان رودخانه، از داده‌های روزانه حوضه آبریز کشکان واقع در استان لرستان استفاده شده است. جهت مدل‌سازی جریان رودخانه کشکان از مدل‌های ماشین بردار پشتیبان، برنامه‌ریزی بیان ژن و شبکه بیزین استفاده شد و نتایج برای بررسی صحت مدل‌های مورد مطالعه با یکدیگر مقایسه گردید. در پژوهش‌های معدودی هر یک از مدل‌های بیان‌شده در پیش‌بینی دبی جریان روزانه مورد بررسی قرار گرفته است؛ اما هدف این پژوهش بررسی هم‌زمان این مدل‌ها در یک حوضه برای پیش‌بینی جریان روزانه رودخانه است.

**مواد و روش‌ها:** در این پژوهش رودخانه کشکان واقع در استان لرستان به‌عنوان منطقه مورد مطالعه انتخاب شده و جریان روزانه مشاهداتی این حوضه در ایستگاه هیدرومتری پلدختر جهت واسنجی و اعتبارسنجی مدل‌ها به‌کار گرفته شد. برای این منظور، در ابتدا ۸۰ درصد از داده‌های جریان روزانه (۱۳۸۳-۱۳۹۰) برای واسنجی مدل‌ها انتخاب شده و ۲۰ درصد داده‌ها (۱۳۹۱-۱۳۹۳) جهت اعتبارسنجی مدل‌ها استفاده شد. برنامه‌ریزی ژن یک فن برنامه‌ریزی خودکار است که راه‌حل مسأله را با استفاده از برنامه‌ریزی رایانه ارائه کرده و عضوی از خانواده برنامه تکاملی است. ماشین بردار پشتیبان نیز یک سیستم یادگیری کارآمد بر مبنای تئوری بهینه‌سازی مقید است. همچنین شبکه بیزین، نمایش معنادار روابط نامشخص مابین پارامترها در یک فرآیند است و گرافی جهت‌دار غیرحلقوی از گره‌ها برای نمایش متغیرهای تصادفی و کمان‌ها برای نمایش روابط احتمالی مابین متغیرها به‌شمار می‌رود. معیارهای ضریب همبستگی، ریشه میانگین مربعات خطا، میانگین قدرمطلق خطا برای ارزیابی و نیز مقایسه عملکرد مدل‌ها در این پژوهش مورد استفاده قرار گرفت.

**یافته‌ها:** نتایج نشان داد هر سه مدل، در ساختاری متشکل از ۱ تا ۵ تأخیر زمانی نتایج بهتری نسبت به سایر ساختارهای تعیین‌شده دارد. همچنین، با توجه به معیارهای ارزیابی مشخص شد که از بین مدل‌های به‌کار رفته مدل

\* مسئول مکاتبه: [reza.dehghani67@yahoo.com](mailto:reza.dehghani67@yahoo.com)

ماشین بردار پشتیبان، بیش‌ترین دقت  $R=0/910$ ، کم‌ترین ریشه میانگین مربعات خطا  $RMSE=0/002 \text{ m}^3/\text{s}$  و کم‌ترین میانگین قدرمطلق خطا  $MAE=0/001 \text{ m}^3/\text{s}$  در مرحله صحت‌سنجی را دارا می‌باشد. نتیجه‌گیری: در مجموع، نتایج نشان داد افزایش تعداد پارامتر مؤثر در مدل‌های مختلف جهت شبیه‌سازی موجب عملکرد بهتر در تخمین دبی جریان می‌شود. همچنین نتایج نشان داد مدل ماشین بردار پشتیبان عملکرد بهتری نسبت به مدل‌های شبکه بیزین و برنامه‌ریزی بیان ژن دارد.

**واژه‌های کلیدی:** برنامه‌ریزی بیان ژن، پیش‌بینی، شبکه بیزین، کشکان، ماشین بردار پشتیبان

### مقدمه

امروزه یکی از مهم‌ترین مسائل جهت مدیریت سیلاب و جلوگیری از صدمات اقتصادی و جانی ناشی از آن، پیش‌بینی صحیح جریان رودخانه‌ها است. تخمین دقیق جریان رودخانه‌ها می‌تواند نقش به‌سزایی در برنامه‌ریزی و مدیریت منابع آب داشته باشد؛ اما فاکتورها و عوامل مختلفی بر روی این پدیده تأثیرگذارند که تحلیل آن‌ها را مشکل می‌سازد. مدل‌های آماری و رگرسیون از معمول‌ترین روش‌های تحلیلی می‌باشند که غالباً با توجه به حل خطی این پدیده‌ها نتایجی همراه با خطا ارائه داده و نمی‌توانند تغییرات زمانی پدیده موردنظر را با دقت قابل‌قبولی مدل کنند. بنابراین انتخاب مدلی که بتواند با استفاده از عوامل تأثیرگذار، جریان رودخانه را به‌طور قابل‌قبولی تخمین زند امری ضروری به‌نظر می‌رسد. امروزه سامانه‌های هوشمند به‌طور گسترده برای پیش‌بینی پدیده‌های غیرخطی مورد استفاده قرار می‌گیرد که مدل ماشین بردار پشتیبان، برنامه‌ریزی بیان ژن و شبکه بیزین از جمله این مدل‌ها هستند. در سال‌های اخیر استفاده از مدل‌های هوشمند در مطالعات پیش‌بینی جریان رودخانه‌ها مورد توجه پژوهشگران قرار گرفته است که از جمله می‌توان به موارد ذیل اشاره نمود: فربودنام و همکاران (۲۰۰۸) در پژوهشی به‌منظور پیش‌بینی جریان روزانه رودخانه ليقوان از مدل برنامه‌ریزی بیان ژن استفاده نمودند

نتایج این پژوهش نشان داد مدل برنامه‌ریزی بیان ژن از دقت بسیار بالایی نسبت به روش شبکه‌های عصبی مصنوعی و مدل‌های سری زمانی برخوردار است (۸). زمانی و همکاران (۲۰۱۴) با استفاده از مدل‌های برنامه‌ریزی بیان ژن و شبکه عصبی مصنوعی به تخمین دبی روزانه رودخانه کارون پرداختند نتایج نشان داد مدل برنامه‌ریزی بیان ژن با توجه به معیارهای ارزیابی، عملکرد مناسب‌تری نسبت به روش شبکه عصبی مصنوعی دارد (۳۳). مشاری و دانشفرز (۲۰۱۴) در پژوهشی عملکرد شبکه بیزین نسبت به سایر مدل‌های هوشمند را جهت پیش‌بینی جریان روزانه رودخانه قوره‌چای موردبررسی قرار دادند و نشان دادند مدل شبکه عصبی مصنوعی دارای بیش‌ترین دقت و مدل‌های برنامه‌ریزی ژنتیک و شبکه بیزین به‌ترتیب در اولویت‌های بعدی قرار دارند (۲۳). احمدی و همکاران (۲۰۱۵) جهت پیش‌بینی جریان روزانه رودخانه باراندوزچای، عملکرد مدل‌های شبکه بیزین و ماشین بردار پشتیبان را مورد تجزیه و تحلیل قرار دادند و به این نتیجه رسیدند مدل ماشین بردار پشتیبان از عملکرد بهتری نسبت به شبکه بیزین برخوردار است (۳). احمدی و همکاران (۲۰۱۵) عملکرد مدل‌های غیرخطی سری زمانی و برنامه‌ریزی ژنتیک در پیش‌بینی جریان روزانه رودخانه باراندوزچای مورد ارزیابی قرار دادند؛ نتایج مطالعه آن‌ها نشان داد که مدل برنامه‌ریزی ژنتیک با درصد خطای کم‌تری

مفهومی Mike را در شبیه‌سازی جریان روزانه حوضه اسکندری به‌کار گرفتند. نتایج نشان داد هر سه مدل دارای عملکرد مناسب در منطقه مورد مطالعه می‌باشند در نهایت مدل ماشین بردار پشتیبان با کرنل تابع پایه شعاعی کم‌ترین خطا را در پیش‌بینی جریان روزانه از خود نشان داد (۱۴). هانگ و همکاران (۲۰۱۴) جریان ماهانه رودخانه ایستگاه هواکسی چین را با استفاده از ماشین بردار پشتیبان پیش‌بینی نمودند و نتایج نشان داد مدل ماشین بردار پشتیبان دقت بالایی در پیش‌بینی جریان ماهانه رودخانه دارد (۱۳). صدیقی و همکاران (۲۰۱۶) جهت پیش‌بینی فرآیند بارش رواناب از مدل‌های شبکه عصبی مصنوعی و ماشین بردار پشتیبان در حوضه آبریز رودک واقع در شمال شرقی تهران با استفاده از اطلاعات تعداد ۹۲ سنجنده مودیس طی سال آماری ۲۰۰۳-۲۰۰۵ نشان دادند مدل ماشین بردار پشتیبان توانایی قابل‌قبولی در تخمین رواناب دارد (۲۸). در پژوهشی دیگر قربانی و همکاران (۲۰۱۶) جهت پیش‌بینی جریان روزانه رودخانه سایپرس تگزاس از مدل‌های ماشین بردار پشتیبان و شبکه عصبی مصنوعی استفاده نمودند و همچنین معیارهای ضریب همبستگی و ریشه میانگین مربعات خطا را جهت ارزیابی مدل‌ها به‌کار بردند و دریافتند مدل ماشین بردار پشتیبان عملکردی مناسبی جهت پیش‌بینی جریان رودخانه داشته و نسبت به مدل شبکه عصبی مصنوعی از دقت مطلوبی برخوردار است (۱۱). نتایج حاصل از پژوهش ادیب و همکاران (۲۰۱۷) که برای پیش‌بینی دبی جریان روزانه حوضه کلسیان واقع در استان مازندران از مدل برنامه‌ریزی بیان ژن در طی دوره آماری ۱۹۷۰-۲۰۰۹ استفاده نمودند که نشان داد با توجه به معیارهای ارزیابی ضریب همبستگی، ریشه میانگین مربعات خطا، میانگین قدرمطلق خطا و میانگین قدرمطلق خطای نسبی مدل برنامه‌ریزی بیان ژن از عملکرد بالایی

جریان رودخانه را پیش‌بینی می‌کند (۲). عیسی‌زاده و همکاران (۲۰۱۶) به بررسی عملکرد توابع کرنل ماشین بردار پشتیبان جهت پیش‌بینی جریان رودخانه حوضه خرخره جای پرداختند و به این نتیجه رسیدند که مدل SVM<sup>۱</sup> از کارایی بهتری نسبت به مدل‌های سری زمانی در پیش‌بینی جریان ماهانه برخوردار است (۷). باتسیس و همکاران (۲۰۱۲) مدل شبکه بیزین را جهت تأثیر تبخیر و تعرق بر بارش- رواناب در حوضه آبخیز شمال کالیفرنیا ارزیابی کردند نتایج پژوهش آن‌ها نشان داد این مدل تقریبی مناسب برای مطالعه وابستگی بارش- رواناب و نیز تخمینی مؤثر برای نتایج میزان جریان بر اساس ورودی‌های مختلف می‌باشد (۴). صادقی‌حصار و همکاران (۲۰۱۲) قابلیت بالای شبکه بیزین را در تخمین رواناب بر اساس آمار و اطلاعات بارندگی ماهانه با الگوریتم یادگیری جستجوی ممنوع در نرم‌افزار Netica مورد بررسی قرار داده و نشان دادند شبکه بیزین روش مفیدی جهت تخمین رواناب است (۲۷). محمدپور و همکاران (۲۰۱۲) مدل ماشین بردار پشتیبان را به‌منظور پیش‌بینی جریان روزانه حوضه آبریز قره‌سو به‌کار بردند و کارایی مدل را با استفاده از معیار ریشه میانگین مربعات خطا مورد ارزیابی قرار دادند که نتایج آن‌ها نشان داد مدل ماشین بردار پشتیبان عملکرد قابل‌قبولی در تخمین جریان رودخانه دارد (۲۲). روشنگر و همکاران (۲۰۱۳) برای پیش‌بینی جریان روزانه رودخانه و نیار از مدل برنامه‌ریزی بیان ژن و شبکه عصبی مصنوعی استفاده کردند نتایج نشان داد مدل برنامه‌ریزی بیان ژن توانایی بالایی در تخمین جریان رودخانه از خود نشان داده است (۲۶). کاکائی‌لفادانی و همکاران (۲۰۱۳) سه مدل ماشین بردار پشتیبان و نروفازی<sup>۲</sup> و مدل هیدرولوژیکی

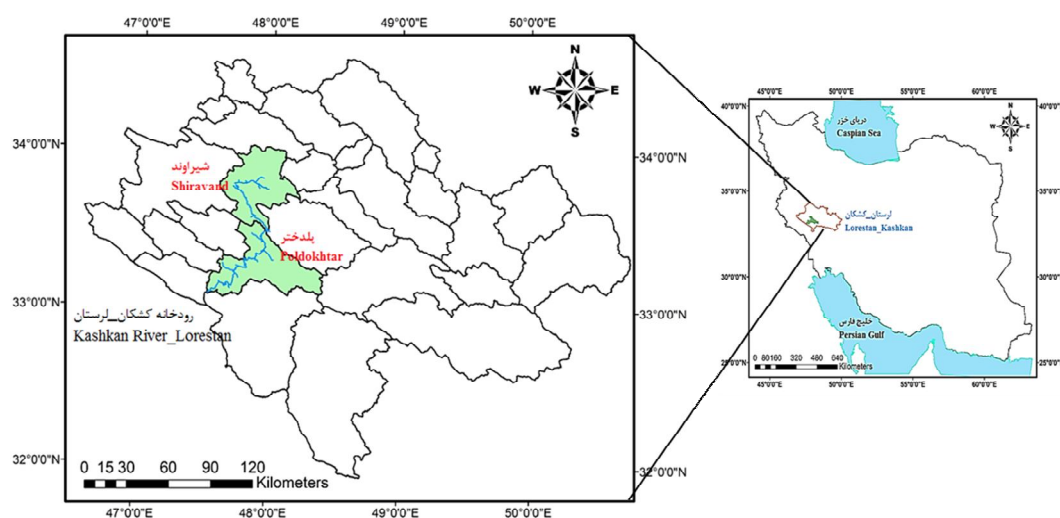
1- Support Vector Machine

2- ANFIS

### مواد و روش‌ها

**منطقه مورد مطالعه:** رودخانه کشکان سیل‌خیزترین رودخانه استان لرستان است. حوزه آبریز کشکان با مساحت ۹۷/۶۶ کیلومترمربع در ناحیه جنوب‌غربی ایران واقع شده است. این حوزه بخش مهمی از سرشاخه‌های پرآب رودخانه کرخه را تشکیل می‌دهد و حدود یک‌سوم خاک لرستان را دربرمی‌گیرد. حوزه آبخیز رودخانه کشکان در تقسیم‌بندی هیدرولوژی ایران جزئی از حوزه آبریز خلیج فارس به حساب می‌آید. این رودخانه در بین طول‌های جغرافیایی  $34^{\circ} 31' 34''$  تا  $33^{\circ} 05' 45''$  درجه تا  $48^{\circ} 12' 6''$  شرقی و عرض‌های جغرافیایی  $33^{\circ} 44' 41''$  تا  $33^{\circ} 05' 45''$  درجه شمالی در استان لرستان واقع شده است. موقعیت منطقه مورد مطالعه در شکل ۱ نشان داده شده است.

جهت پیش‌بینی دبی جریان در مقیاس زمانی روزانه برخوردار است (۱). در مجموع با توجه به پژوهش‌های انجام‌شده و ذکر این نکته که رودخانه کشکان یکی از مهم‌ترین رودخانه‌های استان لرستان و مهم‌ترین منبع تأمین‌کننده آب بخش‌های مختلف و نواحی مجاور خود از لحاظ کشاورزی و شرب می‌باشد و همچنین کاهش جریان این رودخانه مشکلات زیادی در حوزه آبریز ایجاد کرده است؛ بنابراین، اهمیت شبیه‌سازی جریان رودخانه کشکان و اقدامات مدیریتی جهت مدیریت آب آن بیش از پیش ضروری است. بنابراین هدف این پژوهش پیش‌بینی جریان روزانه رودخانه کشکان با استفاده از مدل ماشین بردار پشتیبان و مقایسه نتایج آن با مدل‌های برنامه‌ریزی بیان ژن و شبکه بیزین می‌باشد.



شکل ۱- منطقه مورد مطالعه.

Figure 1. The Study area.

مرتبط با متغیر وابسته  $Y$  که خود تابعی از چند متغیر مستقل  $X$  است، برآورد می‌شود. مشابه سایر مسائل رگرسیونی فرض می‌شود رابطه میان متغیرهای مستقل و وابسته با تابع جبری مانند  $f(x)$  به‌علاوه مقداری اغتشاش (خطای مجاز (۴)) مشخص شود (۳۰).

**ماشین بردار پشتیبان:** ماشین بردار پشتیبان یک سیستم یادگیری کارآمد بر مبنای تئوری بهینه‌سازی مقید است که از اصل استقرای کمینه‌سازی خطای ساختاری استفاده کرده و منجر به یک جواب بهینه کلی می‌گردد (۳۱). در مدل رگرسیون SVM تابعی

$$K(X_j, X) = \phi(X_j)^T \sqrt{b^2 - 4ac} \quad (6)$$

می‌توان از توابع مختلف کرنل برای ساخت انواع مختلف SVM-ε استفاده کرد. انواع توابع کرنل قابل استفاده در مدل SVM رگرسیونی عبارتند از: کرنل چندجمله‌ای<sup>۱</sup> با سه مشخصه هدف و کرنل توابع پایه شعاعی<sup>۲</sup> (RBF) با یک مشخصه هدف، طبق به ترتیب طبق روابط زیر محاسبه می‌گردند.

$$k(x_i, x_j) = (x_i \cdot x_j)^d \quad (7)$$

$$K(x, x_i) = \exp\left(-\frac{\|x - x_i\|^2}{2\sigma^2}\right) \quad (8)$$

در مطالعات بارش-رواناب عمدتاً از مدل RBF استفاده می‌شود (۱۸ و ۱۹). برای انجام فرآیند محاسبات کدی در محیط متلب نوشته شده، که به راحتی در اختیار سایر پژوهشگران قرار می‌گیرد. **برنامه‌ریزی بیان ژن:** روش برنامه‌ریزی بیان ژن، در سال ۱۹۹۹ توسط فریرا ارائه شد (۹). این روش ترکیبی از روش‌های برنامه‌ریزی ژنتیک<sup>۳</sup> (GP) و الگوریتم ژنتیک بوده (GA<sup>۴</sup>) که در آن، کروموزوم‌های خطی و ساده با طول ثابت، مشابه با آنچه که در الگوریتم ژنتیک استفاده می‌شود. در این روش پدیده‌های مختلف با استفاده از مجموعه‌ای از توابع و مجموعه‌ای از ترمینال‌ها، مدل‌سازی می‌شوند. مجموعه توابع، معمولاً شامل توابع اصلی حسابی {/, ×, +, } توابع مثلثاتی یا هر نوع تابع ریاضی دیگر {√, x2, exp, log, sin, cos, ...} و یا توابع تعریف شده که می‌توانند برای تفسیر مدل مناسب باشند. مجموعه ترمینال‌ها، از مقادیر ثابت و متغیرهای

$$f(x) = W^T \cdot \phi(x) + b \quad (1)$$

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

چنانچه W بردار ضرایب و b ثابت مشخصه‌های تابع رگرسیونی و ϕ نیز تابع کرنل باشد، آنگاه هدف پیدا کردن فرم تابعی برای f(x) است. این مهم با آموزش مدل SVM توسط مجموعه‌ای از نمونه‌ها (مجموعه آموزش) محقق می‌شود (۲۱). برای محاسبه w و b لازم است تابع خطا (رابطه ۳) در مدل SVM-ε با در نظر گرفتن شرایط مندرج (قیود) در رابطه ۴ بهینه شود.

$$W^T \cdot \phi(X_i) + b - y_i \leq \varepsilon + \varepsilon_i^*, \quad \frac{1}{2} W^T \cdot W + C \sum_{i=1}^N \varepsilon_i + C \sum_{i=1}^N \varepsilon_i^* \quad (3)$$

$$y_i - W^T \cdot \phi(X_i) - b \leq \varepsilon - \varepsilon_i, \quad \varepsilon_i, \varepsilon_i^* \geq 0, \quad (4)$$

$$i = 1, 2, \dots, N$$

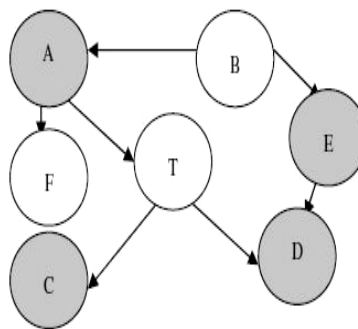
که در آن‌ها، C عددی صحیح و مثبت است که عامل تعیین جریمه در هنگام رخ دادن خطای آموزش مدل است. ϕ تابع کرنل، N تعداد نمونه‌ها و دو مشخصه ε<sub>i</sub>, ε<sub>i</sub><sup>\*</sup> متغیرهای کمبود هستند. در نهایت تابع SVM رگرسیونی را می‌توان به فرم زیر بازنویسی کرد:

$$f(x) = \sum_{i=1}^N \bar{\alpha}_i \phi(x_i)^T \cdot \phi(x) + b \quad (5)$$

که در آن، α<sub>i</sub> میانگین ضرایب لاگرانژ می‌باشد. محاسبه ϕ(x) در فضای مشخصه آن ممکن است بسیار پیچیده باشد (۳۲). برای حل این مشکل روند معمول در مدل SVM رگرسیون انتخاب یک تابع کرنل به صورت رابطه زیر است.

- 1- Polynomial
- 2- Radial Basis Functions (RBF)
- 3- Genetic Programming
- 4- Genetic Algorithm

نمایش می‌دهد (۱۵). در این گراف گره‌ها نشانگر متغیرهای تصادفی گسسته یا پیوسته هستند و کمان‌های جهت‌داری که هر جفت از گره‌ها را به یکدیگر متصل می‌کنند وابستگی بین متغیرها را نمایش می‌دهد. در واقع این شبکه گراف جهت‌داری است که هیچ دوری در آن وجود ندارد (۱۲). نمونه‌ای از گراف ساده در شکل ۲ نشان داده شده است.



شکل ۲- گراف ساده شبکه بیزین.

Figure 2. A graphof simple Bayesian network.

مستقل مسأله تشکیل شده‌اند. برای به‌کارگیری روش برنامه‌ریزی بیان ژن از نرم‌افزار GenXproTools 4.0 استفاده گردید. جهت کسب اطلاعات بیشتر می‌توان به پژوهش قربانی و همکاران (۱۰) مراجعه نمود. **شبکه‌های بیزین:** شبکه‌های بیزین<sup>۱</sup> مدل‌های گرافیکی هستند که برای استدلال در مواردی که پیچیدگی و عدم قطعیت وجود دارد به‌کار می‌روند و یا به‌عبارتی گرافی است که متغیرهای تصادفی و وابستگی آن‌ها را

سپس توزیع‌های احتمال بعدی برای هر گره محاسبه می‌شود که این توزیع احتمال باور نامیده می‌شود (۲۵). **معیارهای ارزیابی:** در این پژوهش به‌منظور ارزیابی دقت و کارایی مدل‌ها، از نمایه‌های ضریب همبستگی (R)، ریشه میانگین مربعات خطا (RMSE) و میانگین قدرمطلق خطا (MAE) طبق روابط زیر استفاده گردید. بهترین مقدار برای این سه معیار به‌ترتیب یک، صفر و صفر است.

$$R = \frac{\sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^N (x_i - \bar{x})^2 \sum_{i=1}^N (y_i - \bar{y})^2}} \quad -1 \leq R \leq 1 \quad (10)$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - y_i)^2} \quad (11)$$

در مورد ساختار شبکه بیزین اگر مجموعه‌ای شامل  $n$  متغیر به‌صورت  $\{X_1, X_2, X_3, \dots, X_n\}$  داشته باشیم که هر  $X_i$  یک متغیر تصادفی باشد و گره‌های والد این متغیر با  $parent(X_i)$  نشان داده شوند، احتمال هر رویداد مانند  $X = (X_1, X_2, X_3, \dots, X_n)$  به‌صورت زیر محاسبه می‌گردد.

$$P(X_1, \dots, X_n) = \prod_{i=1}^n P(X_i | parents(X_i)) \quad (9)$$

بعد از ایجاد شبکه، مقدار روابط بین گره‌های متصل به‌هم به‌دست می‌آید مقدار احتمال شرطی مربوط به هر متغیر در جدولی به‌نام جدول احتمال شرطی قرار می‌گیرد. هر گره یک جدول احتمال شرطی دارد که برای هر ترکیب از مقادیر والد هایش، احتمال شرطی هر کدام از این مقادیر را نگه می‌دارد و

مشاهداتی را به عنوان داده‌های آموزشی انتخاب کرد (۶). با توجه به این که در پژوهش حاضر تأثیر توالی دبی روزهای قبل در پیش‌بینی دبی روز مدنظر بوده، صرفاً از داده‌های نرمال شده دبی جریان با توالی برگشتی تا ۵ روز به عنوان داده‌های آموزشی، به صورت ترکیب‌های مختلف مطابق جدول استفاده شده است که از این پس از آن‌ها به عنوان الگو نام برده خواهد شد و همچنین با توجه به این که مقدار همبستگی متقابل بین داده‌های ورودی و خروجی که مطابق جدول ۲ بیش از ۰/۴۰۰ حاصل شد، ترکیب‌های مختلفی از پارامترهای ورودی به منظور دستیابی به مدل بهینه جهت تخمین جریان روزانه رودخانه کشکان استفاده گردید (۶). بدین منظور از داده‌های ایستگاه هیدرومتری پلدختر واقع در استان لرستان که دارای ۳۶۵۰ رکورد ثبت شده طی بازه زمانی (۱۳۹۳-۱۳۸۳)، در مقیاس زمانی روزانه می‌باشد، استفاده شد. در نهایت تعداد ۲۹۲۰ رکورد برای آموزش و تعداد ۷۳۰ رکورد باقی مانده برای صحت‌سنجی مدل‌های مورد بررسی انتخاب شد. لازم به ذکر است جهت مدل‌سازی ۸۰ درصد داده‌ها برای آموزش و ۲۰ درصد باقی مانده جهت آزمون، به صورت تصادفی، که گستره وسیعی از انواع داده‌ها را پوشش دهد، انتخاب شد (۱۷ و ۲۴). همچنین در جدول ۳ خصوصیات آماری پارامتر دبی جریان نشان داده شده است.

$$MAE = \frac{1}{n} \sum |x_i - y_i| \quad (12)$$

که در آن‌ها،  $x_i$  و  $y_i$  به ترتیب مقادیر مشاهداتی و محاسباتی در گام زمانی  $i$  ام،  $N$  تعداد گام‌های زمانی،  $\bar{x}$  و  $\bar{y}$  نیز به ترتیب میانگین مقادیر مشاهداتی و محاسباتی می‌باشد. علاوه بر معیارهای فوق از نمودارهای پراکنش و سری زمانی مقادیر مشاهداتی-محاسباتی نسبت به زمان نیز جهت تحلیل بیشتر نتایج استفاده شده است.

### نتایج و بحث

یکی از مهم‌ترین مراحل در مدل‌سازی، انتخاب ترکیب مناسبی از متغیرهای ورودی است. از این رو، ابتدا همبستگی متقابل<sup>۱</sup> بین متغیرهای ورودی و خروجی محاسبه گردید و پارامترهای ورودی به منظور دستیابی به مدل بهینه جهت پیش‌بینی جریان رودخانه کشکان انتخاب و در جدول ۱ ارائه شد. در این جدول، ستون‌های  $Q(t-1)$ ،  $Q(t-2)$ ،  $Q(t-3)$ ،  $Q(t-4)$  و  $Q(t-5)$  (به ترتیب جریان رودخانه در زمان‌های  $t-1$ ،  $t-2$ ،  $t-3$ ،  $t-4$  و  $t-5$  به عنوان ورودی و  $Q(t)$  جریان رودخانه در زمان  $t$  به عنوان خروجی مدل در نظر گرفته شد. به منظور آموزش ماهیت سازوکار حاکم بر پدیده نه تنها سبب پیچیدگی الگو و افزایش حافظه درگیر خواهد شد، بلکه سبب کاهش دقت مدل نیز می‌شود. بنابراین در الگوسازی جریان رودخانه نیز باید سعی نمود مؤثرترین داده‌های

جدول ۱- همبستگی متقابل بین پارامترهای ورودی و خروجی.

Table 1. The cross-correlation between input and output parameters.

Q(t-5)	Q(t-4)	Q(t-3)	Q(t-2)	Q(t-1)	Q(t)
0.41	0.45	0.51	0.58	0.7	

1- Cross-correlation

جدول ۲- ترکیب‌های منتخب پارامترهای ورودی مدل‌های ماشین بردار پشتیبان، برنامه‌ریزی بیان زن و شبکه بیزین.

**Table 2. Selected compositions input parameter models, support vector machine and Bayesian network GEP.**

شماره Number	ساختار ورودی Input structure	خروجی Output
1	Q(t-1)	Q(t)
2	Q(t-1), Q(t-2)	Q(t)
3	Q(t-1), Q(t-2), Q(t-3)	Q(t)
4	Q(t-1), Q(t-2), Q(t-3), Q(t-4)	Q(t)
5	Q(t-1), Q(t-2), Q(t-3), Q(t-4), Q(t-5)	Q(t)

جدول ۳- خصوصیات آماری پارامتر مورد استفاده در طی دوره آماری (۱۳۸۳-۱۳۹۳).

**Table 3. The statistical properties of the parameters used in the statistical period.**

چولگی Skewness	انحراف معیار Standard deviation	ماکزیمم Maximum	میانگین Mean	مینیمم Minimum	پارامتر Parameter
1.521	20.129	119	23.435	0.011	آموزش Training
1.859	19.573	111	18.572	0.312	صحت سنجی Testing

C توسط الگوریتم بهینه‌سازی جستجوی شبکه و متغیر  $\gamma$  نیز به صورت آزمون و خطا محاسبه شد. 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 از برنامه اصلاح شده الگوریتم جستجوی شبکه در پژوهش چن و یو (۵) که به نام الگوریتم جستجوی شبکه دو مرحله‌ای معروف است به همراه اعتبارسنجی متقاطع استفاده شد. برای این منظور ابتدا با انتخاب شبکه‌های با ابعاد بزرگ، محدوده مشخصه‌های  $\epsilon$  و  $C$  به‌ازای مقدار ثابت مشخصه  $\gamma$  تعیین شد. سپس با مشخص شدن محدوده مذکور و تقسیم آن به شبکه‌های با ابعاد ریزتر مقادیر دقیق دو مشخصه  $\epsilon$  و  $C$  مشخص شدند. روند مذکور برای دیگر مقادیر  $\gamma$  نیز تکرار شد و بدین طریق مدل‌های متفاوتی با تغییر

نتایج ماشین بردار پشتیبان: به منظور پیش‌بینی جریان روزانه رودخانه کشکان از مدل SVM استفاده شد. در این پژوهش ابتدا به منظور کاهش دامنه تغییرات داده‌های جریان روزانه رودخانه کشکان به همسان‌سازی اطلاعات ورودی و خروجی، بر روی داده‌ها استانداردسازی صورت گرفت. در مرحله بعد مقادیر بهینه مشخصه‌های مدل SVM شامل  $\epsilon$  و  $C$  تعیین می‌گردد. همچنین در این پژوهش تابع کرنل مورد استفاده، تابع RBF انتخاب شد چرا که از دقت بهتری در برآورد جریان روزانه رودخانه‌ها برخوردار است (۱۸ و ۱۹). در این تابع نیز مشخصه  $\gamma$  باید تعیین گردد. بنابراین در حالت کلی برای پیش‌بینی جریان روزانه رودخانه کشکان به وسیله مدل SVM رگرسیون، لازم است که مقادیر بهینه سه مشخصه مذکور محاسبه گردد که بدین منظور دو مشخصه  $\epsilon$  و



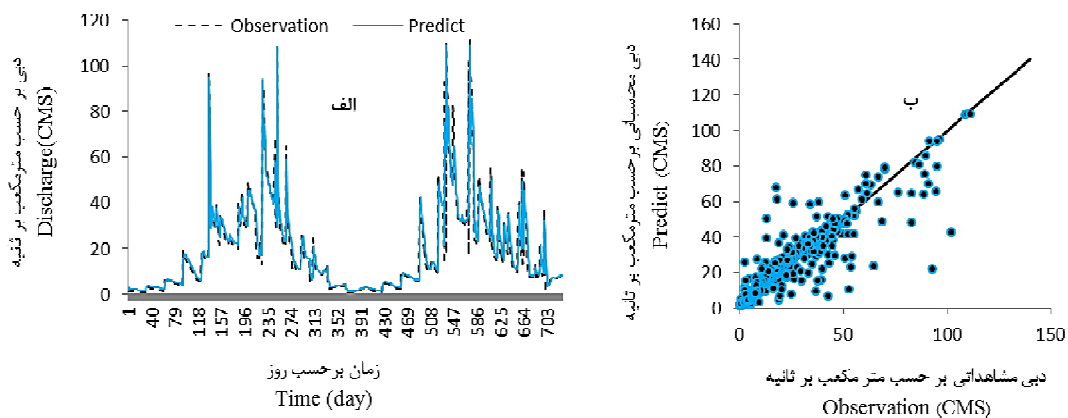
مشاهداتی به غیر از چند نقطه روی خط نیمساز قرار گرفته و این امر دلالت بر برابری مقادیر مشاهداتی و محاسباتی براساس خط  $(y=x)$  است. در شکل (۳-الف) نیز تغییرات مقادیر محاسباتی و مشاهداتی نسبت به زمان نشان داده شد، به طوری که این مدل در تخمین اکثر مقادیر دقت قابل قبولی داشته به گونه ای که این مقادیر را نزدیک به مقدار واقعی شان تخمین زده است. همچنین همان طور که در شکل (۳-ب) قابل مشاهده است مدل ماشین بردار پشتیبان در تخمین برخی مقادیر سیلابی یا پیک در روزهای پربابی عملکرد قابل قبولی ارائه نداده است به گونه ای که این مقادیر مشاهداتی خارج از خط نیمساز واقع شده اند. که این نتایج با پژوهش های قربانی و همکاران (۱۱) و احمدی و همکاران (۳) همخوانی دارد. در تبیین این نتایج می توان بیان نمود با توجه به این که ماشین بردار پشتیبان بر مبنای استفاده از اصل استقرای کمینه سازی خطای ساختاری است بنابراین در شبیه سازی، با به کارگیری روش یادگیری با نظارت در توابع پایه شعاعی باعث می شود تخمین پارامتر از سرعت بالا و خطای کمتری نسبت به سایر کرنل ها داشته باشد و این یکی از ویژگی های ممتاز توابع پایه شعاعی محسوب می شود.

در مقدار  $\gamma$  حاصل شدند. حال می توان از بین مدل های توسعه داده شده مدل با کمترین خطا را تعیین کرده و مشخصه های آن را به عنوان مقادیر بهینه  $\epsilon$ ،  $C$  و  $\gamma$  انتخاب نمود. نتایج حاصل بررسی مدل SVM در جدول ۴ آورده شد. با توجه به جدول مذکور مدل ترکیبی ۵ دارای بیشترین میزان ضریب همبستگی  $R=0.910$  و کمترین میزان ریشه مربعات خطا  $RMSE=0.002 \text{ m}^3/\text{s}$  و کمترین میانگین قدرمطلق خطا  $MAE=0.001 \text{ m}^3/\text{s}$  در مرحله صحت سنجی جواب بهینه تری نسبت به سایر ساختارها از خود نشان داده است. همچنین که در جدول ۴ مشاهده می گردد ماشین بردار پشتیبان عملکرد بالایی در پیش بینی جریان روزانه رودخانه کشکان حتی در صورت استفاده از تنها یک پارامتر ورودی نیز از خود نشان داده است که این امر سبب می شود در مناطقی که نواقص آماری وجود دارد این شبکه با در اختیار داشتن پارامترهای ورودی حداقل مانند دبی جریان در روز گذشته عملکرد قابل قبولی در پیش بینی جریان روزانه رودخانه داشته باشد. در شکل ۳ نمودار بهترین مدل حاصل شده برای داده های بخش صحت سنجی نشان داده شد. همان طور که از شکل (۳-ب) مشخص است بیش تر مقادیر تخمینی و

جدول ۴- تحلیل نتایج مدل ماشین بردار پشتیبان برای ساختارهای مختلف ورودی جریان رودخانه.

Table 4. The analysis of SVM model results for river flow inputs.

متغیرهای مدل SVM			صحت سنجی			آموزش			ساختار Structure
SVM Parameters			Testing			Training			
$\gamma$	$\epsilon$	C	MAE ( $\text{m}^3/\text{s}$ )	RMSE ( $\text{m}^3/\text{s}$ )	R	MAE ( $\text{m}^3/\text{s}$ )	RMSE ( $\text{m}^3/\text{s}$ )	R	
0.87	0.3	10	0.004	0.011	0.858	4.195	10.659	0.848	1
0.436	0.2	10	0.003	0.009	0.873	4.154	10.461	0.854	2
0.28	0.2	10	0.002	0.007	0.876	4.023	10.157	0.863	3
0.25	0.1	10	0.002	0.005	0.868	4.161	10.306	0.859	4
0.22	0.1	10	0.001	0.002	0.910	4.013	10.104	0.871	5



شکل ۳- نمودار مقادیر مشاهداتی و محاسباتی حاصل از مدل ماشین بردار پشتیبان در مرحله صحت‌سنجی.

Figure 3. Observed and predicted values of the SVM model in validation.

همبستگی  $R=0/870$ ، ریشه میانگین مربعات خطا  $RMSE=0/003 \text{ m}^3/\text{s}$  و میانگین قدرمطلق خطا  $MAE=0/001 \text{ m}^3/\text{s}$  نسبت به سایر ساختارها از دقت بیشتری برخوردار است. نمودارهای پراکنش برنامه‌ریزی بیان ژن مربوط به مرحله صحت‌سنجی در شکل (۴-ب) نشان‌دهنده خط برازش مقادیر محاسباتی با بهترین خط برازش  $y=x$  است. همان‌طور که از این شکل مشخص است مقادیر تخمینی و مشاهداتی به‌غیر از چند نقطه که روی خط نیمساز قرار نگرفته در سایر نقاط بر روی خط برازش قرار دارد که این امر دلالت بر برابری مقادیر مشاهداتی و محاسباتی بر اساس خط  $(y=x)$  است. همچنین لازم به ذکر است با توجه به شکل (۴-الف) این مدل در تخمین برخی مقادیر جریان رودخانه خوب عمل کرده و در تخمین مقادیر بیشینه عملکرد قابل‌قبولی از خود نشان نداده است که این نتایج با پژوهش‌های قربانی و همکاران (۱۰)، زمانی و همکاران (۳۳) و فریودنام و همکاران (۸) مطابقت دارد. در تبیین این نتایج می‌توان بیان نمود معادله حاصل از برنامه‌ریزی بیان ژن، از ترکیب تصادفی مجموع پایانه‌ها و توابع به‌دست می‌آید. بنابراین اگر رابطه بین ورودی‌ها و خروجی‌ها خطی باشد، ولی عملگرهای  $\sin$ ،  $\cos$  و... در

نتایج برنامه‌ریزی بیان ژن: استفاده از برنامه‌ریزی بیان ژن به‌دلیل توان انتخاب متغیرهای مؤثر در مدل و حذف متغیرهایی با تأثیر کم‌تر و همچنین توانایی ارائه رابطه صریح جهت تخمین جریان رودخانه کشکان مدنظر قرار گرفت. در این پژوهش از ساختارهای ورودی برای تعیین متغیرهای معنی‌دار استفاده و جهت بررسی نتایج از مجموعه چهار عملگر اصلی استفاده شد. همچنین در این پژوهش جهت ساخت درخت تجزیه و تولید یک جمعیت اولیه از راه‌حلهایی استفاده شد که این راه‌حل‌ها از ترکیب تصادفی مجموعه توابع (عملگرهای ریاضی مورد استفاده در فرمول‌ها) و پایانه‌ها (متغیرهای مسأله و اعداد ثابت) ایجاد می‌شوند نحوه و دلیل انتخاب این نوع عملگرها با توجه به مطالعات خطی و همکاران (۱۶) و قربانی و همکاران (۱۰) صورت گرفت. توابع انتخاب‌شده در این پژوهش شامل مجموعه چهار عملگر اصلی ریاضی است که به‌صورت زیر می‌باشد.

$$F_1: \{+, -, *, /\} \quad (13)$$

نتایج حاصل از مدل برنامه‌ریزی بیان ژن در جدول ۵ نشان داده‌شده است همان‌طور که در این جدول مشخص است ساختار ۵ با بیش‌ترین ضریب

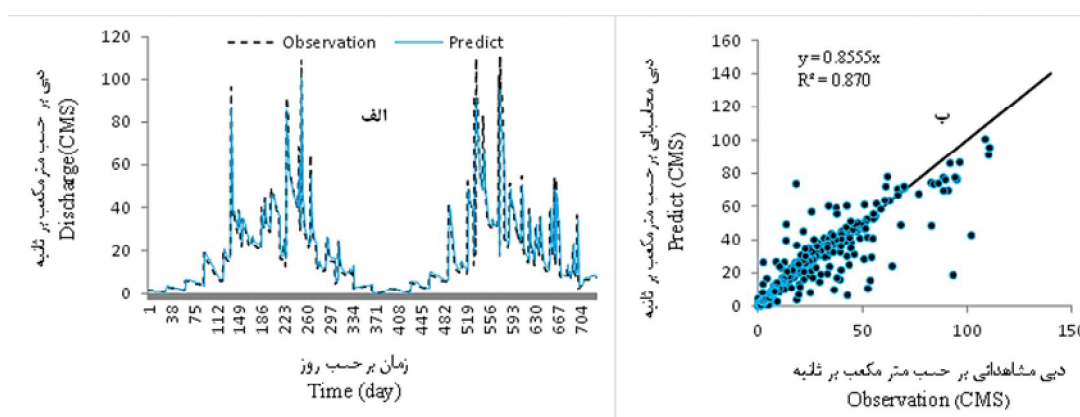
sin، cos و... کاربرد ندارد و نیز با توجه به دقت و سادگی، مدل حاصل از چهار عمل اصلی ریاضی {+، -، \* و /} جهت تخمین جریان روزانه رودخانه کشکان پیشنهاد شد (۱۶).

مجموعه توابع انتخاب شود برنامه‌ریزی بیان ژن در استخراج رابطه از آن عملگرهای انتخابی استفاده می‌کند که این امر باعث کاهش دقت مدل می‌شود که در این پژوهش جهت افزایش دقت مدل عملگرهای

جدول ۵- تحلیل نتایج مدل برنامه‌ریزی بیان ژن برای ساختارهای مختلف ورودی جریان رودخانه.

Table 5. The analysis of GEP model results for river flow inputs.

صحت سنجی Testing			آموزش Training			ساختار Structure
MAE (m <sup>3</sup> /s)	RMSE (m <sup>3</sup> /s)	R	MAE (m <sup>3</sup> /s)	RMSE (m <sup>3</sup> /s)	R	
0.010	0.026	0.849	4.348	10.985	0.838	1
0.003	0.009	0.856	4.182	10.602	0.850	2
0.002	0.007	0.861	3.795	10.846	0.845	3
0.009	0.023	0.865	4.387	10.795	0.844	4
0.001	0.003	0.870	4.083	10.656	0.853	5



شکل ۴- نمودار مقادیر مشاهداتی و محاسباتی حاصل از مدل برنامه‌ریزی بیان ژن در مرحله صحت‌سنجی.

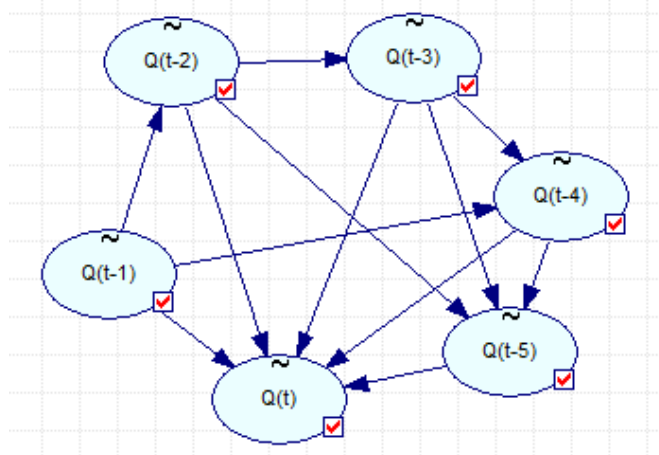
Figure 4. Observed and predicted values of the GEP model in validation.

نیز الگوریتم Clustering برای نمایش میزان تأثیر هر کدام از پارامترها در نظر گرفته شده است. در شکل ۵ مدل ساخته شده شبکه بیزین نشان داده شده است. همان‌طور که در این شکل مشخص است، جریان رودخانه در تأخیرهای یک تا پنج روز تأثیر مستقیم بر روی دبی جریان در زمان موردنظر دارد. همچنین در جدول ۶ نتایج حاصل از ارزیابی مدل شبکه بیزین

نتایج شبکه بیزین: در این پژوهش، از بین الگوریتم‌های K2 و BDeu ساختار شبکه از الگوریتم K2 که مناسب‌ترین پارامتر الگوریتم یادگیری شبکه است، برای ایجاد شبکه بیزین استفاده شد. همچنین از میان روش‌های موجود برای ساخت شبکه، از معمولی‌ترین متد Greedy Thick Thinning جهت الگوریتم یادگیری به‌کار برده شد. برای استنتاج شبکه

دارد که این امر دلالت بر برابری مقادیر مشاهداتی و محاسباتی بر اساس خط  $(y=x)$  است. همچنین لازم به ذکر است با توجه به شکل (۶-الف) این مدل در تخمین مقادیر بیشینه ضعیف عمل کرده است که در این شکل مشهود است. همچنین همان‌طور که در شکل (۶-ب) مشاهده می‌شود در روزهای سیلابی مدل شبکه بیزین عملکرد قابل‌قبولی نداشته است. به‌گونه‌ای که مقادیر مشاهداتی دبی سیلابی را خارج از خط نیمساز قرار گرفته و دلالت بر توانایی پایین شبکه در تخمین نقاط اوج است.

مشاهده می‌شود. همان‌طور که در جدول ۶ مشاهده می‌شود ساختار پنجم توانسته با بیش‌ترین دقت  $R=0.860$  و کم‌ترین ریشه میانگین مربعات خطا  $RMSE=0.068 \text{ m}^3/\text{s}$  و کم‌ترین میانگین قدرمطلق خطا  $MAE=0.003$  در مرحله صحت‌سنجی در اولویت قرار گیرد. در شکل (۶-ب) نمودار پراکنش برنامه‌ریزی بیان ژن مربوط به مرحله صحت‌سنجی نشان‌دهنده خط برازش مقادیر محاسباتی با بهترین خط برازش  $y=x$  است. در این شکل مقادیر تخمینی و مشاهداتی به‌غیر از چند نقطه که روی خط نیمساز قرار نگرفته، در سایر نقاط بر روی خط برازش قرار



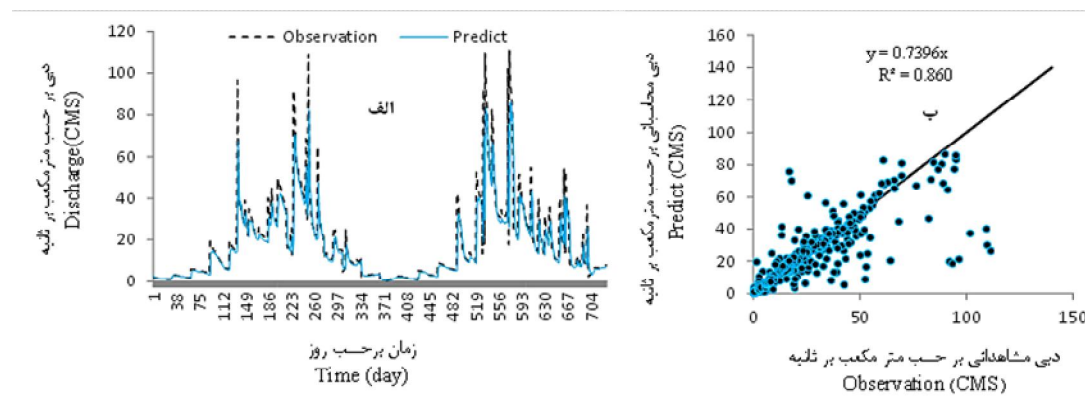
شکل ۵- مدل مفهومی ساخته‌شده برای شبکه بیزین.

Figure 5. Bayesian network conceptual model.

جدول ۶- تحلیل نتایج مدل شبکه بیزین برای ساختارهای مختلف ورودی جریان رودخانه.

Table 6. The analysis of Bayesian Network model results for river flow inputs.

صحت‌سنجی Testing			آموزش Training			ساختار Structure
MAE ( $\text{m}^3/\text{s}$ )	RMSE ( $\text{m}^3/\text{s}$ )	R	MAE ( $\text{m}^3/\text{s}$ )	RMSE ( $\text{m}^3/\text{s}$ )	R	
0.004	0.109	0.849	5.128	11.586	0.838	1
0.003	0.089	0.858	4.685	11.242	0.843	2
0.003	0.078	0.863	4.437	11.067	0.845	3
0.003	0.073	0.863	4.345	11.005	0.846	4
0.003	0.068	0.860	4.262	10.933	0.847	5



شکل ۶- نمودار مقادیر مشاهداتی و محاسباتی حاصل از مدل شبکه بیزین در مرحله صحت‌سنجی.

Figure 6. Observed and predicted values of the BN model in validation.

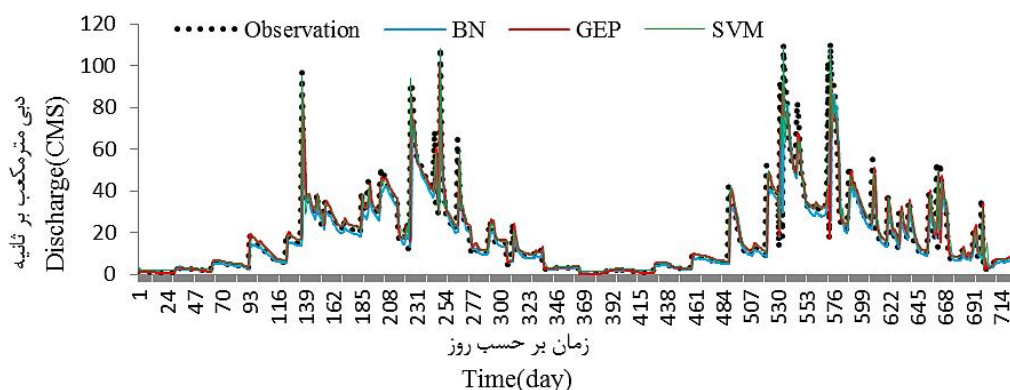
پشتیبان در پیش‌بینی مقادیر حداقل، حداکثر و میانی جریان روزانه رودخانه عملکرد قابل‌قبولی از خود نشان داده است به‌گونه‌ای که مقادیر پیش‌بینی‌شده مدل مقادیر مشاهداتی را به‌خوبی برآورد نموده است. در نهایت اختلاف مقادیر مشاهداتی جریان روزانه رودخانه و محاسباتی بهینه مدل‌ها به‌صورت درصدی از میانگین مقادیر مشاهداتی (مقدار خطا)، محاسبه و نمودار آن نسبت به نمونه داده‌های ثبت‌شده دوره آماری ترسیم گردید (شکل ۸). همان‌طور که در این شکل دیده می‌شود برای هر سه مدل بیش‌تر خطاها در باند  $\pm 1$  قرار گرفته است. که در میان مدل‌های مذکور (ماشین بردار پشتیبان، برنامه‌ریزی بیان ژن و شبکه بیزین)، مدل ماشین بردار پشتیبان کم‌ترین میزان خطا را داشته است.

مقایسه عملکرد مدل‌ها: همان‌طور که در جدول ۷ مشاهده می‌شود با انتخاب جواب بهینه هر کدام از مدل‌ها و مقایسه آن‌ها با یکدیگر مشخص شد که هر سه مدل با دقت خوبی و خطای کم‌تری می‌توانند جریان روزانه رودخانه کشکان را شبیه‌سازی کنند (۲۰ و ۲۹). که از بین مدل‌های به‌کار رفته مدل ماشین بردار پشتیبان، بیش‌ترین دقت  $R=0.910$  و کم‌ترین ریشه میانگین مربعات خطا  $RMSE=0.002 \text{ m}^3/\text{s}$  و کم‌ترین میانگین قدرمطلق خطا  $MAE=0.001 \text{ m}^3/\text{s}$  در مرحله صحت‌سنجی را دارد. در شکل ۷ نتایج هر سه مدل نسبت به مقادیر مشاهداتی در طی زمان نشان داده شد همان‌طور که در شکل ۷ قابل مشاهده است هر سه مدل در تخمین بیش‌تر مقادیر عملکرد قابل‌قبولی دارند، در حالی که مدل ماشین بردار

جدول ۶- تحلیل آماری نتایج مدل‌های ماشین بردار پشتیبان، برنامه‌ریزی بیان ژن و شبکه بیزین برای ساختارهای مختلف ورودی جریان رودخانه.

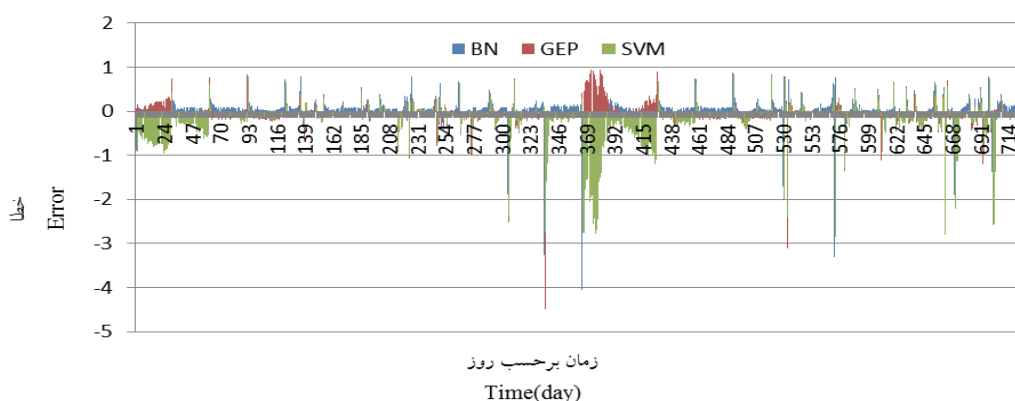
Table 6. The statistical analysis of Support Vector Machine, Gene Expression Programming and Bayesian Network results for river flow inputs.

صحت‌سنجی Testing			آموزش Training			مدل Model
MAE ( $\text{m}^3/\text{s}$ )	RMSE ( $\text{m}^3/\text{s}$ )	R	MAE ( $\text{m}^3/\text{s}$ )	RMSE ( $\text{m}^3/\text{s}$ )	R	
0.001	0.002	0.910	4.013	10.104	0.871	ماشین بردار پشتیبان Support Vector Machine
0.001	0.003	0.870	4.083	10.656	0.853	برنامه‌ریزی بیان ژن Gene Expression Programming
0.003	0.068	0.860	4.262	10.933	0.847	شبکه بیزین Bayesian Network



شکل ۷- نمودار مقادیر مشاهداتی و محاسباتی حاصل از مدل‌های ماشین بردار پشتیبان، برنامه‌ریزی بیان ژن و شبکه یزین در مرحله صحت‌سنجی.

Figure 7. Observed and predicted values of the SVM, GEP and BN model in validation.



شکل ۸- نمودار خطای بهینه هر سه مدل به صورت درصدی از میانگین مقادیر مشاهداتی.

Figure 8. The Graph of optimal error three models in a percentage of mean observed values.

برنامه‌ریزی بیان ژن و ماشین بردار پشتیبان، در ساختاری متشکل از ۱ تا ۵ تأخیر زمانی نتایج بهتری نسبت به سایر ساختارها ارائه می‌دهد. علاوه بر آن مشاهده گردید در مدل ماشین بردار پشتیبان با داشتن حداقل پارامترهای ورودی قادر به پیش‌بینی جریان روزانه با دقت بالایی می‌باشد. این موضوع به‌خصوص در ایستگاه‌هایی که با نواقص آماری مواجه می‌باشند می‌تواند کمک شایانی به برآورد پارامتر موردنظر نماید. همچنین با توجه به معیارهای ارزیابی نتیجه شد که هر سه مدل مورد بررسی، می‌توانند با دقت نسبتاً بالایی میزان جریان روزانه رودخانه را پیش‌بینی نمایند. در این میان، مدل ماشین بردار پشتیبان دقت

### نتیجه‌گیری

در این پژوهش سعی بر آن شد عملکرد مدل‌هایی جهت شبیه‌سازی جریان روزانه رودخانه کشکان واقع در استان لرستان با استفاده از داده‌های روزانه مورد ارزیابی قرار گیرد. مدل‌های به‌کار گرفته‌شده شامل شبکه یزین، برنامه‌ریزی بیان ژن و ماشین بردار پشتیبان است. مقادیر جریان مشاهداتی با جریان پیش‌بینی‌شده در مدل‌های مذکور (شبکه یزین، برنامه‌ریزی بیان ژن و ماشین بردار پشتیبان)، با استفاده از معیارهای ارزیابی مورد مقایسه قرار گرفت. نتایج پژوهش را می‌توان به‌صورت زیر خلاصه نمود: نتایج نشان داد که در هر سه مدل مورد بررسی شامل شبکه یزین،

شعاعی محسوب می‌شود. در مجموع این پژوهش نشان می‌دهد استفاده از مدل ماشین بردار پشتیبان می‌تواند در زمینه پیش‌بینی جریان روزانه رودخانه‌ها مؤثر باشد و همچنین، این مدل می‌تواند به‌نوبه خود برای تسهیل توسعه و پیاده‌سازی استراتژی‌های مدیریت آب‌های سطحی مفید باشد. و گامی در اتخاذ تصمیمات مدیریتی در جهت بهبود کمیت جریان رودخانه‌ها می‌باشد.

بیش‌تر و خطای کم‌تری نسبت به مدل‌های شبکه بیزین و برنامه‌ریزی بیان ژن از خود نشان داده است. همچنین دقت بالای ماشین بردار پشتیبان ناشی از استفاده از اصل استقرای کمینه‌سازی خطای ساختاری است که با به‌کارگیری روش یادگیری با نظارت در توابع پایه شعاعی باعث می‌شود تخمین پارامتر از سرعت بالا و خطای کم‌تری نسبت به سایر کرنل‌ها داشته باشد و این یکی از ویژگی‌های ممتاز توابع پایه

### منابع

1. Adib, A., Mahmoudian Kafshgar Kalaei, M., Mahmoudian Shoushtari, M., and M. Khalili, K. 2017. Using of gene expression programming and climatic data for forecasting flow discharge by considering trend, normality and stationarity analysis. Arabi. J. Geosci. 10: 4. 1-14.
2. Ahmadi, F., Dinpajoh, Y., Fakherifard, A., Khalili, K., and Darbandi, S. 2015. Comparing nonlinear time series models and genetic programming for daily river flow forecasting (Case study: Barandouz-Chai River). Soil and water conservation research. 22: 1. 171-186. (In Persian)
3. Ahmadi, F., Radmanesh, F., and Mirabasi, R. 2015. Comparing the performance of support vector machines and Bayesian networks in predicting daily river flow (Case study: Barandouz-Chai River). Soil and water conservation research. 22: 6. 171-186. (In Persian)
4. Botsis, D., Latinopoulos, P., and Diamantaras, K. 2012. Investigation of The effect of interception and evapotranspiration on the rain fall-run off relationship using Bayesian networks. In: Proceedings of protection and restoration of the environment XI, Thessaloniki.
5. Chen, S.T., and Yu, P.S. 2007. Real-time probabilistic forecasting of flood stages. J. Hydrol. 340: 63-77.
6. Danandeh Mehr, A., and Majdzadeh Tabatabaei, M.R. 2009. I prediction of daily discharge trend of river flow based on genetic programming. J. Water Soil. 24: 2. 325-333. (In Persian)
7. Esazadeh, M., Ahmadzadeh, H., and Ghorbani, M.A. 2016. Assessment of kernel functions performance in river flow estimation using support vector machine. Soil and water conservation research. 23: 3. 171-186. (In Persian)
8. Ferbodnam, N., Ghorbani, M.A., and Alami, M.T. 2008. River flow prediction using genetic programming (Case study: Lighvan River Watershed). J. Soil Water. 19: 1. 107-123. (In Persian)
9. Ferreira, C. 2001. Gene expression programming: a new adaptive algorithm for solving problems. Complex Systems. 13: 2. 87-129.
10. Ghorbani, M.A., Khatibi, R., Asadi, H., and Yousefi, P. 2012. Inter- Comparison of an evolutionary programming model of suspended sediment time-series whit other local model. INTECH. 26: 5. 255-282.
11. Ghorbani, M.A., Khatibi, R., Geol, A., Fazelifard, M.H., and Azani, A. 2016. Modeling river discharge time series using support vector machine and artificial neural networks. Environmental Earth Sciences. 75: 4. 675-685.
12. Heckerman, D. 1997. Bayesian networks for data mining. Data Mining and Knowledge Discovery. 1: 1. 79-119.
13. Huang, S., Chang, J., Huang, Q., and Chen, Y. 2014. Monthly streamflow prediction using modified emd-based support vector machine. J. Hydrol. 511: 4. 764-775.

14. Kakaei Lafadani, E., Moghaddam Nia, A., Ahmadi, A., Jajarmizadeh, M., and Ghafari, M. 2013. Stream flow simulation using svm, anfis and nam models (a case study). *Caspian J. Appl. Sci. Res.* 2: 4. 86-93.
15. Kevin, B., and Nicholson, E. 2010. Bayesian artificial intelligence. Second Edition, United states. 3: 1. 370-450.
16. Khatibi, R., Naghipour, L., Ghorbani, M.A., and Aalami, M.T. 2012. Predictability of relative humidity by two artificial intelligence techniques using noisy data from two Californian gauging stations. *Neural computing and application.* 23: 7. 643-941.
17. Kisi, O., Karahan, M., and Sen, Z. 2006. River suspended sediment modeling using fuzzy logic approach. *Hydrol Process.* 20: 2. 4351-4362.
18. Lin, J.Y., Cheng, C.T., and Chau, K.W. 2006. Using support vector machines for long-term discharge prediction. *Hydrol. Sci. J.* 51: 3. 599-612.
19. Liong, S.Y., and Sivapragasam, C. 2002. Flood stage forecasting with support vector machines. *J. Am. Water Resour.* 38: 4. 173-186.
20. MacKay, D.J.C. 1992. Bayesian interpolation, *Neural Computation.* 4: 1. 415-447.
21. Misra, D., Oommen, T., Agarwal, A., Mishra, S.K., and Thompson, A.M. 2009. Application and analysis of support vector machine based simulation for runoff and sediment yield. *Biosyst. Eng.* 103: 3. 527-535.
22. Mohammadpour, M., Mehrabi, A., and Katouzi, M. 2012. Daily discharge forecasting using support vector machine. *Inter. J. Inf. Elec. Engin.* 2: 5. 769-772.
23. Moshari, K.H., and Daneshfaraz, R. 2014. Comparison of Bayesian networks with other smart models predict river flow in Qvrh tea. Tenth International Congress on Civil Engineering.
24. Nagy, H., Watanabe, K., and Hirano, M. 2002. Prediction of sediment load concentration in rivers using artificial neural network model. *J. Hydraul. Engin.* 128: 3. 558-559.
25. Nguyen, R.T., Prentiss, D., and Shively, J.E. 1998. Rainfall interpolation for Santa Barbara County. UCSB, Department Geography. USA.
26. Roshangar, K., Vojoudi Mehrabani, F., and Alami, M.T. 2013. Forecasting daily stream flows of vaniar river using genetic programming and neural networks approaches. *J. Civil Engin. Urban.* 3: 4. 197-200.
27. Sadeghi Hesar, A., Tabatabaee, H., and Jalali, M. 2012. Monthly rainfall forecasting using bayesian belief networks. *Inter. Res. J. Appl. Bas. Sci.* 3: 11. 2226-2231.
28. Sedighi, F., Vafakhah, M., and Javadi, M. R. 2016. Rainfall-Runoff modeling using support vector machine in snow-affected watershed. *Arab. J. Sci. Engin.* 41: 10. 4065-4076.
29. Tokar, A.S., and Johnson, P.A. 1999. Rainfall-Runoff modeling using artificial neural networks. *J. Hydrol. Engin.* 3: 4. 232-239.
30. Vapnik, V.N. 1995. The nature of statistical learning theory. Springer, New York, Pp: 250-320.
31. Vapnik, V.N. 1998. Statistical learning theory. Wiley, New York, Pp: 250-320.
32. Yoon, H., Jun, S.C., Hyun, Y., Bae, G.O., and Lee, K.K. 2011. A comparative study of artificial neural networks and support vector machines for predicting groundwater levels in a coastal aquifer. *J. Hydrol.* 396: 128-138.
33. Zamani, R., Ahmadi, F., and Radmanesh, F. 2014. Comparison of the gene expression programming, nonlinear time series and artificial neural network in estimating the river daily flow (case study: the Karun river). *J. Soil Water.* 28: 6. 1172-1182. (In Persian)





## **Comparing the performance of support vector machine, gene expression programming and Bayesian networks in predicting river flow (Case study: Kashkan River)**

**\*R. Dehghani<sup>1</sup>, H. Younesi<sup>2</sup> and H. Torabi Podeh<sup>3</sup>**

<sup>1</sup>Ph.D. Student of Water Structure, Faculty of Agriculture, University of Lorestan, Khorramabad, Iran,

<sup>2</sup>Assistant Prof., Dept. of Water Engineering, University of Lorestan, Khorramabad, Iran,

<sup>3</sup>Associate Prof., Dept. of Water Engineering, University of Lorestan, Khorramabad, Iran

Received: 12/23/2016; Accepted: 07/18/2017

### **Abstract**

**Background and Objectives:** River flow prediction is one of the most important key issues in the management and planning of water resources, in particular the adoption of proper decisions in the event of floods and the occurrence of droughts. In order to predict the flow rate of rivers, various approaches have been introduced in hydrology, in which intelligent models are the most important ones.

**Materials and Methods:** In this study, daily data from Kashkan watershed in Lorestan province was used to evaluate the accuracy of models in river flow prediction. Support Vector Machine, Gene Expression Programming and Bayesian network were used to model the daily flow of Kashkan River and the results were compared with each other for the accuracy of the studied models. In a few studies, each of the models presented in the prediction of daily flow has been studied, but the purpose of this study is to simultaneously examine these models in a basin to predict the daily flow of the river. In this research, the Kashkan River in the Lorestan province was selected as the study area and the daily flow of observations of this basin was used at Pul-e-Dokhtar hydrometric station to calibrate and validate the models. For this purpose, at first 80% of the daily flow data (2004-2011) was selected for calibration of the models and 20% of the data (2012-2014) were used to validate the models. Gene expression programming is an automated scheduling method that provides problem solution using computer programming and is part of a family of evolutionary programming. A Support Vector Machine is also an efficient learning system based on optimization theory. Also, the Basin network is a meaningful representation of our uncertain relationships between parameters in a process and a non-circular directional graph of nodes for displaying random variables and arcs to represent potential relationships between variables. The correlation coefficients, root mean square error, mean absolute error was used for evaluation and also comparison of the performance of models in this research.

**Results:** The results showed that all three models have better results in structures of 1 to 5 daily times than other specified structures. In addition to, according to the evaluation criteria, it was found between the models used, the Support Vector Machine model, the highest accuracy of  $R = 0.910$ , the lowest root mean square error of  $RMSE=0.002m^3/s$  and the lowest absolute error value of  $MAE=0.001m^3/s$  at verification stage.

**Conclusions:** The results showed that an increase in the number of effective parameters in different models for simulation results in better performance in the discharge estimation. In addition to, the results showed that the Support Vector Machine model has a better performance than Gene Expression Programming and Bayesian networks.

**Keywords:** Gene expression programming, Forecast, Bayesian network, Kashkan, Support vector Machine

---

\* Corresponding Author; Email: [reza.dehghani67@yahoo.com](mailto:reza.dehghani67@yahoo.com)

*Ar*