

## ارزیابی عملکرد شبکه‌های عصبی مصنوعی و ماشین‌بردار پشتیبان در بازسازی داده‌های ماهانه و فصلی بارش در ایستگاه‌های باران‌سنجی (مطالعه موردی: استان همدان)

مصطفی کدخداحسینی<sup>۱\*</sup>، شایان شامحمدی<sup>۲</sup>، رسول میرعباسی نجف‌آبادی<sup>۳</sup> و حامد نوذری<sup>۴</sup>

### چکیده

بازسازی داده‌های بارش، از اقدامات لازم پیش از پیش‌بینی سیلاب‌ها است. گام نخست در پیش‌بینی سیلاب، پیش‌بینی میزان بارش می‌باشد. پیش‌بینی بارش در یک ایستگاه نیازمند آمار و اطلاعات دقیق و طولانی مدت از مقدار و توزیع زمانی بارندگی در آن ایستگاه است. در پژوهش حاضر از دو مدل هوشمند شبکه عصبی مصنوعی MLP با الگوریتم پس‌انتشار و ماشین‌بردار پشتیبان تحت تابع RBF برای بازسازی داده‌های بارش ماهانه چهار ایستگاه باران‌سنجی واقع در استان همدان، در دوره آماری ۱۳۷۰ تا ۱۳۸۹، استفاده شد. بدین منظور مشخصه‌های مدل SVM شامل  $\epsilon$ ،  $C$  و  $\gamma$  محاسبه شد. داده‌های ناقص ایستگاه هدف ابتدا به کمک داده‌های یک ایستگاه، سپس داده‌های دو ایستگاه و در نهایت داده‌های سه ایستگاه، بازسازی گردید. نتایج نشان داد که با افزایش تعداد ایستگاه‌های دخیل در مرحله آموزش، عملکرد مدل‌ها بهبود می‌یابد. همچنین بر اساس نتایج حاصل شده، در بهترین حالت، مدل شبکه عصبی مصنوعی در بازسازی داده‌های بارش ماهانه و مدل ماشین‌بردار پشتیبان در بازسازی داده‌های بارش فصلی به ترتیب با مقدار خطای  $۱۲/۳۳$  و  $۱۸/۸۲$  میلی‌متر و ضریب تبیین  $۰/۸۷$  و  $۰/۹۴$  نسبت به مدل دیگر از عملکرد بهتری برخوردار بودند.

**واژه‌های کلیدی:** بازسازی، بارش ماهانه، پیش‌بینی، داده‌های گمشده، ضریب تبیین.

**ارجاع:** کدخداحسینی م. شامحمدی ش. میرعباسی نجف‌آبادی ر. و نوذری ح. ۱۳۹۶. ارزیابی عملکرد شبکه‌های عصبی مصنوعی و ماشین‌بردار پشتیبان در بازسازی داده‌های ماهانه و فصلی بارش در ایستگاه‌های باران‌سنجی (مطالعه موردی: استان همدان). مجله پژوهش آب ایران. ۲۶: ۹۱-۱۰۰.

۱- دانش‌آموخته کارشناسی ارشد مهندسی منابع آب، دانشکده کشاورزی، دانشگاه شهرکرد.

۲- استاد گروه مهندسی آب، دانشکده کشاورزی، دانشگاه شهرکرد.

۳- استادیار گروه مهندسی آب، دانشکده کشاورزی، دانشگاه شهرکرد.

۴- استادیار گروه مهندسی آب، دانشکده کشاورزی، دانشگاه بوعلی سینا همدان.

\* نویسنده مسئول: [hosseini\\_mostafa69@yahoo.com](mailto:hosseini_mostafa69@yahoo.com)

تاریخ پذیرش: ۱۳۹۵/۰۱/۳۰

تاریخ دریافت: ۱۳۹۴/۰۶/۲۶

## مقدمه

هیدرولوژیکی و به‌ویژه پیش‌بینی بارش (مایتی و همکاران، ۲۰۱۰؛ لین و همکاران، ۲۰۱۳ و یو و همکاران، ۲۰۱۰b) استفاده شده‌اند. یکی از روش‌های دیگر که به‌طور گسترده در پیش‌بینی مشخصه‌های غیرخطی و پیچیده به‌کاربرده می‌شود، شبکه عصبی مصنوعی است. مدل‌های هوش مصنوعی و شبکه عصبی را اولین بار ماکلوچ و پیتس (۱۹۴۳) مطرح کردند و در دهه‌های اخیر برای پیش‌بینی بارش (هونگ و همکاران، ۲۰۰۹ و یو و همکاران، ۲۰۱۰b، ۲۰۱۰a) مورد استفاده قرار گرفته است. گلابی و همکاران (۱۳۹۲) به مقایسه عملکرد الگوریتم‌های مختلف شبکه عصبی مصنوعی در مدل‌سازی بارندگی فصلی در استان خوزستان پرداختند. در این تحقیق از شبکه‌های MLP<sup>۶</sup> و RBF<sup>۷</sup> و با تغییراتی در لایه‌های میانی، نرون‌ها و الگوریتم‌های آموزش<sup>۸</sup> MOM،<sup>۹</sup> LM<sup>۱۰</sup> و CG<sup>۱۱</sup> برای پیش‌بینی بارش فصلی استفاده شد. نتایج نشان داد که شبکه RBF ضریب همبستگی بالاتری بین بارش مشاهده‌ای و شبیه‌سازی شده دارد. تیزی (۲۰۱۲) از مدل شبکه عصبی برای بازسازی داده‌های بارش ماهانه شهر ایسپارتا بر اساس داده‌های چهار ایستگاه مجاور استفاده کرد. نتایج حکایت از عملکرد مناسب این مدل در برآورد داده‌های بارش ماهانه داشت. کولیپالی و اورا (۲۰۰۷) از حالت‌های مختلف شبکه عصبی مصنوعی برای پر کردن خلاءهای آماری بارش و دمای روزانه در حوضه رودخانه گاتینیو در شمال شرقی کانادا استفاده کردند. نتایج نشان داد که شبکه MLP برای بازسازی داده‌های بارش، حداکثر و حداقل دما به‌خوبی عمل کرده است. خرسندی و همکاران (۲۰۱۱) مدل شبکه عصبی مصنوعی، نسبت نرمال<sup>۱۱</sup>، معکوس وزنی فاصله<sup>۱۲</sup> و مختصات جغرافیایی<sup>۱۳</sup> در بازسازی داده‌های بارش ماهانه استان اصفهان به‌کار گرفتند و نشان دادند که مدل شبکه عصبی مصنوعی عملکرد بهتری در بازسازی داده‌های گمشده دارد. هدف از انجام این پژوهش نیز مقایسه عملکرد شبکه عصبی مصنوعی (ANN) و مدل ماشین‌بردار پشتیبان (SVM) در بازسازی داده‌های ماهانه و فصلی بارش ایستگاه

کامل بودن سری‌های زمانی داده‌های هواشناسی، از اساسی‌ترین مسائل در مطالعات هیدرولوژی و زیست‌محیطی است؛ با این حال ایستگاه‌های هواشناسی اغلب به دلیل سوء نظارت بر قرائت داده‌ها یا مشکلات فنی، دارای خلاءهای آماری هستند (ناردیو و برتی، ۲۰۱۲). از دیرباز برآورد اطلاعات اقلیمی از دست رفته از جمله موضوع‌های مورد توجه هواشناسان، هیدرولوژیست‌ها و کارشناسان محیط‌زیست بوده است (لی و همکاران، ۲۰۱۳ و کاشانی و دین‌پژوه، ۲۰۱۲). داده‌های از دست رفته اغلب با استفاده از مدل‌ها و روش‌های مختلف نظیر استفاده از توابع آماری (راموس و همکاران، ۲۰۰۸)، به کار بردن روش معکوس وزنی فاصله (خسروی و همکاران، ۲۰۱۵؛ تیگوارپو و چاندرامولی، ۲۰۰۵) و استفاده از روش‌های درون‌یابی آماری مانند کریجینگ<sup>۱</sup> (جفری و همکاران، ۲۰۰۱) تخمین زده می‌شوند. به تازگی از مدل‌های هوش مصنوعی ماشین‌بردار پشتیبان<sup>۲</sup> (SVM) و شبکه عصبی مصنوعی<sup>۳</sup> (ANN) نیز برای بازسازی داده‌های بارش استفاده شده است. ماشین‌بردار پشتیبان، از روش‌های یادگیری تحت نظارت<sup>۴</sup> است که برای دسته‌بندی و رگرسیون قابل استفاده است. این روش را وپنیک (۱۹۹۸) بر پایه‌ی تئوری یادگیری آماری<sup>۵</sup> بنا نهاده است. ماشین‌بردار پشتیبان، روشی برای طبقه‌بندی دوتایی در فضای ویژگی‌های دلخواه است و از این‌رو روشی مناسب برای مسائل پیش‌بینی به شمار می‌رود. این روش در اصل یک دسته‌بندی‌کننده دو کلاسه است که کلاس‌ها را توسط یک مرز خطی از هم جدا می‌کند. در این روش نزدیک‌ترین نمونه‌ها به مرز تصمیم‌گیری را بردارهای پشتیبان می‌نامند. این بردارها معادله مرز تصمیم‌گیری را مشخص می‌کنند. الگوریتم‌های شبیه‌سازی هوشمند کلاسیک مانند شبکه‌های عصبی مصنوعی، معمولاً قدر مطلق خطا یا مجموع مربعات خطای داده‌های آموزشی را حداقل می‌کنند؛ ولی مدل‌های SVM، اصل حداقل سازی خطای ساختاری را به کار می‌گیرند (احمدی و همکاران، ۱۳۹۳). به تازگی این مدل‌ها در گستره‌ی وسیعی از مسائل

6- Multi Layer Perceptron

7- Radial Basis Function

8- Momentum

9- Lonberg-Marcorat

10- Conjugate Gradient

11- Normal ratio

12- Inverse distance weighting

13- Geographical coordinate

1- Kriging

2- Support Vector Machine

3- Artificial Neural Network

4- Supervised learning

5- Statistical Learning Theory

باران‌سنجی سرابی در استان همدان است.

درجه و ۳۴ دقیقه تا ۴۹ درجه و ۳۶ دقیقه شرقی واقع شده و دارای اقلیمی سرد است. در این مطالعه، از داده‌های بارش ماهانه ۴ ایستگاه آقاجان‌بلاغی، سرابی، آق‌کهریز و مریانج در خلال سال‌های ۱۳۷۰ تا ۱۳۸۹ استفاده شد. مشخصات این ایستگاه‌ها در جدول ۱ ارائه شده است.

## مواد و روش‌ها

### منطقه مورد مطالعه

استان همدان بین عرض جغرافیایی ۳۳ درجه و ۵۹ دقیقه تا ۳۵ درجه و ۴۵ دقیقه شمالی و طول جغرافیایی ۴۷

جدول ۱- موقعیت جغرافیایی ایستگاه‌های مورد مطالعه و میانگین بارش سالانه

ردیف	ایستگاه	طول جغرافیایی	عرض جغرافیایی	ارتفاع (متر)	متوسط بارش سالانه (میلی‌متر)
۱	سرابی	۴۸° ۱۰' ۵۵"	۳۴° ۵۸' ۱۷"	۱۹۲۵	۴۳۶
۲	آقاجان‌بلاغی	۴۸° ۰۳' ۰۰"	۳۴° ۵۰' ۴۴"	۱۸۰۲	۳۳۲
۳	آق‌کهریز	۴۸° ۲۰' ۰۰"	۳۴° ۵۹' ۳۹"	۱۸۴۴	۳۱۷
۴	مریانج	۴۸° ۲۷' ۲۸"	۳۴° ۴۹' ۴۰"	۱۸۴۱	۴۵۱

ویژگی بستگی دارد؛ به عبارت دیگر، باید با توجه به این پارامترها تابع کرنلی را انتخاب کرد که توانایی آموزش برای ورودی‌های مسأله را داشته باشد. در عمل چهار نوع کرنل خطی<sup>۲</sup> کرنل چند جمله‌ای<sup>۳</sup>، کرنل تانژانت هیپربولیک<sup>۴</sup> و کرنل گوسی<sup>۵</sup> (RBF) به کار گرفته می‌شوند. در جدول ۲ معادلات برخی از کرنل‌های رایج ارائه شده‌اند. در نهایت، تابع تصمیم رگرسیون بردار پشتیبان غیرخطی، به صورت معادله (۲) خواهد بود که کنترل‌کننده میزان نوسان تابع گوسی و همچنین کنترل‌کننده نتایج پیش‌بینی و تعمیم‌دهنده مدل SVM است (یو و همکاران، ۲۰۰۶).

$$f(x_i) = \sum_{i=1}^l (-\partial_i - \partial_i^*) K(x_i, x_j) + b \quad (2)$$

که در این رابطه  $\partial_i$  و  $\partial_i^*$  ضرایب لاگرانژ به منظور بهینه‌سازی و حداکثرسازی تابع،  $K(x_i, x_j)$  تابع کرنل موردنظر و  $b$  از ضرایب مدل SVM رگرسیونی می‌باشد. مشابه با مدل‌های ریاضی و آماری، مدل‌های SVM نیز معایبی دارند. وجود تعداد زیاد متغیر ورودی، اصلی‌ترین مشکل در توسعه این مدل‌هاست که ممکن است مانع یافتن مدل بهینه توسط SVM شود (نوری و همکاران، ۱۳۹۰). در این پژوهش از مدل SVM برای بازسازی

### ماشین بردار پشتیبان

در یک مدل رگرسیونی SVM لازم است وابستگی تابعی متغیر وابسته  $y$  به مجموعه‌ای از متغیرهای مستقل  $x$  تخمین زده شود. فرض بر این است که مانند دیگر مسائل رگرسیونی، رابطه‌ی بین متغیرهای وابسته و مستقل توسط یک تابع معین  $f$  به علاوه یک مقدار اضافی نویز<sup>۱</sup> مشخص شود که در رابطه (۱) نشان داده شده است.

$$y = f(x) + Noise \quad (1)$$

بنابراین موضوع اصلی، پیدا کردن فرم تابع  $f$  است که بتواند به صورت صحیح، موارد جدیدی را که SVM تاکنون تجربه نکرده است، پیش‌بینی کند. این تابع به وسیله آموزش مدل SVM روی یک مجموعه داده به عنوان مجموعه آموزش که شامل فرآیندی برای بهینه‌سازی دائمی تابع خطاست، قابل دستیابی می‌باشد. بر مبنای تعریف این تابع خطا، دو نمونه از مدل‌های SVM شناخته شده‌اند که عبارت‌اند از: الف) مدل‌های رگرسیونی SVM نوع اول که مدل‌های SVM-V نیز نامیده می‌شوند و ب) مدل‌های رگرسیونی SVM نوع دوم که با نام SVM-E شناخته شده هستند. در این مطالعه از مدل SVM-E به دلیل کاربرد گسترده آن در مسائل رگرسیونی استفاده شد. ماشین‌های بردار پشتیبان برای حل مسائل غیرخطی، ابعاد مسأله را از طریق توابع کرنل تغییر می‌دهند. انتخاب کرنل برای SVM به حجم داده‌های آموزشی و ابعاد بردار

2- Linear kernel

3- Polynomial kernel

4- Hyperbolic tangent kernel

5- Radial Base Function kernel

1- Noise

### الگوریتم پس انتشار<sup>۳</sup>

از پرکاربردترین مدل‌های شبکه عصبی، مدل پرسپترون چند لایه است که روی الگوریتم پس انتشار، پایه‌گذاری شده است. اساس آموزش الگوریتم پس انتشار به این صورت است که وزن‌های اولیه ارتباطی به صورت تصادفی انتخاب می‌شود. فرآیند آموزش با استفاده از دو گام زیر اتفاق می‌افتد:

۱- انتشار به جلو<sup>۴</sup>: بردار ورودی  $X_i$  به لایه ورودی فرستاده می‌شود و یک بردار خروجی  $O_j = (O_{1j}, \dots, O_{nj})$  بر اساس وزن‌های جاری ایجاد می‌شود. مقدار  $O_j$  با مقدار خروجی واقعی، یعنی  $B_j$  مقایسه و تابع خطای  $E$  (رابطه (۳)) تولید می‌شود که در این تابع سعی می‌شود تا مربع خطای بین خروجی‌های شبکه و تابع هدف کمینه شود. تضمینی برای این‌که این الگوریتم به کمینه مطلق برسد، وجود ندارد.

$$E = \frac{1}{2} \sum_{i=1}^l \sum_{j=1}^N (B_{ij} - O_{ij})^2 \quad (3)$$

۲- انتشار خطا به عقب<sup>۵</sup>: در این گام خطای به دست آمده از معادله ۳ به وسیله رابطه زیر روی وزن‌ها توزیع می‌شود.

$$\Delta W_{nl} = -\frac{\partial E}{\partial W_{nl}} h \quad (4)$$

$\eta$  ضریب پارامتری است که همگرایی نرخ الگوریتم را کنترل می‌کند. این دو مرحله آن قدر ادامه می‌یابد تا مقدار  $E$  به مقدار کوچک از پیش تعیین شده‌ای همگرا شود که این مقدار  $0.001$ ، و با توجه به این‌که رسیدن به این خطای مطلق، ممکن است انجام نپذیرد، مقدار  $1000$  تکرار برای آموزش شبکه در نظر گرفته شده است. در حقیقت شبکه با توزیع خطای به دست آمده (در هر مرحله) روی وزن‌های قبلی و تکرار این عملیات آموزش می‌بیند.

### داده‌های ورودی و خروجی شبکه

در این مطالعه از داده‌های ۴ ایستگاه باران‌سنجی سرابی، آقاجانبلاغی، مریانج و آق کهریز استفاده شده است. دلیل انتخاب این ایستگاه‌ها نزدیک بودن مسافت هندسی و بالا بودن ضریب تبیین داده‌های بارش ماهانه بین آن‌ها است.

داده‌های گمشده توسط نرم‌افزار Statistica10 و از تابع کرنل غیرخطی RBF با بیش از ۱۰۰ تکرار استفاده شد. تابع کرنل به دلیل این‌که دقت بالاتری در بازسازی داده‌های ناقص بارش داشت، به کار گرفته شد (احمدی و همکاران، ۱۳۹۳). مقادیر بهینه مشخصه‌های مدل SVM شامل  $\epsilon$ ،  $C$  و  $\gamma$  باید تعیین شوند. بدین منظور مشخصه‌های  $\epsilon$  و  $C$  توسط الگوریتم بهینه‌سازی جست‌وجو شبکه<sup>۱</sup> و متغیر  $\gamma$  با آزمون و خطا محاسبه شد. ابتدا شبکه‌هایی با ابعاد بزرگ محدوده مشخصه‌های  $\epsilon$  و  $C$  به ازای مقدار ثابت  $\gamma$  تعیین، سپس این محدوده به ابعاد کوچک‌تر تقسیم می‌شود تا مقادیر دقیق‌تر این متغیرها به دست آیند. این فرآیند برای  $\gamma$  مختلف اجرا و مدل‌های زیادی به ازای تغییر در مقدار  $\gamma$  ایجاد می‌شود که از میان آن‌ها مدلی که دارای کمترین خطا باشد، انتخاب و مشخصه‌های آن به عنوان مقادیر بهینه انتخاب می‌شود.

جدول ۲- برخی توابع کرنل رایج در ماشین‌های بردار پشتیبان (همل، ۲۰۰۹)

نوع تابع	تابع کرنل
خطی	$K(x_i, x_j) = x_i^T \cdot x_j$
چند جمله‌ای	$K(x_i, x_j) = (\gamma x_i^T \cdot x_j + C)^d$
تانژانت هیپربولیک	$K(x_i, x_j) = \tanh(\gamma x_i^T \cdot x_j + C)$
RBF	$K(x_i, x_j) = \exp\left(-\gamma  x_i - x_j ^2\right)$

### شبکه عصبی مصنوعی

یک شبکه عصبی، از نرون‌های مصنوعی تشکیل شده است. نرون یا گره کوچک‌ترین واحد پردازش اطلاعات است که اساس عملکرد شبکه‌های عصبی را تشکیل می‌دهد (منهاج، ۱۳۸۴). هر یک از نرون‌ها، ورودی‌ها را دریافت و پس از پردازش روی آن‌ها، یک سیگنال خروجی تولید می‌کنند. بنابراین هر نرون در شبکه به عنوان مرکز پردازش و توزیع اطلاعات عمل می‌کند. در این مطالعه از شبکه عصبی مصنوعی چند لایه پرسپترون (MLP<sup>۲</sup>) تحت الگوریتم پس انتشار استفاده است.

3- Back Propagation  
4- Forward Propagation  
5- Error back propagation

1- Grid Search  
2- Multi layer perceptron

### نتایج و بحث

در این پژوهش برای مدل‌سازی بارش ماهانه و فصلی ایستگاه باران‌سنجی سرابی با استفاده از مدل‌های شبکه عصبی مصنوعی و ماشین بردار پشتیبان، داده‌های چهارده سال (۷۰-۷۳ و ۸۰-۸۹) برای آموزش و پنج سال (از سال ۷۴ تا سال ۷۹) به‌عنوان داده‌های تست انتخاب شدند. به‌طور کلی، برای آموزش مدل‌ها از ۷۵٪ داده‌ها و برای صحت‌سنجی از ۲۵٪ داده‌ها استفاده شد. برای آموزش مدل‌ها سه حالت در نظر گرفته شده است؛ به‌طوری که برای آموزش مدل‌ها در حالت اول تنها داده‌های یک ایستگاه (آقاجانبلای)، در حالت دوم داده‌های دو ایستگاه (آقاجانبلای و آق‌کهریز) و برای حالت سوم آموزش از داده‌های سه ایستگاه (آقاجانبلای، آق‌کهریز و مریانج) برای بازسازی ایستگاه سرابی استفاده شد. پس از آن داده‌های مدل‌سازی شده با داده‌های مشاهده‌ای مورد مقایسه قرار گرفتند. مقادیر بهینه مشخصه‌های مدل SVM شامل E، C و  $\gamma$  در بهترین حالت (۳) بازسازی بارش ماهانه و فصلی به ترتیب ۰/۱، ۱۰ و ۰/۲۵، ۰/۱ و ۱۰/۲۷۵ و برآورد شد. در بازسازی داده‌های بارش ماهانه ایستگاه سرابی توسط مدل ANN از داده‌های سه ایستگاه مجاور توسط شبکه MLP و الگوریتم آموزش پس انتشار با یک و دو لایه مخفی و با نرون‌های متفاوت آموزش شبکه انجام شد. بدین‌منظور از شاخص‌های آماری میانگین مربعات خطا (RMSE) و ضریب تبیین ( $R^2$ ) استفاده شد. جدول ۴ دقت مدل‌سازی داده‌های بارش ماهانه ایستگاه سرابی با استفاده از مدل‌های شبکه عصبی مصنوعی و ماشین بردار پشتیبان را برای هر سه حالت انجام شده نشان می‌دهد. با توجه به این جدول زمانی که از سه ایستگاه برای آموزش مدل استفاده می‌شود، شبکه عصبی و ماشین بردار پشتیبان دقت بالاتری را از خود نشان می‌دهد. علاوه بر این، شبکه عصبی با اختلاف کمی عملکرد بهتری نسبت به ماشین بردار پشتیبان دارد. اگرچه هر دو روش تقریباً عملکرد مشابهی دارند، زمانی که ورودی به مدل‌ها داده‌های دو ایستگاه بود، ماشین بردار پشتیبان عملکرد بهتری نسبت به شبکه عصبی مصنوعی نشان داد. بنابراین در مناطق و حوضه‌هایی که تعداد ایستگاه‌ها کم باشد، می‌توان از مدل SVM برای پیش‌بینی داده‌های بارش استفاده کرد. همچنین جدول ۵ عملکرد این مدل‌های در بازسازی داده‌های فصلی بارش را نشان

ضریب تبیین میان ایستگاه‌ها در جدول ۳ نشان داده شده است. برای تعیین بهترین الگوی ورودی به شبکه، عوامل مختلفی را که ممکن است در پدیده مؤثر باشند، باید در نظر گرفت. همچنین با توجه به عوامل تأثیرگذار در بازسازی داده‌های بارش ماهانه و فصلی ایستگاه مورد نظر، لایه ورودی شبکه در سه حالت استفاده از داده‌های یک ایستگاه، دو ایستگاه و سه ایستگاه بررسی شده است. بنابراین نرون‌های ورودی برای این حالت‌ها به ترتیب ۱، ۲ و ۳ است. همچنین داده‌های خروجی از شبکه شامل داده‌های ناقص یا گمشده بارش ماهانه در ایستگاه هدف است. برای بازسازی داده‌های بارش در شبکه عصبی از کد نویسی در محیط نرم‌افزار Matlab استفاده شده است.

جدول ۳- ضریب همبستگی میان داده‌های بارش ایستگاه‌های مورد مطالعه

سرابی	مریانج	آق‌کهریز	آقاجانبلای	ایستگاه
۰/۸۸	۰/۸۸	۰/۸۸	۱	آقاجانبلای
۰/۹۰	۰/۹۰	۱	۰/۸۸	آق‌کهریز
۰/۹۳	۱	۰/۹۰	۰/۸۸	مریانج
۱	۰/۹۳	۰/۹۰	۰/۸۸	سرابی

### معیارهای ارزیابی مدل‌ها

برای مقایسه بارش ماهانه مشاهده‌ای و تخمین زده شده توسط مدل‌های ماشین بردار پشتیبان (SVM) و شبکه عصبی مصنوعی (ANN)، از معیارهای ضریب تبیین ( $R^2$ ) و ریشه میانگین مربعات خطا<sup>۱</sup> (RMSE) استفاده شد. این معادلات در روابط (۵) و (۶) نشان داده شده است.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (R_{(pre)i} - R_{(obs)i})^2} \quad (5)$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (R_{(pre)i} - R_{(obs)i})^2}{\sum_{i=1}^n (R_{(obs)i} - R_m)^2} \quad (6)$$

در این روابط  $R_{(pre)i}$  بارش تخمینی در زمان  $i$  و  $R_{(obs)i}$  بارش مشاهده شده در همان زمان و  $R_m$  میانگین مقادیر مشاهداتی هستند. هرچه مقدار ضریب تبیین به ۱ نزدیک‌تر و مقدار RMSE کمتر باشد، نتایج دقیق‌تر خواهد بود.

به‌عنوان داده‌های ورودی به مدل‌ها، بیشتر شده، مقادیر بازسازی شده بارش دقیق‌تر است. نتایج جدول‌های ۴ و ۵ نشان می‌دهد این مدل‌ها در بازسازی داده‌های بارش ماهانه نسبت به بارش فصلی عملکرد بالاتری را دارند.

می‌دهد. همان‌طور که نتایج این جدول نشان می‌دهد در هر سه حالت بازسازی داده‌های فصلی بارش مدل SVM عملکرد بهتر و خطای کمتری نسبت به مدل ANN دارد. به‌طور کلی برای هر دو مدل شبکه عصبی و ماشین‌بردار پشتیبان هرچه تعداد ایستگاه‌های مجاور ایستگاه مبنا،

جدول ۴- ارزیابی عملکرد ترکیب‌های مختلف مدل‌های SVM و ANN در بازسازی داده‌های بارش ماهانه ایستگاه سرابی

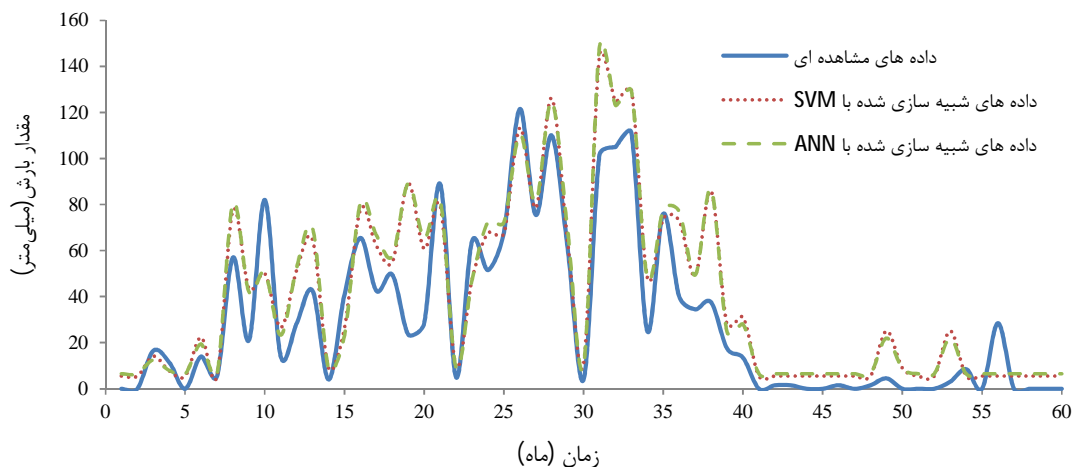
SVM		ANN		ایستگاه هدف	ایستگاه مبنا
R <sup>2</sup>	RMSE (mm)	R <sup>2</sup>	RMSE (mm)		
۰/۸۳۵	۱۸/۱	۰/۸۳	۱۸/۷۹	سرابی	آقاجانبلاغی
۰/۸۵	۱۴/۴۳	۰/۸۴	۱۵/۵۵	سرابی	آق کهریز و آقاجانبلاغی
۰/۸۷	۱۲/۸۹	۰/۸۸	۱۲/۳۳	سرابی	آقاجانبلاغی، آق کهریز و مریانج

جدول ۵- ارزیابی عملکرد ترکیب‌های مختلف مدل‌های SVM و ANN در بازسازی داده‌های بارش فصلی ایستگاه سرابی

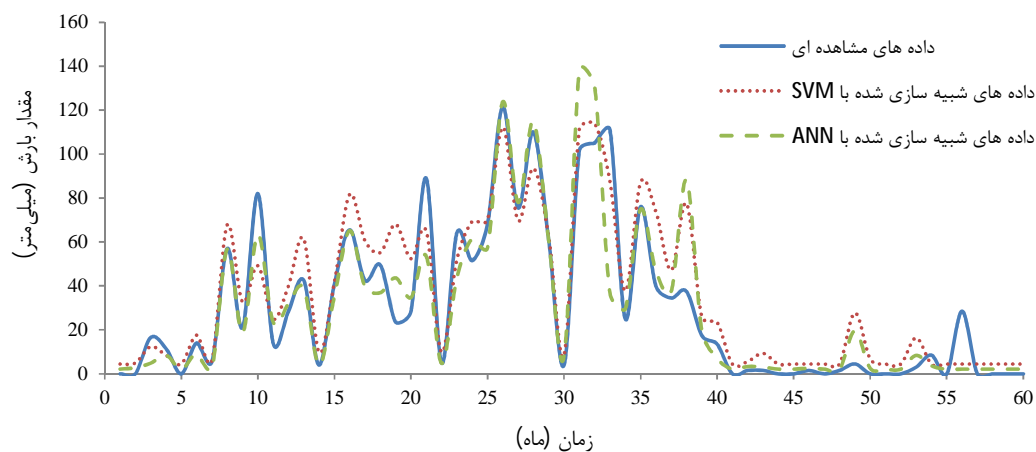
SVM		ANN		ایستگاه هدف	ایستگاه مبنا
R <sup>2</sup>	RMSE (mm)	R <sup>2</sup>	RMSE (mm)		
۰/۸۳	۳۷/۷	۰/۸۴	۴۴/۹	سرابی	آقاجانبلاغی
۰/۸۹	۲۵/۴۷	۰/۹۲	۲۹/۵	سرابی	آق کهریز و آقاجانبلاغی
۰/۹۴	۱۸/۸۲	۰/۹۵	۲۴/۶۲	سرابی	آقاجانبلاغی، آق کهریز و مریانج

بارش ماهانه ایستگاه سرابی با مدل‌های SVM و ANN در مقابل داده‌های مشاهده‌ای را زمانی که از داده‌های سه ایستگاه (آقاجانبلاغی، آق کهریز و مریانج) برای آموزش مدل‌ها استفاده شده است، را نشان می‌دهد. همان‌طور که در این شکل مشاهده می‌شود، حالت سوم نسبت به دو حالت قبل عملکرد بهتری در تخمین داده‌های گمشده دارد و نشان می‌دهد که مدل شبکه عصبی مصنوعی در بازسازی داده‌ها کارایی بهتری دارد. شکل‌های ۴ تا ۶ مقایسه داده‌های شبیه‌سازی شده با مدل‌ها در مقابل داده‌های مشاهده‌ای در بازسازی داده‌های بارش فصلی در هر سه حالت آموزش مدل‌ها را نشان می‌دهد. همان‌طور که این شکل‌ها به‌ترتیب نشان می‌دهند با افزایش تعداد ایستگاه‌های ورودی برای آموزش، عملکرد مدل‌ها افزایش یافته و داده‌های شبیه‌سازی و مشاهده‌ای به هم نزدیک‌تر و مقدار خطا کمتر شده است. همچنین مدل SVM نسبت به مدل ANN عملکرد بهتری را از خود نشان داده است.

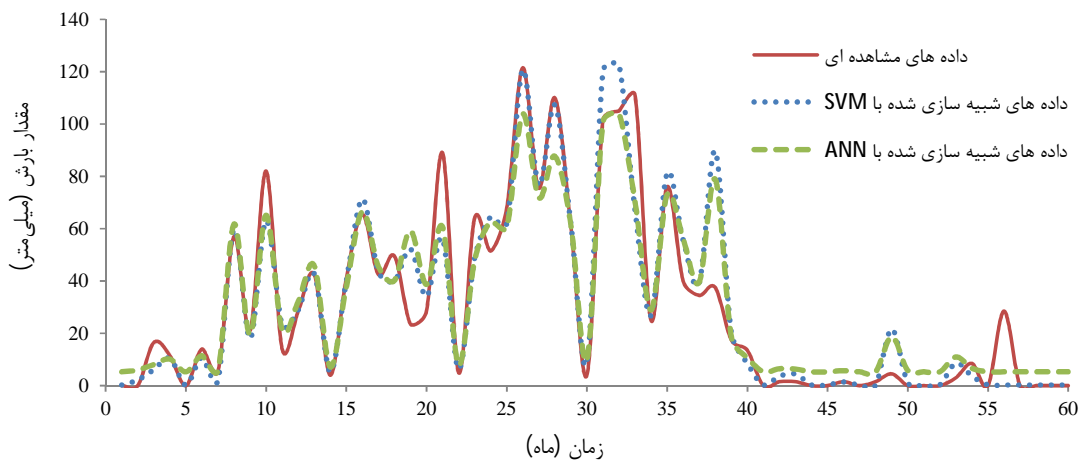
برای ارزیابی و بررسی عملکرد مدل‌های مورد بررسی مقادیر مشاهده شده در مقابل مقادیر پیش‌بینی شده توسط مدل‌ها در شکل‌های ۱ تا ۶ ترسیم شد. در شکل ۱ نتایج شبیه‌سازی داده‌های بارش با ANN و SVM در مقابل داده‌های مشاهده‌ای زمانی که از داده‌های یک ایستگاه (آقاجانبلاغی) برای آموزش مدل‌ها استفاده شده، ارائه گردیده است. این شکل نشان می‌دهد که دو مدل، عملکرد بسیار نزدیکی دارند. شکل ۲ نتایج شبیه‌سازی داده‌های بارش با مدل‌های ANN و SVM در مقابل داده‌های مشاهده‌ای را زمانی که داده‌های دو ایستگاه (آقاجانبلاغی و آق کهریز) برای آموزش مدل‌ها به‌کار گرفته شده نشان می‌دهد. همان‌طور که در این شکل مشاهده می‌شود، با اضافه شدن داده‌های یک ایستگاه دیگر به آموزش مدل‌ها، میزان خطای آن‌ها به میزان قابل توجهی کاهش یافته است. همان‌طور که شکل‌های ۱ و ۲ نشان می‌دهند مدل SVM برآورد دقیق‌تری نسبت به مدل ANN داشته است. شکل ۳ نتایج شبیه‌سازی داده‌های



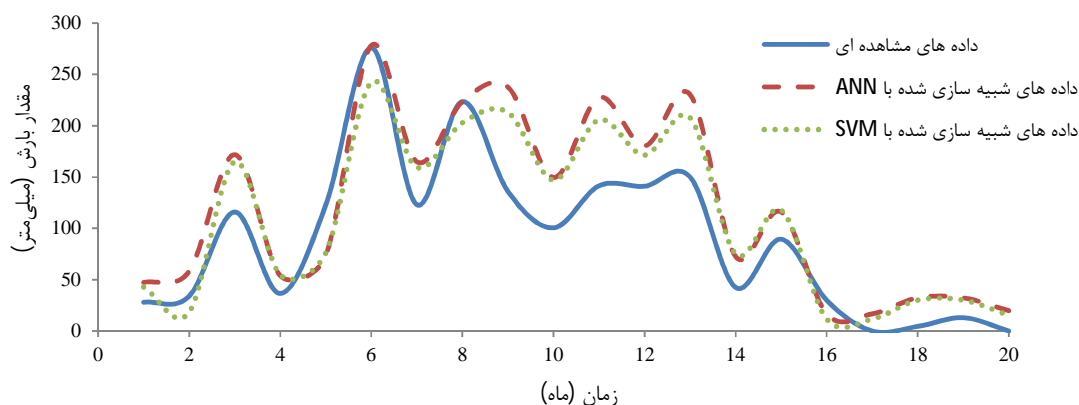
شکل ۱- مقایسه داده‌های بارش ماهانه مشاهده‌ای و تخمینی با شبکه عصبی مصنوعی و ماشین بردار پشتیبان در ایستگاه سرابی (ایستگاه مینا: آقاجان‌بلاغی)



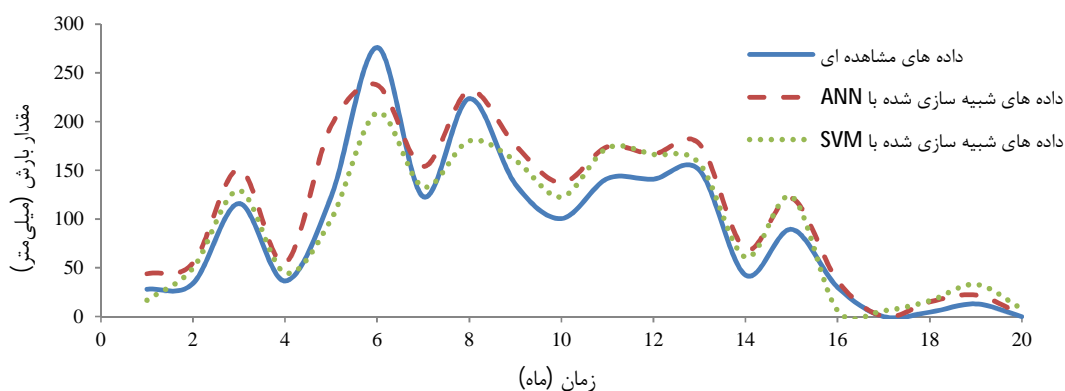
شکل ۲- مقایسه داده‌های بارش ماهانه مشاهده‌ای و تخمینی با شبکه عصبی مصنوعی و ماشین بردار پشتیبان در ایستگاه سرابی (ایستگاه‌های مینا: آقاجان‌بلاغی و آق کهریز)



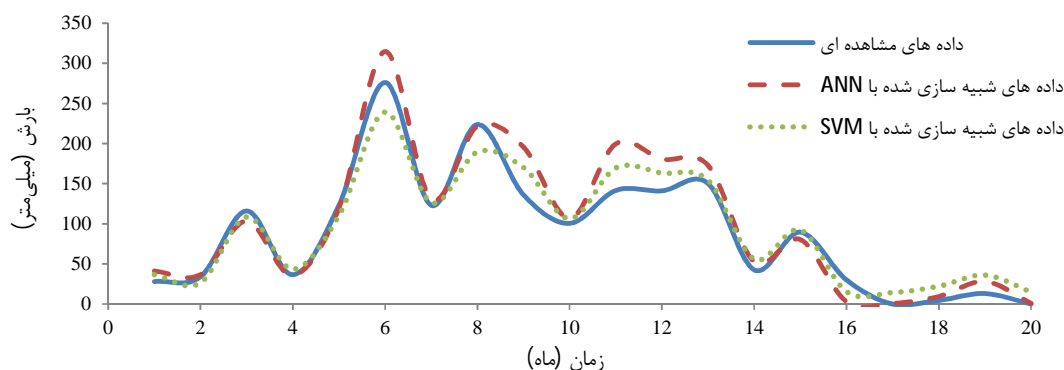
شکل ۳- مقایسه داده‌های بارش ماهانه مشاهده‌ای و تخمینی با شبکه عصبی مصنوعی و ماشین بردار پشتیبان در ایستگاه سرابی (ایستگاه‌های مینا: آقاجان‌بلاغی، آق کهریز و مریانج)



شکل ۴- مقایسه داده‌های بارش فصلی مشاهده‌ای و تخمینی با شبکه عصبی مصنوعی و ماشین بردار پشتیبان در ایستگاه سرابی (ایستگاه مبنا: آقاجان بلاغی)



شکل ۵- مقایسه داده‌های بارش فصلی مشاهده‌ای و تخمینی با شبکه عصبی مصنوعی و ماشین بردار پشتیبان در ایستگاه سرابی (ایستگاه‌های مبنا: آقاجان بلاغی و آق کهریز)



شکل ۶- مقایسه داده‌های بارش فصلی مشاهده‌ای و تخمینی با شبکه عصبی مصنوعی و ماشین بردار پشتیبان در ایستگاه سرابی (ایستگاه‌های مبنا: آقاجان بلاغی، آق کهریز و مریانج)

### نتیجه‌گیری

در این پژوهش عملکرد مدل‌های شبکه عصبی مصنوعی MLP و ماشین بردار پشتیبان RBF در بازسازی مقادیر بارش ماهانه ایستگاه باران‌سنجی سرابی در استان همدان در خلال سال‌های ۱۳۷۰ تا ۱۳۸۹ بررسی شد. در بخش

آموزش شبکه، سه رویکرد در نظر گرفته شد. در حالت اول شبکه به کمک داده‌های ایستگاه آقاجان بلاغی که یکی از ایستگاه‌های مجاور ایستگاه سرابی است، آموزش داده شد. در حالت دوم شبکه به کمک داده‌های ایستگاه‌های آقاجان بلاغی و آق کهریز و در حالت سوم به کمک داده‌های



با مطالعه خرسندی و همکاران (۲۰۱۱) است که نشان دادند مدل ANN در بازسازی داده‌های گمشده سه ایستگاه در مناطق مختلف کشور نسبت به روش‌های نسبت نرمال، مختصات جغرافیایی و معکوس وزنی فاصله، عملکرد بهتری دارند.

• نتایج نشان داد هرچه تعداد ایستگاه‌های ورودی برای آموزش مدل‌ها بیشتر شده، عملکرد مدل‌ها در بازسازی داده‌های بارش ماهانه و فصلی افزایش یافته است. همچنین دقت این مدل‌ها در بازسازی داده‌های بارش ماهانه بهتر از بازسازی داده‌های بارش فصلی بوده است.

### منابع

۱. احمدی ف. رادمنش ف. و میرعباسی نجف‌آبادی ر. ۱۳۹۳. مقایسه روش‌های برنامه‌ریزی ژنتیک و ماشین بردار پشتیبان در پیش‌بینی جریان روزانه رودخانه (مطالعه موردی: رودخانه باراندوزچای). نشریه آب و خاک. ۲۸(۶): ۱۰۱۰-۱۰۲۵.
۲. گلایی م. ر. آخوندعلی ع. م. و رادمنش ف. ۱۳۹۲. مقایسه عملکرد الگوریتم‌های مختلف شبکه عصبی مصنوعی در مدل‌سازی بارندگی فصلی مطالعه موردی: ایستگاه‌های منتخب استان خوزستان. تحقیقات کاربردی علوم جغرافیایی. ۳۰: ۱۵۱-۱۶۹.
۳. منهای م. ۱۳۸۴. مبانی شبکه‌های عصبی. دانشگاه صنعتی امیرکبیر، تهران. ۷۱۶ ص.
۴. نوری ر. خاکپور ا. دهقانی م. و فرخ‌نیا ا. ۱۳۹۰. پیش‌بینی ماهانه جریان با استفاده از ماشین بردار پشتیبان بر مبنای آنالیز مؤلفه‌های اصلی. نشریه آب و فاضلاب. ۱: ۱۱۸-۱۲۳.
5. Coulibaly P. and Evora N. D. 2007. Comparison of neural network methods for infilling missing daily weather records. Journal of Hydrology. 341: 27-41.
6. Hamel L. 2009. Knowledge Discovery with Support Vector Machines. Wiley-Interscience New York. 264 p.
7. Hung N. Q. Babel M. S. Weesakul S. and Tripathi N. K. 2009. An artificial neural network model for rainfall forecasting in Bangkok. Thailand, Hydrology and Earth System Sciences. 13: 1413-1425.
8. Jeffrey S. J. Carter J. O. Moodie K. B. and Beswick A. R. 2001. Using spatial interpolation to construct a comprehensive archive of

سه ایستگاه آقاجانبلاغی و آق‌کهریز و مریانج آموزش داده و سپس داده‌های ایستگاه سرابی به کمک مدل‌های مذکور بازسازی شد. نتایج این مطالعه به صورت زیر ارائه شده است:

- هنگامی که از داده‌های یک ایستگاه برای آموزش مدل‌ها استفاده شد، مدل‌ها عملکرد بسیار نزدیک و تقریباً مشابه را نشان دادند. به‌طور کلی هنگامی که از داده‌های یک و دو ایستگاه برای آموزش در مدل‌سازی بارش ماهانه استفاده شد، مدل ماشین بردار پشتیبان عملکرد بهتری از خود نشان داده است و زمانی که از داده‌های سه ایستگاه برای ورودی به مدل‌ها استفاده شد، مدل شبکه عصبی مصنوعی با اختلاف کمی نسبت به مدل بردار پشتیبان عملکرد بهتری دارد که این نشان می‌دهد چنانچه ایستگاه‌های کافی و مورد قبول برای آموزش مدل‌ها در دسترس باشد، شبکه عصبی مصنوعی عملکرد بهتری دارد و زمانی که تعداد ایستگاه‌های موجود کم باشد، ماشین بردار پشتیبان بهتر عمل می‌کند.
- در بازسازی داده‌های بارش ماهانه، بهترین عملکرد در استفاده از داده‌های سه ایستگاه برای آموزش مدل‌های ANN و SVM به دست آمده که ضریب تبیین و میانگین مربعات خطا برای این الگو به ترتیب ۰/۸۸ و ۱۲/۳۳ میلی‌متر برای مدل ANN و ۰/۸۷ و ۱۲/۸۹ میلی‌متر برای مدل SVM است. این نتیجه با نتایج مطالعه ترزی (۲۰۱۲) که نشان داد، زمانی که از داده‌های ماهانه بارش سه ایستگاه برای آموزش مدل شبکه عصبی استفاده شود نتایج بهتری به دست آمده، مطابقت دارد.
- در بازسازی داده‌های بارش فصلی، بهترین عملکرد در استفاده از داده‌های سه ایستگاه برای آموزش مدل‌های ANN و SVM به دست آمده که ضریب تبیین و میانگین مربعات خطا برای این الگو به ترتیب ۰/۹۵ و ۲۶/۶۲ میلی‌متر برای مدل ANN و ۰/۹۴ و ۱۸/۸۲ میلی‌متر برای مدل SVM است. همچنین در هر سه رویکرد مدل SVM عملکرد بهتری نسبت به مدل ANN را نشان داده است.
- به‌طور کلی باید گفت که مدل‌های ANN و SVM عملکرد مناسبی در بازسازی داده‌های بارش ایستگاه‌های باران‌سنجی دارد که این نتیجه در تطابق

- preprocessing techniques. *Journal of Hydrology*. 389: 146-167.
22. Wu J. Liu M. and Jin L. 2010b. Least square support vector machine ensemble for daily rainfall forecasting based on linear and nonlinear regression, *Advances in Neural Network Research and Applications. Lecture Notes in Electrical Engineering*. 67(1): 55-64.
23. Yu P. S. Chen S. T. and Chang I. F. 2006. Support vector regression for real-time flood stage forecasting. *Journal of Hydrology*. 328: 704-716.
- Australian climate data. *Environmental Modelling & Software*. 16(4): 309-330.
9. Kashani M. H. and Dinpajouh Y. 2012. Evaluation of efficiency of different estimation methods for missing climatological data. *Stochastic Environmental Research and Risk Assessment*. 26: 59-71.
10. Khorsandi Z. Mahdavi M. Salajeghe A. and Eslamian S. S. 2011. Neural network application for monthly precipitation data reconstruction. *Journal of Environmental Hydrology*. 19: 1-12.
11. Khosravi G. Nafarzadegan A. R. Nohegar A. Fathizade H. and Malekian A. 2015. A modified distance-weighted approach for filling annual precipitation gaps: application to different climates of Iran. *Theoretical and Applied Climatology*. 119: 33-42.
12. Li X. Li L. Wang X. and Jiang F. 2013. Reconstruction of hydrometeorological time series and its uncertainties for the Kaidu River Basin using multiple data sources. *Theoretical and Applied Climatology*. 113: 45-62.
13. Lin G. F. Chou Y. C. and Wu M. C. 2013. Typhoon flood forecasting using integrated two-stage support vector machine approach. *Journal of Hydrology*. 486: 334-342.
14. Maity R. Bhagwat P. P. and Bhatnagar A. 2010. Potential of support vector regression for prediction of monthly streamflow using endogenous property. *Hydrological Processes*. 24(7): 917-923.
15. McCulloch W. S. and Pitts W. 1943. A logic calculus of the ideas immanent in nervous activity. *Bull of Math Biophys*. 5: 115-133.
16. Ramos-Calzado P. Gomez-Camacho J. Perez-Bernal F. and Pita-Lopez M. F. 2008. A novel approach to precipitation series completion in climatological datasets: application to Andalusia. *International Journal of Climatology*. 28(11): 1525-1534.
17. Tardivo G. and Berti A. 2012. A dynamic method for gap filling in daily temperature datasets. *Journal of American Meteorological Society*. 51: 1079-1086.
18. Teegavarapua R. S. V. and Chandramouli V. 2005. Improved weighting methods, deterministic and stochastic data-driven models for estimation of missing precipitation records. *Journal of Hydrology*. 312: 191-206.
19. Terzi O. 2012. Monthly Rainfall Estimation Using Data-Mining Process. *Journal Applied Computational Intelligence and Soft Computing*. 20: 1-6.
20. Vapnik V. N. 1998. *Statistical Learning Theory*. Wiley, New York. 768 p.
21. Wu C. L. Chau K. W. and Fan C. 2010a. Prediction of rainfall time series using modular artificial neural networks coupled with data-