

گزارش فنی پیش‌بینی سیلاب‌های ساعتی رودخانه اهرچای با استفاده از روش‌های یادگیری ماشین

محمدتقی ستاری^۱، محمدرضا عبدالله پورآزاد^۲ و رسول میرعباسی نجف‌آبادی^{۳*}

^۱ استادیار، دانشکده کشاورزی، دانشگاه تبریز، آکارسناس ارشد، دانشکده فنی مهندسی، دانشگاه آزاد اسلامی واحد اهر و ^۳ استادیار، دانشکده کشاورزی، دانشگاه شهرکرد

تاریخ پذیرش: ۱۳۹۳/۱۲/۱۸

تاریخ دریافت: ۱۳۹۲/۱۲/۱۱

چکیده

سیل یکی از حوادث طبیعی است که هر ساله خسارات بسیاری در نقاط مختلف جهان به وجود می‌آورد. پیش‌بینی دقیق سیلاب در کاهش خسارات جانی و مالی و مدیریت منابع آب از اهمیت بسزایی برخوردار است. هدف از مطالعه حاضر، مقایسه قابلیت‌های روش‌های رگرسیون ماشین بردار پشتیبان، مدل درختی M5 و مدل رگرسیون خطی در برآورد دبی سیلاب یک و دو ساعت آینده ایستگاه تازه‌کند در رودخانه اهرچای می‌باشد. داده‌های تاریخی دبی-اشل ساعتی ایستگاه تازه‌کند و ۱۴ رویداد مهم سیل برای ایجاد مدل مورد استفاده قرار گرفت. نتایج نشان داد که روش رگرسیون ماشین بردار پشتیبان با ضریب تبیین ۰/۹۶ و جذر میانگین مربعات خطا ($M^3 \cdot s^{-1}$) ۰/۴۷۲ برای سیلاب یک ساعت بعد و $R^2=0/90$ و $RMSE=0/1596 (M^3 \cdot s^{-1})$ برای سیلاب دو ساعت بعد بهترین نتیجه را ارائه نمود. گرچه مدل درختی M5 دقت نسبتاً کمتری نسبت به روش رگرسیون ماشین بردار پشتیبان داشت، ولی به لحاظ ارائه روابط خطی ساده و قابل فهم می‌تواند به‌عنوان یک روش کاربردی در پیش‌بینی دبی سیلاب‌های ساعتی مورد استفاده قرار گیرد.

واژه‌های کلیدی: آذربایجان شرقی، دبی، داده‌کاوی، رگرسیون ماشین بردار پشتیبان، مدل درختی M5

مقدمه

زمینه‌های مختلف و به ویژه علوم مهندسی آب، تمایل به استفاده از روش‌های مبتنی بر داده مورد توجه محققان قرار گرفته است. کارشناسان و مدیران بخش آب عموماً تمایل به استفاده از مدل‌هایی دارند که در عین سادگی هم نتایج دقیقی داشته و هم قابل فهم‌تر و کاربردی‌تر باشد. امروزه استفاده از روش‌های داده‌کاوی از جمله روش‌های ماشین بردار پشتیبان^۱ (SVM)، مدل درختی M5، شبکه‌های عصبی مصنوعی، فازی، برنامه‌ریزی ژنتیک و ... به‌طور گسترده

امروزه پیش‌بینی سیلاب جهت مدیریت منابع آب، کنترل فرسایش، طراحی سازه‌های هیدرولیکی و کاهش خسارات ناشی از سیلاب از اهمیت خاصی برخوردار است. پیش‌بینی‌های دقیق در مقیاس‌های زمانی ساعتی اطلاعات مفیدی در اختیار دست اندرکاران قرار می‌دهد تا با برنامه‌ریزی دقیق‌تر، خسارات احتمالی سیل را به حداقل رسانده و مدیریت مطلوب‌تری بر مخازن سدها داشته باشند. در سال‌های اخیر، با افزایش امکانات سخت‌افزاری و نرم‌افزاری در

¹ Support Vector Machine

شده نشان داد، تغییرات دبی در ۲۴ ساعت گذشته می‌تواند اطلاعات بیشتری در مورد دینامیک پسرقت جریان رودخانه در هفت روز بعد ارائه کند.

Alikhanineghad و همکاران (۲۰۱۳) جهت پیش‌بینی سطح آب‌های زیرزمینی دشت کرمان از روش ماشین بردار پشتیبان به‌عنوان یک روش محاسباتی هوشمند استفاده نمودند. در این تحقیق از داده‌های ماهانه دما، رطوبت، بارندگی و ارتفاع پیژومتریک در خلال سال‌های ۱۹۸۶ تا ۲۰۱۰ استفاده شد. مقایسه نتایج این روش با نتایج شبکه‌های عصبی مصنوعی تابع پایه شعاعی نشان داد که ماشین‌های بردار پشتیبان در پیش‌بینی سطح آب زیرزمینی دارای کارایی بهتر می‌باشد. Seefi و همکاران (۲۰۱۳) از مدل ماشین بردار پشتیبان حداقل مربعات^۳ (LSSVM) برای پیش‌بینی تبخیر-تعرق مرجع در شهر کرمان استفاده کردند.

نتایج مقایسه این روش با دو مدل شبکه‌های عصبی مصنوعی (ANN) و سیستم استنتاج تطبیقی عصبی فازی (ANFIS) نشان داد که مدل‌های ANN و LSSVM از دقت بالاتری نسبت به ANN برخوردار بودند. همچنین تحلیل عدم قطعیت مونت کارلو نتایج این سه روش نشان داد که پیش‌بینی‌های مدل LSSVM عدم قطعیت کمتری نسبت به مدل‌های ANN و ANFIS داشت. Zahiri و Ghorbani (۲۰۱۳) با استفاده از روش‌های تجربی و مدل درختی M5 اقدام به شبیه‌سازی دبی جریان در مقاطع مرکب با استفاده از ۴۰۰ داده دبی-اشل از ۳۰ مقطع مرکب آزمایشگاهی و رودخانه‌ای نمودند. مقایسه آماری نتایج با استفاده از ضریب تبیین (R^2) و ریشه میانگین مربعات خطا (RMSE) نشان از دقت بالای مدل درختی M5 داشت.

Adamowski و Prasher (۲۰۱۲) برای پیش‌بینی جریان روزانه در حوضه کوهستانی Sianji واقع در منطقه هیمالیای هندوستان از دو روش رگرسیون ماشین بردار پشتیبان (SVR) و شبکه‌های موجک (WN) استفاده کردند. مقایسه نتایج نشان داد که گرچه هر دو روش از دقت خوبی در پیش‌بینی جریان روزانه برخوردار بودند، اما دقت روش شبکه‌های

برای طبقه‌بندی و پیش‌بینی پارامترهای مختلف هیدرولوژیک مورد توجه قرار گرفته است.

Bhattacharya و همکاران (۲۰۰۳) قابلیت مدل‌های درختی^۱ و شبکه عصبی مصنوعی^۲ را در مدل‌سازی روابط دبی-اشل با روابط (منحنی‌های) سنتی مقایسه کردند. نتایج حاکی از برتری دو روش تصمیم‌گیری هوشمند درختی و شبکه عصبی بود، با این تفاوت که در جریان‌های با دبی بالا، مدل درختی پیش‌بینی‌های بسیار دقیق‌تری نسبت به شبکه عصبی ارائه داد. Solomatine و Xue (۲۰۰۴) از مدل‌های درختی M5 و شبکه‌های عصبی مصنوعی جهت پیش‌بینی سیلاب در بالادست رودخانه Huai چین استفاده کردند. آن‌ها نتیجه گرفتند، گرچه هر دو مدل نتایج قابل قبولی ارائه می‌دهند، اما مدل درختی به علت آموزش سریع داده‌ها، سادگی و قابل فهم‌تر بودن نسبت به شبکه‌های عصبی مصنوعی، به‌وسیله تصمیم‌گیران ترجیح داده می‌شود.

Bhattacharya و Solomatine (۲۰۰۵) اقدام به مدل‌سازی رابطه دبی-اشل با استفاده از شبکه عصبی مصنوعی و مدل M5 نمودند. نتایج نشان داد که روش درختی M5 با دقت نسبتاً بیشتری رابطه دبی-اشل را مدل‌سازی نمود. Lin و همکاران (۲۰۰۶) اقدام به پیش‌بینی جریان طولانی‌مدت رودخانه لان‌کانگ واقع در کشور چین با استفاده از روش‌های ماشین بردار پشتیبان، شبکه‌های عصبی مصنوعی و مدل سری زمانی ARMA نمودند. نتایج این تحقیق نشان داد که روش ماشین بردار پشتیبان دارای دقت بالاتری بود. Yu و همکاران (۲۰۰۶) از روش ماشین بردار پشتیبان جهت پیش‌بینی سیل ساعتی رودخانه لان‌یانگ در تایوان استفاده نمودند. نتایج حاصل حاکی از دقت مناسب و قابل قبول روش مذکور در پیش‌بینی رویدادهای سیل در یک تا شش ساعت بعد داشت.

Brilly و Stravs (۲۰۰۷) از مدل درختی M5 جهت پیش‌بینی جریان‌های کم هفت روزه در هفت سرشاخه رودخانه ساوا در کشور اسلوانیا استفاده کردند. در این مطالعه از داده‌های سال‌های ۱۹۹۱ تا ۲۰۰۳ استفاده شده بود. ساختار مدل درختی ارائه

¹ Decision tree

² Artificial Neural Networks

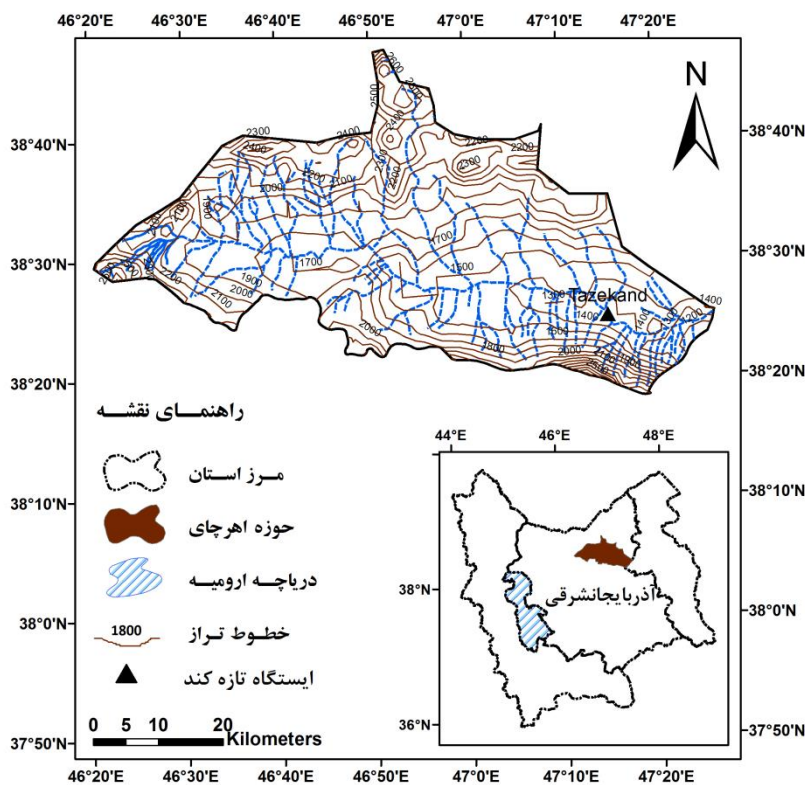
³ Least Square Support Vector Machine

مواد و روش‌ها

منطقه مورد پژوهش: حوزه آبخیز اهرچای واقع در شمال غرب ایران با مساحتی بالغ بر ۲۴۰۰ کیلومتر مربع از زیرحوضه‌های ارس محسوب می‌شود. موقعیت جغرافیایی این حوضه بین $21^{\circ} 46'$ تا 31° عرض 47° طول شرقی و $18^{\circ} 37'$ تا $44^{\circ} 38'$ عرض شمالی واقع شده و از لحاظ ارتفاعی بین ارتفاعات کوه کسبه به ارتفاع ۳۱۴۹ متر و محل ورود به رودخانه قره‌سو به ارتفاع ۸۷۵ متر قرار گرفته است. در این مطالعه از داده‌های ایستگاه تازه‌کند استفاده شد. شکل ۱ موقعیت جغرافیایی حوزه آبخیز اهرچای و ایستگاه تازه‌کند را نشان می‌دهد.

موجک اندکی بیشتر بود. Sattari و همکاران (۲۰۱۳) با استفاده از مدل درختی M5 اقدام به پیش‌بینی جریان تا هفت روز بعد در رودخانه سوهو در ترکیه نمودند.

مقایسه روش مذکور با روش ماشین بردار پشتیبان نشان داد که روش M5 کاربردی‌تر بوده و می‌توان با استفاده از سری زمانی کوتاه‌تر، به نتایج بهتری دست یابد. هدف از تحقیق حاضر، بررسی رویدادهای سیلابی و پیش‌بینی سیلاب‌های ساعتی رودخانه اهرچای در ایستگاه تازه‌کند به کمک روش‌های ماشین بردار پشتیبان (SVM)، مدل درختی M5 و رگرسیون خطی و همچنین، مقایسه عملکرد سه روش مذکور می‌باشد.



شکل ۱- موقعیت جغرافیایی حوزه آبخیز اهرچای و ایستگاه تازه‌کند

قرار گرفت. خصوصیات فیزیکی حوزه آبخیز اهرچای در ایستگاه هیدرومتری تازه‌کند در جدول ۱ و خصوصیات آماری سیلاب‌های حوزه آبخیز اهرچای برای بخش‌های آموزش و آزمون مدل‌ها به ترتیب در جدول ۲ و ۳ ارائه شده است.

مدل درختی M5: درخت‌های تصمیم از نسل جدید تکنیک‌های داده‌کاوی به شمار می‌روند که در دو دهه

در این مطالعه، برای برآورد دقیق سیلاب‌های ساعتی در رودخانه اهرچای از داده‌های ساعتی دبی جریان و سطح جریان (اشل) اردیبهشت ماه ۱۳۸۸ استفاده شد. همچنین از داده‌های ۱۴ رویداد سیل ثبت شده در ایستگاه تازه‌کند جهت ایجاد مدل استفاده شد که ۱۰ رویداد با ۱۷۱ داده برای آموزش و چهار رویداد با ۶۴ داده برای آزمون مدل مورد استفاده

مشابه روش رگرسیون درختی، برای داده‌های بزرگ موثر است. با این وجود مدل درختی در حالت کلی دارای اندازه کوچک‌تر و دقت بیشتری نسبت به رگرسیون درختی است (Emamifar و Rahimi Khob، ۲۰۱۱). نمونه‌ای از ساختار سلسله مراتبی مدل درختی در شکل ۲ نشان داده شده است. همچنین، مثالی از تقسیم‌بندی فضای نمونه‌ها به وسیله الگوریتم مدل درختی M5 در شکل ۳ ارائه شده است.

جدول ۱- خصوصیات فیزیوگرافی حوضه اهرچای در ایستگاه تازه‌کند

مقدار	خصوصیت فیزیوگرافی
۲۰۱۱	مساحت (km ²)
۲۳۱	محیط (km)
۱/۴۴	کشیدگی حوضه (ضریب گراولپوس)
۲/۱	آبدهی ویژه (L.s ⁻¹ .km ⁻²)

اخیر توسعه زیادی یافته‌اند و از آن‌ها نیز می‌توان هم برای کشف و استخراج دانش از یک پایگاه داده‌ها و هم برای ایجاد مدل‌های پیش‌بینی استفاده نمود. درختان تصمیم قادر به تولید توصیفات قابل درک برای انسان، از روابط موجود در یک مجموعه داده هستند و می‌توانند برای دسته‌بندی و پیش‌بینی به‌کار روند (Mahjobi و Tajrishi، ۲۰۱۰). الگوریتم M5 از معمول‌ترین نوع درختان تصمیم‌گیری مورد استفاده جهت دسته‌بندی هستند. مدل درختی اولین بار در سال ۱۹۹۲ توسط کوین لان مطرح شد.

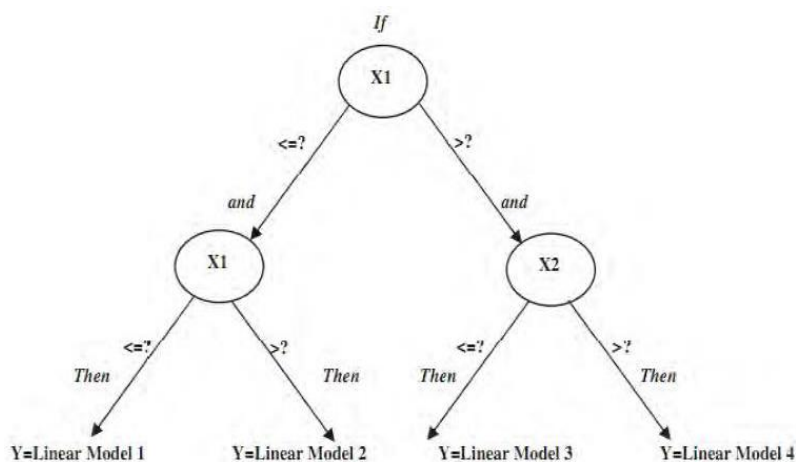
ساختار مدل درختی از یک درخت تصمیم با توابع رگرسیون‌گیری خطی به جای مقادیر نهایی در برگ‌ها، تشکیل می‌شود. مدل درختی M5 یک الگوریتم پیش‌بینی کننده عددی است و گره‌های درخت بر اساس این ویژگی که دارای حداقل خطای انحراف از معیار مورد انتظار باشند، تعیین می‌شوند. مدل درختی

جدول ۲- خصوصیات آماری ایستگاه تازه‌کند در دوره آموزش

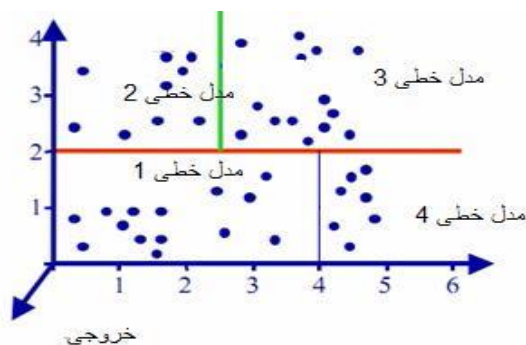
طول کل داده‌ها	میانگین	کمینه	بیشینه	پارامتر	رویداد
۱۳	۶۸	۵۷	۷۵	اشل (cm)	رویداد ۱
۱۳	۴/۴۲	۲/۵۳	۵/۶۳	دبی (m ³ .s ⁻¹)	
۲۰	۷۲/۲۵	۶۳	۸۰	اشل (cm)	رویداد ۲
۲۰	۵/۱۳	۳/۴۹	۶/۷۴	دبی (m ³ .s ⁻¹)	
۲۰	۸۳/۲۵	۶۱	۱۱۲	اشل (cm)	رویداد ۳
۲۰	۸/۱۷	۳/۰۸	۱۶/۲	دبی (m ³ .s ⁻¹)	
۱۴	۶۴/۲۸	۵۷	۷۰	اشل (cm)	رویداد ۴
۱۴	۳/۶۷	۲/۵۷	۴/۶۳	دبی (m ³ .s ⁻¹)	
۱۳	۶۴/۳۰	۵۷	۷۰	اشل (cm)	رویداد ۵
۱۳	۳/۶۷	۲/۵۱	۴/۶۳	دبی (m ³ .s ⁻¹)	
۱۰	۶۶/۲	۶۰	۷۰	اشل (cm)	رویداد ۶
۱۰	۳/۹۸	۳	۴/۶۳	دبی (m ³ .s ⁻¹)	
۲۱	۸۰	۶۴	۹۴	اشل (cm)	رویداد ۷
۲۱	۷/۲۲	۳/۷۵	۱۰/۴	دبی (m ³ .s ⁻¹)	
۲۰	۸۱/۶	۶۰	۱۰۹	اشل (cm)	رویداد ۸
۲۰	۷/۶۹	۲/۹۲	۱۵/۲	دبی (m ³ .s ⁻¹)	
۱۸	۷۷/۵	۶۰	۹۸	اشل (cm)	رویداد ۹
۱۸	۶/۵۴	۲/۹۲	۱۱/۶	دبی (m ³ .s ⁻¹)	
۱۶	۷۲/۶۲	۶۰	۸۶	اشل (cm)	رویداد ۱۰
۱۶	۵/۳۴	۳	۸/۲	دبی (m ³ .s ⁻¹)	

جدول ۳- خصوصیات آماری ایستگاه تازه‌کند در دوره آزمون

طول کل داده‌ها	انحراف معیار	کمینه	بیشینه	پارامتر	رویداد
۱۵	۷۱/۵۳	۶۰	۸۴	اشل (cm)	رویداد ۱۱
۱۵	۵/۰۹	۲/۹۲	۷/۶۹	دبی ($m^3.s^{-1}$)	
۱۴	۶۵/۲۱	۵۹	۷۲	اشل (cm)	رویداد ۱۲
۱۴	۳/۸۳	۲/۸۵	۵/۰۲	دبی ($m^3.s^{-1}$)	
۱۹	۷۳/۵۲	۵۷	۸۸	اشل (cm)	رویداد ۱۳
۱۹	۵/۵۷	۲/۵۷	۸/۷۲	دبی ($m^3.s^{-1}$)	
۱۲	۶۳/۵	۵۷	۶۸	اشل (cm)	رویداد ۱۴
۱۲	۳/۵	۲/۵	۴/۲۵	دبی ($m^3.s^{-1}$)	



شکل ۲- ساختار سلسله مراتبی مدل درختی M5



شکل ۳- تقسیم‌بندی فضای نمونه‌ها به وسیله الگوریتم مدل درختی M5

طبقه‌بندی دوتائی در فضای ویژگی‌های دلخواه است و از این روشی مناسب برای مسائل پیش‌بینی به شمار می‌رود (Pai و Hong، ۲۰۰۷). ماشین بردار پشتیبان در اصل یک دسته‌بندی کننده دو کلاسی است که کلاس‌ها را به وسیله یک مرز خطی از هم جدا می‌کند. در این روش نزدیک‌ترین نمونه‌ها به مرز تصمیم‌گیری

ماشین بردار پشتیبان: ماشین بردار پشتیبان یکی از روش‌های یادگیری تحت نظارت است که هم برای دسته‌بندی و هم رگرسیون قابل استفاده است. این روش توسط Vapnik (۱۹۹۸) بر پایه تئوری یادگیری آماری^۱ بنا نهاده شده است. SVM روشی برای

¹ Statistical Learning Theory

به این پارامترها تابع کرنلی را انتخاب نمود که توانایی آموزش برای ورودی‌های مساله را داشته باشد. در عمل چهار نوع کرنل خطی، کرنل چند جمله‌ای، کرنل تانژانت هیپربولیک و کرنل گوسی (RBF) به کار گرفته می‌شوند. در جدول ۴ معادلات برخی کرنل‌های رایج ارائه شده‌اند.

در نهایت، تابع تصمیم رگرسیون بردار پشتیبان غیرخطی به صورت معادله زیر خواهد بود (Yu و همکاران، ۲۰۰۶).

$$f(x_i) = \sum_{i=1}^l (-\partial_i - \partial_i^*) K(x_i, x_j) + b \quad (4)$$

در این تحقیق، برای پیش‌بینی سیلاب ساعتی در رودخانه اهرچای و در ایستگاه هیدرومتری تازه‌کند از نرم‌افزار WEKA¹ که در دانشگاه Waikato نیوزیلند توسعه یافته است، استفاده شد. این نرم‌افزار حاوی مجموعه‌ای از به‌روزترین الگوریتم‌های یادگیری ماشینی و ابزارهای پیش‌پردازش داده‌ها می‌باشد.

جدول ۴- توابع کرنل رایج در ماشین‌های بردار پشتیبان (Hamel, ۲۰۰۹)

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

رگرسیون خطی: تحلیل رگرسیون این امکان را فراهم می‌آورد تا تغییرات متغیر وابسته را از طریق متغیرهای مستقل پیش‌بینی و سهم هر یک از متغیرهای مستقل را در تبیین متغیر وابسته تعیین شود. در واقع مدل‌های بر مبنای رگرسیون یک نگاشت خطی یا غیرخطی از یک متغیر وابسته و یک یا چند متغیر مستقل ایجاد می‌کنند. عوامل مهم در ایجاد یک رابطه رگرسیونی عبارتند از، (۱) انتخاب مناسب‌ترین تعداد متغیر مستقل، (۲) پی بردن به نوع همبستگی بین متغیرهای مستقل و وابسته و (۳) انتخاب موثرترین متغیرهای مستقل، هنگامی که تعداد

را بردارهای پشتیبان می‌نامند. این بردارها معادله مرز تصمیم‌گیری را مشخص می‌کنند. الگوریتم‌های شبیه‌سازی هوشمند کلاسیک مانند شبکه‌های عصبی مصنوعی معمولاً قدرمطلق خطا یا مجموع مربعات خطای داده‌های آموزشی را حداقل می‌کنند، ولی مدل‌های SVM اصل حداقل‌سازی خطای ساختاری را به کار می‌گیرند (Hamel, ۲۰۰۹).

Vapnik (۱۹۹۸) با استفاده از مفهوم حاصل‌ضرب داخلی نشان داد که ابتدا می‌توان بردار ورودی x را با یک تبدیل غیرخطی به یک فضای با بعد زیاد انتقال داد و در آن فضا حاصل‌ضرب داخلی را انجام داد و ثابت کرد که اگر یک کرنل متقارن، شرایط قضیه مرسر را داشته باشد، اعمال این کرنل در فضای ورودی با بعد کم می‌تواند حاصل‌ضرب را به شدت کاهش دهد. اگر بردار ورودی x_i به وسیله تابع غیرخطی $\Phi(x)$ داخل فضای ویژه نگاشت شود، تابع تصمیم به شکل زیر خواهد بود.

$$f(w x) = w \cdot \Phi(x) + b \quad (1)$$

که در آن، w و b بردارهای پارامتر تابع می‌باشند. مسئله رگرسیون غیرخطی می‌تواند مانند مسئله بهینه‌سازی به صورت زیر بیان شود.

$$\min \frac{1}{2} \sum_{ij} (\partial_i - \partial_i^*)(\partial_j - \partial_j^*) \langle \Phi(x_i), \Phi(x_j) \rangle + \varepsilon \sum_{i=1}^l (\partial_i - \partial_i^*) - \sum_{i=1}^l y_i (\partial_i - \partial_i^*) \quad (2)$$

با محدودیت‌های زیر بیان شود (Hamel, ۲۰۰۹).

$$\sum_{ij} (\partial_i - \partial_i^*) = 0$$

$$0 \leq \partial_i \leq C, \quad i = 1, 2, \dots, l \quad (3)$$

$$0 \leq \partial_i^* \leq C, \quad i = 1, 2, \dots, l$$

که در آن، ε تولرانس خطا، ∂_i و ∂_i^* ضرائب لاگرانژ و $\langle \Phi(x_i), \Phi(x_j) \rangle$ ضرب داخلی توابع $\Phi(x_i)$ و $\Phi(x_j)$ می‌باشند. اطلاعات کمی در مورد انتخاب تابع غیرخطی مناسب $\Phi(x)$ در دسترس می‌باشد. به علاوه، محاسبه $\Phi(x_i)$ و $\Phi(x_j)$ در فضای ویژه بسیار پیچیده می‌باشد. ماشین‌های بردار پشتیبان برای حل مسائل غیرخطی، ابعاد مساله را از طریق توابع کرنل $K(x_i, x_j) = \langle \Phi(x_i), \Phi(x_j) \rangle$ تغییر می‌دهند. انتخاب کرنل برای SVM به حجم داده‌های آموزشی و ابعاد بردار ویژگی بستگی دارد. به عبارت دیگر، باید با توجه

¹ Waikato Environment for Knowledge Analysis

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (Q_i - \hat{Q}_i)^2}{n}} \quad (8)$$

که در آن، Q_i مقدار مشاهداتی، \hat{Q}_i مقدار محاسباتی، n تعداد مشاهدات، \bar{Q} میانگین داده‌های مشاهداتی و \hat{Q} میانگین مقادیر محاسباتی می‌باشند.

نتایج و بحث

در این تحقیق برای پیش‌بینی سیلاب ساعتی، ابتدا رویدادهای سیلابی مربوط به دوره‌های آموزش و آزمون در ایستگاه هیدرومتری تازه‌کند مشخص شد (شکل ۴). برای مشخص کردن پارامترهای ورودی مدل و تعریف سناریوهای مختلف، همبستگی زمانی بین دبی جریان و سطح جریان (اشل) در ساعات قبل و ساعت جاری و ترکیب اشل در ساعات قبل محاسبه شد (جدول ۵).

چنان که از جدول ۵ استنباط می‌شود، مقادیر جریان (Q) حداقل تا شش ساعت قبل و سطح جریان یا اشل (E) تا هفت ساعت قبل از همبستگی بالایی برخوردار است. بر اساس ضرایب همبستگی با تاخیرهای زمانی ساعتی، ترکیب‌های مختلفی از پارامترها جهت پیش‌بینی دبی سیلاب یک ساعت بعد $Q(t+1)$ و دو ساعت بعد $Q(t+2)$ رودخانه اهرچای در ایستگاه هیدرومتری تازه‌کند طبق جدول ۶ در قالب ۲۲ سناریو ارائه شد.

زیادی متغیر مستقل موجود باشد. در رگرسیون خطی، متغیر وابسته y_i ترکیب خطی‌ای از ضرایب (پارامترها) است (که لازم نیست نسبت به متغیرهای مستقل خطی باشند). مدل رگرسیون خطی با N متغیر مستقل x_i و ضرایب خطی β_0 و β_i به صورت زیر تعریف می‌شود.

$$y_i = \beta_0 + \sum_{i=1}^N \beta_i x_i + \varepsilon_i, \quad i=1, \dots, N \quad (5)$$

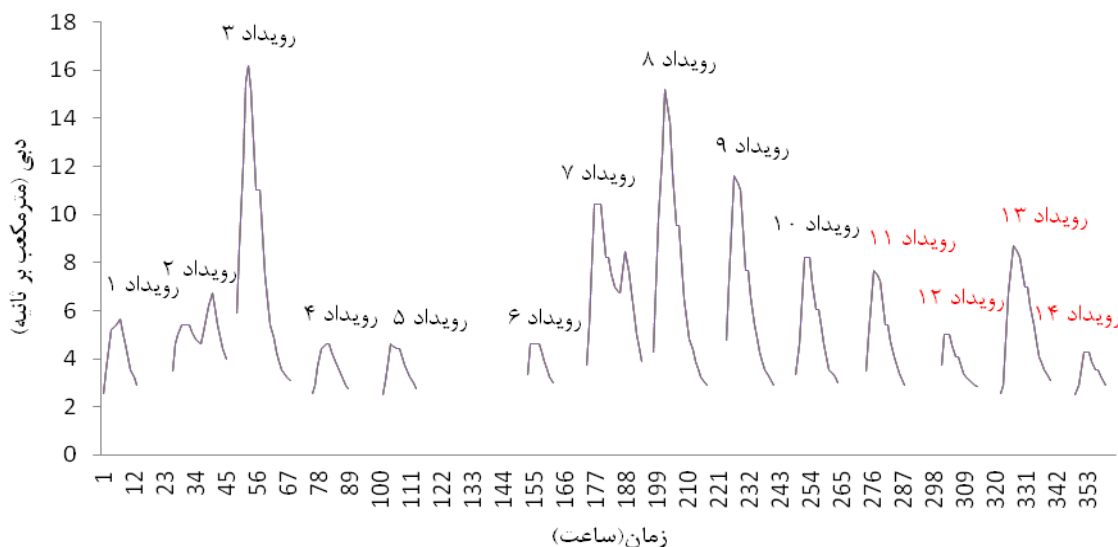
که در آن، ε_i مقدار خطای مشاهده i ام و y_i متغیر وابسته می‌باشد. با داشتن مجموعه‌ای از مشاهدات می‌توان ضرایب رگرسیون را تخمین و مدل را به دست آورد.

$$y_i = \hat{\beta}_0 + \sum_{i=1}^N \hat{\beta}_i X_i + e_i \quad (6)$$

که در آن، $e_i = y_i - \hat{y}_i$ مانده دارد. روش رایج برای تخمین پارامترها، روش حداقل مربعات می‌باشد که در آن پارامترها معادله رگرسیون با کمینه کردن مجموع مانده‌ها به دست می‌آیند.

معیارهای ارزیابی: جهت ارزیابی مدل‌های مورد استفاده در این مطالعه، از شاخص‌های ضریب تبیین (R^2) و جذر میانگین مربعات خطا (RMSE) استفاده شده که به ترتیب از روابط زیر قابل محاسبه می‌باشند.

$$R^2 = \left[\frac{\sum_{i=1}^n (Q_i - \bar{Q})(\hat{Q}_i - \bar{\hat{Q}})}{\sqrt{\sum_{i=1}^n (Q_i - \bar{Q})^2 \sum_{i=1}^n (\hat{Q}_i - \bar{\hat{Q}})^2}} \right]^2 \quad (7)$$



شکل ۴- رویدادهای سیلابی مربوط به دوره‌های آموزش و آزمون در ایستگاه هیدرومتری تازه‌کند

جدول ۵- همبستگی زمانی دبی و اشل در ساعات قبل برای ایستگاه تازه‌کند

ضریب همبستگی	تاخیر ساعتی	ورودی
۰/۹۶	۱	اشل
۰/۸۹	۲	
۰/۷۹	۳	
۰/۶۸	۴	
۰/۵۸	۵	
۰/۴۹	۶	
۰/۹۷	۱	دبی
۰/۹۱	۲	
۰/۸۴	۳	
۰/۷۶	۴	
۰/۶۸	۵	
۰/۶۰	۶	
۰/۵۴	۷	

جدول ۶- سناریوهای مورد استفاده برای پیش‌بینی دبی سیلاب ساعتی در ایستگاه تازه‌کند

سناریو	ورودی‌ها
S1	Q(t)
S2	Q(t), Q(t-1)
S3	Q(t), Q(t-1), Q(t-2)
S4	Q(t), Q(t-1), Q(t-2), Q(t-3)
S5	Q(t), Q(t-1), Q(t-2), Q(t-3), Q(t-4)
S6	Q(t), Q(t-1), Q(t-2), Q(t-3), Q(t-4), Q(t-5)
S7	Q(t), Q(t-1), Q(t-2), Q(t-3), Q(t-4), Q(t-5), Q(t-6)
S8	E(t)
S9	E(t), E(t-1)
S10	E(t), E(t-1), E(t-2)
S11	E(t), E(t-1), E(t-2), E(t-3)
S12	E(t), E(t-1), E(t-2), E(t-3), E(t-4)
S13	E(t), E(t-1), E(t-2), E(t-3), E(t-4), E(t-5)
S14	E(t), E(t-1), E(t-2), E(t-3), E(t-4), E(t-5), E(t-6)
S15	E(t), E(t-1), E(t-2), E(t-3), E(t-4), E(t-5), E(t-6), E(t-7)
S16	E(t), Q(t)
S17	E(t), E(t-1), Q(t), Q(t-1)
S18	E(t), E(t-1), E(t-2), Q(t), Q(t-1), Q(t-2)
S19	E(t), E(t-1), E(t-2), E(t-3), Q(t), Q(t-1), Q(t-2), Q(t-3)
S20	E(t), E(t-1), E(t-2), E(t-3), E(t-4), Q(t), Q(t-1), Q(t-2), Q(t-3), Q(t-4)
S21	E(t), E(t-1), E(t-2), E(t-3), E(t-4), E(t-5), Q(t), Q(t-1), Q(t-2), Q(t-3), Q(t-4), Q(t-5)
S22	E(t), E(t-1), E(t-2), E(t-3), E(t-4), E(t-5), E(t-6), Q(t), Q(t-1), Q(t-2), Q(t-3), Q(t-4), Q(t-5), Q(t-6)

ساعتی با روش درخت تصمیم M5 که مربوط به سناریوی ۲۲ می‌باشد، در جدول ۷ ارائه شده است. با توجه به جدول ۷ در روش مدل درخت تصمیم M5 سناریوی S22 شامل ترکیب داده‌های دبی و اشل

در این تحقیق، در طی فرایند مدل‌سازی از همه سناریوهای تعریف شده استفاده شد و بر اساس معیارهای ارزیابی سناریوهای منتخب در هر دو روش مشخص شد. بهترین نتایج پیش‌بینی دبی سیلاب

دبی سیلاب یک ساعت بعد $Q(t+1)$ در شکل ۵ و روابط خطی متناظر با آن در جدول ۸ ارائه شده است. همان‌گونه که در جدول ۸ آورده شده است، مدل M5 می‌تواند رویدادهای سیلابی ایستگاه تازه‌کند اهرچای در یک ساعت بعد را با کمک پنج رابطه خطی پیش‌بینی نماید.

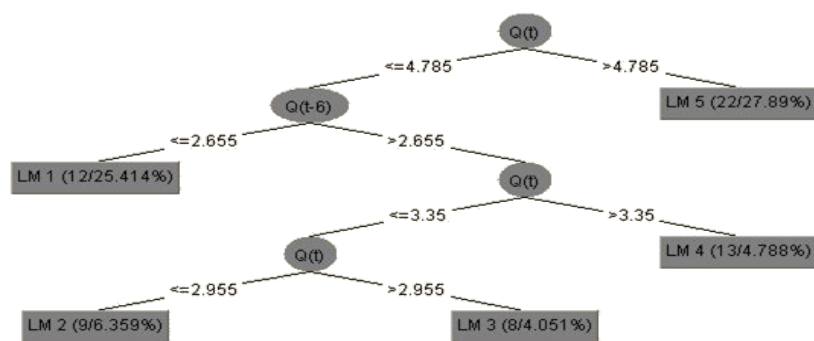
ساعت جاری و دبی و اشل تا شش ساعت قبل مربوط به رویداد ۱۳ با $R^2=0/93$ و $(m^3.s^{-1})$ و $RMSE=0/2946$ برای سیلاب یک ساعت بعد و رویداد ۱۴ با $R^2=0/89$ و $RMSE=0/0883 (m^3.s^{-1})$ برای سیلاب دو ساعت بعد، بهترین نتایج را ارائه نموده‌اند. ساختار درختی مدل M5 جهت پیش‌بینی

جدول ۷- نتایج بهترین سناریوی (شماره ۲۲) پیش‌بینی دبی سیلاب ساعتی با روش درخت تصمیم M5

Q(t+2)		Q(t+1)		سناریو	رویداد
RMSE	R ²	RMSE	R ²		
۰/۰۹۶۱	۰/۸۹	۰/۰۳۷۸	۰/۹۵	S22	رویداد ۱
۰/۲۱۵۰	۰/۶۱	۰/۰۴۶۰	۰/۸۴	S22	رویداد ۲
۱/۹۲۷۰	۰/۸۷	۰/۵۰۰۳	۰/۹۵	S22	رویداد ۳
۰/۱۱۵۰	۰/۸۱	۰/۰۳۴۰	۰/۹۳	S22	رویداد ۴
۰/۱۴۴۳	۰/۷۲	۰/۰۴۱۵	۰/۹۰	S22	رویداد ۵
۰/۳۰۷۳	۰/۵۳	۰/۰۸۵۹	۰/۸۲	S22	رویداد ۶
۰/۶۷۳۸	۰/۷۲	۰/۲۲۴۴	۰/۸۹	S22	رویداد ۷
۱/۲۴۶۹	۰/۸۹	۰/۴۶۱۹	۰/۹۴	S22	رویداد ۸
۰/۳۴۲۴	۰/۹۳	۰/۲۱۳۶	۰/۹۵	S22	رویداد ۹
۰/۸۵۷۹	۰/۷۵	۰/۱۵۱۶	۰/۹۲	S22	رویداد ۱۰
۰/۲۹۵۲	۰/۸۲	۰/۲۴۹۱	۰/۸۴	S22	رویداد ۱۱
۰/۰۸۹۴	۰/۸۹	۰/۰۸۶۸	۰/۷۹	S22	رویداد ۱۲
۰/۴۷۸۵	۰/۸۷	۰/۲۹۴۶	۰/۹۳	S22	رویداد ۱۳
۰/۰۸۸۳	۰/۸۹	۰/۰۳۰۲	۰/۸۶	S22	رویداد ۱۴

و دبی همان ساعت و شش ساعت قبل می‌باشد، در رویداد ۱۱ با $R^2=0/96$ و $RMSE=0/0472 (m^3.s^{-1})$ برای پیش‌بینی دبی سیلاب یک ساعت بعد و $R^2=0/90$ و $RMSE=0/1596 (m^3.s^{-1})$ برای پیش‌بینی دبی سیلاب دو ساعت بعد، بهترین نتایج را ارائه داده است.

در جدول ۹ بهترین نتایج مربوط به پیش‌بینی دبی سیلاب در یک ساعت و دو ساعت آتی با روش ماشین بردار پشتیبان ارائه شده است. از جدول ۹ استنباط می‌شود که جهت پیش‌بینی دبی سیلاب با روش ماشین بردار پشتیبان، سناریوی S22 که شامل ترکیب داده‌های اشل همان ساعت و شش ساعت قبل



شکل ۵- ساختار درختی ارائه شده به وسیله مدل M5 برای سناریوی S22

جدول ۸- روابط خطی ارائه شده به وسیله مدل درخت تصمیم M5 برای سناریوی S22

LM num: 1	LM num: 2	LM num: 3	LM num: 4	LM num: 5
$Q(t+1) =$	$Q(t+1) =$	$Q(t+1) =$	$Q(t+1) =$	$Q(t+1) =$
$0.4396 * Q(t)$	$0.4407 * Q(t-1)$	$0.4407 * Q(t-1)$	$0.1462 * Q(t)$	$0.9098 * Q(t)$
$+ 0.5304 * Q(t-1)$	$- 0.2068 * Q(t-2)$	$- 0.2054 * Q(t-2)$	$+ 0.4582 * Q(t-1)$	$- 0.389 * Q(t-1)$
$- 0.3988 * Q(t-2)$	$+ 0.1595 * Q(t-3)$	$+ 0.1595 * Q(t-3)$	$- 0.2393 * Q(t-2)$	$+ 0.2458 * Q(t-3)$
$- 0.0304 * Q(t-3)$	$- 0.1543 * Q(t-6)$	$- 0.1543 * Q(t-6)$	$+ 0.1595 * Q(t-3)$	$- 0.2376 * Q(t-6)$
$+ 0.61 * Q(t-6)$	$+ 2.6412$	$+ 2.6596$	$- 0.1543 * Q(t-6)$	$+ 2.0925$
$+ 0.6495$			$+ 2.3009$	

جدول ۹- نتایج بهترین سناریوی (شماره ۲۲) پیش‌بینی دبی سیلاب ساعتی با ماشین بردار پشتیبان

Q(t+2)		Q(t+1)		سناریو	رویداد
RMSE	R ²	RMSE	R ²		
۰/۰۴۸۸	۰/۹۱	۰/۰۳۴۳	۰/۹۴	S 22	رویداد ۱
۰/۱۸۸۱	۰/۵۹	۰/۰۶۱۴	۰/۸۳	S 22	رویداد ۲
۲/۴۳۷۹	۰/۸۰	۰/۴۷۶۳	۰/۹۵	S 22	رویداد ۳
۰/۰۵۵۹	۰/۸۲	۰/۰۲۳۶	۰/۹۳	S 22	رویداد ۴
۰/۰۷۳۵	۰/۸۷	۰/۰۳۳۶	۰/۹۱	S 22	رویداد ۵
۱/۷۹۳۵	۰/۶۱	۰/۱۱۶۳	۰/۸۰	S 22	رویداد ۶
۰/۹۲۳۱	۰/۶۷	۰/۲۳۶۱	۰/۸۹	S 22	رویداد ۷
۱/۷۰۱۳	۰/۸۴	۰/۳۶۲۴	۰/۹۶	S 22	رویداد ۸
۰/۵۵۰۷	۰/۹۰	۰/۱۳۸۱	۰/۹۷	S 22	رویداد ۹
۰/۵۲۶۶	۰/۷۷	۰/۱۴۱۹	۰/۸۴	S 22	رویداد ۱۰
۰/۱۵۹۶	۰/۹۰	۰/۰۴۷۲	۰/۹۶	S 22	رویداد ۱۱
۰/۱۰۴۴	۰/۸۹	۰/۰۵۲۳	۰/۹۰	S 22	رویداد ۱۲
۰/۳۸۹۴	۰/۸۸	۰/۰۹۱۵	۰/۹۶	S 22	رویداد ۱۳
۰/۰۵۷۲	۰/۷۹	۰/۰۱۶۷	۰/۹۳	S 22	رویداد ۱۴

برای یک ساعت بعد و دو ساعت بعد به ترتیب در شکل ۶-الف و ۶-ب نشان داده شده است. چنان‌که از شکل ۶ استنباط می‌شود، روش‌های درخت تصمیم M5 و ماشین بردار پشتیبان با دقت نسبتاً خوبی نسبت به روش رگرسیون خطی قادر به پیش‌بینی سیلاب ساعتی می‌باشند. مقادیر مشاهداتی و پیش‌بینی شده سیلاب با استفاده از روش‌های مختلف در دوره آزمون رویداد ۱۱ برای یک ساعت بعد و دو ساعت بعد به ترتیب در شکل ۷-الف و ۷-ب نشان داده شده است. همان‌گونه که در این شکل مشاهده می‌شود، ماشین بردار پشتیبان بیشترین دقت و رگرسیون خطی کمترین دقت را در پیش‌بینی دبی سیلاب یک و دو ساعت آینده نشان داده‌اند. به علاوه این که در هر سه روش مورد بررسی دقت نتایج برای پیش‌بینی دبی سیلاب یک ساعت آینده بیشتر از پیش‌بینی سیلاب دو ساعت آینده می‌باشد. همچنین، بررسی

برای سنجش قابلیت روش‌های یادگیری ماشین با استفاده از روش رگرسیون خطی اقدام به مدل‌سازی دبی سیلاب شد. نتایج حاصل از مدل‌سازی دبی سیلاب در دوره‌های آموزش و آزمون با استفاده از رگرسیون خطی در جدول ۱۰ ارائه شده است. چنان‌که در جدول ۱۰ مشخص شده است، در پیش‌بینی رویدادهای سیلابی ساعتی ایستگاه تازه‌کند با مدل رگرسیون خطی ساده، سناریوی S22 در رویداد ۱۳ برای پیش‌بینی دبی سیلاب یک ساعت بعد با $R^2=0/86$ و $RMSE=0/2954 (m^3.s^{-1})$ و برای پیش‌بینی دبی سیلاب دو ساعت بعد با $R^2=0/57$ و $RMSE=0/9811 (m^3.s^{-1})$ بهترین نتایج را ارائه داده است.

مقادیر مشاهداتی و پیش‌بینی شده دبی سیلاب با استفاده از روش‌های مختلف در دوره آزمون رویداد ۱۳

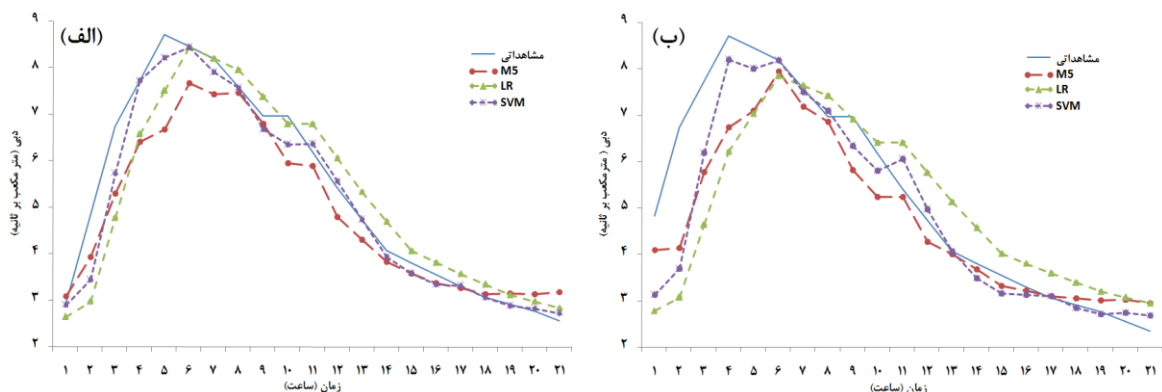
هیدروگراف‌های سیلاب در شکل‌های ۶ و ۷ نشان می‌دهد که دقت هر سه روش مورد بررسی در بازوی پایین‌رونده هیدروگراف بیشتر از نقطه اوج هیدروگراف و بازوی بالارونده هیدروگراف می‌باشد. مقادیر مشاهداتی و پیش‌بینی شده سیلاب با استفاده از مدل SVM در دوره آزمون رویداد ۱۱ برای یک ساعت بعد و دو ساعت بعد به ترتیب در شکل ۸-الف و ۸-ب نشان داده شده است. همان‌گونه که از این شکل پیداست، مدل SVM در پیش‌بینی دبی سیلاب ساعتی از دقت قابل قبولی برخوردار است و با افزایش دوره زمانی

پیش‌بینی از دقت نتایج کاسته می‌شود.

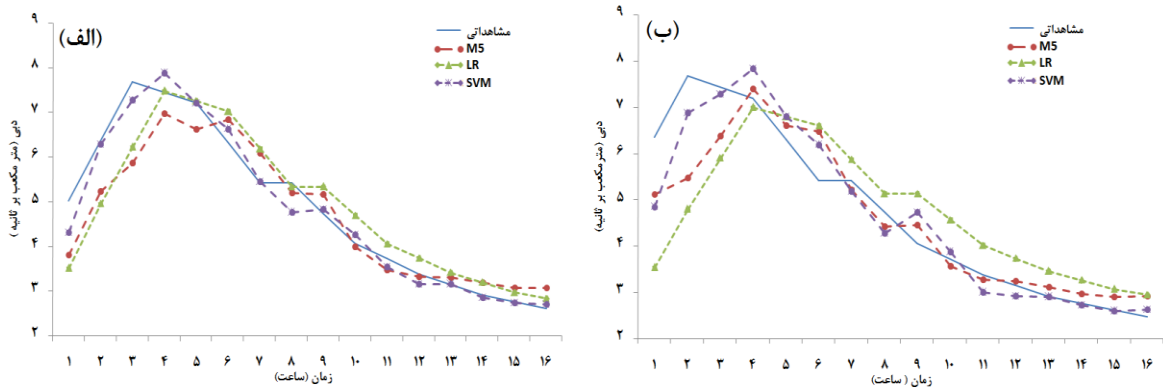
نتایج مطالعه حاضر با نتایج محققان دیگر از قبیل Yu و همکاران (۲۰۰۶) در پیش‌بینی دبی سیلاب ساعتی رودخانه لان‌یانگ در تایوان، Lin و همکاران (۲۰۰۶) در پیش‌بینی جریان طولانی‌مدت رودخانه لان‌کانگ در چین، Adamowski و Prasher (۲۰۱۲) در پیش‌بینی جریان روزانه حوضه کوهستانی Sianji در هندوستان و Sattari و همکاران (۲۰۱۳) در پیش‌بینی جریان رودخانه Sohu در ترکیه کاملاً منطبق می‌باشد.

جدول ۱۰- نتایج بهترین سناریوی (شماره ۲۲) پیش‌بینی سیلاب ساعتی با رگرسیون خطی (LR)

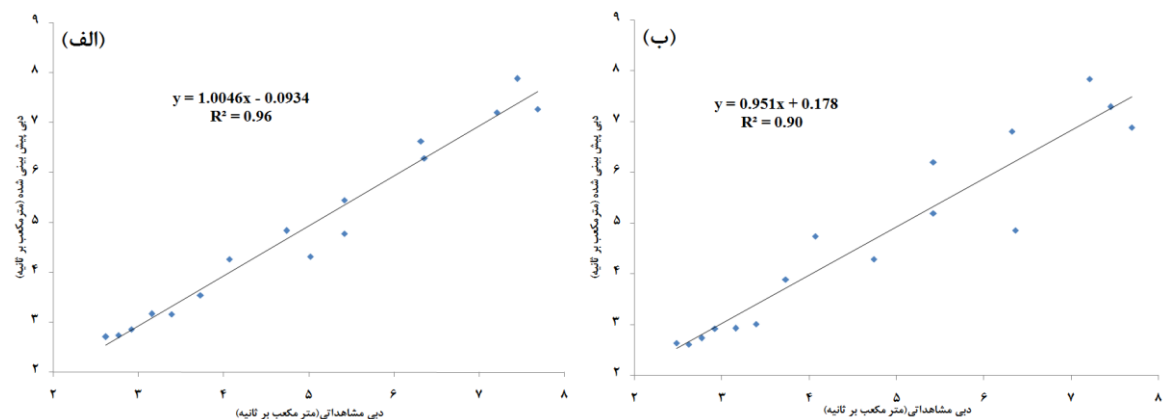
Q(t+2)		Q(t+1)		سناریو	رویداد
RMSE	R ²	RMSE	R ²		
۰/۴۳۳۲	۰/۳۲	۰/۱۴۰۸	۰/۷۲	S 22	رویداد ۱
۰/۲۹۷۵	۰/۲۰	۰/۱۱۰۸	۰/۶۰	S 22	رویداد ۲
۴/۳۴۴۷	۰/۵۷	۱/۳۵۸۵	۰/۸۶	S 22	رویداد ۳
۰/۲۳۴۵	۰/۳۳	۰/۰۶۸۱	۰/۷۵	S 22	رویداد ۴
۰/۲۵۹۱	۰/۲۵	۰/۰۷۹۲	۰/۶۹	S 22	رویداد ۵
۰/۴۱۳۶	۰/۲۰	۰/۱۵۱۹	۰/۵۵	S 22	رویداد ۶
۱/۶۱۲۴	۰/۳۲	۰/۵۳۹۸	۰/۷۲	S 22	رویداد ۷
۳/۳۳۴۱	۰/۶۳	۱/۰۹۳۳	۰/۸۷	S 22	رویداد ۸
۱/۹۱۶۵	۰/۶۰	۰/۶۷۷۷	۰/۸۵	S 22	رویداد ۹
۱/۰۹۷۱	۰/۴۵	۰/۳۵۹۶	۰/۸۰	S 22	رویداد ۱۰
۰/۷۵۷۵	۰/۵۲	۰/۲۷۱۵	۰/۸۰	S 22	رویداد ۱۱
۰/۱۳۵۴	۰/۶۵	۰/۰۸۳۹	۰/۷۶	S 22	رویداد ۱۲
۰/۹۸۱۱	۰/۵۷	۰/۲۹۵۴	۰/۸۶	S 22	رویداد ۱۳
۰/۱۶۴۹	۰/۲۲	۰/۰۵۴۱	۰/۶۷	S 22	رویداد ۱۴



شکل ۶- مقایسه مقادیر مشاهداتی و پیش‌بینی شده سیلاب در دوره آزمون برای رویداد ۱۳: الف- سیلاب یک ساعت بعد و ب- سیلاب دو ساعت بعد



شکل ۷- مقایسه مقادیر مشاهداتی و پیش‌بینی شده در دوره آزمون برای رویداد ۱۱، الف- سیلاب یک ساعت بعد و ب- سیلاب دو ساعت بعد



شکل ۸- نمودار پراکنش مقادیر مشاهداتی و پیش‌بینی شده در دوره آزمون برای رویداد سیلابی ۱۱ با مدل SVM، الف- سیلاب یک ساعت بعد و ب- سیلاب دو ساعت بعد

نتیجه‌گیری

در این مطالعه جهت پیش‌بینی دبی رویدادهای سیلاب ساعتی رودخانه اهرچای در ایستگاه هیدرومتری تازه‌کند از روش‌های ماشین بردار پشتیبان، مدل درختی M5 و رگرسیون خطی ساده استفاده شد. برای ایجاد مدل‌های مذکور از داده‌های تاریخی دبی-اشل ساعتی ایستگاه تازه‌کند و ۱۴ رویداد مهم سیل در این ایستگاه استفاده شد. نتایج بررسی‌ها نشان داد که روش ماشین بردار پشتیبان در پیش‌بینی سیلاب یک و دو ساعت آینده به ترتیب با $R^2=0/96$ و $R^2=0/90$ از دقت بالاتری نسبت به دو روش دیگر برخوردار بود، ولی روش درخت تصمیم M5 با ارائه روابط و قوانین ساده و قابل فهم‌تر، قابلیت خوبی در

پیش‌بینی سیلاب ساعتی از خود نشان داد. به‌طوری‌که مدل M5 توانست تنها با پنج رابطه خطی، میزان سیلاب یک و دو ساعت بعد ایستگاه تازه‌کند را به ترتیب با $R^2=0/93$ و $R^2=0/87$ پیش‌بینی کند. همچنین، نتایج نشان داد که از بین سه روش مورد بررسی روش رگرسیون خطی کمترین دقت را داشت و در هر سه روش با افزایش دوره زمانی پیش‌بینی سیلاب از یک ساعت به دو ساعت، دقت پیش‌بینی‌ها کاهش می‌یابد. با توجه به قابلیت‌های روش‌های ماشین بردار پشتیبان و مدل درختی M5 پیشنهاد می‌شود که این روش‌ها در پیش‌بینی سیلاب و پارامترهای هواشناسی مختلف در دیگر حوضه‌های کشور مورد استفاده قرار گیرد.

منابع مورد استفاده

1. Adamowski, J. and S.O. Prasher. 2012. Comparison of machine learning methods for runoff forecasting in mountainous watersheds with limited data. Journal of Water and Land Development,

- 17(7-8): 89–97.
2. Alikhanineghad, M., A.J. Kamali and B. Nematollahi. 2013. Ground water forecasting using support vector machine, case study: Kerman watershed. 2nd International Conference on Plant, Water, Soil and Weather Modeling, may 8-9 2013, Kerman, Iran (in Persian).
 3. Bhattacharya, B. and D.P. Solomatine. 2003. Neural networks and M5 model trees in modeling water level-discharge relationship for an Indian river. In European Symposium on Artificial Neural Networks, Bruges (Belgium), 23–25 April; 407–412.
 4. Bhattacharya, B. and D.P. Solomatine. 2005. Neural networks and M5 model trees in modeling water level–discharge relationship. *Neurocomputing*, 63: 381–396.
 5. Emamifar, S. and A. Rahimi Khob. 2011. Evaluation of M5 model tree and experimental model Angstrom for estimating radiation reaching the Earth's surface. 1st National Conference on Agricultural Meteorology and Water Management, College of Agriculture and Natural Resources, Tehran University, Tehran, Iran (in Persian).
 6. Hamel, L.H. 2009. Knowledge discovery with support vector machines. Wiley, 262 pages.
 7. Lin, J.Y., C.T. Cheng and K.W. Chau. 2006. Using support vector machines for long term discharge prediction. *Hydrological Sciences Journal*, 51(4): 599–612.
 8. Mahjobi, A. and M. Tajrishi. 2010. Comparison of artificial neural network algorithms and decision trees in predicting changes in river salinity, case study: Karun River. 4th Conference and Exhibition on Environmental Engineering, Tehran Iran (in Persian).
 9. Pai, P.F. and W.C. Hong. 2007. A recurrent support vector regression model in rainfall forecasting. *Hydrological Process*, 21: 819-827.
 10. Sattari, M.T., M. Pal, H. Apaydin and F. Ozturk. 2013. M5 model tree application in daily river flow forecasting in Sohu stream, Turkey. *Water Resources*, 40(3): 233–242.
 11. Seefi, A., M. Mirlatifi and H. Reahi. 2013. Introduction and application of Least Square Support Vector Machine (LSSVM) for simulation of reference evapotranspiration and uncertainty analysis of results, A case study of the Kerman city. *Journal of Irrigation and Water Engineering*, 13: 67-79 (in Persian).
 12. Solomatine, D.P. and Y. Xue. 2004. M5 model trees compared to neural networks: application to flood forecasting in the upper reach of the Huei River in china. *Journal of Hydrologic Engineering*, 9(6): 491-501.
 13. Stravs, L. and M. Brilly. 2007. Development of a low-flow forecasting model using the M5 machine learning method. *Hydrological Sciences Journal*, 52(3): 466-477.
 14. Vapnik, V.N. 1998. Statistical learning theory. Wiley, 736 pages.
 15. Yu, P.S., S.T. Chen and I.F. Chang. 2006. Support vector regression for real-time flood stage forecasting. *Journal of Hydrology*, 328: 704-716.
 16. Zahiri, A.R. and K.H. Ghorbani. 2013. Flow discharge prediction in compound channels by using decision model tree M5. *Journal of Water and Soil Conservation*, 20(3): 113-132 (in Persian).