



## بررسی کارایی مدل‌های هیبریدی ماشین بردار پشتیبان در تخمین جریان رودخانه‌های حوضه آبریز کرخه

رضا دهقانی<sup>۱</sup>، حسن ترابی پوده<sup>۲\*</sup>، حجت‌اله یونسی<sup>۳</sup>، بابک شاهی نژاد<sup>۴</sup>

۱- دانشجوی دکتری سازه‌های آبی، دانشگاه لرستان

۲- دانشیار گروه مهندسی آب، دانشکده کشاورزی، دانشگاه لرستان

۳- استادیار گروه مهندسی آب، دانشکده کشاورزی، دانشگاه لرستان

۴- استادیار گروه مهندسی آب، دانشکده کشاورزی، دانشگاه لرستان

تأیید نهایی مقاله: ۱۳۹۹/۰۲/۱۷

وصول مقاله: ۱۳۹۸/۱۰/۰۶

### چکیده

در این پژوهش کاربرد مدل‌های هیبریدی ماشین بردار پشتیبان - موجک و ماشین بردار پشتیبان - بی‌زین جهت برآورد دبی رودخانه‌های حوضه آبریز کرخه بر اساس آمار آبدهی روزانه ایستگاه‌های هیدرومتری واقع در بالادست سد کرخه طی دوره آماری ۱۳۸۷-۱۳۹۷ مورد بررسی و ارزیابی قرار گرفته است. معیارهای ضریب همبستگی، ریشه میانگین مربعات خطا و میانگین قدر مطلق خطا برای ارزیابی و عملکرد مدل‌ها مورد استفاده قرار گرفت. نتایج نشان داد ساختارهای ترکیبی نتایج قابل قبولی در مدل‌سازی دبی رودخانه ارائه می‌نمایند. همچنین مقایسه‌ی مدل‌ها نشان داد مدل هیبریدی ماشین بردار پشتیبان-موجک دقت بهتری در پیش‌بینی جریان از خود نشان داده است. در مجموع نتایج نشان داد استفاده از مدل هیبریدی ماشین بردار پشتیبان می‌تواند در زمینه‌ی پیش‌بینی دبی روزانه مفید باشد.

**کلمات کلیدی:** شبکه‌ی بی‌زین، ماشین بردار پشتیبان، موجک، حوضه آبریز کرخه

## ۱- مقدمه

مهم‌ترین مسائل جهت مدیریت سیلاب و جلوگیری از صدمات اقتصادی و جانی ناشی از آن، برآورد دقیق جریان رودخانه‌ها است. بر این اساس دستیابی به روش‌های مطمئن پیش‌بینی جریان رودخانه‌ها به‌منظور برنامه‌ریزی در بهره‌برداری به‌موقع از منابع آب از اهمیت روزافزونی برخوردار است (دهقانی و همکاران، ۱۳۹۶). لذا انتخاب مدلی که بتواند با استفاده از عوامل تأثیرگذار، جریان رودخانه را به‌طور قابل قبولی تخمین زند امری ضروری به نظر می‌رسد. امروزه سامانه‌های هوشمند به‌طور گسترده برای پیش‌بینی پدیده‌های غیرخطی مورد استفاده قرار می‌گیرد. یکی از روش‌هایی که در سال‌های اخیر در زمینه‌ی هیدرولوژی مورد توجه قرار گرفته است مدل هیبریدی ماشین بردار پشتیبان با تبدیل موجک و شبکه‌ی بیزین می‌باشد مدل ماشین بردار پشتیبان یک سیستم یادگیری کارآمد بر مبنای تئوری بهینه‌سازی مقید است (دهقانی و همکاران، ۱۳۹۶). استفاده از موجک به‌عنوان یکی از تبدیل‌های ریاضی نوین و کارآمد در زمینه پردازش سیگنال و تجلیل سری‌های زمانی است به کمک تبدیل موجک که قابلیت تجزیه سری زمانی به چند زیر سری زمانی با مقیاس‌های متفاوت دارد، می‌توان رفتار کوچک‌مقیاس و بزرگ‌مقیاس یک فرآیند هیدرولوژیکی را مورد تحلیل قرار داد که باعث افزایش دقت تحلیل مورد نظر خواهد شد (صالحی و همکاران، ۱۳۹۷). در سال‌های اخیر استفاده از مدل‌های هوشمند و هیبریدی در مطالعات پیش‌بینی جریان رودخانه‌ها مورد توجه محققین قرار گرفته است که از جمله می‌توان به موارد ذیل اشاره نمود: هانگ و همکاران<sup>۱</sup> (۲۰۱۴) جریان ماهانه‌ی رودخانه‌ی ایستگاه هواکسی<sup>۲</sup> چین را با استفاده از ماشین بردار پشتیبان پیش‌بینی نمودند و نتایج نشان داد مدل ماشین بردار پشتیبان دقت بالایی در پیش‌بینی جریان ماهانه‌ی رودخانه دارد. قربانی و همکاران (۲۰۱۸) جهت پیش‌بینی جریان ماهانه‌ی رودخانه‌ی اجیدیر<sup>۳</sup> واقع در ترکیه عملکرد مدل‌های هوش مصنوعی هیبریدی را مورد تجزیه و تحلیل

۱- Hung et al.,

۲- Huaxian

۳- Egirdir

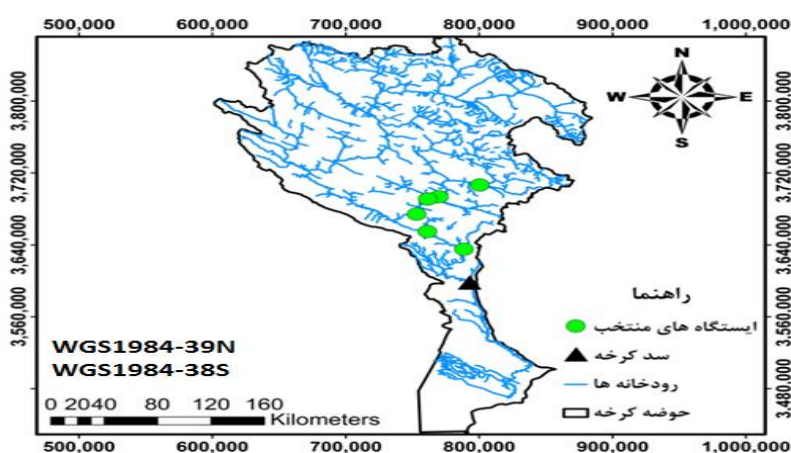
نتایج نشان داد مدل ترکیبی شبکه‌ی عصبی مصنوعی با الگوریتم شب‌تاب از عملکرد بهتری نسبت به سایر مدل‌های ترکیبی برخوردار است. قربانی و همکاران (۲۰۱۸) عملکرد مدل‌های هیبریدی شبکه عصبی مصنوعی و ماشین بردار پشتیبان را در تخمین دبی رودخانه‌ی زرینه‌رود واقع در ایران مورد بررسی قرار دادند نتایج نشان داد که مدل هیبریدی شبکه‌ی عصبی مصنوعی دقت مناسب‌تری نسبت به ماشین بردار پشتیبان دارد. در مجموع با توجه به پژوهش‌های انجام شده و ذکر این نکته که رودخانه‌های حوضه‌ی آبریز کرخه از مهم‌ترین حوضه‌های آبریز کشور و مهم‌ترین منبع تأمین‌کننده‌ی آب بخش‌های مختلف و نواحی مجاور خود از لحاظ کشاورزی و شرب می‌باشد و همچنین کاهش جریان رودخانه‌های این حوضه‌ی آبریز مشکلات زیادی در حوضه آبریز ایجاد کرده است، از سوی دیگر با توجه به اینکه مدل ماشین بردار پشتیبان عملکرد مناسبی نسبت به سایر روش‌های هوشمند مرسوم از جمله شبکه عصبی مصنوعی در پیش‌بینی جریان دارد جهت افزایش دقت و کاهش میزان خطا از ترکیب مدل مذکور با تبدیل موجک و شبکه‌ی بیزین استفاده شد (قربانی و همکاران، ۲۰۱۶). لذا هدف این پژوهش پیش‌بینی و مقایسه‌ی کارایی جریان روزانه رودخانه‌های حوضه‌ی آبریز کرخه با استفاده از مدل‌های هیبریدی ماشین بردار پشتیبان - موجک و ماشین بردار پشتیبان - بیزین می‌باشد.

## ۲- موارد و روش‌ها

### - منطقه‌ی مورد مطالعه

حوضه‌ی آبریز کرخه با مساحت ۵۱۶۴۰ کیلومتر مربع در جنوب غرب ایران در محدوده‌ی ۳۰ درجه و ۵۸ دقیقه تا ۳۴ درجه و ۵۶ دقیقه عرض شمالی و طول ۴۶ درجه و ۶ دقیقه تا ۴۹ درجه و ۱۰ دقیقه شرقی قرار دارد. حوضه‌ی آبریز کرخه جزئی از حوضه‌ی آبریز خلیج فارس بوده که از شمال به حوضه‌ی رودخانه‌های سیروان، سفیدرود، قره چای، از غرب به حوضه‌ی مرزی ایران و عراق، از جنوب به قسمتی از مرز غربی کشور و از سمت شرق به رودخانه‌ی دز محدود می‌شود. رودخانه‌ی کرخه طولی بالغ بر ۹۰۰ کیلومتر داشته و به‌عنوان سومین رودخانه‌ی بزرگ کشور بر اساس متوسط آبدهی سالانه (۸/۵ میلیارد

مترمکعب) است (شرکت مدیریت منابع آب ایران، ۱۳۹۵). در شکل (۱) ایستگاه‌های منتخب حوضه‌ی آبریز کرخه از شرکت آب منطقه‌ای لرستان و سازمان آب و برق خوزستان در دسترس بود، اخذ گردید نشان داده شده است.



شکل (۱) منطقه‌ی مورد مطالعه

Fig(1) The case study

#### - 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 (واپنیک، ۱۹۹۵). در مدل رگرسیون SVM تابعی مرتبط با متغیر وابسته  $Y$  که خود تابعی از چند متغیر مستقل  $X$  است، برآورد می‌شود. در مدل SVM همانند مسائل رگرسیونی رابطه‌ی میان متغیرهای مستقل و وابسته به صورت رابطه‌ی (۲) بیان می‌گردد که در این معادله  $f(x)$  تابع SVM به علاوه خطای مجاز مشخص شود (واپنیک، ۱۹۹۸).

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

$$y=f(x)+noise \quad (۲)$$

چنانچه  $W^T$  ترانهاده بردار ضرایب و  $b$  ثابت مشخصه‌های تابع رگرسیونی و  $\phi$  نیز تابع کرنل باشد، آنگاه هدف پیدا کردن فرم تابعی برای  $f(x)$  است. این مهم با آموزش مدل SVM توسط مجموعه‌ای از داده‌ها (مجموعه آموزش) محقق می‌شود (میسرا و همکاران، ۲۰۰۹). برای محاسبه  $W$  و  $b$  لازم است تابع خطا (معادله‌ی ۳) در مدل SVM- $\epsilon$  با در نظر گرفتن شرایط مندرج (قیود) در معادله‌ی (۴) و (۵) کمینه شود (هامل، ۲۰۰۹).

$$\frac{1}{2}W^T.W+C\sum_{i=1}^N\epsilon_i+C\sum_{i=1}^N\epsilon_i^* \quad (۳)$$

$$W^T.\phi(X_i)+b-y_i\leq\epsilon+\epsilon_i^* \quad (۴)$$

$$y_i-W^T.\phi(X_i)-b\leq\epsilon+\epsilon_i, \epsilon_i, \epsilon_i^*\geq 0, i=1,2,\dots,N \quad (۵)$$

در معادلات بالا  $C$  عددی صحیح و مثبت است، که عامل تعیین جریمه در هنگام رخ دادن خطای آموزش مدل است.  $\phi$  تابع کرنل،  $N$  تعداد نمونه‌ها و دو مشخصه  $\epsilon_i^*$  و  $\epsilon_i$  متغیرهای کمبود هستند. در نهایت تابع SVM رگرسیونی را می‌توان به فرم زیر بازنویسی کرد:

$$f(x)=\sum_{i=1}^N\bar{\alpha}_i\phi(x_i)^T.\phi(x)+b \quad (۶)$$

در معادله‌ی ۶،  $\bar{\alpha}_i$  میانگین ضرایب لاگرانژ می‌باشد. محاسبه‌ی  $\phi(x)$  در فضای مشخصه‌ی آن ممکن است بسیار پیچیده باشد (یون و همکاران<sup>۱</sup>، ۲۰۱۱). برای حل این مشکل روند معمول در مدل SVM رگرسیون انتخاب یک تابع کرنل به صورت رابطه‌ی زیر است.

$$K(X_j,X)=\phi(X_j)^T\sqrt{b^2-4ac} \quad (۷)$$

می‌توان از توابع مختلف کرنل برای ساخت انواع مختلف SVM- $\epsilon$  استفاده کرد. پرکاربردترین توابع کرنل چندجمله‌ای<sup>۲</sup> و کرنل توابع پایه شعاعی<sup>۳</sup> (RBF) و کرنل خطی،

۱- Yoon et al.,

۲- Polynomial

۳- Radial Basis Functions (RBF)

به ترتیب طبق روابط ۸-۱۰ محاسبه می‌گردند (باساک و همکاران، ۲۰۰۷؛ واپنیک و چرونسکی، ۱۹۹۱). در این پژوهش از این سه تابع کرنل استفاده شده است. قابل ذکر است فرآیند محاسبات ماشین بردار پشتیبان بر اساس کدنویسی در محیط متلب انجام، و پارامترهای توابع کرنل از طریق سعی و خطا بهینه گردیدند.

$$K(x, x_j) = (t + x_i \cdot x_j)^d \quad (۸)$$

$$K(x, x_j) = \exp\left(-\frac{\|x - x_j\|^2}{2\sigma^2}\right) \quad (۹)$$

$$k(x, x_j) = x_i \cdot x_j \quad (۱۰)$$

#### - تبدیل موجک

تبدیل موجک به‌عنوان روشی جایگزین برای تبدیل فوریه‌ی زمان کوتاه ارائه شده است و هدف از ارائه‌ی آن، غلبه بر مشکلات مربوط به قدرت تفکیک‌پذیری فرکانس در تبدیل فوریه‌ی زمان کوتاه است. در تبدیل موجک همانند تبدیل فوریه‌ی زمان کوتاه، سیگنال موردنظر به پنجره‌هایی تقسیم شده و تبدیل موجک بر روی هرکدام از این پنجره‌ها به‌صورت جداگانه انجام می‌گیرد (واپنیک، ۱۹۹۸). اما مهم‌ترین تفاوت آن‌ها در این است که در تبدیل موجک علاوه بر اینکه قدرت تفکیک فرکانس‌های یک سیگنال یا طول پنجره، متناسب با نوع فرکانس تغییر می‌کند، هم‌زمان عرض پنجره یا مقیاس فرکانس نیز متناسب با نوع فرکانس تغییر می‌کند. به عبارت دیگر، در تبدیل موجک به جای فرکانس، مقیاس وجود دارد. یعنی تبدیل موجک، نوعی تبدیل زمان - مقیاس است. بر همین اساس با استفاده از تبدیل موجک، در مقیاس‌های بالا سیگنال منبسط شده و جزئیات سیگنال قابل تجزیه و تحلیل است و در مقیاس‌های پایین سیگنال منقبض شده و کلیات سیگنال قابل بررسی می‌باشد (وانگ و همکاران، ۲۰۰۰). یک موجک به معنای موج کوچک، بخشی یا پنجره‌ای از سیگنال اصلی است که انرژی آن در زمان متمرکز شده است. با استفاده از تبدیل یا آنالیز موجک می‌توان یک سیگنال یا سری زمانی مادر را به موجک‌هایی با سطح تفکیک و مقیاس‌های مختلف تجزیه کرد. بنابراین موجک‌ها نمونه‌های انتقال یافته و

تفکیک شده سیگنال مادر هستند که نوساناتی در یک طول متناهی داشته و شدیداً میرا هستند. بر اساس این ویژگی مهم تبدیل موجک، می‌توان سری‌های زمانی نامانا و گذرا را به صورت موضعی مورد تجزیه و تحلیل قرارداد (شین و همکاران، ۲۰۰۵).

تبدیل موجک به دو صورت پیوسته و گسسته تعریف می‌شود.

تبدیل موجک پیوسته<sup>۱</sup> (CWT)

تبدیل موجک پیوسته‌ی تابع  $f(t)$  به صورت روابط (۱۱) و (۱۲) تعریف می‌شود (واپنیک، ۱۹۹۸).

$$CWT_f^\psi(s, \tau) = \Psi_f^\psi(s, \tau) = \frac{1}{\sqrt{|s|}} \int_{-\infty}^{+\infty} f(t) \psi^* \left( \frac{t-\tau}{s} \right) dt = \langle f(t), \psi_{s,\tau}(t) \rangle \quad (11)$$

$$\psi_{s,\tau}(t) = \frac{1}{\sqrt{|s|}} \psi \left( \frac{t-\tau}{s} \right) \quad (12)$$

رابطه‌ی (۱۲) یک رابطه با دو متغیر  $s$  و  $\tau$  است که  $s$  پارامتر مقیاس<sup>۲</sup> (عکس فرکانس) و  $\tau$  پارامتر انتقال<sup>۳</sup> می‌باشند. علامت \* نیز نشان‌دهنده‌ی مزدوج مختلط است.  $\psi$  تابع پنجره یا موجک مادر و  $\frac{1}{\sqrt{|s|}} \psi \left( \frac{t-\tau}{s} \right)$  موجک‌های حاصل از انتقال و تغییر مقیاس موجک مادر هستند (وانگ و همکاران، ۲۰۰۰). واژه‌ی مادر به این سبب به کار می‌رود که تمامی نسخه‌های انتقال‌یافته و مقیاس شده (موجک‌های دختر)، همگی از این تابع به دست می‌آیند. یعنی موجک مادر یک الگو برای سایر پنجره‌هاست. علامت (...) نیز نشان‌دهنده‌ی ضرب برداری دو تابع در فضای سیگنال است.

- شبکه بیزین

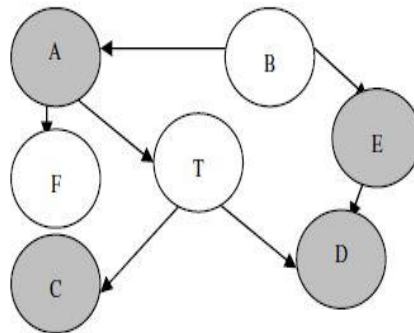
شبکه‌های بیزین مدل‌های گرافیکی هستند که برای استدلال در مواردی که پیچیدگی و عدم قطعیت وجود دارد به کار می‌روند و یا به عبارتی گرافی است که متغیرهای تصادفی و

۱- Continues Wavelet Transform

۲- Scaling

۳- Translation

وابستگی آنها را نمایش می‌دهد (کوبین و نیچولسون، ۲۰۱۰). 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 (هکرمن، ۱۹۹۷). نمونه‌ای از گراف ساده در شکل (۲) نشان داده شده است.



شکل (۲) گراف ساده شبکه‌ی بیزین  
Fig(2) A graph of simple Bayesian network

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

$$P(x_1, \dots, x_n) = \prod_{i=1}^n P(x_i | \text{parents}(X_i)) \quad (13)$$

بعد از ایجاد شبکه، مقدار روابط بین گره‌های متصل به هم به دست می‌آید مقدار احتمال شرطی مربوط به هر متغیر در جدولی بنام جدول احتمال شرطی قرار می‌گیرد (نوبین و همکاران، ۱۹۹۸).



## - معیارهای ارزیابی

در این تحقیق به منظور ارزیابی دقت و کارایی مدل‌ها، از نمایه‌های ضریب تبیین ( $R^2$ )، ریشه میانگین مربعات خطا (RMSE) و میانگین قدر مطلق خطا (MAE) طبق روابط زیر استفاده گردید (قربانی و دهقانی، ۱۳۹۵). بهترین مقدار برای این سه معیار به ترتیب یک، صفر و صفر می‌باشد.

$$R^2 = \frac{\sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^N (x_i - \bar{x})^2 \sum_{i=1}^N (y_i - \bar{y})^2}} \quad -1 \leq R \leq 1 \quad (14)$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - y_i)^2} \quad (15)$$

$$MAE = \frac{1}{n} \sum |x_i - y_i| \quad (16)$$

در روابط بالا،  $x_i$  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.

## ۳- 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 رودخانه‌های حوضه‌ی آبریز کرخه انتخاب و در جدول (۲) ارائه شد. در این جدول، ستون‌های  $Q(t-1)$ ،  $Q(t-2)$ ،  $Q(t-3)$  و  $Q(t-4)$  (به ترتیب جریان رودخانه در زمان‌های  $t-1$ ،  $t-2$ ،  $t-3$ ،  $t-4$ ) به عنوان ورودی و  $Q(t)$  (جریان رودخانه در زمان  $t$ ) به عنوان خروجی مدل در نظر گرفته شد. با توجه به اینکه در تحقیق حاضر تأثیر توالی دبی روزهای قبل در پیش‌بینی دبی روز مدنظر بوده، صرفاً از داده‌های نرمال شده دبی جریان با توالی برگشتی تا ۴ روز به عنوان داده‌های آموزشی، به صورت ترکیب‌های مختلف مطابق جدول استفاده شده است که از این

1- Cross-correlation

پس از آن‌ها به‌عنوان الگو نام‌برده خواهد شد و همچنین با توجه به اینکه مقدار همبستگی متقابل بین داده‌های ورودی و خروجی که مطابق جدول ۱ بیش از ۰/۷۵۰ حاصل شد، ترکیب‌های مختلفی از پارامترهای ورودی به‌منظور دستیابی به مدل بهینه جهت تخمین جریان روزانه حوضه‌ی آبریز کرخه استفاده گردید (داننده مهر و مجدزاده، ۱۳۸۹). بدین منظور از داده‌های ایستگاه‌های هیدرومتری چم انجیر، مادیان رود، افرینه، کشکان، پل زال و جلوگیر واقع در حوضه‌ی آبریز کرخه که دارای ۳۶۵۰ رکورد ثبت‌شده طی بازه‌ی زمانی (۱۳۸۷-۱۳۹۷)، در مقیاس زمانی روزانه می‌باشد، استفاده شد. در نهایت تعداد ۲۹۲۰ رکورد برای آموزش و تعداد ۷۳۰ رکورد باقی‌مانده برای صحت‌سنجی مدل‌های موردبررسی انتخاب شد. لازم به ذکر است جهت مدل‌سازی ۸۰ درصد داده‌ها برای آموزش و ۲۰ درصد باقی‌مانده جهت آزمون، به‌صورت تصادفی، که گستره‌ی وسیعی از انواع داده‌ها را پوشش دهد، انتخاب شد (کیسی و همکاران، ۲۰۰۶؛ ناگای و همکاران، ۲۰۰۲).

جدول (۱) همبستگی متقابل بین پارامترهای ورودی و خروجی

Tab (1) The cross-correlation between input and output parameters

ایستگاه	Q(t-1)	Q(t-2)	Q(t-3)	Q(t-4)
چم انجیر	۰/۹۴۰	۰/۹۲۱	۰/۸۹۳	۰/۸۵۴
مادیان رود	۰/۸۹۰	۰/۸۶۴	۰/۸۱۴	۰/۷۵۴
افرینه	۰/۹۱۴	۰/۸۹۲	۰/۸۶۵	۰/۸۱۳
کشکان	۰/۹۲۵	۰/۹۰۴	۰/۸۷۶	۰/۸۳۴
پل زال	۰/۹۲۰	۰/۸۹۶	۰/۸۸۵	۰/۸۲۲
جلوگیر	۰/۹۲۴	۰/۸۹۷	۰/۸۸۰	۰/۸۲۶

جدول (۲) ترکیب‌های منتخب پارامترهای ورودی

Tab (2) Selected compositions input parameter models

شماره	ساختار ورودی	خروجی
۱	Q(t-1)	Q(t)
۲	Q(t-1), Q(t-2)	Q(t)
۳	Q(t-1), Q(t-2), Q(t-3)	Q(t)
۴	Q(t-1), Q(t-2), Q(t-3), Q(t-4)	Q(t)

### - نتایج مدل هیبریدی ماشین بردار پشتیبان - موجک

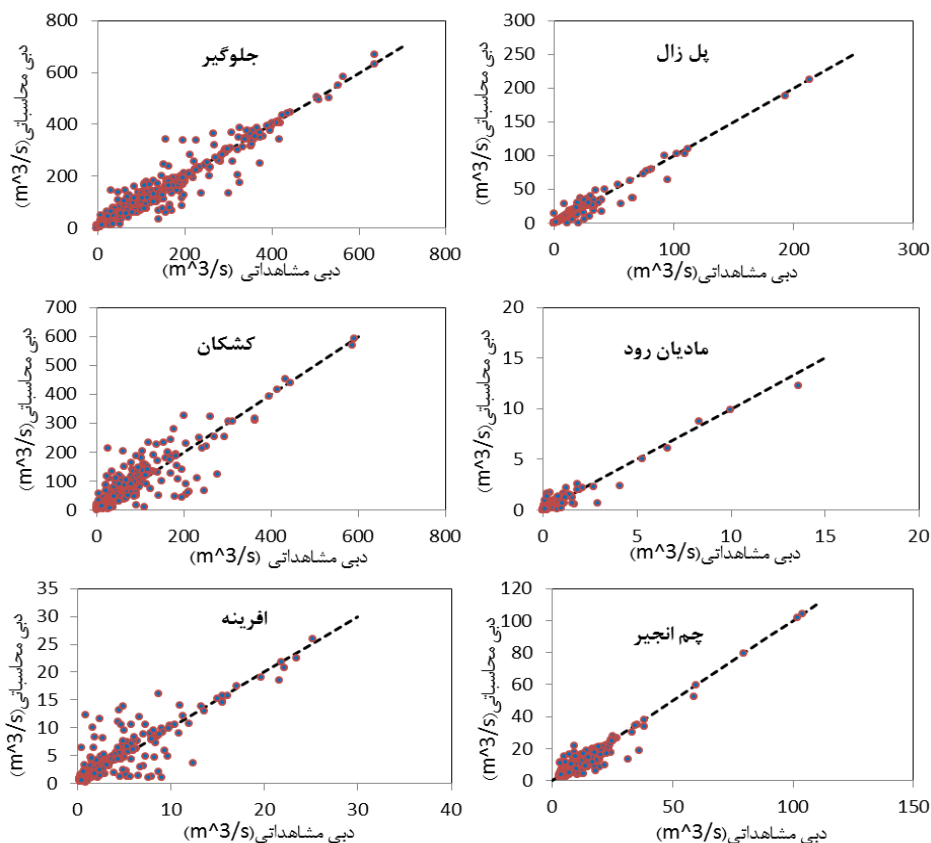
به منظور بررسی نتایج حاصل از مدل ترکیبی، ابتدا با استفاده از تبدیل موجک، پارامترهای ورودی به زیر سیگنال‌ها تجزیه و سپس زیر سیگنال‌های مزبور، به‌عنوان ورودی به مدل ماشین بردار پشتیبان وارد می‌شوند. به‌منظور کاهش دامنه‌ی تغییرات داده‌های جریان روزانه‌ی رودخانه‌های حوضه‌ی آبریز کرخه به همسان‌سازی اطلاعات 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 توابع کرنل مختلفی از جمله خطی، چندجمله‌ای و توابع پایه‌ی شعاعی مورد استفاده و تابع RBF انتخاب شد چرا که از دقت بهتری در برآورد جریان روزانه‌ی رودخانه‌ها برخوردار است (لین و همکاران، ۲۰۰۶). در این تابع نیز مشخصه  $\gamma$  بایستی تعیین گردد. بنابراین در حالت کلی برای پیش‌بینی جریان روزانه‌ی رودخانه‌ها به‌وسیله‌ی مدل SVM، لازم است که مقادیر بهینه سه مشخصه‌ی مذکور محاسبه گردد که بدین منظور دو مشخصه  $\epsilon$  و  $C$  توسط الگوریتم بهینه‌سازی جستجوی شبکه و متغیر  $\gamma$  نیز به‌صورت آزمون و خطا محاسبه شد. البته قابل‌ذکر است که الگوریتم بهینه‌سازی جستجوی شبکه بسیار کند عمل می‌کند و زمان محاسباتی زیادی را به خود اختصاص می‌دهد. برای حل این مشکل از برنامه‌ی اصلاح‌شده الگوریتم جستجوی شبکه در پژوهش چن و یو (۲۰۰۷) که به نام الگوریتم جستجوی شبکه‌ی دومرحله‌ای معروف است به همراه اعتبار سنجی متقاطع استفاده شد. برای این منظور ابتدا با انتخاب شبکه‌های با ابعاد بزرگ، محدوده‌ی مشخصه‌های  $\epsilon$  و  $C$  به ازای مقدار ثابت مشخصه  $\gamma$  تعیین شد. سپس با مشخص شدن محدوده‌ی 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 در مقدار  $\gamma$  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  $\epsilon$ ،  $C$  و  $\gamma$  انتخاب نمود. و بدین ترتیب مدل

ترکیبی ماشین بردار پشتیبان موجک حاصل می‌گردد. یکی از نکات مهم و اساسی در این پژوهش توابع موجک مختلف مورد بررسی قرار گرفت و موجک کلاه مکزیکی از عملکرد بهتری نسبت به سایر توابع برخوردار بود. در جدول (۳) نتایج حاصل از مدل ترکیبی در ایستگاه‌های منتخب نشان داده شده است. در این جدول مدل هیبریدی در ایستگاه‌های چم انجیر، کشکان و جلوگیر از دقت بیشتر و خطای کمتری برخوردار است. در شکل (۳) نمودار پراکنش مدل هیبرید ماشین بردار پشتیبان-موجک مربوط به مرحله‌ی 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 گرفته، که این امر دلالت بر برابری مقادیر مشاهداتی و محاسباتی بر اساس خط  $(y=x)$  است. همچنین همان‌طور که در این شکل مشهود است مدل هیبریدی ماشین بردار پشتیبان-موجک در پیش‌بینی مقادیر حداکثر و حداقل از عملکرد قابل قبولی برخوردار است به گونه‌ای که این مقادیر را نزدیک به مقدار واقعی‌شان تخمین زده است که با نتایج پژوهش ژو و همکاران (۲۰۱۶) همخوانی دارد.

جدول (۳) تحلیل نتایج مدل هیبریدی ماشین بردار پشتیبان-موجک برای ایستگاه‌های منتخب

**Tab (3)** The analysis of WSVM model results for river flow inputs.

ایستگاه	آموزش			صحت سنجی		
	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
چم انجیر	۰/۰۴۸	۰/۰۴۸	۰/۹۴۲	۰/۰۲۳	۰/۰۳۱	۰/۹۶۴
مادیان رود	۰/۰۹۵	۰/۰۸۴	۰/۸۸۷	۰/۰۷۵	۰/۰۶۸	۰/۹۱۵
افرینه	۰/۰۷۴	۰/۰۷۹	۰/۹۲۱	۰/۰۵۸	۰/۰۵۸	۰/۹۳۶
کشکان	۰/۰۶۳	۰/۰۴۴	۰/۹۳۲	۰/۰۳۷	۰/۰۲۱	۰/۹۵۶
پل زال	۰/۰۸۸	۰/۰۸۱	۰/۸۹۴	۰/۰۷۴	۰/۰۶۳	۰/۹۲۷
جلوگیر	۰/۰۸۴	۰/۰۷۷	۰/۹۱۴	۰/۰۶۳	۰/۰۵۱	۰/۹۴۲



شکل (۳) نمودار پراکنش مقادیر مشاهده‌ای و محاسباتی حاصل از مدل هیبریدی ماشین بردار پشتیبان- موجک در مرحله‌ی صحت‌سنجی

Fig(3) Scatter figure for calculation and observational amounts obtained by WSV

- نتایج مدل هیبریدی ماشین بردار پشتیبان - بیزین

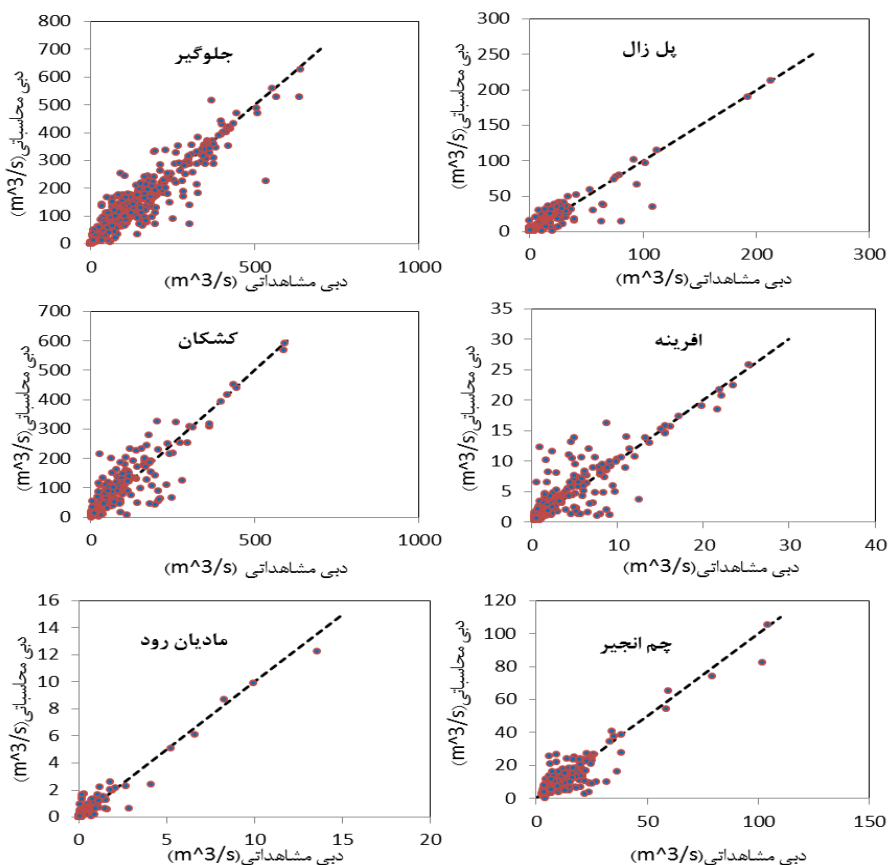
به منظور مقایسه نتایج مدل ترکیبی ماشین بردار پشتیبان-موجک از مدل ماشین بردار پشتیبان-بیزین استفاده شد. در این مدل برای پیش‌بینی جریان روزانه رودخانه‌ها به‌وسیله مدل SVM، لازم است که مقادیر بهینه سه مشخصه مذکور محاسبه گردد که بدین منظور دو مشخصه  $\epsilon$  و  $C$  توسط قانون احتمالات بیزی و متغیر  $\gamma$  نیز به‌صورت آزمون و خطا

محاسبه شد نتایج حاصل بررسی مدل‌ها در جدول (۴) آورده شده است. در این جدول مدل ماشین بردار پشتیبان - بیزین در ایستگاه کشکان از دقت بیشتر و خطای کمتری برخوردار است. در شکل (۴) نمودار بهترین مدل حاصل شده برای داده‌های بخش صحت سنجی نشان داده شده است. همان‌طور که در این شکل مشخص است بیشتر مقادیر تخمینی و مشاهداتی به غیر از یک نقطه روی خط نیمساز قرار گرفته و این امر ناشی از برابری مقادیر مشاهداتی و محاسباتی می‌باشد.

جدول (۴) تحلیل نتایج مدل ماشین بردار پشتیبان - بیزین برای ایستگاه‌های منتخب

Tab (4) The analysis of BN-SVM model results for river flow inputs.

صحت سنجی			آموزش			ایستگاه
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	
۰/۰۲۲	۰/۰۴۵	۰/۹۵۱	۰/۰۳۷	۰/۰۵۴	۰/۹۳۶	چم انجیر
۰/۰۸۲	۰/۰۹۸	۰/۹۰۱	۰/۰۹۶	۰/۱۱۷	۰/۸۷۵	مادیان رود
۰/۰۶۴	۰/۰۸۵	۰/۹۲۰	۰/۰۵۳	۰/۰۷۴	۰/۹۱۴	افرینه
۰/۰۳۲	۰/۰۵۸	۰/۹۴۵	۰/۰۴۱	۰/۰۶۸	۰/۹۲۶	کشکان
۰/۰۷۸	۰/۰۹۳	۰/۹۱۰	۰/۰۸۶	۰/۱۰۵	۰/۸۸۴	پل زال
۰/۰۵۷	۰/۰۸۴	۰/۹۲۳	۰/۰۷۴	۰/۰۹۵	۰/۸۸۶	جلوگیر



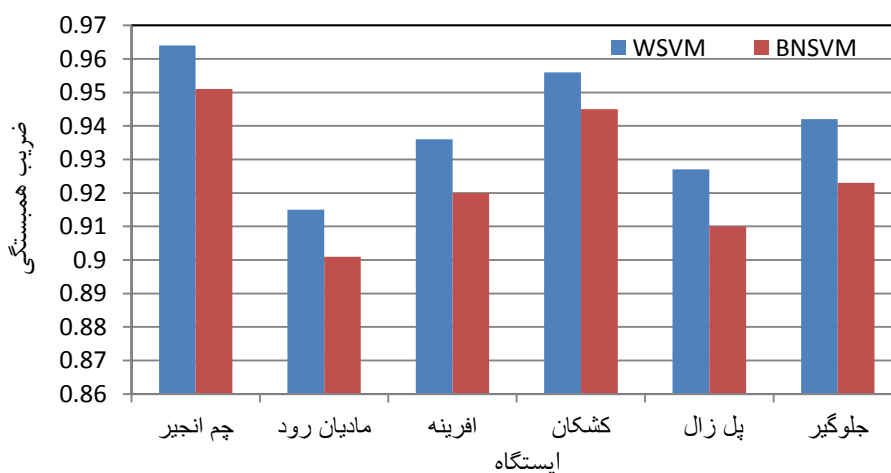
شکل (۴) نمودار پراکنش مقادیر مشاهداتی و محاسباتی حاصل از مدل ماشین بردار پشتیبان-بیزین در مرحله‌ی صحت‌سنجی

Fig(4) Scatter figure for calculation and observational amounts obtained by BN-SVM

#### - مقایسه عملکرد مدل‌ها

در ادامه با انتخاب جواب بهینه هر کدام از مدل‌های هیبریدی هوش مصنوعی و مقایسه‌ی آنها با یکدیگر مشخص گردید، هر دو مدل با دقت خوبی توانسته‌اند دبی حوضه‌ی آبریز کرخه را شبیه‌سازی کنند در شکل (۴) نمودار مقادیر مشاهداتی و محاسباتی حاصل از مدل‌های ماشین بردار پشتیبان- موجک و ماشین بردار پشتیبان- بیزین در مرحله‌ی

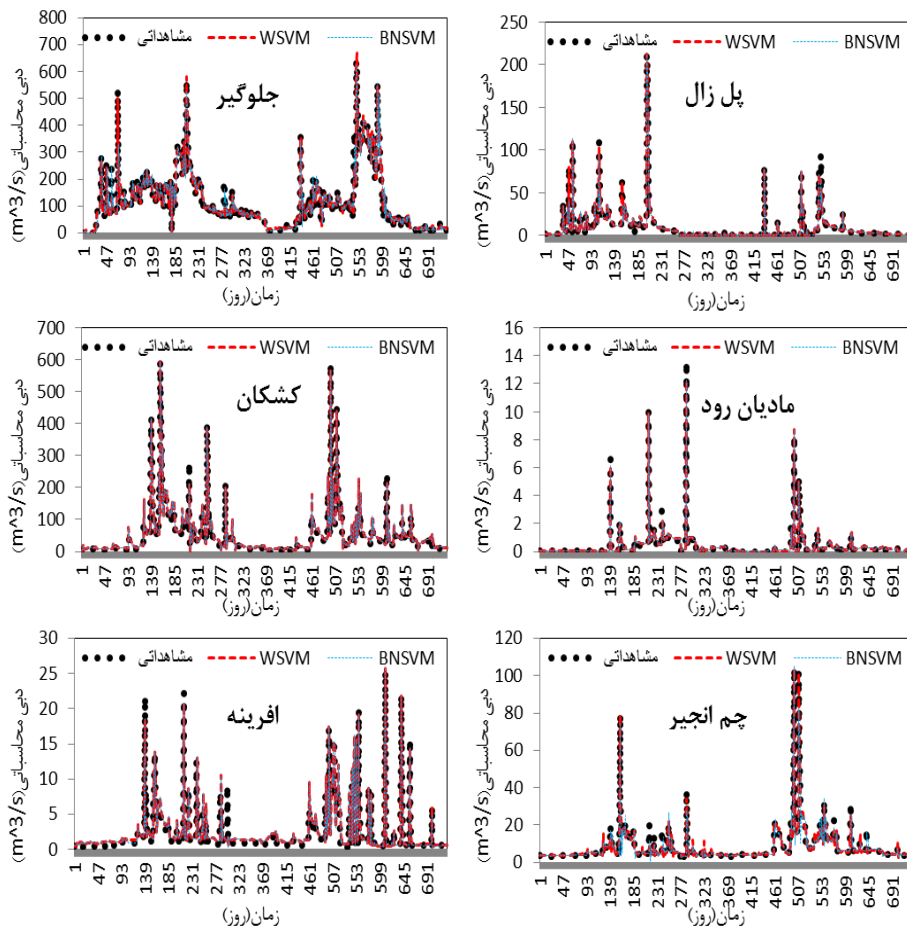
صحت‌سنجی نشان داده شده است در این شکل مدل ماشین بردار پشتیبان\_ موجک در تخمین مقادیر حداقل و حداکثر ایستگاه‌های مورد بررسی عملکرد مناسبی از خود نشان داده است در تبیین این نتایج می‌توان بیان نمود تبدیل موجکی با جدا نمودن سیگنال به فرکانس‌های بالا و پائین ویژگی‌های چند مقیاسی سیگنال را در اختیار داشته و دقت مدل را تا حد قابل‌توجهی، بالا می‌برد (شاهی‌نژاد و دهقانی، ۱۳۹۷). همچنین طبق شکل (۶) مدل ماشین بردار پشتیبان- بیزین در تخمین مقادیر حداقل نتایج قابل قبولی دارد. از سوی دیگر طبق معیارهای ارزیابی که در شکل (۵) نشان داده شد مدل هیبریدی ماشین بردار پشتیبان- موجک خطای کمتر و دقت بیشتری در ایستگاه‌های مورد بررسی بدلیل همپوشانی داده‌های محاسباتی با مشاهداتی از خود نشان داده است که این امر در ایستگاه‌های جلوگیر و پل زال به دلیل وجود آورد ایستگاه‌های بالادست مشهود است. همچنین در ایستگاه‌های چم انجیر و کشکان هر دو مدل مورد بررسی در برآورد دبی رودخانه‌های حوضه‌ی آبریز کرخه کارایی مناسبی دارند.



شکل (۵) میزان ضریب همبستگی مدل‌های مورد بررسی در بخش صحت‌سنجی

Fig(5) Correlation coefficient of models in accuracy section





شکل (۶) نمودار مقادیر مشاهداتی و محاسباتی حاصل از مدل‌های ماشین بردار پشتیبان-موجک و ماشین بردار پشتیبان-بیزین در مرحله صحت‌سنجی

Fig(6) Observed and predicted values of the WSVM and BN-SVM model in validation

#### ۴- نتیجه‌گیری

در این تحقیق سعی بر آن شد عملکرد مدل‌هایی جهت شبیه‌سازی جریان روزانه‌ی رودخانه‌های حوضه‌ی آبریز کرخه با استفاده از داده‌های روزانه مورد ارزیابی قرار گیرد.

مدل‌های به کار گرفته شده شامل مدل‌های هیبریدی ماشین بردار پشتیبان\_ موجک و ماشین بردار پشتیبان\_ بیزین است. مقادیر جریان مشاهداتی با جریان پیش‌بینی شده در مدل‌های مذکور (مدل هیبریدی ماشین بردار پشتیبان\_ موجک و ماشین بردار پشتیبان\_ بیزین)، با استفاده از معیارهای ارزیابی مورد مقایسه قرار گرفت. نتایج تحقیق را می‌توان به صورت زیر خلاصه نمود: نتایج نشان داد که در هر دو مدل مورد بررسی شامل مدل هیبریدی ماشین بردار پشتیبان\_ موجک و ماشین بردار پشتیبان\_ بیزین، در ساختاری متشکل از ۱ تا ۴ تأخیر زمانی نتایج بهتری نسبت به سایر ساختارها ارائه می‌دهد. همچنین با توجه به معیارهای ارزیابی نتیجه شد که هر دو مدل مورد بررسی، می‌توانند با دقت نسبتاً بالایی میزان جریان روزانه رودخانه را پیش‌بینی نمایند. در این میان، مدل هیبرید ماشین بردار پشتیبان\_ موجک دقت بیشتر و خطای کمتری نسبت به مدل ماشین بردار پشتیبان\_ بیزین از خود نشان داده است. در مجموع این تحقیق نشان می‌دهد استفاده از مدل ترکیبی ماشین بردار پشتیبان\_ موجک می‌تواند در زمینه پیش‌بینی جریان روزانه رودخانه‌ها مؤثر باشد. همچنین، این مدل می‌تواند به نوبه 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.

## ۵- منابع

- 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: Barandoozchay River), *Soil and water conservation research*, 22(6), 171-186, (in Persian).
- Basak, D., Pal, S., and Patranabis, D.C. (2007). Support vector regression, *Neural Inf Process*, 11(2), 203-225.
- Chen S.T., and Yu, P.S. (2007). Real-time probabilistic forecasting of flood stages, *Journal of Hydrology*, 340(2), 63-77.
- Chen, S.T., and Yu, P.S. (2007). Real-time probabilistic forecasting of flood stages, *Journal of Hydrology*, 340(2), 63-77.
- Danandeh Mehr, A., and Majdzadeh Tabatabaei, M.R. (2009). I prediction of daily discharge trend of river flow based on genetic programming, *Journal of Water and Soil*, 24(2), 325-333, (In Persian).
- Daneshavr Vosoughi, F., Monafian, V. (2018). Application of Hybrid Supportive-Wavelet Machine and Neural Network Combined Models in Predicting Groundwater Level in Ardabil Plain, *Hydrogeomorphology*, 5(17), 45-64.
- Dehghani, R., Younesi, H., Torabi Poudeh, H. (2016). Comparing the performance of Support Vector Machines, Gene Expression Programming and Bayesian networks in predicting river flow (Case study: Kashkan River), *Journal of Soil and Water Conservation Research*, 24(4), 161-177. (in Persian).
- horbani, M.A., Deo, R., Karimi, V., Yassen, Z.M., Terzi, O. (2018). Implementation of a hybrid MLP-FFA model for water level prediction of Lake Egirdir, Turkey, Stochastic, *Environmental Research and Risk Assessment*, 32(6), 1683-1697.
- 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.
- Ghorbani, M.A., Khatibi, R., Karimi, V., Yaseen, Z.M., Zounemat-Kermani, M. (2018). Learning from multiple models using artificial intelligence to

improve model prediction accuracies: Application to River Flows, *Water Resour Management*, 32(13), 4201-4215.

Hamel, L. (2009). Knowledge discovery with support vector Machines, hoboken, N.J. John Wiley.

Hekerman, D. (1997). Bayesian Networks for Data Mining, *Data Mining and Knowledge Discovery*, 1(1), 79-119.

Huang, S., Chang, J., Huang, Q., and Chen, Y. (2014). Monthly streamflow prediction using modified emd-based support vector machine, *Journal of Hydrology*, 511(4), 764-775.

Iran Water Resources Management Company. (2015). *Water Monthly*, 4(2), 24-60.

Kisi, O., and Guven, A. (2010). A machine code-based genetic programming for suspended sediment concentration estimation, *Advanced in Engineering Software*, 41(7), 939-945.

Kisi, O., and Shiri, J. (2012). River suspended sediment estimation by climatic variables implication: Comparative study among soft computing techniques, *Computers & Geosciences*, 43(4), 73-82.

Kisi, O., Karahan, M., and Sen, Z. (2006). River suspended sediment modeling using fuzzy logic approach, *Hydrology of Process*, 20(2), 4351-4362.

Korb, K. B., Nicholson, A. E. (2010). *Bayesian artificial intelligence*, second edition, CRC Press, 360-452.

Lin, J.Y., Cheng, C.T., and Chau, K.W. (2006). Using support vector machines for long-term discharge prediction, *Hydrolog Science Journal*, 51(3), 599-612.

Liong, S.Y., and Sivapragasam, C. (2002). Flood stage forecasting with support vector machines, *Journal Am Water Resour*, 38(4), 173-186.

Misra, D., Oommen, T., Agarwa, 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.

Nagy, H., Watanabe, K., and Hirano, M. (2002). Prediction of sediment load concentration in rivers using artificial neural network model, *Journal of Hydraulics Engineering*, 128(3), 558-559.

- Nguyen, R.T., Prentiss, D., Shively, J.E. (1998). *Rainfall interpolation for Santa Barbara County*, UCSB, Department Geography, USA.
- Salehi, M., Radmanesh, F., Zarei, H., Mansouri, B., Solgi, A. (2016). A combined Time Series – Wavelet Model For Prediction of Ground Water Level (Case Study: Firuzabad Plain), *Irrigation Science and Engineering*, 41(4), 1-16, (in Persian).
- Shahinejad, B., Dehghani, R. (2018). Comparison of wavelet neural network models, support vector machine and gene expression programming in estimating the amount of oxygen dissolved in rivers, *Iranian Water Resources Research*, 14(3), 226-238, (in Persian).
- Shin, S., Kyung, D., Lee, S., Taik & Kim, J., and Hyun, J. (2005). An application of support vector machines in bankruptcy prediction model, *Expert Systems with Applications*, 28(4), 127-135.
- Vapnik, V., and Chervonenkis, A. (1991). The necessary and sufficient conditions for consistency in the empirical risk minimization method, *Pattern Recognition and Image Analysis*, 1(3), 283-305.
- Vapnik, V.N. (1995). *The Nature of Statistical Learning Theory*. Springer, New York.
- Vapnik, V.N. (1998). *Statistical learning theory*, Wiley, New York.
- Wang, D., Safavi, A.A., and Romagnoli, J.A. (2000). Wavelet-based adaptive robust M-estimator for non-linear system identification, *AIChE Journal*, 46(4), 1607-1615.
- 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, *Journal of Hydrol*, 396(4), 128–138.