

پیش بینی فصلی خشکسالی هواشناسی با استفاده از ماشین های بردار پشتیبان

محسن ناصری^۱

بنفشه زهرائی^۲

علیرضا نیکبخت شهبازی^۱

(دریافت ۸۹/۱/۱۵ پذیرش ۹۰/۶/۶)

چکیده

در تحقیقات مختلف، پارامترهای هواشناسی متفاوتی در پیش بینی دوره های کم بارش مورد توجه قرار گرفته اند. در این تحقیق نمایه بارش استاندارد شده (SPI) برای ۶ سناریوی فصل (پاییز، زمستان، بهار، پاییز+ زمستان، زمستان+ بهار و پاییز تا بهار) محاسبه شده و متغیرهای هواشناسی پیش بینی کننده دمای هوا (در سطح ۳۰۰، ۵۰۰، ۷۰۰ و ۸۵۰ میلی بار) و ارتفاع ژئوپتانسیل (در سطح ۳۰۰، ۵۰۰، ۷۰۰ و ۸۵۰ میلی بار) در محدوده طول و عرض جغرافیایی ۰ تا ۶۰ درجه شمالی و ۰ تا ۹۰ درجه شرقی، در سالهای (۱۳۵۴-۱۳۸۶) برای پیش بینی پدیده خشکسالی هواشناسی مورد استفاده قرار گرفت. در این مدل پیش بینی، بازه زمانی پیش بینی کننده بین ماههای اکتبر تا آوریل برای SPI پیش بینی شده در همان بازه زمانی قرار دارد. نمایه بارش استاندارد شده در حوضه های مورد مطالعه (حوضه سدهای طالقان و ماملو) بر اساس بارش متوسط حوضه ها که به روش میانگین معکوس فاصله وزندار محاسبه شده، تخمین زده شده است. یکی از روشهای یادگیری آماری با استفاده از ناظر به نام ماشین بردار پشتیبان (SVM) برای تدوین مدل پیش بینی SPI استفاده شد. با استفاده از تکنیک آماری مبتنی بر آنتروپی مشترک اطلاعات، نقاط مؤثر بر بارش حوضه سدهای تهران در فصل بهار بیشتر در جنوب، جنوب غربی و شمال غربی کشور و در فصل پائیز، شمال، شمال غربی و جنوب و در زمستان در شمال غربی و غرب کشور تشخیص داده شدند. نتایج مدل SVM در اکثر موارد پیش بینی، دقت مناسب داشت. این روش می تواند در پیش بینی رفتارهای غیرخطی داده های هواشناسی با طول دوره آماری کوتاه مورد استفاده قرار گیرد. این دقت برای دسته بندی SPI فصلهای پاییز و بهار بیشتر از سایر سناریوها است.

واژه های کلیدی: خشکسالی هواشناسی، ماشین بردار پشتیبان، نمایه بارش استاندارد شده، استان تهران

Seasonal Meteorological Drought Prediction Using Support Vector Machine

AliReza Nikbakht Shahbazi¹

Banafsheh Zahraie²

Mohsen Nasseri³

(Received Apr. 3, 2010 Accepted Aug. 27, 2011)

Abstract

In various researches, implementation of meteorological parameters in drought prediction is studied. In the current work, meteorological drought classes based on Standardized Precipitation Index (SPI) for six seasonal scenarios (autumn, winter, spring, autumn + winter, winter + spring, and autumn + winter + spring) and meteorological predictors contained ground and sea surface temperature, weather temperature (at 300, 500, 700 and 850 mi bar) and geopotential height (at 300, 500, 700 and 850 mi bar) wide of North (0, 60) and East (0, 90) was applied in prediction models based on data from 1975 to 2005. In these models, temporal range of meteorological predictors is between October to April month on the same predicted SPI. SPI was calculated based on mean precipitation at seasonal time scale in the main watershed of Tehran (Taleghan, Mamloo) by

1. Ph.D. of Hydrology and Water Resources, Dept. of Water Sciences and Eng., Sciences and Research Branch, Islamic Azad University, Tehran, Iran (Corresponding Author) (+98 21) 55380865 nikhakhta@gmail.com
2. Assoc. Prof., of Civil Eng., Tehran University, Tehran
3. Ph.D. Candidate of Water Eng., Dept. of Civil Eng., Tehran University, Tehran

- ۱- دکترای هیدرولوژی و منابع آب، دانشگاه آزاد اسلامی، واحد علوم و تحقیقات، گروه علوم و مهندسی آب، تهران، ایران (نویسنده مسئول) ۵۵۳۸۰۸۶۵ (+۹۸ ۲۱) nikhakhta@gmail.com
- ۲- دانشیار دانشکده مهندسی عمران، پردیس دانشکده های فنی، دانشگاه تهران
- ۳- دانشجوی دکترای مهندسی آب، دانشکده مهندسی عمران، دانشگاه تهران

Inverse Weighted Distance method. The well known statistical supervised machine learning method, support vector machine (SVM), is applied to predict SPI. Regarding to selected data points, the effective regions on Tehran precipitation are southern, southwestern and northwestern of Iran in spring, northern and northwestern in autumn and northwestern and western in winter. SVM depicted accurate results in prediction of SPI, spatially prediction of SPI in all scenarios, and it can be proposed as a very suitable statistical learning method in investigating of nonlinear behavior of meteorological phenomena with a short samples. The predicted SPI in spring and autumn are more accurate than the other scenarios.

Keywords: Climatological Drought, Support Vector, Standardized Precipitation Index (SPI), Tehran Province.

۱- مقدمه

قدرتمند و مهم معرفی گردید از پیش قدر یادگیری^{۱۰} که از تئوری یادگیری آماری استخراج شده استفاده می‌کند به گونه‌ای که از آغاز معرفی این روش، از آن در کاربردهای مختلفی استفاده شده است [۱۰].

تاکنون در زمینه‌های مختلف هیدرولوژی از روش SVM استفاده شده است که برخی از موارد آن در ادامه آورده شده است. لیونگ و همکاران^{۱۱} از SVM برای پیش‌بینی سیلاب استفاده نمودند [۱۱]. کوی^{۱۲} و چان^{۱۳} از SVM برای تعیین ساختار تابع شعاعی در شبکه‌ها استفاده کردند. آنها از این شبکه برای مدل‌سازی بین بارش و رواناب استفاده نمودند [۱۲]. یو و همکاران^{۱۴} روشی را برای تلفیق تئوری آشفستگی و SVM برای پیش‌بینی رواناب روزانه پیشنهاد نمودند [۱۳]. بری^{۱۵} و هان^{۱۶} از روش SVM برای پیش‌بینی رواناب با تأکید بر انتخاب و شناسایی ساختار مدل مناسب و پارامترهای مربوطه استفاده نمودند [۱۴]. دیبیک و همکاران^{۱۷} قابلیت‌های SVM را در خصوص پیش‌بینی‌های هیدرولوژیکی نشان داده‌اند. آنها با استفاده از این روش در زمینه دسته‌بندی داده‌های سنجش از دور و مدل‌سازی بارندگی و رواناب و مقایسه آن با روش‌های شبکه عصبی مصنوعی به نتایج خوبی برای پیش‌بینی رسیده‌اند [۱۵]. تریپاتی و همکاران^{۱۸} از تکنیک SVM برای برآورد آماری بارندگی ماهانه در هند استفاده کردند. نتایج نشان داده است که مدل SVM به‌عنوان گزینه برتر نسبت به سایر روش‌های معمول جهت پیش‌بینی بارش به روش آماری می‌تواند مورد استفاده قرار گیرد [۱۶]. وانگ و همکاران^{۱۹} از مدل SVM برای پیش‌بینی میزان آلاینده‌های هوا در یک سری زمانی بر اساس پایش آلاینده‌های هوای موجود در هنگ کنگ استفاده نمودند

شاخصهای زیادی به‌منظور پیش‌بینی خشکسالی ارائه شده است که از آن جمله می‌توان به شاخص درصد از نرمال^۱ که مفهوم اساسی آن تقسیم بارش واقعی بر بارش نرمال است، شاخص دهکها^۲، شاخص رطوبتی گیاه^۳، شاخص خشکسالی احیایی^۴، شاخص پالم^۵ و شاخص استاندارد شده بارش^۶ اشاره نمود. در میان شاخصهای خشکسالی، شاخص پالم و SPI کاربرد وسیعی در تعیین و درجه‌بندی خشکسالی زراعی و هواشناسی در اکثر کشورها دارد. این شاخص تنها به دلیل نیاز به آمار بارندگی ماهانه که در اکثر ایستگاههای کشور در دسترس می‌باشد، قابلیت درجه‌بندی و ارزیابی خشکسالی در مقیاسهای زمانی مختلف کوتاه مدت، میان مدت و درازمدت و قدرت پایش نزدیک به زمان واقعی را دارد.

نویسندگان متعددی نقش متغیرهای هواشناسی نظیر دما و فشار سطح دریا را بر روی خشکسالی و بارندگی فصلی بررسی نموده‌اند و برای پیش‌بینی شاخص SPI از مدل‌های مختلفی نظیر زنجیره مارکوف و یا شبکه‌های عصبی مصنوعی استفاده نموده‌اند [۵-۱]. برای بررسی ارتباط آماری بین متغیرهای هواشناسی، خشکسالی و بارش، روشهای متفاوتی مورد توجه قرار گرفته است که از جمله می‌توان به شبکه‌های عصبی مصنوعی، ماشین بردار پشتیبان^۷، خوشه‌بندی و زنجیره مارکوف اشاره نمود [۶-۹]. در این میان ماشین‌های بردار پشتیبان، سیستم‌های یادگیری خاصی هستند که از یک فضای فرضیه به نام فضای مشخصه^۸، شامل توابع خطی با بعد زیاد استفاده کرده و می‌توانند برای پیش‌بینی (دسته‌بندی و رگرسیون) مورد استفاده قرار گیرند. این سیستم‌ها توسط الگوریتم یادگیری که بر اساس تئوری بهینه‌سازی استوار است، آموزش داده می‌شوند. این تئوری که توسط وپنیک و همکاران^۹ به‌عنوان روشی

¹⁰ Learning Bias

¹¹ Liang et al.

¹² Choy

¹³ Chan

¹⁴ Yu et al.

¹⁵ Bray

¹⁶ Han

¹⁷ Dibike et al.

¹⁸ Tripathi et al.

¹⁹ Wang et al.

¹ Percent of Normal

² Deciles

³ Crop Moisture Index

⁴ Reclamation Drought Index

⁵ Palmer Drought Severity Index

⁶ Standardized Precipitation Index (SPI)

⁷ Support Vector Machine (SVM)

⁸ Feature Space

⁹ Vapnik et al.

[۱۷]. بهزاد و همکاران برای پیش‌بینی رواناب با زمان پیش‌دید یک روزه در رودخانه بختیاری ایران از SVM استفاده نمودند [۱۸]. لین و همکاران^۱ از تکنیک SVM برای پیش‌بینی جریان مؤثر مخزن استفاده کردند. چن و همکاران^۲ از ماشین بردار پشتیبان پشتیبان برای ریز مقیاس کردن بارش روزانه استفاده کردند و با روش آنالیز چند معیاره مقایسه نمودند و نشان دادند نتایج پیش‌بینی‌های حاصل از SVM دقیق‌تر است [۱۹].

نتایج گویای این واقعیت است که روش SVM خیلی سریع‌تر از شبکه‌های عصبی مصنوعی متداول آموزش می‌بیند. همچنین SVM پیش‌بینی دقیق‌تری را نسبت به روشهای شبکه عصبی مصنوعی بازگشتی نشان داده است [۲۰]. کیسی^۳ و سیمن^۴ برای پیش‌بینی جریان ماهانه رودخانه از روش SVM و تابع تبدیل موجی استفاده نمودند. بر اساس معیارهای ارزیابی MSE به‌دست آمده نتایج پیش‌بینی‌ها دقیق‌تر بوده است [۲۱]. نوری و همکاران از سه روش برای انتخاب ورودی‌های مدل SVM و ANN برای پیش‌بینی ماهانه جریان استفاده نمودند و بر اساس آن معیار ارزیابی جدیدی برای مدل‌های هوشمند پیش‌بینی ارائه نمودند. نتایج گویای برتری قابل ملاحظه مدل SVM در پیش‌بینی‌هاست [۲۲].

مدل‌های پیش‌بینی که تاکنون برای پیش‌بینی بارش، خشکسالی و یا رواناب فصلی و سالانه در سطح ملی توسعه داده شده‌اند عمدتاً بر پایه مدل‌های سری زمانی نظیر مدل‌های زنجیره مارکوف و یا تکنیک‌های شبکه عصبی (از نوع مدل پرسپترون چند لایه یا MLP) هستند. از برتری‌های SVM نسبت به سایر روشها آن است که با داده‌های آموزشی کمتر و متغیرهای کمتری کار می‌کند اما نسبت به تغییر متغیرها بسته به داده‌های آموزشی حساس است. در این روش در فضای ورودی داده‌ها فقط از نقاطی استفاده می‌شود که در حاشیه اطمینان ماشین بردار پشتیبان قرار دارند و خطاهای بیشتر از حد معینی که با آزمون و خطا تعیین می‌شود، متحمل نمی‌شوند. ضمناً زمان اجرای مدل در روش SVM نسبت به سایر روشها کوتاه‌تر است [۱۰-۱۲].

هدف از این پژوهش پیش‌بینی زمان وقوع و تداوم خشکسالی با توجه به شاخص SPI با استفاده از داده‌های هواشناسی موجود ثبت شده در ایستگاههای زمینی و ماهواره‌ای و جستاری در خصوص تعیین مناسب‌ترین پیش‌بینی‌کننده‌های هواشناسی بود. در این پژوهش الگوهای تغییرات متغیرهای دما و ارتفاع معادل فشار در سطوح مختلف جو در محدوده جغرافیایی مؤثر بر بارش منطقه مورد

تحقیق و بررسی قرار گرفت و سپس نقش این متغیرها در پیش‌بینی دراز مدت بارندگی در محدوده مورد تحقیق تعیین شد. در انتها، بازه زمانی مناسب برای برقراری ارتباط آماری بین پارامترهای هواشناسی و متغیر پیش‌بینی شونده (SPI) تعیین شد. وجه تمایز این پژوهش نحوه انتخاب پیش‌بینی‌کننده‌ها و تعریف سناریوهای مختلف قابل کاربرد در برنامه‌ریزی منابع آب و پیش‌بینی شاخص خشکسالی با زمان پیش‌دید قابل توصیه است.

۲- مواد و روشها

هدف ماشین‌های یادگیرنده، یافتن یک تابع پیوسته با مقدار حقیقی است که بتواند به ازای مقادیر ورودی پیش‌بینی دقیقی از مقدار خروجی داشته باشد، به‌طور کلی تئوری یادگیری با استفاده از داده‌های آموزشی انجام می‌پذیرد. در این صورت مسئله یادگیری شامل انتخاب تابعی از بین توابع $f(x, \alpha), \alpha \in \mathcal{R}$ خواهد شد که بتواند پاسخ ناظر (هر بردار ورودی \mathbf{x} و بر اساس تابع توزیع شرطی $P(y | x)$ که البته ثابت ولی نامعلوم است، مقدار y را باز می‌گرداند) را به بهترین وجه پیش‌بینی نماید. این گزینش بر اساس مجموعه‌ای از داده‌های آموزشی $(x_1, y_1), \dots, (x_m, y_m)$ است که بر اساس احتمال شرطی $P(x, y) = P(x)P(y | x)$ انتخاب می‌شوند.

به‌منظور انتخاب بهترین تخمین موجود برای پاسخ یک ناظر می‌توان اختلاف (خطا) $L(y, f(x, \alpha))$ موجود بین پاسخ ناظر به بردار ورودی x و پاسخ تابع که توسط ماشین یادگیر به‌دست آمده را محاسبه نمود. اگر مقدار افت قابل پیش‌بینی که توسط تابع خطای^۵ زیر داده می‌شود در نظر گرفته شود.

$$R(\alpha) = \int L(y, f(x, \alpha)) dP(x, y) \quad (1)$$

هدف، یافتن تابع $f(x, \alpha_0)$ است به‌طوری که مقدار $R(\alpha)$ (بر روی دسته توابع $f(x, \alpha), \alpha \in \Lambda$) کمینه شود. این درحالی است که تابع توزیع احتمالی مجهول بوده و تنها اطلاعات موجود، مجموعه داده‌های آموزشی است [۲۳].

۲-۱- مدل رگرسیون ماشین بردار پشتیبان (SVR)

این روش از جمله روشهای نسبتاً جدیدی است که در سالهای اخیر کارایی خوبی نسبت به روشهای قدیمی‌تر برای دسته‌بندی و رگرسیون از جمله شبکه‌های عصبی چند لایه پرسپترون^۶ نشان داده است. وپنیک برای ساختن شکل رگرسیونی ماشین بردار پشتیبان از

¹ Lin et al.

² Chen et al.

³ Kasi

⁴ Cimen

⁵ Risk Functional

⁶ Regression Model

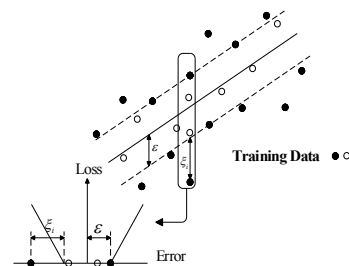
⁷ Multy Layer Perceptron (MLP)

تابع خطای جدیدی استفاده نمود که به نام تابع خطای ϵ -insensitive خوانده می شود و به صورت زیر تعریف می گردد

$$L(y, f(x, \alpha)) = \begin{cases} 0 & \text{if } |y - f(x, \alpha)| \leq \epsilon \\ |y - f(x, \alpha)| - \epsilon & \text{otherwise} \end{cases} \quad (2)$$

با توجه به رابطه ۲، می توان دریافت که خطاهایی با مقادیر کمتر از ϵ در نظر گرفته نشده اند (شکل ۱). به عبارت دیگر در این تابع، خطاهایی در محدوده کمتر از ϵ متحمل جریمه نمی شوند. از این محدوده به نام لوله ϵ یاد می شود و در مسائل چند بعدی ناحیه ϵ -insensitive شکلی نظیر یک دال خواهد داشت و یا به طور کلی این محدوده بین دو آبر صفحه^۱ موازی قرار گرفته است. برای توسعه الگوریتم در ابتدا باید تخمین یک تابع خطی بررسی شود. کلیه توابع خطی به شکل کلی زیر می باشند

$$f(x) = \langle w, x \rangle + b, \quad w, x \in X, \quad b \in R. \quad (3)$$



شکل ۱- تابع خطای ϵ -insensitive [۱۷]

در رابطه ۳، $\langle \cdot, \cdot \rangle$ مشخص کننده ضرب داخلی دو بردار (w بردار وزن و x فضای ورودی) در فضای هیلبرت است. هدف روند یادگیری، یافتن تابع f با کمترین خطا بر اساس داده های مستقل و با توزیع یکنواخت $(x_1, y_1), \dots, (x_m, y_m) \subseteq X \times Y$ است که الگوریتم SVRE-نامیده می شود. برای این منظور به کمینه سازی تابع خطای عمومی شده R_{reg} و بر اساس تابع خطای ϵ -insensitive پرداخته شد. R_{reg} را می توان بر اساس شکل باز شده R_{emp}^{ϵ} بازنویسی کرد

$$R_{emp}^{\epsilon}[f] = \frac{1}{m} \sum_{i=1}^m |y_i - f(x_i)|_{\epsilon} \quad (4)$$

$$R_{reg}[f] = \frac{1}{2} \|w\|^2 + C \cdot R_{emp}^{\epsilon}[f]$$

که در این رابطه

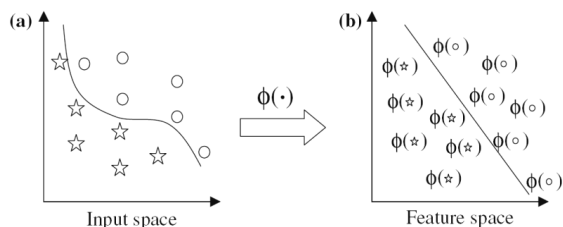
^۱ Hyperplane

R_{emp}^{ϵ} خطای آموزش را در قالب تابع خطای ϵ -insensitive محاسبه می نماید و ضریب C ثابتی است که به نوعی مقدار $\|w\|^2$ را با توجه به پیچیدگی تابع تعیین می کند. کمینه سازی عبارت (۴) نشان دهنده ایده اصلی تئوری یادگیری آماری است که بیان می کند برای دستیابی به کمترین خطا واقعی، کنترل پیچیدگی مدل و همچنین خطای مربوط به داده های آموزش ضروری است. عمومیت پذیری با استفاده از داده های محدود امکان پذیر است اگر و تنها اگر تابع تخمین ظرفیت محدودی داشته باشد. مسئله کمینه سازی خطای هموار شده پس از تبدیل به یک تابع کمینه ساز و ساخت عبارت لاگرانژی مورد نظر به یک مسئله بهینه سازی مقید محدب تبدیل می گردد که با میل دادن مشتقات جزئی آن نسبت به متغیرهای تصمیم قابل حل است [۱۰]. پس از حل مسئله بهینه سازی بالا مقادیر w و f به صورت زیر به دست می آیند

$$w = \sum_{i=1}^m (\alpha_i^* - \alpha_i) x_i \quad (5)$$

$$f(x) = \sum_{i=1}^m (\alpha_i^* - \alpha_i) \langle x_i, x \rangle + b \quad (6)$$

قبل از تقسیم خطی برای اینکه ماشین بتواند داده های با پیچیدگی بالا را پیش بینی کند داده ها به وسیله تابع ϕ به فضای ابعاد خیلی بالاتر منتقل می شوند (شکل ۲). برای حل مسئله با ابعاد خیلی بالا از قضیه دوگانگی لاگرانژ برای تبدیل مسئله کمینه سازی مورد نظر به فرم دوگانی آن استفاده می گردد که در آن به جای تابع پیچیده ϕ که به فضایی با ابعاد بالا می رود، تابع ساده تری به نام تابع هسته (کرنل) که ضرب برداری تابع ϕ است ظاهر می شود. از توابع هسته مختلفی از جمله هسته های خطی، شعاعی، چندجمله ای و سیگموئید می توان استفاده نمود. بنابراین کافی است در مسائل غیرخطی از کرنل مقادیر ورودی به جای خود تابع استفاده شود. با توجه به تئوری توضیح داده شده دقت در تعیین پارامترهای هموارساز C میزان ϵ و مقدار پارامترهایی که ترم موجود در تابع کرنل است، در کاهش خطای مسئله تأثیر بسیار زیادی دارد.



شکل ۲- مفهوم نگاشت از فضای ورودی به فضای مشخصه [۱۵]

۲-۲- نمایه بارش استاندارد شده و انتخاب پارامترهای

پیش بینی

یکی از پارامترهای معتبر در خصوص تعیین رژیم بارش، نمایه بارش استاندارد شده است. مک کی و همکاران^۱ این پارامتر را به منظور تعریف و پایش هواشناختی بارش ارائه داده‌اند [۲۴]. این شاخص امکان مشخص کردن بی‌سابقه بودن یک خشکسالی یا یک ترسالی در مقیاس زمانی مشخص برای هر منطقه‌ای از زمین که دارای سابقه آمار تاریخی باشد را می‌دهد. نمایه SPI به دلیل اینکه فرم بدون بعد داشته و نیاز به اطلاعات گسترده‌ای برای محاسبه ندارد یکی از پرکاربردترین نمایه‌های ارزیابی خشکسالی هواشناسی است.

SPI بیانگر مقدار عددی انحراف معیاری است که یک مقدار مشخص بارش از میانگین دارد. توزیع گاما، چگونگی سری‌های زمانی اقلیمی بارش را به خوبی بیان می‌کند. به عبارتی در بازه زمانی پایه‌ای که پارامترهای گاما برای آن بازه محاسبه شده است، SPI توزیع نرمال استاندارد با امید ریاضی صفر و انحراف معیار یک خواهد داشت. وجود یک شاخص با امید ریاضی و واریانس ثابت برای ممکن شدن مقایسه معنی‌دار بین ایستگاهها و مناطق مختلف بسیار مفید است. SPI در هر مکان عددی نرمال‌سازی شده است، زیرا مبین توزیع تناوب و تکرار بارش در یک ایستگاه را در بر دارد. به علاوه، این شاخص در زمان نیز نرمال است. چون در هر بازه زمانی مورد نظر برای مقاصد مختلف قابل اندازه‌گیری است. به علاوه فارغ از اینکه در چه مکانی و در چه بازه زمانی SPI محاسبه می‌شود، SPI نشانگر احتمال تجمع نسبت به یک بازه زمانی پایه است که پارامترهای گاما برای آن محاسبه شده‌اند. می‌توان با داشتن یک سری زمانی از بارش ماهانه در یک مکان مشخص، SPI را برای هر ماهی از آمار تاریخی برای n ماه اخیر (n=1,2,3,...,12,..., 48) به دست آورد. بنابراین SPI برای بازه‌های زمانی متفاوت، به عنوان مثال ۳ ماه تا هر مقداری مثل ۴۸ ماه به دست می‌آید. در جدول ۱ اقلیم‌های متناظر با حدود شاخصهای SPI با توجه به پیشنهاد مک کی آمده است [۲۳].

انتخاب پارامترهای مؤثر ورودی مدل در فرایند مدلسازی حائز اهمیت بالایی است. ابزارهای مختلفی در تحلیل‌های آماری به این منظور وجود دارد. یکی از این روشها در این عرصه، شاخص اطلاعات متقابل^۲ است [۲۵]. این الگوریتم بر اساس عدم قطعیت در سری‌های گسسته پایه‌ریزی شده است. این شاخص آن‌گونه که در ضریب همبستگی تشابه مورد ارزیابی قرار می‌گیرد، کاهش عدم

قطعیت در شناسایی یک متغیر x را با داشتن متغیر y دیگر بیان می‌کند. این نمایه شاخص اطلاعات متقابل نام دارد [۲۵ و ۲۶].

جدول ۱- مقادیر SPI و وضعیت اقلیمی متناظر با آن [۲۳]

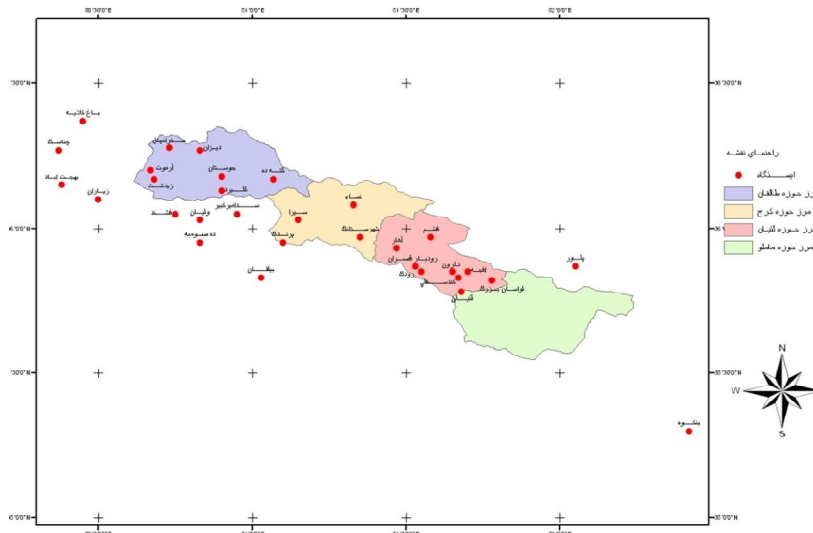
نمایه مربوطه	حدود SPI	اقلیم
VW	+۲	فوق‌العاده مرطوب
W	۱/۹۹ تا ۱/۵	خیلی مرطوب
NW	۱ تا ۱/۴۹	مرطوب متوسط
N	-۰/۹۹ تا -۰/۹۹	نزدیک به نرمال
ND	-۱ تا -۱/۴۹	خشک ملایم
D	-۱/۵ تا -۱/۹۹	خیلی خشک
VD	-۲ و کمتر	فوق‌العاده

۳-۲- منطقه مورد مطالعه و اطلاعات مورد استفاده

محدوده مورد مطالعه، حوضه آبریز سدهای ماملو و طالقان است که تأمین کننده آب تهران است. اطلاعات ۳۱ ایستگاه باران‌سنجی در این محدوده برای تحلیل بارش و خشکسالی هواشناسی مورد استفاده قرار گرفته است. متوسط بارندگی در تهران ۳۴۰ میلی‌متر و حجم آب تجدید شونده ۴/۶ میلیارد مترمکعب و حجم ریزشهای جوی ۶/۵۶ میلیارد متر مکعب در سال است. شکل ۳ حوضه‌های مربوطه و موقعیت ایستگاههای مورد استفاده را نشان می‌دهد. طول آمارای ایستگاههای مورد استفاده از سال ۱۳۵۴-۵۵ تا ۱۳۸۴-۸۵ است. سامانه‌های مؤثر بر اقلیم مناطق مرکزی و غرب ایران را می‌توان به چهار گروه عمده مرکز پر فشار سیبری (Siberia) (Azores, 40° N, 25° W)، مرکز کم فشار آزر (East of Mediterranean, 35° N, 50° E) و مرکز کم فشار سودان (Sudan, 10° N, 30° E) تقسیم‌بندی کرد [۳]. شکل ۴ محدوده جغرافیایی در نظر گرفته شده برای بررسی ارتباط اطلاعات هواشناسی با بارش فصلی در حوضه‌های آبریز مورد مطالعه را نشان می‌دهد. همان‌طور که در این شکل ملاحظه می‌شود، کلیه مناطقی که به نحوی در شکل‌گیری سامانه‌های بارش مؤثر هستند در محدوده مورد بررسی در نظر گرفته شده‌اند. با توجه به گسترده بودن محدوده مورد مطالعه، منطقه نمایش داده شده در شکل ۴ به منطقه‌های جغرافیایی با ابعاد پنج درجه عرض جغرافیایی در پنج درجه طول جغرافیایی دسته‌بندی شده‌اند. تعداد منطقه‌های مورد بررسی در تحقیق ۱۸۰ مورد است. اطلاعات هواشناسی مورد استفاده در این پژوهش که برای هریک از منطقه‌های جغرافیایی مورد استفاده در طول دوره تاریخی ۳۱ ساله مورد استفاده قرار گرفته‌اند، شامل دمای هوا در سطح زمین، دمای هوا در سطوح ۳۰۰، ۵۰۰، ۷۰۰ و ۸۵۰ میلی‌بار و ارتفاع

¹ McKee et al.

² Mutual Information Index (MI)



شکل ۳- موقعیت ایستگاه‌های مورد استفاده نسبت به حوضه‌های مورد بررسی

پیش‌بینی‌کننده^۲ با زمان پیش دید^۳ مناسب برای پیش‌بینی دوره‌های کم بارش خواهند بود. با استفاده از ۱۱۳۴۰ سری زمانی (۹ متغیر هواشناسی \times ۱۸۰ منطقه جغرافیایی \times ۷ ماه از سال) تحلیل زمانی- مکانی پارامترهای هواشناسی انجام می‌گیرد.

جدول ۲- شماره مشخص‌کننده متغیرهای هواشناسی

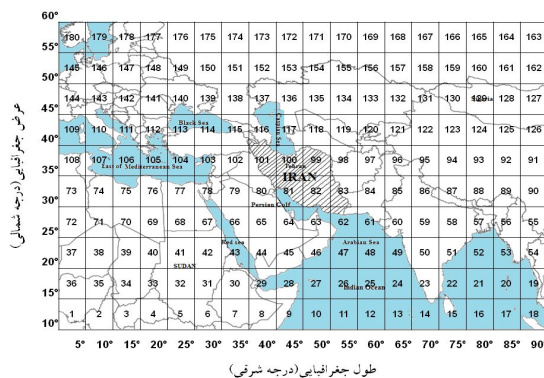
علامت	پارامترهای هواشناسی	شماره
AT1000	دما در سطح زمین	۱
AT300	دما در سطح ۳۰۰ میلی بار	۲
AT500	دما در سطح ۵۰۰ میلی بار	۳
AT700	دما در سطح ۷۰۰ میلی بار	۴
AT850	دما در سطح ۸۵۰ میلی بار	۵
GH300	ارتفاع معادل فشار در سطح ۳۰۰ میلی بار	۶
GH500	ارتفاع معادل فشار در سطح ۵۰۰ میلی بار	۷
GH700	ارتفاع معادل فشار در سطح ۷۰۰ میلی بار	۸
GH850	ارتفاع معادل فشار در سطح ۸۵۰ میلی بار	۹

۳- فرایند مدلسازی

برای بررسی تغییرات بارندگی و توسعه مدل پیش‌بینی برای محدوده مورد مطالعه که توسط سیستم برنامه‌ریزی و مدیریت منابع آب قابل بهره‌برداری باشد، ۶ سناریو مختلف طراحی شد. این سناریوها شامل مجموع بارش فصل بهار (SPI1)، مجموع بارش فصل پاییز (SPI2)، مجموع بارش فصل زمستان (SPI3)، مجموع بارش فصلهای پاییز + زمستان (SPI4)، مجموع بارش فصلهای زمستان + بهار (SPI5)، مجموع بارش فصلهای پاییز + زمستان +

² Predictor
³ Leadtime

معادل فشار^۱ در سطوح ۳۰۰، ۵۰۰، ۷۰۰ و ۸۵۰ میلی‌بار است. این متغیرها بر اساس تحقیقات گذشته و تأثیر بیشتر این متغیرها بر سامانه‌های بارش‌زا در محدوده مورد مطالعه انتخاب شدند. در جدول ۲ متغیرهای مورد استفاده و علامت مشخصه نظیر هر یک آورده شده است.



شکل ۴- گستره محدوده ۱۸۰ گانه جغرافیایی (متوسط ماهانه متغیرهای هواشناسی از سایت NCAR در هر محدوده اخذ شده است)

به‌منظور بررسی ارتباط بارش متوسط و شاخص SPI در شش سناریو ذکر شده با متغیرهای هواشناسی مذکور، این متغیرها در ماه‌های آوریل تا اکتبر که تقریباً ۶ ماهه دوم سال آبی را پوشش می‌دهند در نظر گرفته شدند. به این ترتیب در صورت تشخیص ارتباط آماری مناسب، این متغیرها قابل استفاده به‌عنوان

¹ Geopotential Height

بهار (SPI6) بودند. این اطلاعات در روی دو حوضه ماملو و طالقان به منظور تخمین میانگین بارش در هر یک از سناریوهای بالا در هر سال به کار رفت. این اطلاعات از مؤسسه تحقیقات آب و وزارت نیرو، بخش آمار و اطلاعات دریافت شد. در انتخاب محدوده مورد مطالعه سعی شد تا کلیه مناطقی که به نحوی بر آب و هوای ایران و به خصوص مناطق غرب و مرکزی ایران مؤثر هستند، در نظر گرفته شوند.

۳-۱- محاسبه نمایه بارش استاندارد شده

با توجه به گستره مکانی داده‌های مورد استفاده نیاز است تا میانگین بارش برای محاسبه SPI در ایستگاههای مورد نظر و در ماه و سال مورد نظر تعیین شود. به منظور تخمین مکانی بارش، مدل معکوس فاصله وزندار^۱ به دلیل روانی و سهولت استفاده و نیز تناسب آن با کوهستانی بودن بخش قابل توجهی از محدوده مورد مطالعه، انتخاب شد و سپس اطلاعات به دست آمده به مقادیر SPI تبدیل شد.

۳-۲- انتخاب متغیرهای پیش‌بینی کننده و پیش‌بینی شونده

شاخص اطلاعات متقابل بر روی متغیرهای هواشناسی و نمایه بارش استاندارد شده برای هر سناریو انجام گرفت. از معیارهای مهم که در ادبیات انتخاب پارامترهای مدل‌های رگرسیونی کاربرد گسترده‌ای یافته، شاخص اطلاعات متقابل و یا شاخص MI است. این معیار به منظور انتخاب تأخیرهای مناسب از یک پدیده زمان‌مند به منظور مدل‌سازی در متون مورد تأکید قرار گرفته و عملکرد آن تا حدود قابل توجهی در ادبیات فنی موضوع تبیین شده است [۲۵] و [۲۶]. در این تحقیق، شاخصهای MI به دست آمده از کل متغیرهای هواشناسی به ترتیب نزولی مرتب گردید و شماره منطقه‌هایی که دارای بیشترین MI است به دست آمد. شاخصهای MI بزرگ‌تر برای تعیین نقطه پارامترهای پیش‌بینی کننده انتخاب شد. با توجه به اینکه هر نقطه پارامتر با یک سناریو بیشترین ارتباط را داشته است، آن نقطه پارامترها به عنوان پیش‌بینی کننده همان سناریو انتخاب

می‌شوند. فقط سناریوهایی در این مرحله انتخاب شده‌اند که ارتباط آماری متغیرهای هواشناسی آنها با SPI از میانگین شاخص MI برای هر متغیر بیشتر باشد. نقطه پارامترهایی که بیشترین همبستگی را با نمایه بارش استاندارد شده داشتند به عنوان ورودی مدل پیش‌بینی برای هر سناریو استفاده شدند. از میان ۱۸۰ منطقه جغرافیایی مورد بررسی ۷۴ منطقه جغرافیایی برای سناریوهای مختلف انتخاب شدند که متغیرهای پیش‌بینی کننده انتخاب شده برای سناریوهای منتخب در جدول ۳ آورده شده است.

۳-۳- پیش‌بینی خشکسالی با استفاده از ماشین بردار پشتیبان

روش کار برای آموزش و واسنجی مدل به این صورت بود که ابتدا داده‌های موجود برای هر سناریو به طور تصادفی به ۲ دسته (۷۵ درصد اطلاعات به منظور آموزش و واسنجی مدل ماشین بردار پشتیبان و تعیین پارامترها و پارامترهای تابع کرنل و ۲۵ درصد برای آزمون و تست مدل) تقسیم بندی شدند. از متغیرهای هواشناسی انتخاب شده در هر سناریو حوضه به عنوان ورودی مدل پیش‌بینی و از نتایج نمایه بارش استاندارد شده به عنوان خروجی مدل رگرسیونی ماشین بردار پشتیبان استفاده شد. سپس تلاشی مبتنی بر سعی و خطا برای یافتن پارامترهای بهینه مدل آغاز شد. در این تحقیق از نرم افزار LIBSVM برای مدل SVR استفاده شد که در این نرم افزار از توابع مختلف کرنل اعم از خطی، چند جمله‌ای، سیگموئید و شعاعی و با تکرار اجرای برنامه و انتخاب بازه‌های مناسب برای پارامترهای مدل ($0 < \epsilon < 1$, $1 < C < 1000000$) و پارامترهای تابع کرنل نتایج شامل نوع کرنل، پارامترها به همراه میزان خطای میانگین مربعات^۲ به دست آمده، تعیین شده است [۲۷]. پس از بررسی نتایج برای هر سناریو فصل -حوضه مدلی که کمترین خطا را دارا بود، انتخاب شد. سپس پیش‌بینی مقادیر واقعی نمایه بارش استاندارد شده انجام شد که نتایج در جدول ۴ ارائه شده است. نتایج آزمون مدل و مقایسه آن با مقادیر واقعی در شکل ۵ نشان داده شده است.

¹ Inverse Weighted Distance (IWD)

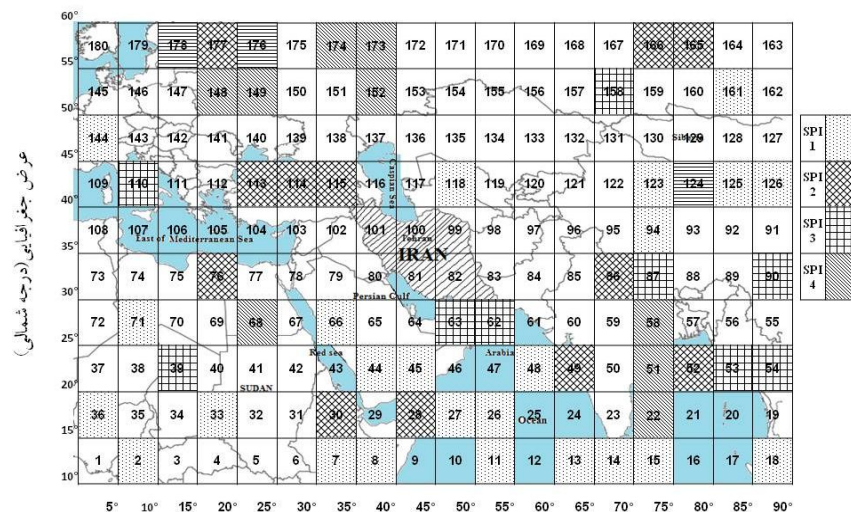
² Mean Square Error (MSE)

جدول ۳- متغیرهای هواشناسی انتخاب شده برای سناریوهای مختلف (اعداد نشان دهنده مناطق جغرافیایی مورد نظر)

فصل SPI	حوضه آبریز	متغیرهای پیش‌بینی کننده مورد استفاده با توجه به شاخص MI
۱- بهار	ماملو	AT1000(118,15), AT300(129), AT500(144), AT700(70,131), AT850(7), GH850(8,45)
	طالقان	AT700(123,106,158), GH500(26,27,29), GH700(13,14)
۲- پاییز	ماملو	AT850(49,86), GH850(165,166)
۳- پاییز+زمستان	طالقان	AT1000(49), AT500(68), AT700(74), GH300(152), GH700(22), GH850(174)
۴- زمستان+بهار	طالقان	GH500(138), GH700(47,115), GH850(152)

جدول ۴- نتیجه نهایی پارامترهای مدل SVR

سناریو SPI	حوضه	MSE	C×1000	ε	تابع کرنل	پارامتر
بهار	ماملو	۰/۰۸	۶۵۰	۰/۳۵	خطی	-
بهار	طالقان	۰/۱۴	۳۴۵	۰/۷	خطی	-
پاییز	ماملو	۰/۴	۶۸	۰/۱۴	شعاعی	-۰/۹
پاییز+زمستان	طالقان	۰/۵۶	۱۵۴	۰/۷	خطی	-
زمستان+بهار	طالقان	۰/۲۳	۶۶	۰/۰۶	شعاعی	۱/۱



طول جغرافیایی (درجه شرقی)

شکل ۵- منطقه‌هایی که متغیرهای هواشناسی بیشترین ارتباط آماری را با SPI فصل‌های مربوطه داشته‌اند با هاشور متفاوت مشخص شده‌اند.

آموزش به دست آمد. مشخصات مدل‌های شبکه‌های عصبی MLP به کار رفته که در مجموع و برای کلیه سناریوها در مرحله تست خطای کمتری نسبت به سایر مدل‌ها داشته‌اند در جدول ۵ آورده شده است. در شبکه‌های توسعه داده شده تابع تحریک منتخب سیگموئید بود و شبکه‌ها به صورت برخط آموزش دیدند. ضمناً با استفاده از رویکرد سعی و خطا سعی در نزدیک شدن به بهترین ساختار و پارامترها شد و بهترین نتایج ارائه گردید. در مرحله تست مدل، نمایه بارش استاندارد شده برای هشت سال شبیه‌سازی شد و با میزان مشاهده شده آن مقایسه شد. با بررسی نتایج حاصله مشاهده شد که پارامتر همبستگی مطلوبی از تست شبکه حاصل نمی‌شود. با استفاده از روابط ۷ تا ۹ مقادیر NMSE، MSE و ERROR برای مقایسه پیش‌بینی‌های انجام گرفته توسط شبکه‌های عصبی MLP و ماشین بردار پشتیبان محاسبه شد، که نتایج حاصله در جدول ۵ ارائه گردیده‌اند.

$$MSE = \frac{\sum_{i=1}^n (X_p - X_o)^2}{n} \quad (7)$$

به منظور مقایسه پیش‌بینی‌های انجام گرفته به وسیله مدل رگرسیون ماشین بردار پشتیبان و بررسی عدم قطعیت در پیش‌بینی‌ها با استفاده از همان ورودی‌های انتخاب شده برای مدل SVR، پیش‌بینی این بار با استفاده از شبکه‌های عصبی پرسپترون انجام شد. برای این منظور با استفاده از نرم‌افزار Neurosolution مدل‌های مختلف شبکه‌های عصبی توسعه داده شد و آموزش و تست مدل شبکه برای هر یک از شبکه‌ها انجام گرفت. شبکه به کار گرفته شده یک شبکه پرسپترون چندلایه بود که به صورت پیش‌خور^۱ عمل می‌نمود و در سه لایه (یک لایه ورودی، یک لایه میانی و یک لایه خروجی) طراحی شده بود. نرون‌های لایه ورودی به شبکه وارد می‌شد و خروجی شامل یک نرون بود که همان SPI می‌باشد. روش آموزش شبکه پس انتشار خطا است. برای آموزش شبکه از متغیرهای هواشناسی استفاده شده برای ماشین بردار پشتیبان استفاده شد. ۷۵ درصد مجموع داده‌ها برای آموزش و ۲۵ درصد آن برای تست مدل استفاده شد. میزان میانگین مربع خطا (MSE) و میانگین مربع خطای نرمال شده (NMSE) پس از ۱۰۰۰ سیکل

^۱ Feed Forward

$$NMSE = \frac{\sum_{i=1}^n (X_p - X_o)^2}{n.S^2} \quad (8)$$

$$Error = \frac{|X_o - X_p|}{X_o} \quad (9)$$

نتایج گویای عملکرد مناسب عملکرد مدل ماشین بردار پشتیبان نسبت به شبکه‌های عصبی MLP در اکثر پیش‌بینی‌ها به‌خصوص در سناریو پاییز حوضه ماملو است (شکل ۶-ج). در جدول ۷ حداقل زمان پیش‌دید برای هر سناریو و حوضه آبریز با توجه به متغیرهای هواشناسی انتخاب شده، مشخص شده است. همان‌طور که ملاحظه می‌شود زمان پیش‌دید پیش‌بینی نمایه بارش استاندارد شده برای حوضه ماملو و طالقان برای هر فصل متفاوت است. زمان پیش‌دید سناریو فصل بهار برای کلیه حوضه‌های مورد تحقیق برابر ۵ ماه است. این در حالی است که زمان پیش‌دید در سناریو فصل زمستان بین ۲ تا ۴ ماه بوده و در سناریوهایی که با فصل پاییز آغاز می‌شود، ۲ تا ۳ ماه است. بدیهی است که هر چه زمان پیش‌دید افزایش یابد میزان خطای مدل نیز افزایش می‌یابد. اما در مدل پیش‌بینی ایجاد شده بر اساس ماشین‌های بردار پشتیبان، نتایج حاصل برای استفاده در بازه زمانی مورد نظر با کمترین خطا ایجاد شده است.

شکل ۷ نشان‌دهنده کلیه پیش‌بینی‌های انجام گرفته توسط مدل‌ها در یک نمودار است. همان‌طور که ملاحظه می‌شود نیمساز نمودار و مقادیر حدی قابل صرف‌نظر کردن بر اساس خطا با توجه به تغییرات SPI ترسیم شده است، به‌طوری که مقادیر داخل محدوده، دارای خطای قابل صرف‌نظر کردن، فرض می‌شود. در این صورت ۱۷ نقطه از کلیه پیش‌بینی‌های صورت گرفته توسط ماشین بردار پشتیبان خارج از این محدوده قرار می‌گیرد، در حالی که ۳۴ نقطه از پیش‌بینی‌های شبکه‌های عصبی خارج از این محدوده قرار دارد. نتایج شکل ۷ گویای برتری کلی پیش‌بینی‌های صورت گرفته توسط ماشین بردار پشتیبان است.

که در این روابط X_p مقادیر پیش‌بینی شده، X_o مقادیر مشاهده شده، n تعداد داده‌ها و S^2 واریانس داده‌های مشاهداتی است. همان‌طور که مشاهده می‌شود خطای پیش‌بینی SPI واقعی در شبکه‌های عصبی MLP بسیار بالا است زیرا داده‌های مورد استفاده برای آموزش محدود می‌باشند در حالی که در روش ماشین بردار پشتیبان تعداد داده‌های آموزشی عامل محدود کننده‌ای نیست. در ضمن هماهنگی بین مقادیر پیش‌بینی شده دو مدل وجود ندارد.

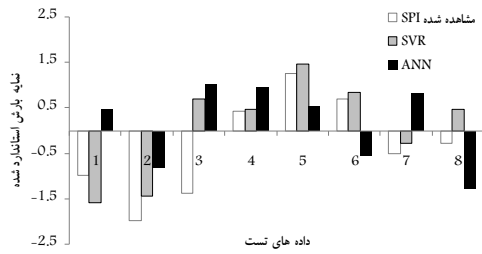
در جدول ۶، نتایج بهترین مدل‌های توسعه داده شده به‌وسیله مدل‌های بردار پشتیبان و شبکه عصبی همراه با مدل‌های معادل آن با استفاده از روش دیگر آورده شده است. نتایج گویای برتری شبکه‌های عصبی MLP در سناریوی بهار در حوضه ماملو است. در این سناریو عملکرد بهتر شبکه‌های عصبی (MLP) نشان‌دهنده عملکرد بهتر مدل‌های شبکه‌های عصبی در آموزش داده‌ها و پیش‌بینی است. علت دقت بیشتر شبکه‌های عصبی در این سناریو می‌تواند در تفاوت استفاده از پیش‌بینی‌کننده‌های مورد استفاده در هر یک از سناریوها و تعداد و نوع متغیرهای پیش‌بینی‌کننده باشد. در شکل ۶ بهترین مقادیر SPI پیش‌بینی شده توسط مدل‌ها با مقادیر مشاهده شده، مقایسه شده است. همان‌طور که مشهود است بهترین مقادیر پیش‌بینی متعلق به سناریوی بهار حوضه ماملو است.

جدول ۵- مشخصات مدل‌های مختلف شبکه عصبی MLP و خطاهای به‌دست آمده در سناریوها

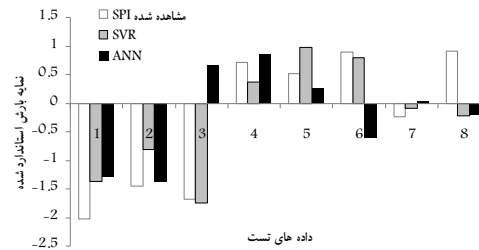
MSE	NMSE	Error	تعداد نرون‌ها	تعداد لایه‌های پنهان	روش آموزش شبکه	SPI سناریو
۱/۱۵	۰/۸۷	۲/۲	۱۰	۱	مونتم	بهار- ماملو
۰/۱۵	۰/۱۷	۰/۳۲	۱۳	۱	مونتم	بهار- طالقان
۰/۸۷	۱/۵۴	۱/۳	۹	۱	مونتم	پاییز- ماملو
۰/۷	۰/۳۶	۰/۹	۷	۱	مونتم	پاییز+ زمستان- طالقان
۱/۳	۰/۸۶	۱/۹	۱۰	۱	مونتم	زمستان+ بهار- طالقان

جدول ۶- کمترین و بیشترین خطای پیش‌بینی در سناریوهای منتخب مدل ANN در مقایسه با مدل SVM

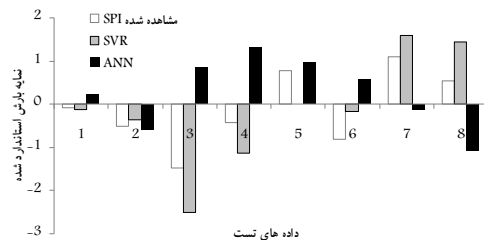
خطای پیش‌بینی SPI در مرحله آزمون مدل‌ها						
مدل SVR			مدل ANN			سناریو SPI
MSE	NMSE	Error	MSE	NMSE	Error	
۰/۳	۰/۳۳	۰/۸	۰/۱۵	۰/۱۷	۰/۳۲	بهار- طالقان
۰/۷۳	۱/۲۹	۱/۱	۰/۸۷	۱/۵۴	۱/۳	پاییز- ماملو



(الف)

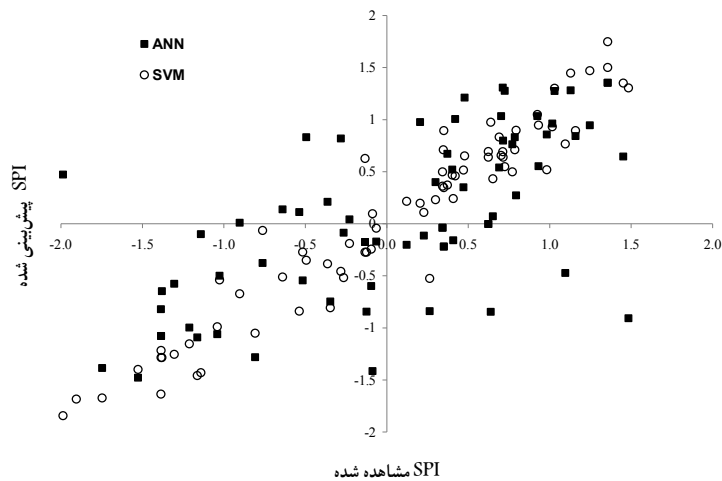


(ب)



(ج)

شکل ۶- مقایسه مقادیر پیش‌بینی شده و مشاهده شده نمایه بارش استاندارد شده از SVR (الف) فصل بهار حوضه ماملو، (ب) فصل بهار حوضه طالقان، (ج) فصل پاییز حوضه ماملو



شکل ۷- مقایسه مقادیر مشاهده شده و پیش‌بینی شده SPI در کلیه سناریو ها توسط مدل شبکه‌های عصبی و ماشین بردار پشتیبان

جدول ۷- پیش‌بینی نمایه بارش استاندارد شده برای هر سناریو با زمان پیش دید معین

حوضه آبریز	زمان پیش دید مربوط به هر سناریو (ماه)	
	SPI ₅	SPI ₁₂
ماملو	۵	۲
طالقان	۵	۴

۴- نتایج و بحث

نتایج نشان می‌دهد که با توجه به نقطه پارامترهای (پیش‌بینی کننده‌ها) انتخاب شده، محدوده‌های دارای ارتباط مؤثر با بارش محدوده مورد مطالعه در فصل بهار بیشتر در جنوب، جنوب غربی و شمال غربی کشور و در فصل پاییز، در شمال، شمال غربی و جنوب و در زمستان، در شمال غربی و غرب کشور قرار دارند. متغیرهای پیش‌بینی‌کننده مورد استفاده به ترتیب بیشتر در سطوح ۸۵۰، ۷۰۰، ۵۰۰ و ۱۰۰ میلی‌بار سطح جو قرار گرفته‌اند. ۶۰ درصد داده‌های انتخاب شده برای کلیه سناریوها متغیر ارتفاع معادل فشار است که نشان دهنده تأثیر بیشتر این متغیر بر نمایه بارش استاندارد شده است. متغیرهای هواشناسی پیش‌بینی‌کننده ارتفاع معادل فشار در سطح ۸۵۰ میلی‌بار و ۷۰۰ میلی‌بار در کلیه سناریوها به کار رفته است و بیشترین تعداد متغیری است که در پیش‌بینی‌ها به کار رفته است. پس از این دو متغیر، متغیرهای ارتفاع معادل فشار و دما در سطح ۵۰۰ میلی‌بار و دما در سطح زمین بیشترین تعداد پیش‌بینی‌کننده را شامل می‌شود. برای سناریوهایی که با فصل پاییز شروع می‌شوند متغیرهای هواشناسی سطوح ۸۵۰ و ۷۰۰ میلی‌بار بر نمایه بارش استاندارد شده با استفاده از شاخص MI مؤثر شناخته شده‌اند. در حالی که برای سناریوهایی که با فصل زمستان آغاز می‌شوند تقریباً کلیه سطوح بر نمایه بارش استاندارد شده مؤثر شناخته شده‌اند. می‌توان نتیجه گرفت که سامانه‌های مؤثر بر SPI برای فصلهای زمستان پیچیده‌تر و با دخالت کلیه سطوح در نظر گرفته شده در پژوهش حاضر است. درصد قابل توجهی از پیش‌بینی‌کننده‌های انتخاب شده در ماه می و ژوئن قرار گرفته‌اند که امکان پیش‌بینی با زمان پیش‌دید حداقل ۳ ماهه و حداکثر ۵ ماهه را برای سناریوهای مختلف فراهم می‌کند.

از بین سناریوهای منتخب، در ۳ سناریو بهار، پاییز و زمستان + بهار کمترین مقدار MSE برای آموزش داده‌ها مشاهده شده است. در این بین نیز با توجه به معیار ارزیابی NMSE بیشترین دقت پیش‌بینی در مرحله آزمون مدل برای سناریوهای بهار در حوضه‌های ماملو و طالقان و زمستان + بهار در حوضه طالقان حاصل شده است. نتایج نشان می‌دهد مدل پیش‌بینی با اعتمادپذیری قابل توصیه برای برنامه‌ریزی بهره‌برداری از منابع آب در حوضه‌های مورد مطالعه، قابل استفاده است.

نزدیک‌ترین پیش‌بینی مقدار نمایه بارش استاندارد شده با توجه به معیارهای ارزیابی مدل در سال آبی ۸۰-۷۹ که همراه با خشکسالی بوده است برای سناریوهای بهار، زمستان، زمستان + بهار انجام شده است. همچنین در سال آبی ۷۷-۷۸ برای سناریو پاییز و پاییز + زمستان و در سال آبی ۷۸-۷۹ برای سناریوهای بهار، زمستان + بهار، SPI پیش‌بینی شده به مقدار واقعی آن

نزدیک‌تر است. مقایسه خطای پیش‌بینی نشان داد که روش ماشین بردار پشتیبان از قابلیت اطمینان بالاتری نسبت به شبکه‌های عصبی برخوردار است. به طوری که در سالهایی که با خشکسالی شدید همراه بوده است پیش‌بینی با دقت قابل توصیه و با زمان پیش‌دید پنج ماهه در فصل بهار صورت گرفته است. در مرحله آزمون مدل در سناریوی بهار حوضه ماملو و در سناریوی زمستان + بهار در حوضه طالقان هر کدام در یک سال مقدار SPI پیش‌بینی شده شرایط خیلی مرطوب را نشان می‌دهد در حالی که مقدار مشاهده شده آن شرایط خشکسالی هواشناسی شدید را نشان می‌دهد. علت این امر می‌تواند عدم کفایت داده‌های آموزشی و متغیرهای هواشناسی مورد استفاده برای شرایط مرطوب باشد. بر مبنای نتایج حاصل از پیش‌بینی SPI با استفاده از ماشین بردار پشتیبان می‌توان متغیرهایی که بهترین نتیجه را در مدل بردار پشتیبان با زمان پیش‌دید مشخص می‌دهند، تعیین نمود. بر این اساس جدول پیشنهادی استفاده از متغیرهای مورد نظر برای پیش‌بینی هر سناریو تعیین و ارائه گردید.

احتمال اینکه در مدل ماشین بردار پشتیبان نیز مانند شبکه‌های عصبی مصنوعی MLP پدیده بیش‌آموزشی^۱ در مورد داده‌های آموزشی اتفاق بیفتد وجود دارد. البته این مسئله نیاز به تحقیقات بیشتر و همچنین استفاده از سایر متغیرهای هواشناسی مؤثر بر سامانه‌های بارش‌زا در کشور دارد. پیشنهاد می‌شود از سایر روشهای بهینه‌سازی برای انتخاب متغیرهای پیش‌بینی‌کننده استفاده شود و به‌عنوان ورودی مدل SVR یا MLP مورد استفاده قرار گیرد.

۵- نتیجه‌گیری

در این پژوهش نمایه بارش استاندارد شده در حوضه سدهای طالقان و ماملو بر اساس بارش متوسط حوضه‌ها که به‌روش میانگین معکوس فاصله وزندار محاسبه شده، تخمین زده شد. ماشین بردار پشتیبان برای تدوین مدل پیش‌بینی SPI استفاده شد. با استفاده از تکنیک آماری مبتنی بر آنتروپی مشترک اطلاعات، نقاط مؤثر بر بارش در حوضه سدهای تهران در فصل بهار بیشتر در جنوب، جنوب غربی و شمال غربی کشور و در فصل پاییز در شمال، شمال غربی و جنوب و در زمستان در شمال غربی و غرب کشور تشخیص داده شدند. نتایج مدل SVM در اکثر موارد، پیش‌بینی با دقت مناسبی را نشان داد. این روش می‌تواند در پیش‌بینی رفتارهای غیر خطی داده‌های هواشناسی یا طول دوره آماری کوتاه مورد استفاده قرار گیرد. این دقت برای دسته‌بندی SPI فصلهای پاییز و بهار بیشتر از سایر سناریوها است.

^۱ Overfitting

- 1- Ebrahimi, R., Zahraie, B., and Nasser, M. (2011). "Mid-term prediction of meteorological drought using fuzzy inference system." *J. of Water and Wastewater*, 78, 112-125 (In Persian)
- 2- Nazemossadat, M.J. (1998). "Persian Gulf sea surface temperature as a drought diagnostic for southern parts of Iran." *J. of Drought News Network*, 10, 12-14.
- 3- Nazemossadat, M.J., and Cordery I., (2000). "On the relationship between ENSO and autumn rainfall in Iran." *International J. of Climatology*, 1, 42-67.
- 4- Karamouz, M., Fallahi, M., Nazif, S., and Rahimi Farahani, M. (2009). "Long lead rainfall prediction using statistical downscaling and artificial neural network modeling." *Scintia Iranica*, 16, 165-172.
- 5- Zahraie, B., and Karamouz, M. (2004). "Seasonal precipitation prediction using large scale climate signals." *Proceedings of EWRI-2004 Conference*, Salt lake City, USA.
- 6- Loukas A., and Vasiliades, L. (2004). "Probabilistic analysis of drought spatiotemporal characteristics in thessaly region." *Greece. Natural Hazards and Earth System Sciences*, 4, 719-731.
- 7- Paulo, A. A., Ferreira, E., Coelho, C., and Pereira, L.S. (2005). "Drought class transition analysis through Markov and loglinear models, an approach to early warning." *Agricultural Water Management*, 77, 59-81.
- 8- Moreira, E.E., Paulo, A.A., Pereira, L.S., and Mexia, J.T. (2006). "Analysis of SPI drought class transitions using loglinear models." *J. Hydrol.*, 331, 349-359.
- 9- Zahraie, B., and Roozbahani, A., (2007). "Climate signal clustering using genetic algorithm for precipitation forecasting: A case study of southeast of Iran." *Proceedings of the World Environmental and Water Resources Congress (ASCE)*, Tampa, Florida, USA.
- 10- Vapnik, V. N., and Cortes, C. (1995). "Support vector networks." *Machine Learning*, 20, 273-297.
- 11- Liong, S.-Y., and Sivapragasam, C. (2002). "Flood stage forecasting with support vector machines." *J. of the American Water Resources Association*, 38 (1), 173-196.
- 12- Choy, K.Y., and Chan, C.W. (2003). "Modelling of river discharges and rainfall using radial basis function networks based on support vector regression." *International J. of Systems Science*, 34(14-15), 763-773.
- 13- Yu, X., Liong, S.-Y., and Babovic, V. (2004). "EC-SVM approach for realtime hydrologic forecasting." *J. of Hydroinformatics*, 6, 209-223.
- 14- Bray, M., and Han, D. (2004). "Identification of support vector machines for runoff modeling." *J. of Hydroinformatics*, 6 (4), 265-280.
- 15- Dibike, Y.B., Velickov, S., Solomatine, D., and Abbott, M.B. (2001). "Model induction with support vector machines: Introduction and applications." *J. of Computing in Civil Engineering*, 15 (3), 208-216.
- 16- Tripathi, Sh., Srinivas, V. V., and Nanjundiah, R. S. (2006). "Downscaling of precipitation for climate change scenarios: A support vector machine approach." *J. of Hydrology*, 330, 62-640.
- 17- Wang, W. C., and Men, W. L. (2008). "Online prediction model based on support vector machine." *Neurocomputing*, 71, 550-558.
- 18- Behzad, M., Asghari, K., Eazi, M., and Palhang, M. (2009). "Generalization performance of support vector machines and neural networks runoff modeling." *Expert System with Applications*, 36, 7624-7629.
- 19- Chen, S-T., Yu, P.Sh., and Tang, H. Y. (2010). "Statistical downscaling of daily precipitation using support vector machines and multivariate analysis." *J. of Hydrology*, 385, 13-23.
- 20- Lin, G., Chen, G., Huang, P., and Chou, Y. (2009). "Support vector machine-based models for hourly reservoir inflow forecasting during typhoon-warning periods." *J. of Hydrology*, 372, 17-29.
- 21- Kisi, O., and Cimen, M. (2011). "A wavelet-support vector machine conjunction model for monthly streamflow forecasting." *J. of Hydrology*, 399(1-2), 132-140.

- 22- Nooria, R. A.R., Karbassia, A., Moghaddamniac, D., Hand, M.H., Zokaei-Ashtianie, A., Farokhniab, F., and Ghafari Goushehc, M. (2011). "Assessment of input variables determination on the SVM model performance using PCA, Gamma test, and forward selection techniques for monthly stream flow prediction ." *J. of Hydrology*, 401 (3-4), 177-189.
- 23- McKee, T.B., Doesken, N.J., and Kleist, J. (1993). "The relationship of drought frequency and duration to time scales." *In: Proceedings of the Eighth Conference on Applied Climatology. Am. Meteor. Soc.*, Boston, 179-184.
- 24- Pai, P.-F., and Hong, W.-C. (2007). "A recurrent support vector regression model in rainfall forecasting," *Hydrological Processes*, 21, 819-827.
- 25- Witten Ian H., and Eibe, F. (2005). *Data mining: Practical machine learning tools and techniques*, Morgan Kaufmann Pub., Amsterdam.
- 26- Peng, H.C., Long, F., and Ding, C. (2005). "Feature selection based on mutual information: criteria of max-dependency, max-relevance, and min-redundancy." *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27, 1226-1232.
- 27- Chang, C.-C., and Lin, C.-J. (2009). "LIBSVM: A library for support vector machines." <<http://www.csie.ntu.edu.tw/~cjlin/libsvm>>. (Version 2.91, November 2009).