

Evaluating the efficiency of SVM, LS-SVM and SVM-GOA models in simulating the Flood peak discharge at the Poldokhtar station

Abstract

In order to control and minimize the damaging impacts of floods, flood modeling or simulation is a fundamental solution. Identifying effective models for this purpose is crucial in watershed management. This study evaluates the accuracy of support vector machine models combined with the support vector machine (SVM), Grasshopper algorithm (SVM-GOA) and least square support vector machine (LS-SVM) in simulating the flood peak discharge of Poldokhtar station in the Karkheh basin. For this study, 74 flood events from 2009 to 2016 at the Poldokhtar station and data from 13 daily rainfall stations in the upstream area for the same period were utilized. Subsequently, 52 events were allocated for training, and 22 for validation. The comparison of results was conducted using three statistical indicators: Correlation coefficient (R^2), Root mean square error (RMSE), Nash efficiency (Ns), and Standard error (SE). Additionally, uncertainty analysis was performed using two indexes: ARIL and POC. The results indicate the relative superiority of the LS-SVM_model with $SE=0.407$, $RMSE=110.16$, $NS=0.91$ and $R^2=0.92$ compared to the SVM_model with $SE=0.5$, $RMSE=137.70$, $NS=0.87$ and $R^2=0.88$ and SVM-GOA model with $SE=0.519$, $RMSE=144.53$, $NS=0.83$ and $R^2=0.9$. The study's overall conclusion is that the LS-SVM model is more accurate, faster, and easier to implement compared to the SVM and SVM-GOA models. As a result, it can be confidently preferred over the SVM and SVM-GOA models due to its significant advantages. The research emphasizes the critical importance of precise flood modeling and simulation in watershed management for mitigating the destructive impact of floods.

Key words: *flood modeling, support vector machine, Grasshopper algorithm, Karkheh basin, Poldokhtar station.*

ارزیابی کارایی مدل‌های SVM، LS-SVM و SVM-GOA در شبیه‌سازی دبی اوج سیل ایستگاه پل

دختر

چکیده:

مدل‌سازی یا شبیه‌سازی سیل یکی از راهکارهای اساسی برای مدیریت و کاهش اثرات مخرب این پدیده بوده و شناسایی مدل‌هایی کارآمد بدین منظور، یکی از مهم‌ترین ارکان در مدیریت حوضه‌های آبریز است. در این پژوهش دقت مدل‌های ماشین بردار کلاسیک (SVM)، ماشین بردار پشتیبان تلفیق شده با الگوریتم ملخ (GOA-SVM) و حداقل مربعات ماشین بردار پشتیبان (LS-SVM) در شبیه‌سازی دبی اوج سیل ایستگاه پل دختر در حوضه کرخه، مورد ارزیابی قرار گرفته است. بدین منظور از آمار ۷۴ واقعه سیل در محدوده سال‌های ۱۳۸۸ تا ۱۳۹۵ در ایستگاه پل دختر و بارش روزانه ۱۳ ایستگاه باران‌سنجی در حوضه آبریز بالادست این ایستگاه استفاده شده است. از این تعداد، ۵۲ واقعه برای آموزش و ۲۲ واقعه نیز برای صحت‌سنجی مدل‌ها انتخاب شد. مقایسه نتایج به کمک چهار شاخص آماری ضریب تبیین (R^2)، جذر میانگین مربعات خطا (RMSE)، خطای استاندارد (SE)، ضریب نش (NS) و همچنین تحلیل عدم قطعیت به کمک دو شاخص متوسط طول بازه نسبی (ARIL) و درصد پوشش (POC) صورت گرفت. نتایج حاکی از برتری نسبی مدل LS-SVM با $SE=0/407$ ، $R^2=0/519$ ، $RMSE=110/16$ ، $NS=0/91$ و $R^2=0/92$ نسبت به مدل SVM با $SE=0/5$ ، $RMSE=137/70$ ، $NS=0/87$ و $R^2=0/88$ و مدل SVM-GOA با $SE=0/519$ ، $RMSE=144/53$ ، $NS=0/83$ و $R^2=0/9$ است. متوسط مدت زمان اجرای مدل LS-SVM در حد چند ثانیه و این زمان در مدل SVM-GOA در حد چند ساعت است. از سوی دیگر تنظیم پارامترهای مدل SVM کلاسیک بصورت دستی نیز مستلزم صرف زمان زیادی است. لذا مدل LS-SVM به دلیل دارا بودن پارامترهای قابل تنظیم کمتر نسبت به مدل‌های SVM و SVM-GOA، از لحاظ اجرایی از سهولت بیشتری برخوردار است. لذا می‌توان با قطعیت و اختلافی چشمگیر مدل LS-SVM را نسبت به دو مدل دیگر در ارجحیت قرار داد.

کلمات کلیدی: الگوریتم ملخ، حوضه کرخه، پل دختر، مدل‌سازی سیل، ماشین بردار پشتیبان.

مقدمه

سیل یکی از ویرانگرترین بلایای طبیعی است که می‌تواند تاثیر عمیقی بر اکوسیستم و زندگی انسان داشته باشد و باعث آسیب زیست محیطی و زیان‌های اقتصادی به مناطق مسکونی، کشاورزی و منابع آب شود (Choubin et al., 2019). سیل در مقیاس ملی، عامل بیشترین تعداد مرگ و میر و خسارات و از جمله حوادث طبیعی شایع در ایران پس از زلزله می‌باشد (گلشن و همکاران، ۱۳۹۷). در میان سوانح طبیعی بیشترین خسارت در بخش‌های کشاورزی، شیلات، مسکن و زیرساخت‌ها، متعلق به سیل بوده و در اغلب سال‌های گذشته حدود هفتاد درصد از اعتبارات سالیانه "طرح کاهش اثرات بلایای طبیعی" و ستاد حوادث غیر مترقبه، صرف جبران خسارات ناشی از این پدیده شده‌است (کریمی فیروزجائی، ۱۳۹۷). تغییر کاربری اراضی، توسعه ساخت و سازها و همچنین دخل و تصرف غیر مجاز در بستر و حریم رودخانه از مهم‌ترین عوامل تشدید کننده خسارات ناشی از سیلاب می‌باشند (کریمی فیروزجائی، ۱۳۹۷). مدل‌سازی یا شبیه‌سازی سیل یکی از راهکارهای اساسی جهت کنترل و کاهش اثرات مخرب این پدیده بوده و شناسایی مدل‌هایی کارآمد بدین منظور، یکی از مهم‌ترین ارکان در مدیریت حوضه‌های آبریز است. روش‌های گوناگونی به منظور پیش‌بینی سیل در سراسر جهان وجود دارد که در این میان تکنیک‌های یادگیری ماشین به واسطه عملکرد خوب در مواجهه با مجموعه داده‌های بزرگ، بسیار محبوب هستند (Ranasinghe and Ilmini., 2020). از جمله پژوهش‌های صورت گرفته پیرامون این موضوع می‌توان به پژوهش Agnihotri et al. (2022) با هدف پیش‌بینی سیل در ایستگاه Matijuri واقع در حوضه رودخانه باراک در آسام هندوستان به کمک مدل‌های سیستم

استنتاج فازی عصبی - تطبیقی (ANFIS) و مدل ترکیبی ANFIS و الگوریتم بهینه‌سازی کلونی مورچه‌ها (ACO) اشاره نمود. تجزیه و تحلیل نتایج این پژوهش نشان داد که مدل ANFIS-ACO در مقایسه با مدل ANFIS مستقل، از دقت و اطمینان بیشتری برخوردار است. (Ke et al., 2021) از چهار مدل هوش مصنوعی شامل درخت تصمیم (DT)، شبکه عصبی پرسپترون چندلایه (MLP)، جنگل تصادفی (RF) و SVM برای پیش‌بینی دبی سیل در استان شانشی چین استفاده نمودند. نتایج این پژوهش نشان می‌دهد که همه مدل‌ها می‌توانند عملکرد خوبی در حوضه‌های نیمه مرطوب برای پیش‌بینی کوتاه‌مدت داشته باشند. با این حال، چهار مدل هوش مصنوعی دقت نسبی کمتری در حوضه‌های نیمه خشک دارند و تنها مدل SVM می‌تواند به پیش‌بینی رضایت بخشی دست یابد. با افزایش زمان پیش‌بینی، عملکرد مدل‌های فوق بسیار متفاوت است. (Tabbussum and Qayoom, 2021) در پژوهشی عملکرد مدل‌های ANN، منطبق فازی و ANFIS را در پیش‌بینی سیل مورد ارزیابی قرار دارند. نتایج این پژوهش حاکی از عملکرد بهتر مدل ANFIS نسبت به دو مدل دیگر است. (Sahoo and Ghose, 2022) به پیش‌بینی دبی سیل در ایستگاه مونیخال در آسام هند به کمک دو مدل SVM-GOA و SVM-GOA پرداختند. نتایج این پژوهش نشان داد که مدل توسعه‌یافته SVM-GOA می‌تواند به عنوان یک روش کارآمد نسبت به SVM برای پیش‌بینی دقیق سیل مورد استفاده قرار گیرد. شریفی گرمدره و همکاران (۱۳۹۸) به ارزیابی کارایی مدل‌های SVM، شبکه عصبی مصنوعی (ANN) و رگرسیون چند متغیره غیرخطی در حوضه آبخیز دریاچه نمک پرداختند و دریافتند که روش مدلسازی ماشین بردار پشتیبان نسبت به دو روش دیگر از دقت بیشتری در پیش‌بینی دبی‌های حداکثر لحظه‌ای برخوردار است. (Shada et al., 2022) از دو مدل SVM و ماشین بردار پشتیبان ترکیبی با تبدیل موجک (WSVM) به منظور پیش‌بینی سیل در رودخانه Achankovil در کرالا استفاده نمودند. بدین صورت که پس از حذف نویز سیگنال ورودی (بارندگی و سطح آب) به کمک تبدیل موجک، اجزای موثر سیگنال ورودی به مدل‌های SVM و WSVM معرفی گردید. عملکرد این مدل‌ها با ترکیب شبکه عصبی مصنوعی و تبدیل موجک (WANN) نیز مورد مقایسه قرار گرفت. نتایج این مطالعه حاکی از توانایی بالای مدل WSVM در پیش‌بینی سیل بوده و نشان می‌دهد که مدل WSVM در مقایسه با مدل WANN دارای عملکرد بهتری بوده است. (Sahoo et al., 2022) به پیش‌بینی سیل در حوضه رودخانه باراک به کمک مدل MLP و SVM با بهینه‌ساز گرگ خاکستری (GWO) پرداختند و کارایی مدل‌های MLP-GWO و SVM-GWO با مدل‌های ساده MLP و SVM را مورد ارزیابی قرار دادند. نتایج این پژوهش نشان داد که SVM-GWO در طول دوره‌های آزمایشی عملکرد بهتری نسبت به MLP-GWO، SVM و MLP دارد. (Singh et al., 2022) در مطالعه‌ای به بررسی دقت مدل SVM ادغام شده با الگوریتم بهینه‌سازی کلونی زنبور عسل (ABC) به منظور پیش‌بینی جریان ماهانه در دو ایستگاه Salebhata و Sundargarh در حوضه رودخانه ماهانادی هند پرداختند. مقایسه نتایج مدل ترکیبی و مدل SVM معمولی به کمک معیارهای آماری کمی، نشان داد که مدل ترکیبی SVM-ABC داده‌های پیچیده هیدرولوژیکی را بهتر پردازش می‌کند و قابلیت تعمیم و دقت پیش‌بینی بالاتری دارد.

با بررسی مطالعات صورت گرفته می‌توان به اهمیت پیش‌بینی دقیق سیل، به منظور کنترل خسارات ناشی از آن پی برد. ایران نیز از جمله کشورهایی است که به دلیل تغییرات زیاد آب و هوایی و با داشتن آب و هوای نیمه خشک آسیب پذیر بوده و طی دو دهه گذشته با چندین سیل ویرانگر در منطقه روبرو بوده است. استان لرستان یکی از استان‌های سیل خیز کشور می‌باشد و در بین نقاط مختلف استان، شهرستان پلدختر بدلیل وسعت زیاد حوضه بالادست، حجم بالای سیلاب حوضه و همچنین وضعیت پوشش گیاهی حوضه و تغییر کاربری اراضی و وضعیت هیدروژئومورفولوژی آن (این حوضه دارای واحدهای کوهستانی مرتفع با ساختمان سنگی و شیب بسیار زیاد و فرسایشی است و به هنگام بارشهای جوی، جریان آب سطحی بلافاصله بر روی دامنه آنها ایجاد میشود)، یکی از سیل خیزترین نقاط می‌باشد. این منطقه دارای پتانسیل سیل خیزی بسیار بالایی است و شاهد سیل‌های قابل توجهی بوده است (اسدی و همکاران، ۱۴۰۱). طبق نتایج موسوی و همکاران در سال ۱۳۹۷، مساحت تحت پوشش جنگل در این منطقه طی دوره ی ۳۱ ساله ۱۹۸۶ الی ۲۰۱۷ به یک سوم کاهش یافته است و تغییرات عمده ای در کاربری اراضی صورت گرفته است. بطوریکه مساحت بخش مسکونی منطقه در این بازه زمانی به میزان ۵ برابر افزایش یافته است. همچنین تمرکز و ساخت و سازهای مسکونی در چند دهه

گذشته در حاشیه رودخانه گسترش یافته است. لذا با توجه به توسعه فیزیکی شهر و خطر وقوع سیل در این محدوده، این منطقه به عنوان منطقه مورد مطالعه انتخاب شد. در این راستا، با توجه به تحقیق‌های صورت گرفته می‌توان گفت که در میان مدل‌های ذکر شده، مدل SVM روش مناسب برای مدل‌سازی فرایند‌های غیر خطی، پیچیده و پویا مانند سیل معرفی شده است که دارای دقت قابل قبولی در پیش‌بینی سیل می‌باشد. لذا با توجه به دقت قابل قبول این روش و با توجه به عملکرد موفق الگوریتم‌های فراکاوشی در افزایش دقت این مدل، در این تحقیق از دو مدل هیبریدی LS-SVM^۱ و SVM-GOA^۲ جهت شبیه‌سازی دبی اوج سیل ایستگاه پل دختر، استفاده شده و دقت این مدل‌ها و همچنین ماشین بردار کلاسیک (SVM) مورد ارزیابی قرار می‌گیرد.

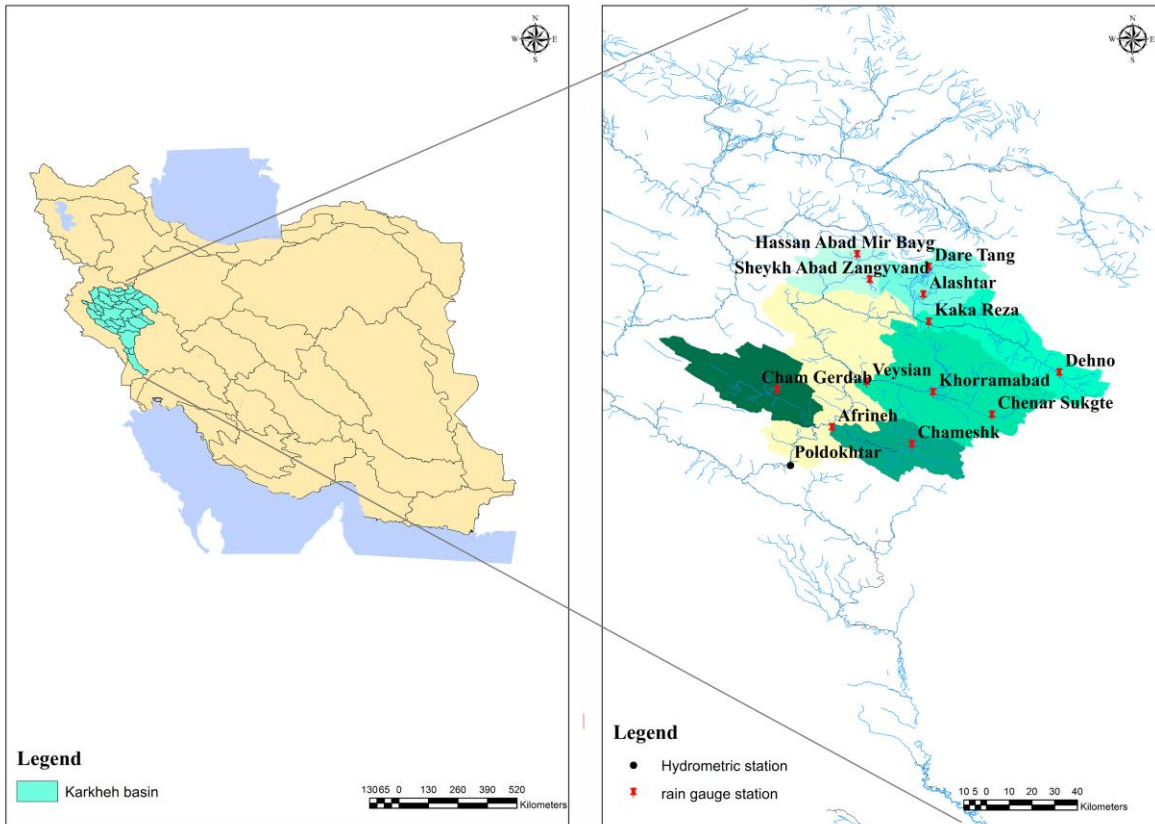
مواد و روش‌ها

منطقه مورد مطالعه

حوضه آبریز کرخه با مساحت ۵۱۳۳۷ کیلومتر مربع جزئی از حوضه آبریز خلیج فارس به شمار می‌رود. این حوضه شامل قسمت‌هایی از استان‌های کرمانشاه، همدان، کردستان، ایلام، لرستان، مرکزی و خوزستان بوده و بخش عمده آن در سه استان کرمانشاه، لرستان و همدان قرار دارد. حوضه کرخه از شمال به حوضه رودخانه‌های سیروان، سفیدرود و قره‌چای، از مغرب به حوضه‌های مرزی ایران و عراق، از جنوب به قسمتی از مرز غربی و از قسمت شرقی به حوضه رودخانه دز محدود می‌گردد (گزارش سازمان برنامه و بودجه، ۱۳۹۱). حوضه آبریز کشکان با مساحت ۹۲۷۵/۶۶ کیلومتر مربع، یکی از زیرحوضه‌های مهم حوضه آبریز کرخه بوده و در مختصات جغرافیایی ۴۷ درجه و ۱۲ دقیقه تا ۴۸ درجه و ۵۹ دقیقه طول شرقی و ۳۳ درجه و ۸ دقیقه تا ۳۴ درجه و ۲ دقیقه عرض شمالی در استان لرستان واقع شده است (نصری و همکاران، ۱۴۰۰). رودخانه کشکان به طول حدود ۲۷۰ کیلومتر به عنوان یکی از سرشاخه‌های مهم و پرآب رودخانه کرخه، در جنوب غربی پلدختر در منطقه‌ای بنام کل سفی به رودخانه سیمره پیوسته و رودخانه کرخه را به وجود می‌آورد (آزادی و همکاران، ۱۳۹۹) به منظور مطالعه دقیق‌تر و انتخاب صحیح ایستگاه‌های باران‌سنجی مورد مطالعه، نواحی موثر در شکل‌گیری سیل ایستگاه پل دختر به شش زیر حوضه مجزا (شکل ۱) تقسیم شد و سپس یک ایستگاه هیدرومتری در محدوده پل دختر و دوازده ایستگاه باران‌سنجی در بالادست آن (حوضه آبریز کشکان) با مشخصات و موقعیت مکانی مندرج در جدول و شکل ۱ در این محدوده انتخاب گردید.

۱. Least-Squares Support Vector Machine

۲. Support Vector Machine Models Combined With Grasshopper Optimization Algorithm



شکل ۱. موقعیت مکانی ایستگاه‌های مورد مطالعه

جدول ۱. مشخصات ایستگاه‌های مورد مطالعه

ردیف	ایستگاه	طول جغرافیایی (درجه)	عرض جغرافیایی (درجه)	ارتفاع (متر)
۱	چمشک	۳۳/۲۳	۴۸/۲۰	۱۴۷۵
۲	چم گرداب	۳۳/۴۵	۴۷/۶۷	۴۰۳
۳	حسن آباد میربیگ	۳۳/۹۸	۴۷/۹۸	۱۶۰۷
۴	ویسان	۳۳/۴۸	۴۸/۰۲	۱۰۲۰
۵	شیخ آباد زنگی وند	۳۳/۸۸	۴۸/۰۳	۱۷۳۷
۶	خرم آباد	۳۳/۴۴	۴۸/۲۸	۱۱۴۷/۸
۷	الشتر	۳۳/۸۲	۴۸/۲۵	۱۵۶۷/۱
۸	پل دختر	۳۳/۱۵	۴۷/۷۲	۷۱۳/۵
۹	آفرینه	۳۳/۳۰	۴۷/۸۸	۸۰۵
۱۰	دهنو	۳۳/۵۲	۴۸/۷۸	۱۷۷۰
۱۱	کاکا رضا	۳۳/۷۲	۴۸/۲۷	۱۵۳۰
۱۲	دره تنگ	۳۳/۹۳	۴۸/۲۷	۱۸۰۰
۱۳	چنار سوخته	۳۳/۳۵	۴۸/۵۲	۱۴۲۰

مدل ماشین بردار پشتیبان کلاسیک (SVM)

ماشین بردار پشتیبان سیستم یادگیری کارآمدی است که در سال ۱۹۶۵ توسط محقق روسی به نام ولادیمیر وپنیک^۳ به منظور رسیدن به یک جواب بهینه با استفاده از اصل استقرایی کمینه‌سازی خطای ساختاری ارائه شد (Cristianini and Shawe-Taylor., 2000). پیش‌بینی‌های ثمربخش این مدل ناشی از دو ویژگی برجسته قابلیت عمومی سازی عالی و سازگاری با داده‌های پراکنده و کم می‌باشد (Behzad et al., 2010). این مدل با چهار نوع کرنل خطی، چند جمله‌ای، تانژانت هیپربولیک و گوسی، تابع خطای زیر (رابطه ۱) را با در نظر گرفتن یکسری محدودیت‌ها (رابطه ۲) کمینه می‌کند (نوذری و توکلی، ۱۳۹۷).

$$\frac{1}{2} W^T W + C \sum_{i=1}^N \xi_i + C \sum_{i=1}^N \xi_i^* \quad (\text{رابطه ۱})$$

$$W^T \phi(x_i) + b - y_i \leq \varepsilon + \xi_i^*$$

$$y_i - W^T \phi(x_i) - b \leq \varepsilon + \xi_i \quad (\text{رابطه ۲})$$

$$\xi_i^*, \xi_i \geq 0$$

در روابط بالا C ثابت گنجایش، W بردار ضرایب، ξ_i و ξ_i^* ضرایب کمبود، b ضریب ثابت، N الگوهای آموزش مدل و ϕ تابع کرنل است.

حداقل مربعات ماشین بردار پشتیبان (LS-SVM)

مدل حداقل مربعات ماشین بردار پشتیبان (LS-SVM) که توسط وندوال^۴ و سویکنس^۵ معرفی شده است، محدودیت‌های مشابهی با مدل SVM کلاسیک دارد، اما نسبت به این مدل به دلیل پیچیدگی محاسباتی کمتر، دقت و سرعت بالاتر، در ارجحیت است (سیفی، ۱۳۸۹). مدل رگرسیون (LS-SVM) از رابطه ۳ پیروی کرده که در آن، w و b به ترتیب مقادیر وزن‌ها و آریبی تابع رگرسیون هستند و از طریق حداقل سازی تابع هدف در رابطه ۴ تعیین می‌شوند (پوررضا بیلندی و همکاران، ۱۳۹۳).

$$y(x_i) = w^T \cdot \phi(x_i) + b \quad (\text{رابطه ۳})$$

$$\min_{w,e,b} j(w, e) = \frac{1}{2} w^T w + \frac{\gamma}{2} \sum_{i=1}^N e_i^2 \quad (\text{رابطه ۴})$$

۳. Vladimir Vapnik

۴. Vandewalle

۵. Suykens

با محدودیت:

$$y_i = w^T \varphi(x_i) + b + e_i \quad (\text{رابطه ۵})$$

در روابط فوق φ تابع کرنل γ پارامتر تنظیم کننده خطا و e_i خطای داده‌های آموزشی می‌باشد.

الگوریتم بهینه سازی ملخ (GOA)

الگوریتم بهینه‌سازی ملخ یک روش بهینه‌سازی فراابتکاری الهام گرفته شده از رفتار گروهی ملخ‌ها می‌باشد که در سال ۲۰۱۷ توسط صارمی و همکاران ارائه گردیده است. از مزیت‌های این الگوریتم می‌توان به مشارکت کلیه عوامل جستجو در به روزرسانی موقعیت هر عامل جستجو، توجه ویژه در راستای جلوگیری از به دام افتادن در نقاط بهینه محلی، ایجاد تعادل بین قابلیت جستجو سراسری و محلی و عدم نیاز به اطلاعات گرادیان فضای جستجو، اشاره نمود (کریمی و رضائی آهوانویی، ۱۴۰۰). در این الگوریتم به منظور به روزرسانی موقعیت ملخ‌ها از رابطه زیر استفاده می‌گردد.

$$X_i^d = T_d + c \left[\sum_{j=1}^n c \left(\frac{ub_d - lb_d}{2} \right) \cdot s(|X_j^d - X_i^d|) \frac{X_j - X_i}{d_{i,j}} \right] \quad (\text{رابطه ۶})$$

در این رابطه X_i موقعیت ملخ نام، X_j موقعیت ملخ زام، $d_{i,j}$ فاصله بین i و j زمین ملخ، ub و lb به ترتیب بیانگر حد بالا و حد پایین متغیرهای تصمیم، T بهترین موقعیت به دست آمده تاکنون و پارامتر c نیز معرف یک ضریب کاهشی برای محدود کردن منطقه خنثی، دافعه و جاذبه می‌باشد. در این الگوریتم موقعیت ملخ‌ها در گروه بصورت رابطه ۷ تعریف می‌شود:

$$X_i = S_i + G_i + A_i \quad (\text{رابطه ۷})$$

در این رابطه S_i ، G_i و A_i به ترتیب اثر برهم کنش اجتماعی، گرانش و باد بر حرکت ملخ‌ها می‌باشد (کریمی و رضائی آهوانویی، ۱۴۰۰). در این پژوهش از الگوریتم GOA به منظور بهینه‌سازی پارامترهای مدل SVM، تحت عنوان مدل SVM-GOA استفاده شده است.

معیارهای ارزیابی

چهار شاخص آماری ضریب تبیین (R^2)، جذر میانگین مربعات خطا (RMSE)، خطای استاندارد (SE) و ضریب نش (NS) معیار ارزیابی کارایی و دقت مدل‌های مورد استفاده در این پژوهش قرار گرفت.

$$R^2 = \frac{[\sum_{i=1}^n (Q_i^s - \bar{Q}_i^s)(Q_i^o - \bar{Q}_i^o)]^2}{\sqrt{\sum_{i=1}^n (Q_i^s - \bar{Q}_i^s)^2 \sum_{i=1}^n (Q_i^o - \bar{Q}_i^o)^2}} \quad (\text{رابطه ۸})$$

$$RMSE = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (Q_i^o - Q_i^s)^2} \quad \text{رابطه ۹}$$

$$SE = \frac{RMSE}{\bar{x}} \quad \text{رابطه ۱۰}$$

$$NS = 1 - \frac{\sum_{i=1}^n (Q_i^o - Q_i^s)^2}{\sum_{i=1}^n (Q_i^o - \bar{Q}_i^o)^2} \quad \text{رابطه ۱۱}$$

در این روابط، Q_i^o دبی مشاهداتی، Q_i^s دبی شبیه‌سازی، \bar{Q}_i^o میانگین دبی مشاهداتی، \bar{Q}_i^p میانگین دبی شبیه‌سازی و n تعداد داده‌ها می‌باشد.

تحلیل عدم قطعیت

در تکرار چندباره مدل‌های مبتنی بر یادگیری ماشین، بدون آنکه تغییری در ورودی و یا ساختار مدل ایجاد شود، خروجی‌های متفاوتی حاصل می‌شود و بر همین اساس نمی‌توان به نتایج حاصل از یکبار اجرای مدل اعتماد نمود، لذا به منظور برآورد عدم قطعیت در خروجی مدل‌های مورد استفاده از دو شاخص متوسط طول بازه نسبی (ARIL) و درصد پوشش (POC) استفاده گردید.

$$ARIL = \frac{1}{N} \times \left(\sum_{i=1}^N \frac{Q_i^+ - Q_i^-}{Q_i^o} \right) \quad \text{رابطه ۱۲}$$

$$POC = \frac{1}{N} \sum_{i=1}^N \delta_i, \delta_i = \begin{cases} 1 & \text{if } Q_i^- \leq Q_i^o \leq Q_i^+ \\ 0 & \text{otherwise} \end{cases} \quad \text{رابطه ۱۳}$$

در این روابط، Q_i^+ و Q_i^- حد بالا و پایین محدوده اطمینان در گام زمانی i ، N تعداد گام‌های زمانی و Q_i^o دبی مشاهداتی می‌باشد.

روش‌شناسی

در این پژوهش که به منظور مقایسه و ارزیابی مدل‌های SVM، LS-SVM و SVM-GOA در مدلسازی سیل در ایستگاه هیدرومتری پل دختر انجام شده است، مراحل زیر به ترتیب صورت گرفت. در ابتدا آمار و اطلاعات مربوط به سیل و بارش در حوضه‌ی بالادست ایستگاه هیدرومتری پلدختر در مقیاس روزانه از سازمان‌های مربوطه دریافت شده و سپس یک دوره آماری مشترک بین ایستگاه‌های باران سنجی و هیدرومتری مورد مطالعه انتخاب شد (محدوده سال‌های ۸۸ تا ۹۵). در گام بعد تعداد ۷۴ واقعه سیل با دبی حداکثر لحظه‌ای بیش از ۶۰ متر مکعب بر ثانیه، در این محدوده زمانی شناسایی گردید و به تعیین مقادیر بارش در روز وقوع سیل در هر یک از ایستگاه‌های باران سنجی بالادست پرداخته شد. به دلیل تاثیر بارش روزهای قبل از وقوع سیل بر شکل‌گیری آن که البته تحت تاثیر زمان تمرکز حوضه است، به منظور تحلیل دقیق‌تر علاوه بر بارش در روز وقوع سیل از بارش پنج روز ماقبل آن نیز استفاده گردید و نهایتاً پس از انتخاب و تعیین مقادیر ورودی و خروجی، مدل‌های مورد نظر بصورت زیر اجرا شدند.

مدل LS-SVM

مدلسازی توسط این مدل به وسیله نرم افزار MATLAB صورت گرفته و از بارش شش روزه (بارش در روز وقوع سیل و پنج روز ماقبل آن) در ایستگاه‌های بالادست به عنوان ورودی و از دبی حداکثر سیل در ایستگاه پل دختر به عنوان خروجی این مدل استفاده شد و آموزش آن با نسبت ۷۰ به ۳۰ و به صورت کاملاً تصادفی صورت پذیرفت. در فرآیند آموزش، از توابع کرنل پایه شعاعی (RBF)، خطی (LIN) و چندجمله‌ای (POLY) استفاده شد. و سپس به شبیه‌سازی سیل در دوره آموزش و صحت‌سنجی پرداخته شد و در نهایت شاخص‌های آماری R^2 ، RMSE، SE و NS میان دبی‌های مشاهده‌ای و شبیه‌سازی شده توسط این مدل در دو مرحله آموزش و صحت‌سنجی و به تفکیک برای هریک از توابع کرنل مورد نظر محاسبه شد.

مدل SVM

مدلسازی توسط این مدل به وسیله نرم افزار MATLAB صورت گرفت و به منظور ارزیابی و مقایسه بهتر مدل‌های مورد نظر در این پژوهش، مدلسازی به کمک همان الگوی انتخابی از مقادیر ورودی و خروجی تعیین شده در مدل LS-SVM و همچنین تابع کرنل RBF صورت گرفت. در فرآیند آموزش، پارامترهای مدل بصورت کاملاً دستی و به وسیله سعی و خطا در جهت افزایش دقت مدلسازی تنظیم شدند، در این راستا مدل SVM ۵۰۰ مرتبه اجرا شد.

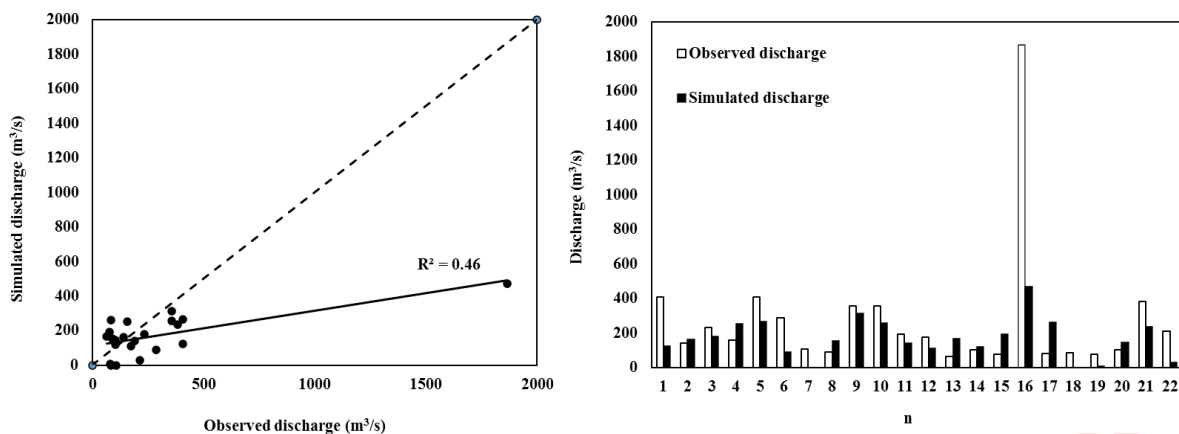
مدل SVM-GOA

در این مرحله دقت و کارایی مدل ماشین بردار تلفیق شده با الگوریتم GOA مورد ارزیابی قرار گرفت. لذا، ابتدا مدل SVM و الگوریتم بهینه‌سازی GOA در نرم افزار MATLAB کدنویسی شدند و سپس دو مدل در جهت بهینه‌سازی پارامترهای ماشین بردار پشتیبان به کمک الگوریتم GOA، تلفیق شدند. در این مرحله نیز مدلسازی به کمک همان الگوی انتخابی از مقادیر ورودی و خروجی تعیین شده در بخش قبل و همچنین تابع کرنل RBF صورت گرفت. در ادامه مدل SVM-GOA با اندازه جمعیت ۵۰، ۱۰۰ مرتبه مورد تکرار قرار گرفت.

نتایج و بحث

مدل LS-SVM

پس از آماده سازی مدل LS-SVM مطابق با بخش روش‌شناسی، به شبیه‌سازی سیل در دوره آموزش و صحت‌سنجی پرداخته شد. نتایج مربوط به محاسبه شاخص‌های آماری میان دبی‌های مشاهده‌ای و شبیه‌سازی شده توسط این مدل در دو مرحله آموزش و صحت‌سنجی و به تفکیک برای هریک از توابع کرنل مورد نظر در شکل و جدول ۲ ارائه شده است.



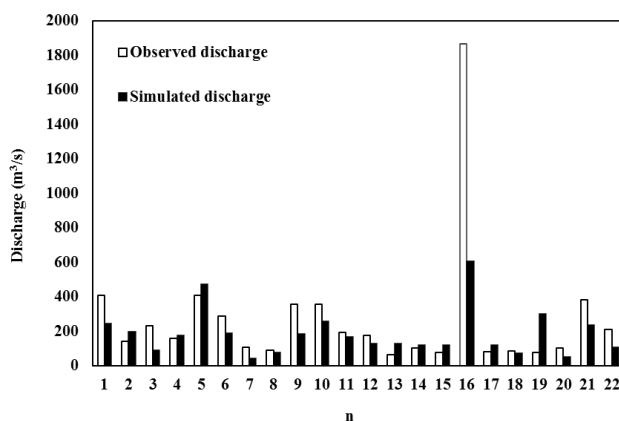
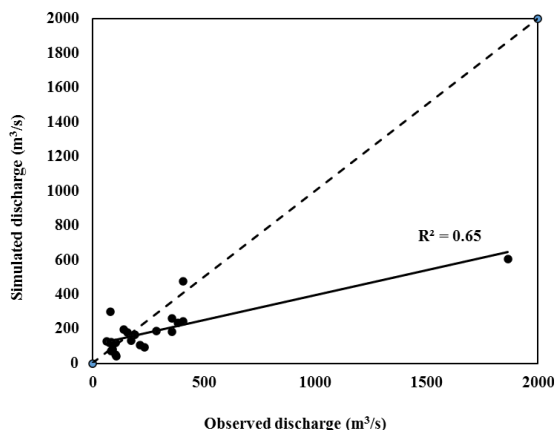
شکل ۲. مقایسه سیل مشاهداتی و برآورد شده توسط مدل LS-SVM در مرحله صحت سنجی

جدول ۲. نتایج مربوط به مدل‌سازی سیل در مرحله آموزش و صحت‌سنجی به وسیله مدل LS-SVM

صحت سنجی				آموزش				ساخار مدل
R ²	RMSE (m ³ /s)	SE	NS	R ²	RMSE (m ³ /s)	SE	NS	کرنل
۰/۴۶	۳۲۱/۸۳۹	۱/۱۸۹	۰/۲۴	۰/۹۸۶	۴۳/۵۲۵	۰/۱۸۹	۰/۹۳	RBF
۰/۲۸	۳۳۶/۲۳۸	۱/۳۴۲	۰/۱۸	۰/۹۵۵	۸۰/۰۱۹	۰/۳۴۸	۰/۷۸	LIN
۰/۱۶	۳۵۷/۷۳۶	۱/۳۲۱	۰/۰۷	۱	.	.	.	POLY

همانطور که ملاحظه می‌شود در بین سه تابع کرنل مورد استفاده، نتایج حاصل از انتخاب کرنل RBF، در مقایسه با دو کرنل دیگر در دو مرحله آموزش و صحت‌سنجی، از دقت بیشتری برخوردار است. اما علی‌رغم نتایج به دست آمده در مرحله آموزش، مقادیر شاخص‌های ارزیابی در مرحله صحت‌سنجی، حاکی از دقت پایین مدل LS-SVM در پیش‌بینی سیل بوده است. لذا در ادامه در جهت بهبود و ارتقا نتایج حاصل از مرحله قبل، دو سناریو به صورت زیر مطرح گردید:

- سناریو اول: در این سناریو، مقادیر همبستگی بین دبی پیک سیل و سری‌های زمانی بارش در روز وقوع سیل و پنج روز ماقبل آن محاسبه شد. مقادیر همبستگی در روز وقوع سیل در حداکثر مقدار خود قرار داشته و در روزهای قبل به شدت کاهش می‌یابند، لذا می‌توان تاثیر بارندگی در روزهای قبل را بر حداکثر سیل نادیده گرفت و تنها از بارش در روز وقوع سیل در فرآیند مدل‌سازی استفاده کرد. در ادامه مدل LS-SVM با در نظر گرفتن این سناریو دوباره اجرا گردید. نتایج مربوط به محاسبه شاخص‌های آماری در مراحل آموزش و صحت‌سنجی و همچنین نمودار مقادیر مشاهداتی و پیش‌بینی شده توسط این مدل در جدول و شکل ۳ ارائه شده است.



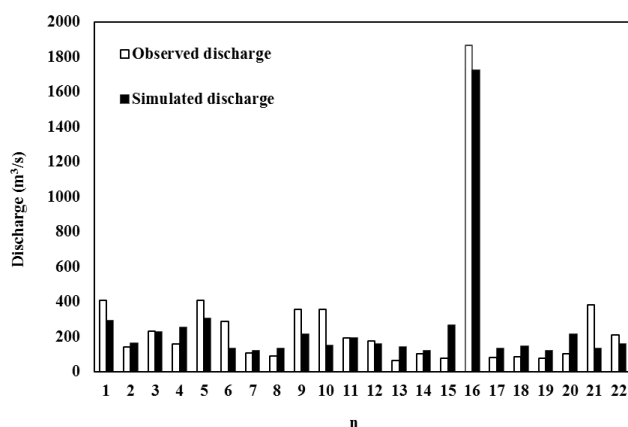
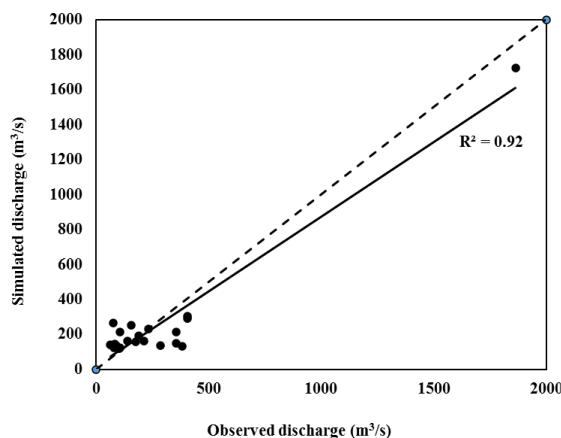
شکل ۳. مقایسه سیل مشاهداتی و برآورد شده توسط مدل LS-SVM در مرحله صحت سنجی

جدول ۳. نتایج مربوط به مدل‌سازی سیل در مرحله آموزش و صحت‌سنجی به وسیله مدل LS-SVM

صحت سنجی				آموزش				ساخار مدل
R ²	RMSE (m ³ /s)	SE	NS	R ²	RMSE (m ³ /s)	SE	NS	کرنل
۰/۶۵	۲۸۹/۸۳۰	۱/۰۷۰	۰/۳۹	۰/۹۹	۱۱/۲۵۳	۰/۰۴۹	۰/۹۹	RBF
۰/۵۴	۳۱۰/۰۷	۱/۱۴۵	۰/۳۱	۰/۵۴	۲۴۵/۷۵	۱/۰۷۰	۰/۰۶	LIN
۰/۵۲	۲۶۱/۸۵	۰/۹۶۷	۰/۴۹	۱	۰	۰	۰	POLY

همانطور که ملاحظه می‌شود در این قسمت نیز کرنل RBF از دقت بیشتری برخوردار بوده و در مجموع در این سناریو، اگرچه مقدار R^2 با بهبود بیشتری روبرو شد اما میزان بهبود در مقادیر SE و RMSE تنها ۱۳ درصد بود. با توجه به مقادیر بالای شاخص SE، سناریو دوم به صورت زیر ارزیابی شد.

■ سناریو دوم: در این بخش مقادیر همبستگی میان دبی حداکثر سیل و بارش در روز وقوع سیل در تمامی ایستگاه‌های بالادست محاسبه شد و در نتیجه معیار مدل‌سازی، بارش در ایستگاه‌هایی با بیشترین مقدار همبستگی در نظر گرفته شد. علاوه بر این به ترتیب هریک از ایستگاه‌های باران‌سنجی درون حوضه بالادست حذف شده و تاثیر آن بر دقت نتایج نهایی مورد بررسی قرار گرفت. در صورتی که به دنبال حذف یک ایستگاه دقت مدل کاهش نمی‌یافت، بارش آن به کلی حذف شده و در غیر اینصورت در روند مدل‌سازی باقی می‌ماند. این کار نهایتاً باعث شناسایی چهار ایستگاه چمشک، ویسان، خرم آباد و الشتر با بیشترین میزان تاثیر بر دقت مدل‌سازی گردید. در ادامه مدل LS-SVM مشابه مرحله قبل و این بار با در نظر گرفتن دو سناریو فوق اجرا شد که نتایج آن در جدول و شکل ۴ آمده است.



شکل ۴. مقایسه سیل مشاهداتی و برآورد شده توسط مدل LS-SVM در مرحله صحت سنجی

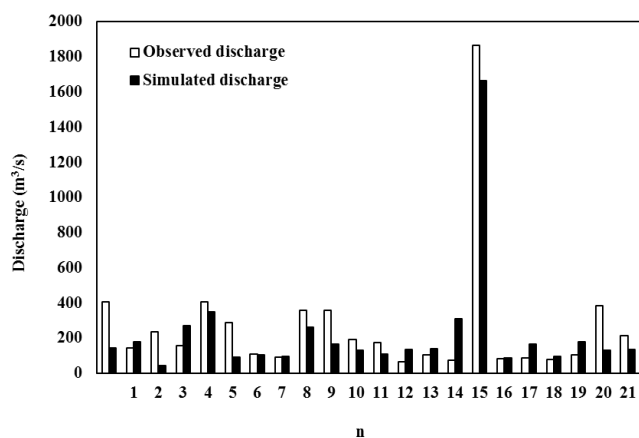
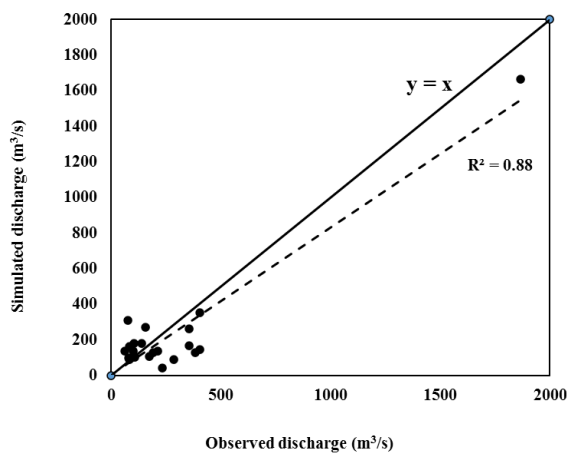
جدول ۴- نتایج مربوط به مدلسازی سیل در مرحله آموزش و صحت‌سنجی به وسیله مدل LS-SVM

صحت سنجی				آموزش				ساخار مدل
R ²	RMSE (m ³ /s)	SE	NS	R ²	RMSE (m ³ /s)	SE	NS	کرنل
۰/۹۲	۱۱۰/۱۶۱	۰/۴۰۷	۰/۹۱	۰/۹۵	۷۳/۴۰۳	۰/۳۱۹	۰/۹۵	RBF
۰/۴۶	۲۴۳/۹۴۱	۱/۰۶۲	۰/۴۲	۰/۵۳	۲۸۲/۲۵۲	۱/۰۴۲	۰/۵۵	LIN
۰/۸۷	۳۵۸/۱۷۸	۱/۳۳۳	۰/۰۷	۰/۹۷	۵۳/۲۹۴	۰/۳۳۳	۰	POLY

در نتیجه استفاده از سناریو دوم میزان بهبود در مقادیر SE و RMSE به ۶۷ درصد ارتقاء داده شده و R² در محدوده ۰/۹۲ قرار گرفت. همانطور که ملاحظه می‌شود، دقت مدلسازی بصورت چشم‌گیری افزایش یافته و این موضوع نشان می‌دهد که لزوماً با استفاده از سری زمانی بارش با مقیاس ۵ روزه و آمار کلیه ایستگاه‌های بالادست، دقت مدلسازی افزایش نیافته و استفاده از سری‌های اضافی با ایجاد اختلال در فرآیند آموزش، دقت را کاهش می‌دهد. استفاده از آمار بارش چهار ایستگاه چمشک، ویسان، خرم آباد و الشتر، علاوه بر ارتقای عملکرد مدل LS-SVM در مدلسازی سیلاب، موجب شناسایی سه محدوده با بیشترین میزان تأثیر بر دقت شبیه‌سازی گردید.

مدل SVM

پس از آماده سازی مدل SVM مطابق با بخش روش‌شناسی، به شبیه‌سازی سیل در دوره آموزش و صحت‌سنجی پرداخته شد. نتایج مربوط به محاسبه شاخص‌های آماری میان دبی مشاهداتی و شبیه‌سازی شده توسط این مدل در بهترین حالت شبیه‌سازی در دو مرحله آموزش و صحت‌سنجی در شکل و جدول ۵ ارائه شده است. همچنین میانگینی از شاخص‌های آماری محاسبه شده در تکرارهای گوناگون (۵۰۰ تکرار) در جدول ۶ ارائه شده است.



شکل ۵. مقایسه سیل مشاهداتی و برآورد شده توسط مدل SVM در مرحله صحت سنجی

جدول ۵. بهترین نتیجه مدل‌سازی سیل در مرحله آموزش و صحت‌سنجی به وسیله مدل SVM

صحت سنجی				آموزش				ساخار مدل
R^2	RMSE (m ³ /s)	SE	NS	R^2	RMSE (m ³ /s)	SE	NS	کرنل
۰/۸۸	۱۳۷/۷۰	۰/۵	۰/۸۷	۰/۸۸	۱۱۳/۷۳	۰/۴۹	۰/۸۸	RBF

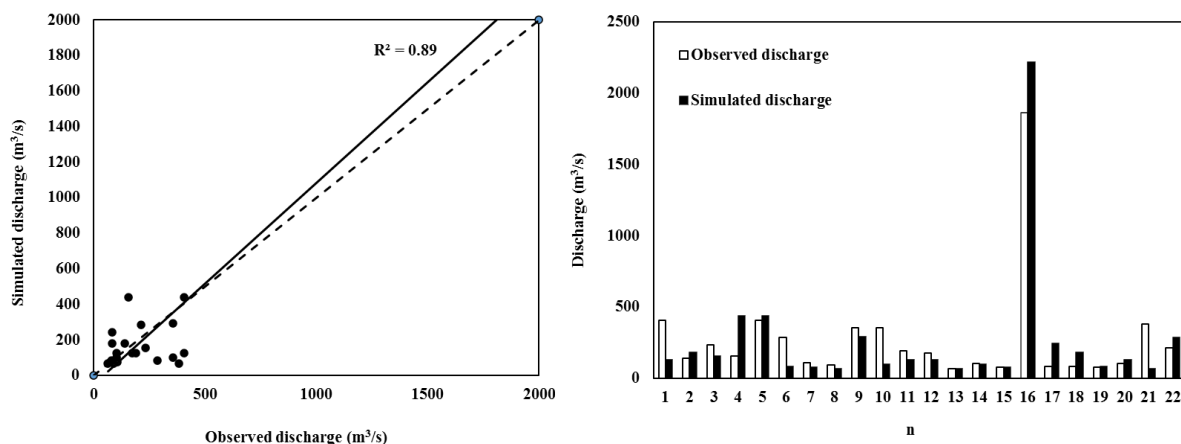
جدول ۶. میانگین نتایج مدل‌سازی سیل در مرحله آموزش و صحت‌سنجی به وسیله مدل SVM

صحت سنجی				آموزش				ساخار مدل
\bar{R}^2	\overline{RMSE} (m ³ /s)	\overline{SE}	\overline{NS}	\bar{R}^2	\overline{RMSE} (m ³ /s)	\overline{SE}	\overline{NS}	کرنل
۰/۶۲	۱۹۸/۴۸۱	۰/۷۴	۰/۷۳	۰/۵۹	۱۹۲/۲۷	۰/۸۴	۰/۶۵	RBF

همانطور که ملاحظه می‌شود، مدل LS-SVM با $SE=0/407$ ، $RMSE=110/16$ ، $NS=0/91$ و $R^2=0/92$ نسبت به بهترین نتیجه مدل SVM با $SE=0/5$ ، $RMSE=137/70$ ، $NS=0/87$ و $R^2=0/88$ عملکرد نسبتاً بهتری دارد. اما باید به این نکته توجه داشت که میانگین شاخص‌های آماری در اجراهای مختلف مدل SVM در جدول ۶ نتایج رضایت بخشی را نشان نمی‌دهند.

مدل SVM-GOA

پس از آماده سازی مدل SVM-GOA مطابق با بخش روش‌شناسی، به شبیه‌سازی سیل در دوره آموزش و صحت‌سنجی پرداخته شد. نتایج مربوط به محاسبه شاخص‌های آماری میان دبی مشاهداتی و شبیه‌سازی شده توسط این مدل در دو مرحله آموزش و صحت‌سنجی در شکل و جدول ۷ ارائه شده است.



شکل ۶. مقایسه سیل مشاهداتی و برآورد شده توسط مدل SVM-GOA در مرحله صحت سنجی

جدول ۷. نتایج مربوط به مدلسازی سیل در مرحله آموزش به وسیله مدل SVM-GOA

صحت سنجی			ساختار مدل			
R ²	RMSE (m ³ /s)	SE	NS	تکرار	اندازه جمعیت	کرنل
۰/۸۹	۱۴۴/۵۳۳	۰/۵۱۹	۰/۸۳	۵۰	۱۰۰	RBF

در مجموع، نتایج حاکی از برتری نسبی مدل LS-SVM با $SE=0/407$ ، $RMSE=110/16$ ، $NS=0/91$ و $R^2=0/92$ نسبت به مدل SVM-GOA با $SE=0/519$ ، $RMSE=144/53$ ، $NS=0/83$ و $R^2=0/9$ است. همچنین مدل SVM با $SE=0/5$ ، $RMSE=137/70$ و $NS=0/87$ و $R^2=0/88$ است. متوسط مدت زمان اجرای مدل LS-SVM در حد چند ثانیه و این زمان در مدل SVM-GOA در حد چند ساعت است. از سوی دیگر تنظیم پارامترهای مدل SVM کلاسیک بصورت دستی نیز مستلزم صرف زمان زیادی است. همچنین مدل LS-SVM به دلیل دارا بودن پارامترهای قابل تنظیم کمتر نسبت به دو مدل دیگر، از لحاظ اجرایی از سهولت بیشتری برخوردار است و همه این عوامل موجب می‌شوند که با قطعیت و اختلافی چشمگیر مدل LS-SVM را نسبت به دو مدل دیگر در ارجحیت دانست.

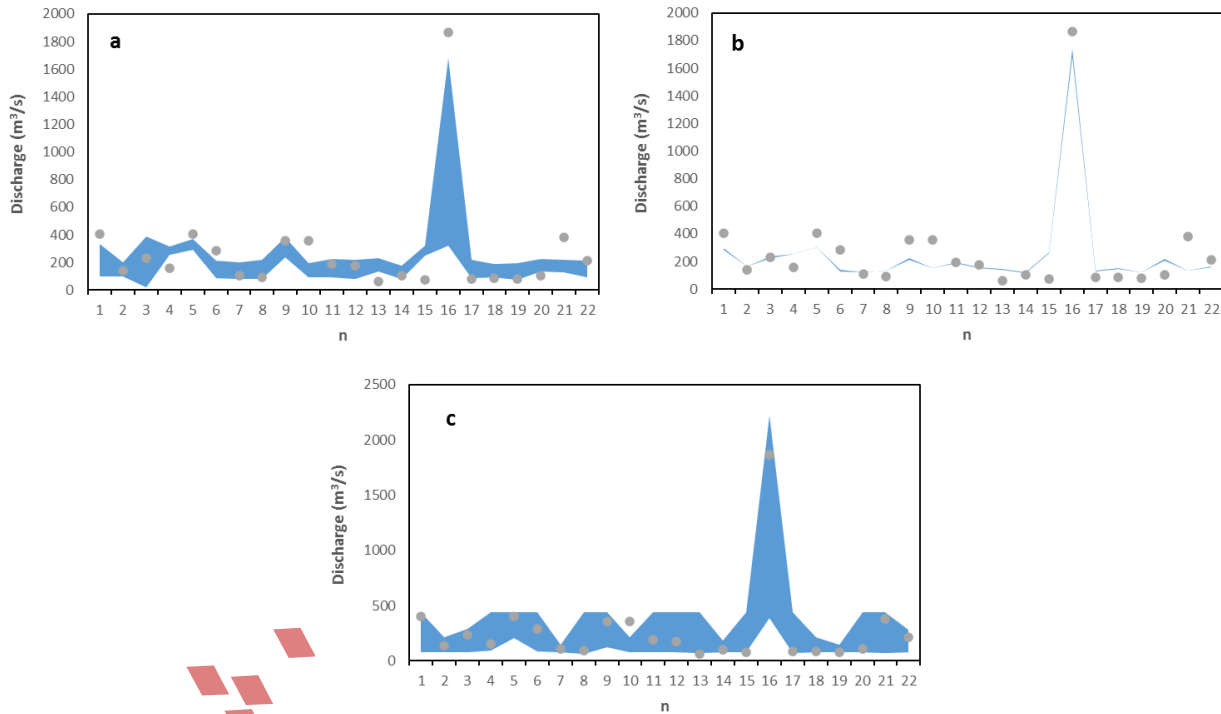
اگر چه نتایج نشان دهنده برابری نسبی دو مدل SVM و SVM-GOA است، اما مدت زمان اجرای بالای مدل SVM-GOA خصوصاً در مسائلی مشابه این پژوهش، موجب می‌شود تا رسیدن به اهدافی نظیر تحلیل حساسیت، نیازمند ماه‌ها زمان باشد و از این نظر می‌توان مدل SVM را نسبت به مدل SVM-GOA در اولویت قرار داد.

تحلیل عدم قطعیت

در این بخش نتایج خروجی مدل‌ها برای شرایط ورودی یکسان، محاسبه شده و سه مدل SVM، LS-SVM و SVM-GOA از نظر میزان تصادفی و یا غیر تصادفی بودن خروجی، مورد ارزیابی و مقایسه قرار گرفت. بدین منظور هریک از مدل‌های LS-SVM و SVM-GOA. ۲۰ مرتبه و مدل SVM، ۵۰۰ مرتبه با ورودی و ساختار یکسان اجرا شده و پارامترهای ARIL و POC مطابق روابط ۱۲ و ۱۳ برآورد شدند که نتایج آن در جدول ۸ و شکل ۷ ارائه شده است.

جدول ۸. نتایج مربوط به تحلیل عدم قطعیت

SVM		LS-SVM		SVM-GOA	
ARIL	POC	ARIL	POC	ARIL	POC
۰/۸۵	۰/۷۲	۰/۰۳۱	۰/۰۹	۱/۸۹	۰/۸۶



شکل ۷. ترسیم باند اطمینان در مدل‌های (a) SVM، (b) LS-SVM و (c) SVM-GOA

مقادیر کوچکتر پارامتر ARIL، نشان دهنده عرض کمتر باند عدم قطعیت و در نتیجه پایین بودن عدم قطعیت می‌باشد، لذا مدل LS-SVM با ARIL=0.031 نسبت به مدل SVM با ARIL=0.85 و SVM-GOA با ARIL=1.89 دارای عدم قطعیت کمتری است. این موضوع به وضوح در شکل ۷ قابل مشاهده است. پارامتر POC نیز مقادیر بین صفر تا یک را شامل می‌شود و هر چه به یک نزدیک‌تر می‌شود نشان می‌دهد که داده‌های مشاهده‌ای بیشتری در باند اطمینان قرار دارند. در نتیجه اگرچه مدل SVM-GOA با POC=0.86، نسبت به LS-SVM با POC=0.09 و SVM با POC=0.72 دارای داده‌های مشاهداتی بیشتری در محدوده مورد نظر است، اما با لحاظ پارامتر ARIL و همچنین شکل ۷، میتوان مدل LS-SVM را دارای عدم قطعیت کمتری نسبت به دو مدل دیگر دانست. همچنین نتایج حاکی از عدم قطعیت پایین‌تر مدل SVM در مقایسه با مدل SVM-GOA است.

نتیجه گیری

در این پژوهش که به منظور مقایسه و ارزیابی مدل های SVM، LS-SVM و SVM-GOA در مدل سازی سیل ایستگاه هیدرومتری پل دختر صورت گرفت، نتایج زیر حاصل شد:

- الزام استفاده از سری زمانی بارش با مقیاس ۵ روزه و استفاده از آمار کلیه ایستگاه های بالادست، موجب افزایش دقت مدل سازی نخواهد شد و استفاده از سری های اضافی با ایجاد اختلال در فرآیند آموزش، دقت را کاهش می دهد. می توان نتیجه گرفت که پیچیدگی اضافه لزوماً منجر به بهبود عملکرد مدل های هیدرولوژیکی نمی شود و این عملکرد بسته به متغیر هیدرولوژیکی در نظر گرفته شده یا شرایط هیدرولوژیکی می تواند متفاوت باشد (Rene Orth et al. 2015). مدل های با پیچیدگی بالا، با تولید خطاهای تصادفی یا بایاس در داده ها، واریانس بالا ایجاد می کنند و باعث ایجاد خطا می شوند (García-Callejas and Araújo, 2016).
- در مجموع، نتایج محاسبه شاخص های آماری، حاکی از برتری نسبی مدل LS-SVM نسبت به مدل های SVM و SVM-GOA است.
- استفاده از آمار بارش چهار ایستگاه چمشک، ویسان، خرم آباد و الشتر، علاوه بر ارتقای عملکرد مدل LS-SVM در مدل سازی سیلاب، موجب شناسایی سه محدوده با بیشترین میزان تاثیر بر دقت شبیه سازی گردید.
- در این تحقیق متوسط مدت زمان اجرای مدل LS-SVM در حدود چند ثانیه و این زمان در مدل SVM-GOA در حدود چند ساعت بود. از سوی دیگر تنظیم پارامترهای مدل SVM کلاسیک بصورت دستی نیز مستلزم صرف زمان زیادی است لذا مدل LS-SVM به دلیل دارا بودن پارامترهای قابل تنظیم کمتر نسبت به مدل های SVM و SVM-GOA از سهولت بیشتری برای اجرا، برخوردار بود.
- اگر چه نتایج نشان دهنده برابری نسبی دو مدل SVM-GOA و SVM است اما مدت زمان اجرای بالای مدل SVM-GOA خصوصاً در مسائلی مشابه این پژوهش، موجب می شود تا رسیدن به اهدافی نظیر تحلیل حساسیت، نیازمند ماه ها زمان باشد و از این نظر می توان مدل SVM را نسبت به مدل SVM-GOA در اولویت قرار داد.
- خروجی مدل LS-SVM برای ورودی های یکسان، ثابت بوده و تصادفی نیست. این در حالی است که خروجی مدل SVM-GOA دارای باند اطمینان عریض تری بوده و نشان دهنده تصادفی بودن خروجی آن در شبیه سازی سیلاب می باشد.

"هیچگونه تعارض منافع بین نویسندگان وجود ندارد"

منابع

- اسدی، معصومه؛ حیدری، زهرا و امامی، کامیار. (۱۴۰۱). ارزیابی روند توسعه فیزیکی شهر پلدختر به سمت مناطق سیل خیز. *جغرافیا و مخاطرات محیطی*، ۱۱(۲)، ۴۷۷-۴۹۸.
- آزادی، فهیمه؛ صدوق، سید حسن؛ قهرودی، منیژه و شهابی، هیمین. (۱۳۹۹). پهنه‌بندی حساسیت خطر سیل در حوضه آبخیز رودخانه کشکان با استفاده از دو مدل WOE و EBF. *جغرافیا و مخاطرات محیطی*، ۹(۱)، ۴۵-۶۰.
- پوررضا بیلندی، محسن؛ خاشعی سیوکی، عباس و صادقی طبس، صادق. (۱۳۹۳). پیش بینی رواناب روزانه با مدل حداقل مربعات ماشین بردار پشتیبان (LS-SVM)، نشریه پژوهش‌های حفاظت آب و خاک، ۲۱(۶)، ۲۹۳-۴۰۳.
- سازمان برنامه و بودجه استان خوزستان. (۱۳۹۱). مطالعات آمایش استان خوزستان-گزارش پشتیبان، جلد هفتم. <https://www.mpo-kz.ir>
- سیفی، اکرم (۱۳۸۹). توسعه یک سیستم خبره برای پیش‌بینی تبخیر و تعرق مرجع روزانه با استفاده از ماشین بردار پشتیبان (SVM) و مقایسه نتایج آن با ANFIS، ANN و روش‌های تجربی. پایان نامه کارشناسی ارشد. به راهنمایی سید مجید میر لطیفی. تهران: دانشگاه تربیت مدرس، دانشکده علوم کشاورزی.
- شریفی گرمدره، ابراهیم؛ وفاخواه، مهدی؛ اسلامیان، سیدسعید. (۱۳۹۸). ارزیابی کارایی سامانه‌های ماشین بردار پشتیبان و شبکه عصبی مصنوعی در تحلیل منطقه‌ای سیلاب (مطالعه موردی: حوضه آبخیز دریاچه نمک). *علوم آب و خاک*، ۳۳(۱)، ۳۶۶-۳۵۱.
- کرمی، حجت و رضائی آهوانویی، علیرضا. (۱۴۰۰). برآورد ضریب آبدهی سرریزهای کلید پیاپویی انحنادار با استفاده از ترکیب رگرسیون بردار پشتیبان و الگوریتم‌های ملخ و کرم شب تاب. *نشریه علمی پژوهشی مهندسی آبیاری و آب ایران*، ۱۲(۲)، ۱۸۶-۲۰۲.
- کریمی فیروزجائی، محمد؛ عبدالهی کاکرودی، عطا اله و جلوخانی نیارکی، محمدرضا. (۱۳۹۷). تهیه نقشه خطر سیل مبتنی بر انرژی جریان با استفاده از سیستم اطلاعات جغرافیایی مطالعه موردی: رودخانه نکا. *پژوهش‌های ژئومورفولوژی کمی*، ۴(۵)، ۱۷۵-۱۵۹.
- گلشن، محمد؛ اسمعیل عوری، اباذر و خسروی؛ خه بات. (۱۳۹۷). ارزیابی حساسیت به سیل حوضه آبخیز تالار با استفاده از مدل نسبت فراوانی احتمالاتی. *مخاطرات محیط طبیعی*، ۷(۱۵)، ۱۶-۱.
- موسوی، میرنجف؛ سارلی، رضا؛ خداداد، مهدی. (۱۳۹۷). آشکارسازی تغییرات کاربری اراضی و پوشش گیاهی شهرستان پلدختر با استفاده از تصاویر ماهواره لندست. *فصلنامه مطالعات محیطی هفت حصار*، ۲۶(۷)، ۱۱۵-۱۰۳.
- نصری، منوچهر؛ قبادی، مرتضی و سپه‌وند، علیرضا. (۱۴۰۰). اولویت بندی کیفی رودخانه‌های حوضه کشکان در استان لرستان جهت توسعه آبی‌پروری (ماهیان سردآبی) با استفاده از ترکیب روش‌های VIKOR و FUZZY AHP. *توسعه آبی‌پروری (علوم زیستی)*، ۱۵(۳)، ۱۱۱-۱۲۶.
- نوذری، حامد و توکلی، فاطمه. (۱۳۹۷). پیش‌بینی جریان با استفاده از مدل ماشین بردار پشتیبان بر مبنای سری‌های زمانی دبی و بارش در ایستگاه‌های بالادست (مطالعه موردی: ایستگاه هیدرومتری تله زنگ). *مدل سازی در مهندسی*، ۱۶(۵۴)، ۹۵-۱۰۴.

References

- Agnihotri, A., Sahoo, A., & Diwakar, M. K. (2022). Flood Prediction Using Hybrid ANFIS-ACO Model: A Case Study BT - *Inventive Computation and Information Technologies* (S. Smys, V. E. Balas, & R. Palanisamy (eds.); pp. 169-180). Springer Nature Singapore.

- Asadi, M., Haidary, Z., Emami, K. (2022). Assessing the Physical Development Trend of Poldokhtar City Towards Flood-Prone Areas, *Journal of Geography and Environmental Hazards*, 11(2), pp. 159-174. (In Persian). doi: 10.22067/geoeh.2021.72169.1101
- Azadi, F., Sadough, S. H., Ghahroudi, M., Shahabi, H. (2020). Zoning of Flood Risk in Kashkan River basin using Two Models WOE and EBF, *Journal of Geography and Environmental Hazards*, 9(1), pp. 45-60. (In Persian). doi: 10.22067/geo.v9i1.83090
- Behzad, M., Asghari, K., & Coppola, E. A. (2010). Comparative Study of SVMs and ANNs in Aquifer Water Level Prediction. *Journal of Computing in Civil Engineering*, 24(5), 408–413. [https://doi.org/10.1061/\(ASCE\)CP.1943-5487.0000043](https://doi.org/10.1061/(ASCE)CP.1943-5487.0000043)
- Bilondi, M. P., Siuki, A. K., & Tabas, S. S. (2015). Daily rainfall-runoff modeling with Least Square Support Vector Machine (LS-SVM). *Water and Soil Conservation*, 21(6), 293–304. (In Persian) <https://www.magiran.com/paper/1401516> LK - <https://www.magiran.com/paper/1401516>
- Choubin, B., Moradi, E., Golshan, M., Adamowski, J., Sajedi-Hosseini, F., & Mosavi, A. (2019). An ensemble prediction of flood susceptibility using multivariate discriminant analysis, classification and regression trees, and support vector machines. *Science of The Total Environment*, 651, 2087–2096. <https://doi.org/10.1016/j.scitotenv.2018.10.064>
- Cristianini, N., & Shawe-Taylor, J. (2000). *An Introduction to Support Vector Machines and Other Kernel-based Learning Methods*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511801389>
- García-Callejas, D., & Araújo, M. (2016). The effects of model and data complexity on predictions from species distributions models. *Ecological Modelling*, 326, 4-12, <https://doi.org/10.1016/j.ecolmodel.2015.06.002>.
- Golshan, M., Ouri, A. E., & Khosravi, K. (2018). Flood Susceptibility assessments Using Frequency Ratio model in Talar Watershed. *Journal of Natural environment hazards*, 7(15), 1–16. (In Persian) <https://www.magiran.com/paper/1814600>
- Nasri M, Ghobadi M, Sepahvand A. (2021). Qualitative Prioritization of Rivers of Kashkan Basin in Lorestan Province for Aquaculture (Coldwater Fishes) Development Using Fuzzy AHP and VIKOR methods. *JAD*, 15 (3) :111-126. (In Persian)
- karami, H & Rezaei Ahvanooei, A. (2021). Estimating Discharge Coefficient of Curved Piano Key Overflows Using Combination of Support Vector Regression and Grasshopper and Firefly Algorithms. *Irrigation and Water Engineering*, 12(2), 186–202. (In Persian) <https://doi.org/10.22125/iwe.2021.142103>
- karimi firoozjaei, M., abdolaho kakroodi, A & jolodar niyaraki, M. (2018). Preparing a flood risk map based on the flow energy by using geographical information system Case study: Nekaroud River. *Quantitative Geomorphological Research*, 5(4), 159–175. (In Persian). https://www.geomorphologyjournal.ir/article_78069.html
- Ke, Z., Jiefan, N. I. U., Xi, L. I., & Lijun, C. (2021). Comparison of artificial intelligence flood forecasting models in China's semi-arid and semi-humid regions. *Water Resources Protection*, 37(1), 28–35. <http://szybhen/article/abstract/bh20210105>
- Mousavi, M., Sareli, R., Khodadad, M. (2019). Detection of land use and vegetation changes in Poldokhtar Using Landsat Satellite Images. *Haft Hesar J Environ Stud*, 7 (26), 103-115. (In Persian)
- Nozari, H., tavakoli, F. (2018). Stream Flow Prediction Using Support Vector Machine Based on Discharge and Precipitation Time series on Upstream Stations (Case Study: Taleh Zang Hydrometric Station), *Journal of Modeling in Engineering*, 16(54), 95-104. (In Persian) doi:

10.22075/jme.2017.11363.1112

- Orth, R., Staudinger, M., Seneviratne., S., Seibert, J., & Zappa, M. (2015) Does model performance improve with complexity? A case study with three hydrological models. *Journal of Hydrology*, 523, 147-159. <https://doi.org/10.1016/j.jhydrol.2015.01.044>.
- Plan and budget organization of Khuzestan province. (2011). Studies of the provincial training program - supporting report, the seventh volume. (In Persian). <https://www.mpo-kz.ir>
- Ranasinghe, R. M., & Ilmini, W. (2020). *Introducing a LSTM based Flood Forecasting Model for the Nilwala river basin with a Mobile Application-a Review*.
- Sahoo, A., & Ghose, D. K. (2022). Flood Forecasting Using Hybrid SVM- GOA Model: A Case Study, *Intelligent Systems* (S. K. Udgata, S. Sethi, & X.-Z. Gao (eds.); pp. 407–416). Springer Nature Singapore.
- Sahoo, A., Samantaray, S., & Ghose, D. K. (2022). Multilayer perceptron and support vector machine trained with grey wolf optimiser for predicting floods in Barak river, India. *Journal of Earth System Science*, 131(2), 85. <https://doi.org/10.1007/s12040-022-01815-2>
- Seifi, A. (2010). *Developing of expert system to prediction of daily evapotranspiration by support vector machine and compare results to ANN, ANFIS and experimental method* [Department of Water Engineering, Tarbiat Modarres University, Tehran, Iran]. (In Persian)
- Shada, B., Chithra, N. R., & Thampi, S. G. (2022). Hourly Flood Forecasting Using Hybrid Wavelet-SVM. *Journal of Soft Computing in Civil Engineering*, 6(2), 1–20. <https://doi.org/10.22115/scce.2022.317761.1383>
- Shrifi Garmdareh, E., Vafakhah, M., & Eslamian, S. (2019). Assessment the Performance of Support Vector Machine and Artificial Neural Network Systems for Regional Flood Frequency Analysis (A Case Study: Namak Lake Watershed). *jwss*, 23 (1), 351-366. (In Persian)
- Singh, U. K., Kumar, B., Gantayet, N. K., Sahoo, A., Samantaray, S., & Mohanta, N. R. (2022). *A Hybrid SVM-ABC Model for Monthly Stream Flow Forecasting BT - Advances in Micro-Electronics, Embedded Systems and IoT* (V. V. S. S. Chakravarthy, W. Flores-Fuentes, V. Bhateja, & B. N. Biswal (eds.); pp. 315–324). Springer Nature Singapore.
- Tabbussum, R., & Qayoom, D. (2021). Performance evaluation of artificial intelligence paradigms—artificial neural networks, fuzzy logic, and adaptive neuro-fuzzy inference system for flood prediction. *Environmental Science and Pollution Research*, 28. <https://doi.org/10.1007/s11356-021-12410-1>

Evaluating the efficiency of SVM, LS-SVM and SVM-GOA models in simulating the Flood peak discharge at the Poldokhtar station

EXTENDED ABSTRACT

Introduction

Floods are among the most devastating natural disasters, with profound impacts on ecosystems, human life, and causing environmental damage and economic losses to residential areas, agriculture, and water resources. (Choubin et al., 2019). Among natural disasters, floods cause the most damage in agriculture, fisheries, housing, and infrastructure sectors. Iran is also vulnerable to climate change due to its semi-arid climate and has experienced several devastating floods in the region over the past two decades. Various methods for flood forecasting exist worldwide, with machine learning techniques being particularly popular for their effectiveness in handling large datasets (Ranasinghe and Ilmini., 2020). In this context, based on the conducted research, it can be stated that among the models mentioned, the Support Vector Machine (SVM) model has been identified as a suitable method for modeling non-linear, complex, and dynamic processes like floods, demonstrating acceptable accuracy in flood forecasting. Therefore, considering the satisfactory accuracy of this method and the successful performance of meta-exploratory algorithms in enhancing its accuracy, two hybrid models, LS-SVM and SVM-GOA, were employed to simulate the peak discharge of the Poldokhtar station flood. The accuracy of these models, along with the classical Support Vector Machine (SVM), was assessed.

Methodology

In the research conducted to compare and evaluate the SVM, LS-SVM, and SVM-GOA models in flood modeling at the Poldokhtar hydrometry station, the following steps were undertaken. Initially, statistics and information regarding floods and precipitation in the upstream basin of the Poldokhtar hydrometric station were obtained daily from relevant organizations. Subsequently, a common statistical period was selected between the studied rain gauge and hydrometric stations (years 2009 to 2016). In the next phase, 74 flood events with instantaneous maximum discharge exceeding 60 cubic meters per second were identified within this timeframe, and the precipitation amount on the day of each flood event was determined at each upstream rain gauge station. Considering the impact of rainfall in the days leading up to a flood event on its occurrence, influenced by the time of concentration of the basin, a more detailed analysis included not only the rainfall on the day of the flood but also the rainfall from the five days prior to it. Finally, after selecting and determining the input and output values, the desired models were implemented.

Results and Discussion

In this research, which was conducted in order to compare and evaluate SVM, LS-SVM and SVM-GOA models in the flood modeling of Poldokhtar hydrometry station, the following results were obtained:

- Necessarily, the use of rainfall time series with a 5-day scale and the use of statistics from all upstream stations will not increase the accuracy of modeling, and the use of additional series will reduce the accuracy by disrupting the training process. It can be inferred that increased complexity does not necessarily result in improved performance of hydrological models, and the performance may vary based on the hydrological variable or conditions considered
- Overall, the results of calculating statistical indices demonstrate the relative superiority of the LS-SVM model compared to the SVM and SVM-GOA models.
- The utilization of rainfall statistics from the four stations of Chameshk, Veysian, Khorramabad, and Aleshtar not only enhanced the performance of the LS-SVM model in flood modeling but also

facilitated the identification of three key areas with the most significant impact on simulation accuracy.

- In this study, the average execution time of the LS-SVM model was approximately a few seconds, whereas the SVM-GOA model took several hours to run. Additionally, manually setting parameters for the classic SVM model is a time-consuming process. Therefore, the LS-SVM model was easier to implement due to its fewer adjustable parameters compared to the SVM-GOA and SVM models.
- Although the results show the relative equality of the SVM-GOA and SVM models, but the high execution time of the SVM-GOA model, especially in problems similar to this research, causes it to take months to achieve goals such as sensitivity analysis. From this point of view, the SVM model can be prioritized over the SVM-GOA model.
- The output of the LS-SVM model for the same inputs is constant and not random. Meanwhile, the output of SVM-GOA model has a wider confidence band and it shows the randomness of its output in flood simulation.

Key words: Flood modeling, Support vector machine, Grasshopper algorithm, Karkheh basin, Poldokhtar station.

پایان نامه
مستشاری
پدیده