

مقاله علمی-پژوهشی

رویکردی جدید در روندیابی سیلاب بر مبنای تلفیق تئوری بیز، ماشین بردار پشتیبان و الگوریتم بهینه‌سازی فراابتکاری

زهرا خرمی پور^۱، مهدی ولیخان انارکی^۲، سعید فرزین^{۳*}

تاریخ دریافت: ۱۴۰۲/۱۰/۲۵ تاریخ پذیرش: ۱۴۰۳/۰۴/۰۷

چکیده:

روندیابی سیل از طریق رودخانه‌ها، به‌منظور کنترل و جلوگیری از خسارات ناشی از سیل، از اقدامات مهم به شمار می‌آید. پیش‌بینی هیدروگراف سیلاب به‌طور کلی به دو روش هیدرولوژیکی (ماسکینگام و عددی) و یا روش‌های هیدرولیکی (حل معادلات سنت-ونانت) انجام می‌شود. روش‌های عددی به علت کم بودن پارامترها، هزینه کم محاسبات و در نهایت دقت مناسب، به‌صورت گسترده مورد توجه قرار دارد. در این مطالعه، به بررسی کاربرد هوش مصنوعی برای روندیابی جریان رودخانه پرداخته شده است. به این منظور، از الگوریتم ترکیبی، استراتژی تکاملی انطباق ماتریس کوواریانس (CMAES) و ماشین بردار بیزین (RVM) استفاده شده است. همچنین، نتایج آن با خروجی‌های حاصل از الگوریتم‌های طبقه‌بندی درخت تصمیم (CART)، الگوریتم تقویت گرادین مضعاف XGBoost، ماشین بردار بیزین ساده (RVM) و جنگل تصادفی (RF) مورد مقایسه قرار گرفته است. در نهایت برای ارزیابی عملکرد و کارایی الگوریتم‌ها، از معیارهای ارزیابی میانگین مطلق خطای (MAE)، ریشه میانگین مربعات خطا (RMSE) و ضریب تعیین (R^2) استفاده شده است. محاسبات روندیابی سیل در بازه ملاثانی اهواز (به‌عنوان مطالعه موردی) برای دوره مطالعاتی ۲۰۱۱-۲۰۱۷ انجام شده است. نتایج به‌دست‌آمده، حاکی از موفقیت مناسب الگوریتم ترکیبی RVM-CMAES به‌عنوان ابزاری قدرتمند در بحث پیش‌بینی هیدروگراف، برای مدیریت سیلاب با معیارهای ارزیابی، $MAE=19/69$ ، $RMSE=29/37$ و $R^2=0/93$ در دوره آموزش و $MAE=19/16$ ، $RMSE=27/83$ و $R^2=0/94$ در دوره آزمون می‌باشد. علاوه بر این نمودارهای تیلور و سری زمانی هیدروگراف روندیابی شده حاکی از برتری الگوریتم ترکیبی نسبت به سایر الگوریتم‌ها بوده است. الگوریتم پیشنهادی دارای پتانسیل بالا برای استفاده در سامانه‌های هشدار سیل و سایر مسائل مربوط به مدیریت منابع آب می‌باشد.

واژه‌های: پیش‌بینی سیلاب، هیدروگراف، روندیابی سیلاب، روش‌های یادگیری ماشین

مقدمه:

و انجام اقدامات پیشگیرانه، پیش‌بینی سیل به‌عنوان یک اقدام غیر سازه‌ای امری حیاتی است (Yuan et al., 2020; pal et al., 2022). در سال‌های اخیر، اثرات منفی مخاطرات طبیعی مانند سیل ناگهانی، افزایش یافته است (Bui et al., 2019). درک نرخ اوج جریان و به‌کارگیری روش‌های پیش‌بینی کننده، مانند آنالیز هیدروگراف، می‌تواند به حداقل رساندن مشکلات و پیشگیری از خسارات ناشی از سیلاب، مدیریت و کنترل سیلاب رودخانه بر اساس شناخت ویژگی‌های سیل کمک کند. روندیابی سیلاب بر اساس مدل‌های فیزیکی (بر اساس قانون بقای جرم) و مدل‌های داده‌کاوی قابل انجام است (Safavi, 2011). مدل‌های داده‌کاوی به دلیل توانایی آن‌ها در ایجاد روابط مبتنی بر داده‌ها بدون در نظر گرفتن روابط فیزیکی محبوبیت پیدا کرده‌اند. این مدل‌ها با برقراری رابطه بین متغیرهای هیدرولوژیکی از طریق آموزش و برازش داده‌ها، دقت را

سیلاب ممکن است ناشی از بارندگی، شکست سدها، خاک‌ریزها و یا ترکیبی از این عوامل شکل گیرد. بررسی تغییرات به وجود آمده در موج سیلاب در مسیر یک رودخانه، روندیابی سیلاب نامیده می‌شود. به همین دلیل برای به حداقل رساندن خسارات ناشی از سیل

۱- دانشجوی کارشناسی ارشد، گروه مهندسی آب و سازه‌های هیدرولیکی، دانشکده مهندسی عمران، دانشگاه سمنان، سمنان، ایران
۲- دانش‌آموخته دکتری، گروه مهندسی آب و سازه‌های هیدرولیکی، دانشکده مهندسی عمران، دانشگاه سمنان، سمنان، ایران
۳- دانشیار، گروه مهندسی آب و سازه‌های هیدرولیکی، دانشکده مهندسی عمران، دانشگاه سمنان، سمنان، ایران
(* نویسنده مسئول: Email: saeed.farzin@semnan.ac.ir)

کرده‌اند. داده‌های مد نظر را با استفاده از روش‌های، درخت رگرسیون، ماشین بردار پشتیبان و ANFIS مورد آموزش قرار داده‌اند. نتایج به‌دست‌آمده توسط معیارهای ارزیابی R^2 ، MSE و دیاگرام تیلور مورد مقایسه قرار گرفته‌اند. نتایج حاصل، نشان‌دهنده موفقیت ANFIS در روندیابی سیلاب در این مطالعات را نشان می‌دهد (Sarigöl and Katipoğlu, 2023). برای پیش‌بینی و جلوگیری از خسارات ناشی از سیلاب در آنکارا از مدل‌های مختلفی مانند RF، K-NN، GBoost، BTree و SVM استفاده کرده‌اند. نتایج حاصل از معیارهای ارزیابی، نشان‌دهنده آن است که K-NN دارای کمترین میزان خطا و همچنین بهترین عملکرد در پیش‌بینی سیلاب را دارد (Katipoğlu and Sarigöl, 2023). برای به دست آوردن اطلاعات سیل در پایین‌دست به‌منظور کاهش سیل از ۶ مدل مانند SVR، GPR، RFR، MLP، LSTM و GRU استفاده کرده‌اند. نتایج حاصل از تحلیل و بررسی خروجی مدل‌ها نشان‌دهنده موفقیت مدل GRU برای مسیریابی سیل رودخانه یانگ تسه است (Zhou and Kang, 2023).

با توجه به بررسی‌های انجام شده در زمینه روندیابی سیلاب، به علت اینکه سیل می‌تواند منجر به زیان‌های بسیاری شود؛ مدیریت، پیش‌بینی و روندیابی سیلاب از ضروری‌ترین گام‌ها در این زمینه می‌باشد که با توجه به مروری بر تحقیقات پیشین، برای به دست آوردن ابرپارامترها روش‌های یادگیری ماشین از سعی و خطا استفاده شده است. درحالی‌که این روش‌ها دارای تعداد زیادی پارامتر هستند و تعیین آن‌ها با استفاده از روش سعی و خطا نیاز به هزینه و زمان محاسباتی زیادی دارد. لذا، برای بالا بردن سرعت انجام مدل‌سازی و افزایش دقت، در مطالعه حاضر از روش ترکیبی RVM-CMAES استفاده شده است. این روش به داده‌های هیدرولیکی جریان و هندسه رودخانه نیاز ندارد و بر اساس داده‌های مشاهداتی جریان رودخانه در پایین‌دست و بالادست رودخانه روندیابی سیل را انجام می‌دهد. در روش ارائه شده، پارامترهای بهینه‌الگوریتم RVM توسط الگوریتم بهینه‌سازی CMAES به دست می‌آیند. این امر باعث بالا رفتن دقت، عدم نیاز به سعی و خطا و در نتیجه بالا رفتن سرعت در روندیابی جریان رودخانه می‌شود. روش پیشنهادی می‌تواند جهت پیش‌بینی سیل در سامانه‌های هشدار سیل در مناطق نزدیک به رودخانه‌ها به کار گرفته شود.

مواد و روش‌ها

مراحل روش مورد استفاده در این مطالعه برای روندیابی سیلاب در بازه ملاثانی - اهواز، به مراحل زیر تقسیم می‌شوند: ۱- آماده‌سازی داده‌ها (به‌طور کلی دو نوع متغیر داریم، متغیرهای وابسته و متغیرهای مستقل که در مطالعه روندیابی سیلاب، هیدروگراف بالادست را

افزایش می‌دهند. ایجاد رابطه آماری بین ورودی هیدرولوژیکی و خروجی بدون در نظر گرفتن صریح روابطی که در میان فرآیندهای فیزیکی وجود دارد؛ توسعه الگوریتم‌های ترکیبی جدید یادگیری ماشین و چالش برآورد پارامترها و داده‌های اصلی را برطرف کرده است. که این امر موجب، بهبود دقت، کاهش هزینه‌های محاسباتی، همگرایی سریع و سازگاری با مجموعه داده‌های گسترده می‌شود (Tayfur., 2017). در سال‌های اخیر مطالعاتی در این زمینه انجام شده است که در ادامه بررسی می‌شوند.

از شبکه عصبی (ANN)، الگوریتم ژنتیک (GA)، بهینه‌ساز کلونی مورچه (ACO) و ازدحام ذرات (PSO) به پیش‌بینی هیدروگراف سیلاب استفاده کرده‌اند. نتایج نشان داد که روش‌های ذکر شده، هیدروگراف پایین‌دست را با دقت خوبی روندیابی کرده‌اند (Tayfur et al., 2007). به‌منظور روندیابی سیلاب و استخراج هیدروگراف پایین‌دست برای کاهش خسارات ناشی از سیل، از روش‌های برنامه‌ریزی غیرخطی (NLP)، و رگرسیون خطی (MLR) استفاده کرده‌اند. نتایج به‌دست‌آمده توسط این دو روش را با روش ماسکینگام مقایسه کرده‌اند. نتایج نشان می‌دهد، خروجی‌های ناشی از دو روش NLP و ML، نتایجی نزدیک به روش ماسکینگام داشته‌اند (Zare et al., 2014). به‌منظور پیش‌بینی جریان رودخانه از روش شبکه عصبی مصنوعی ANN استفاده کرده‌اند. نتایج به‌دست‌آمده، نشان‌دهنده عملکرد مناسب شبکه عصبی مصنوعی برای تشخیص خطر سیل در رودخانه می‌باشد (Elsafi, 2014). از روش ماسکینگام خطی و الگوریتم بهینه‌سازی ازدحام ذرات (PSO) برای روندیابی سیلاب در بازه ملاثانی و اهواز استفاده کرده‌اند. این روش در مقایسه با رویکردهای هیدرولوژیکی سنتی کارایی بهتری ارائه داده است (Norouzi and Bazargan, 2019). پیش‌بینی سیلاب‌هایی که به‌طور مکرر در مناطق کوهستانی خشک رخ می‌دهد دارای محدودیت داشتن داده می‌باشد. در این مطالعه با استفاده از مدل ANN و Muskingum به پیش‌بینی سیلاب در مناطق کوهستانی خشک پرداخته‌اند. نتایج نشان‌دهنده برتری ANN برای پیش‌بینی سیل در مناطق کوهستانی خشک با داده‌های رواناب کمیاب می‌باشد (Yuan et al., 2020). از دو روش طبقه‌بندی مبتنی بر شیء و RVM به‌منظور نظارت بر مناطق سیلابی برای جلوگیری از سیلاب و تعیین خسارات ناشی از سیل استفاده کرده‌اند. نتایج به‌دست‌آمده حاکی از دقت بالا و عملکرد مناسب RVM برای پیش‌بینی و تعیین حدود آب در مناطقی که خطر سیلاب وجود دارد را نشان می‌دهد (Sharifi, 2020). برای پیش‌بینی هیدروگراف سیلاب از ANN و برای تولید داده‌های مصنوعی از HEC-RAS استفاده کرده‌اند. نتایج نشان‌دهنده دقت زیاد و یکی بودن هیدروگراف پیش‌بینی شده توسط ANN بوده است (Tawfik, 2023). به‌منظور محاسبات و برآورد سیل در مناطق پایین‌دست، از داده‌های هیدروگراف میانگین روزانه سیل استفاده

ارتباط چندگانه RVS استفاده می‌کند. RVM می‌تواند عدم قطعیت را طبقه‌بندی کند و همچنین مشکل داده‌های کوچک را که قابل حل نیستند را با توجه به محدودیت‌ها و شرایط موجود هموار و حل کند. این روش با ایجاد مدل‌هایی با ساختار و روش پارامتر سازی که با محتوای اطلاعاتی داده در ارتباط هستند، از پیچیدگی جلوگیری می‌کند. مزیت RVM نسبت به SVM استفاده از تعداد کمتری تابع هسته پایه (KF)، برای ساخت مدل رگرسیونی است. در این الگوریتم برای پیش‌بینی تابع $f(x)$ در نقطه تصادفی x و در زمان t از مدل خطی بر اساس روابط ۲-۴ استفاده می‌شود.

$$t_i = f(x_i) + \varepsilon_i \quad (2)$$

در معادله ۲، ε_i مؤلفه‌ای برای اندازه‌گیری نویز است که میانگین آن صفر می‌باشد.

$$y(x) = \sum_{i=1}^m \omega_i \varphi_i(x) \quad (3)$$

در معادله ۳، با فرض خطی بودن مدل، $y(x)$ یک تابع مجهول است، $\varphi_i(x)$ تابع ادغام شده از توابع پایه شناخته شده است و ω_i بردار تشکیل‌شده از وزن ترکیبی خطی است (Gayathri et al., 2016; Wang et al., 2022). پیش‌بینی RVM بر اساس تابع زیر (معادله ۴) انجام می‌شود.

$$\sum_{i=1}^N \omega_i \cdot k(x \cdot x_i) + \omega_0 \quad (4)$$

در رابطه ۴، ω_i بردارهای وزن و k تابع هسته هستند. شکل ۱، دیاگرام ساختار RVM را نمایش می‌دهد.

الگوریتم بهینه‌ساز CMAES:

CMAES استراتژی تکامل انطباق ماتریس کوواریانس یک بهینه‌ساز پیشرفته برای عملکردهای جعبه‌سیاه پیوسته است (Hansen and Ostermeier, 2001). الگوریتم‌های بهینه‌ساز بیزی برای بهینه‌سازی مسائل با ابعاد کم مورد استفاده قرار می‌گیرد درحالی‌که بهینه‌ساز CMAES برای ارزیابی بهینه‌سازی مسائل با ابعاد زیاد استفاده می‌شود (Loshchilov, 2013). این الگوریتم راه‌حل‌ها را با استفاده از یک توزیع نرمال چند متغیره، ارزیابی می‌کند. همچنین می‌تواند در مقایسه با دیگر بهینه‌سازها، برای مسائل پیچیده‌تر استفاده شود و نتایج دقیق‌تری ارائه دهد (Gelbart et al., 2014; Hansen et al., 2003). پارامترهای الگوریتم مورد نظر بر اساس روابط ۵-۷ محاسبه می‌شود.

$$x_k^{(g+1)} = (x)_\mu^{(g)} + \sigma^{(g)} \cdot B^{(g)} \cdot D^{(g)} \cdot z_k^{(g+1)} \quad k=1, \dots, \lambda \quad (5)$$

$$(x)_\mu^{(g)} = \frac{1}{\mu} \sum_{i \in I_{sel}^g} x_i^{(g)} \quad (6)$$

$$(z)_\mu^{(g+1)} = \frac{1}{\mu} \sum_{i \in I_{sel}^{g+1}} z_i^{(g+1)} \quad (7)$$

به‌عنوان متغیر مستقل به مدل می‌دهیم) ۲- تقسیم‌بندی داده‌ها به دو دوره آموزش و آزمون ۳- آموزش و آزمون الگوریتم‌های یادگیری ماشین.

در مطالعه حاضر مشابه با مطالعه (El-Haddad et al., 2021)، ۷۰ درصد داده‌ها به دوره آموزش و ۳۰ درصد را به دوره آزمون اختصاص داده شده است.

الگوریتم درخت طبقه‌بندی و رگرسیون (CART)

الگوریتم درخت رگرسیون طبقه‌بندی (CART)، شکلی از تقسیم‌بندی بازگشتی باینری هستند که در آن هر گره قرار است به دو گروه تقسیم شود و این تقسیم کردن تا زمانی که انشعاب دیگری نتوان ساخت، ادامه دارد. این الگوریتم دارای پارامترهایی می‌باشد که برای بهبود مدل، می‌تواند تغییر کند. طبقه‌بندی شامل انتخاب مکرر بهینه‌ترین ویژگی‌ها، تقسیم‌بندی داده‌ها، ساخت درخت تصمیم‌گیری دودویی و در نهایت استخراج یک درخت تصمیم که امکان طبقه‌بندی را فراهم می‌کند، می‌باشد. درخت تصمیم CART برای آنکه تصمیم بگیرد چگونه درخت را انتخاب کند، از معیاری به نام معیار شاخص جینی (GINI) استفاده می‌کند. جینی شاخص ناخالصی مجموعه نمونه را اندازه‌گیری می‌کند. این شاخص طبق معادله ۱، محاسبه می‌شود (Breiman, 2017; Byeon, 2015; Lam et al., 2019).

$$G = \sum_{i=1}^C P_{(i)} \cdot [1 - P_{(i)}] \quad (1)$$

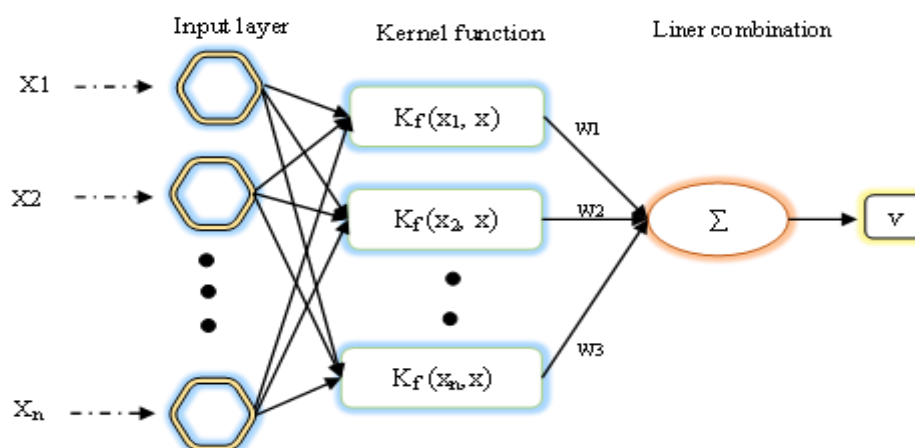
در رابطه ۱: G ضریب جینی، $P_{(i)}$ احتمال یک نمونه است که در دسته i طبقه‌بندی می‌شود (Chen et al., 2023).

الگوریتم XGBoost:

XGBoost یک روش درخت تقویت شده مقیاس‌پذیر است که برای مدل‌سازی و پیش‌بینی مسائل مختلف طراحی شده است. XGBoost یک الگوریتم یادگیری گروهی جدید مبتنی بر درخت است. XGBoost، از سازوکار پیش‌مرتب‌سازی برای یافتن نقطه تقسیم بهینه استفاده می‌کند. همه ویژگی‌ها بر اساس مقادیر ویژگی از پیش مرتب شده‌اند و در اولین پیمایش ذخیره می‌شوند. XGBoost، از محاسبات موازی پشتیبانی می‌کند اما از موازی‌سازی مشخصه پشتیبانی نمی‌کند. تابع ضرر را به فرم بسط تیلور مرتبه دوم با استفاده از هر دو مشتق اول و دوم، برای بهبود سرعت بهینه‌سازی، بهینه می‌کند. این الگوریتم از مشکل پیش‌برازش جلوگیری می‌کند (Chen et al., 2016; Han et al., 2022).

ماشین بردار بیزین (RVM)

ماشین بردار بیزین (RVM) در سال ۱۹۹۹ (Tipping, 1999) ارائه شده است. RVM روش توسعه‌یافته SVM است که از بردارهای



شکل ۱- دیاگرام ساختار RVM

می‌شود. در نهایت عامل جستجو، بهترین موقعیت و بهترین عملکرد تعیین می‌شود. با توجه به اینکه محاسبه پارامترهای ماشین یادگیری RVM به صورت سعی و خطا تعیین نشده است، دقت محاسبات بیشتر می‌شود.

معیارهای ارزیابی الگوریتم‌ها:

در مطالعه حاضر، نتایج الگوریتم‌های استفاده شده برای پیش‌بینی جریان سیلاب، توسط معیارهای ارزیابی از جمله، میانگین مطلق خطا (MAE)، ریشه میانگین مربعات خطا (RMSE)، و ضریب تعیین (R^2)، ارزیابی می‌شود. روابط معیارهای ارزیابی ذکر شده به شرح معادلات ۸-۱۰ می‌باشد (Valikhan Anaraki et al., 2019, Ehteram et al., 2018, Azad et al., 2019)

$$MAE = \sum_{i=1}^n \frac{|num_{i,exp} - num_{i,est}|}{N} \quad (8)$$

$$RMSE = \left(\sqrt{\frac{\sum_{i=1}^n (num_{i,exp} - num_{i,est})^2}{N}} \right) \quad (9)$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (Flow_{i,exp} - Flow_{i,est})^2}{\sum_{i=1}^n (Flow_{i,exp} - Flow_{i,exp})^2} \quad (10)$$

در معادله ۱۰، $Flow_{i,exp}$ مقادیر مشاهده شده، $Flow_{i,est}$ مقادیر تخمین شده، و n تعداد کل داده‌ها می‌باشد.

محدوده مطالعاتی و داده‌های مورد استفاده:

رودخانه کارون پرآب‌ترین رودخانه کشور ایران، با طولی بیش از ۹۰۰ کیلومتر است. حوضه آبریز این رودخانه با مساحتی بالغ بر ۶۷۰۰۰ کیلومتر مربع می‌باشد. ایستگاه مورد مطالعه در این تحقیق، ایستگاه هیدرومتری ملاثانی در طول جغرافیایی ۴۸ درجه و ۵۳ دقیقه

در روابط (۵-۷)، g بیان‌کننده نسل فعلی، I_{sel}^g مجموعه‌ای شاخص‌های افراد منتخب نسل g و $\sigma^{(g)}$ اندازه مرحله جهانی می‌باشد.

جنگل تصادفی (RF):

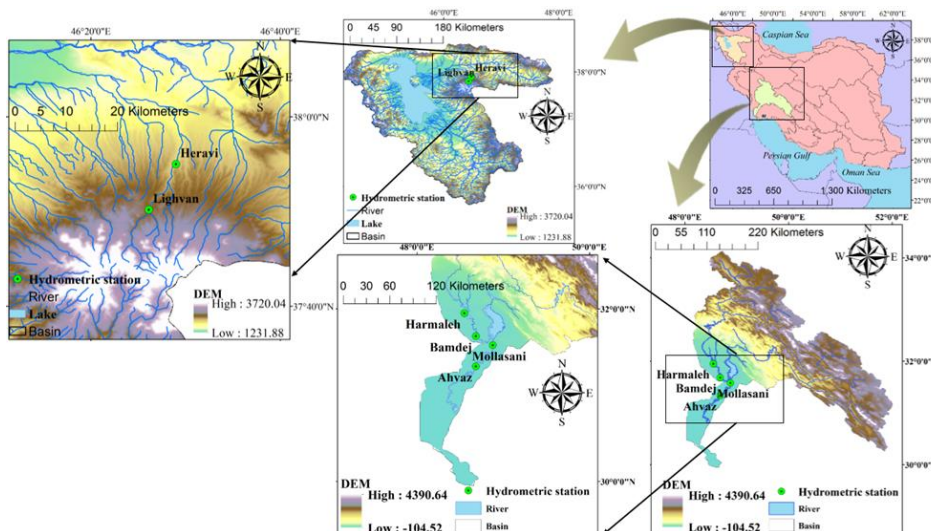
بریمن (Breiman., 2001) مدل RF را به‌عنوان یک الگوریتم یادگیری مؤثر با استفاده از مجموعه‌ای از درختان تصمیم معرفی کرد. RF به‌طور کلی در دو بخش مختلف یعنی رگرسیون و طبقه‌بندی استفاده می‌شود. RF با استفاده از تعداد زیادی از درخت‌های رگرسیون در مقایسه با الگوریتم درخت تصمیم ساده باعث افزایش دقت می‌شود. هدف از درختان عبارت است از تقسیم یک مجموعه داده به زیرمجموعه‌های با ناهمگنی کمتر و با توجه به یک متغیر پاسخ. با استفاده از متغیرهای پیش‌بینی کننده مجموعه داده‌ها در گره‌ها تا زمانی که یک معیار توقف وجود داشته باشد، تقسیم می‌شود (Sultana et al., 2018; Gholizadeh et al., 2020).

الگوریتم هیبریدی RVM - CMAES:

الگوریتم ماشین یادگیری RVM دارای سه پارامتر، α ، β و $coef1$ (عرض تابع کرنل) می‌باشد. درحالی‌که روش خاصی برای تعیین مقدار این پارامترها وجود ندارد، اما تأثیر زیادی بر نتیجه و دقت ماشین یادگیری RVM می‌گذارند. بنابراین در پژوهش حاضر به‌جای جستجوی تصادفی این پارامترها از الگوریتم بهینه‌ساز CMAES استفاده می‌شود. به این منظور، ابتدا پارامترهای α ، β و $coef1$ به‌طور تصادفی به هر یک از عوامل جستجوی CMAES اختصاص داده می‌شوند. سپس RVM-CMAES برای هر عامل جستجو (α ، β و $coef1$) اجرا می‌شود. در ادامه تابع مورد نظر برای هر یک محاسبه

حوضه مورد مطالعاتی و همچنین بازه‌های انتخاب شده برای پیش‌بینی هیدروگراف جریان آورده شده است.

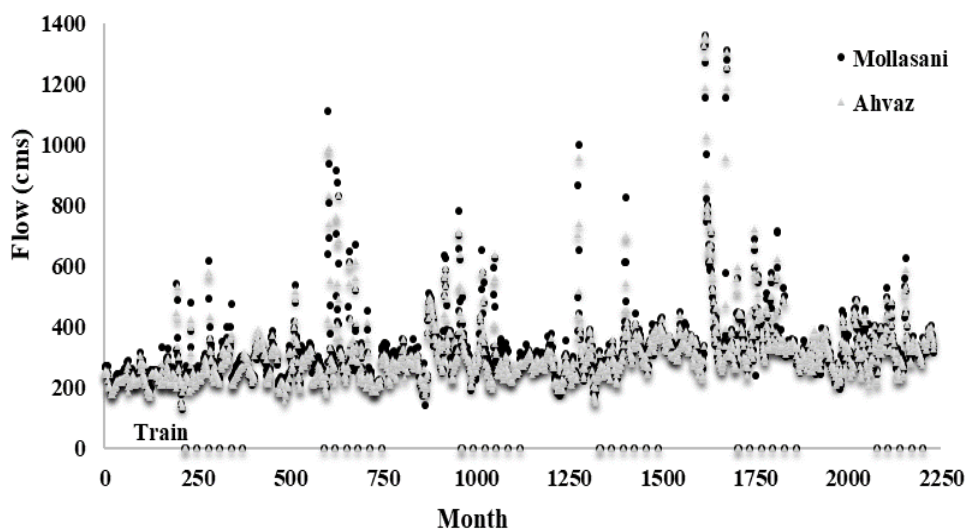
شرقی، عرض جغرافیایی ۳۱ درجه و ۳۵ دقیقه شمالی و ارتفاع ۱۸ متر از سطح دریا، در سال ۱۳۴۴ تأسیس شده و در فاصله ۳۵ کیلومتری شمال شهر اهواز واقع شده است. در شکل ۲، موقعیت قرارگیری



شکل ۲- موقعیت قرارگیری محدوده مطالعاتی (Valikhan Anaraki et al., 2023)

این وسیله می‌توان هیدروگراف جریان خروجی را بر اساس جریان ورودی در بالادست پیش‌بینی کرد و نسبت به هشدار سیل اقدام نمود. در شکل ۳، طبق راهنمای مشخص شده در نمودار، نقاط روشن مربوط به ایستگاه اهواز و نقاط تیره مربوط به ایستگاه ملاثانی می‌باشد.

در شکل ۳، نتایج هیدروگراف ورودی و خروجی بازه ملاثانی- اهواز نمایش داده شده است. با توجه به این شکل، دبی حداکثر و حجم هیدروگراف ورودی (ملاثانی) با عبور جریان از بازه رودخانه مقداری کاهش می‌یابد. با این وجود دبی حداکثر و حجم سیلاب در ایستگاه خروجی (اهواز) همچنان قابل توجه است. لذا، نیاز به روش‌های دقیق برای روندیابی سیل در این بازه از رودخانه است. به



شکل ۳- هیدروگراف ملاثانی و اهواز

نتایج و بحث

به ترتیب ۲۳/۳۹۰، ۲۵/۸۵۴، ۲۴/۱۰۴، ۳۹/۲۷۳ و ۲۹/۳۷۸ به دست آمد. مقدار R^2 نیز به ترتیب برابر با ۰/۹۶۱۳، ۰/۹۵۲، ۰/۹۵۸، ۰/۹۳۷ و ۰/۹۳۸ محاسبه شد. بنابراین، در دوره آموزش دقت الگوریتم‌های بررسی شده قابل قبول است. لیکن برای انتخاب بهترین الگوریتم، باید از داده‌های جدید جهت آزمون الگوریتم‌ها استفاده کرد. بر اساس نتایج دوره آزمون در جدول ۱، الگوریتم ترکیبی RVM-CMAE از دقت بیشتری نسبت به سایر الگوریتم‌ها برخوردار بود. مقادیر معیارهای MAE، RMSE و R^2 برای الگوریتم RVM-CMAE به ترتیب ۲۹، ۲۶ و ۹۲ درصد بهتر از سایر الگوریتم‌ها تخمین زده شده است.

در این بخش نتایج حاصل از روندیابی سیل توسط روش الگوریتم‌های یادگیری ماشین، برای بازه ملاثانی- اهواز مورد بررسی قرار گرفته است. جدول ۱، مقادیر دقت به دست آمده از ارزیابی الگوریتم‌های استفاده شده را نشان می‌دهد. بر اساس نتایج این جدول در دوره آموزش، مقدار معیار MAE برای الگوریتم‌های RF، CART، XGBoost، RVM و RVM-CMAE به ترتیب برابر با ۱۵/۴۳۳، ۱۶/۶۲۸، ۲۷/۸۶۶، ۲۹/۳۷۸ و ۱۹/۶۹۲ بود. معیار RMSE برای الگوریتم‌های RF، CART، XGBoost، RVM و RVM-CMAE

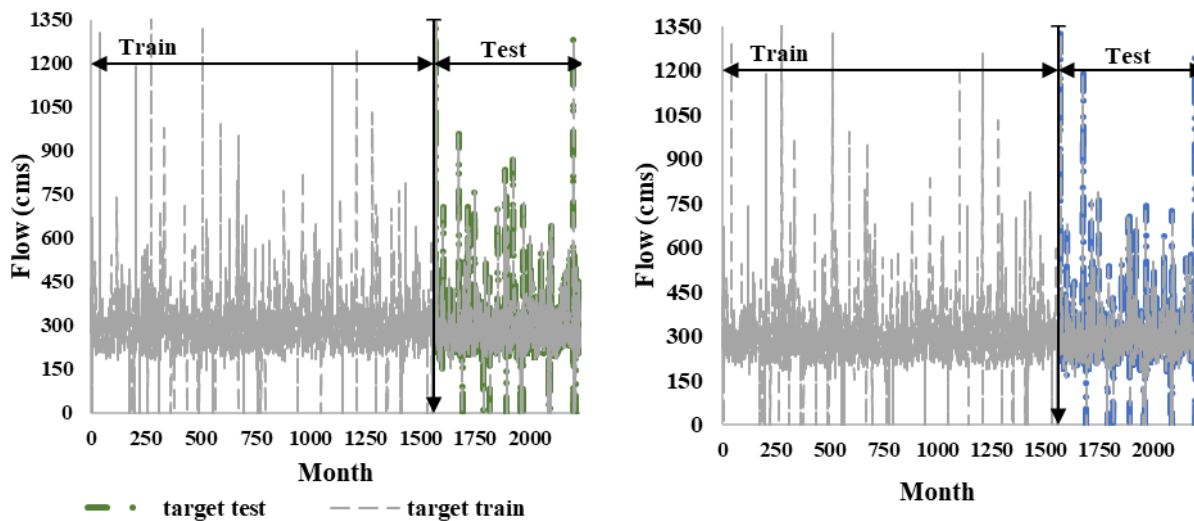
جدول ۱- نتایج معیارهای ارزیابی در دوره آموزش و آزمون

دوره	مدل	MAE	RMSE	R^2
آموزش	CART	۱۵/۴۳۳	۲۳/۳۹۰	۰/۹۶
	RF	۱۷/۹۱۳	۲۵/۸۵۴	۰/۹۵
	XGBoost	۱۶/۶۲۸	۲۴/۱۰۴	۰/۹۶
	RVM	۲۷/۸۶۶	۳۹/۲۷۳	۰/۹۴
	RVM_CMAES	۱۹/۶۹۲	۲۹/۳۷۸	۰/۹۴
آزمون	CART	۲۲/۱۸۷	۳۲/۷۱۹	۰/۳۶
	RF	۲۰/۰۰۷	۲۹/۲۵۵	۰/۹۴
	XGBoost	۲۰/۸۴۴	۳۰/۹۲۹	۰/۹۳
	RVM	۲۶/۸۴۰	۳۷/۶۸۴	۰/۹۴
	RVM_CMAES	۱۹/۱۶۹	۲۷/۸۳۲	۰/۹۴

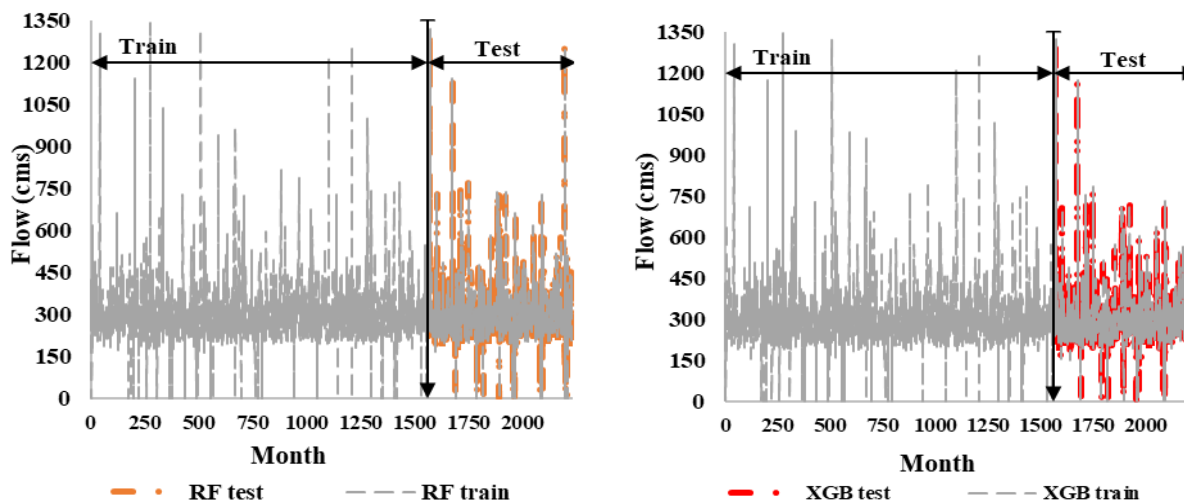
هیدروگراف مشاهداتی برخوردار بوده است. این موضوع صحت نتایج جدول ۱ را تأیید می‌کند.

در شکل ۵، نمودار R^2 نتایج مدل‌سازی شده روش‌های مدل‌سازی در دوره آزمون نمایش داده شده است. این نمودار برآزش یک مدل رگرسیونی را ارزیابی می‌کند. ضریب تعیین (R^2) نشان‌دهنده نسبتی از واریانس برحسب متغیر وابسته است که از متغیرهای مستقل قابل پیش‌بینی باشد. نمودارهای R^2 نشان‌دهنده ارتباط بین پیش‌بینی‌های مدل و داده‌های واقعی است. هرچه مقدار R^2 به یک نزدیک‌تر باشد، مدل بهتری داریم که با داده‌ها تطابق بیشتری دارد. در حالی که اعداد نزدیک به صفر به معنای عدم تطابق مدل با داده‌های واقعی هستند. با توجه به مقادیر به دست آمده R^2 در شکل ۵ مشاهده می‌شود، بیشترین مقدار R^2 در دوره آزمون مربوط به روش پیشنهادی RVM-CMAES با مقدار ۰/۹۴۳۲ و کمترین مقدار به دست آمده مربوط به روش CART با مقدار ۰/۹۲۰۹ می‌باشد. این ارزیابی می‌تواند به انتخاب مدل مناسب برای پیش‌بینی جریان کمک کند.

در شکل ۴، نتایج سری زمانی مدل‌سازی شده هیدروگراف پایین دست بازه ملاثانی- اهواز در مقابل هیدروگراف مشاهداتی نمایش داده شده است. این نمودار مربوط به تحلیل سری زمانی هستند که نشان می‌دهند عملکرد مدل‌های مختلف پیش‌بینی، از جمله XGBoost، RF، CART و RVM-CMAE در مقایسه با داده‌های آزمون و داده‌های آموزش چگونه است. به طور کلی هر نمودار شامل دو خط است که یکی برای داده‌های آزمون و دیگری برای داده‌های آموزش است. اختلاف بین این دو خط نشان‌دهنده عملکرد مدل در پیش‌بینی داده‌های جدید باشد. در نمودارهای آموزش، مدل‌ها به داده‌های آموزشی تطابق بیشتری دارند، اما در نمودارهای آزمون، عملکرد مدل‌ها در پیش‌بینی داده‌های جدید نشان‌دهنده توانایی واقعی آنها است. با توجه به این شکل، در دوره آموزش، هیدروگراف مدل شده از برآزش خوبی نسبت به هیدروگراف مشاهداتی برخوردار است. در حالی که در دوره آزمون، هیدروگراف حاصل از الگوریتم RVM-CMAE نسبت به سایر الگوریتم‌های بررسی شده از تطابق بیشتری با



(الف) (ب)

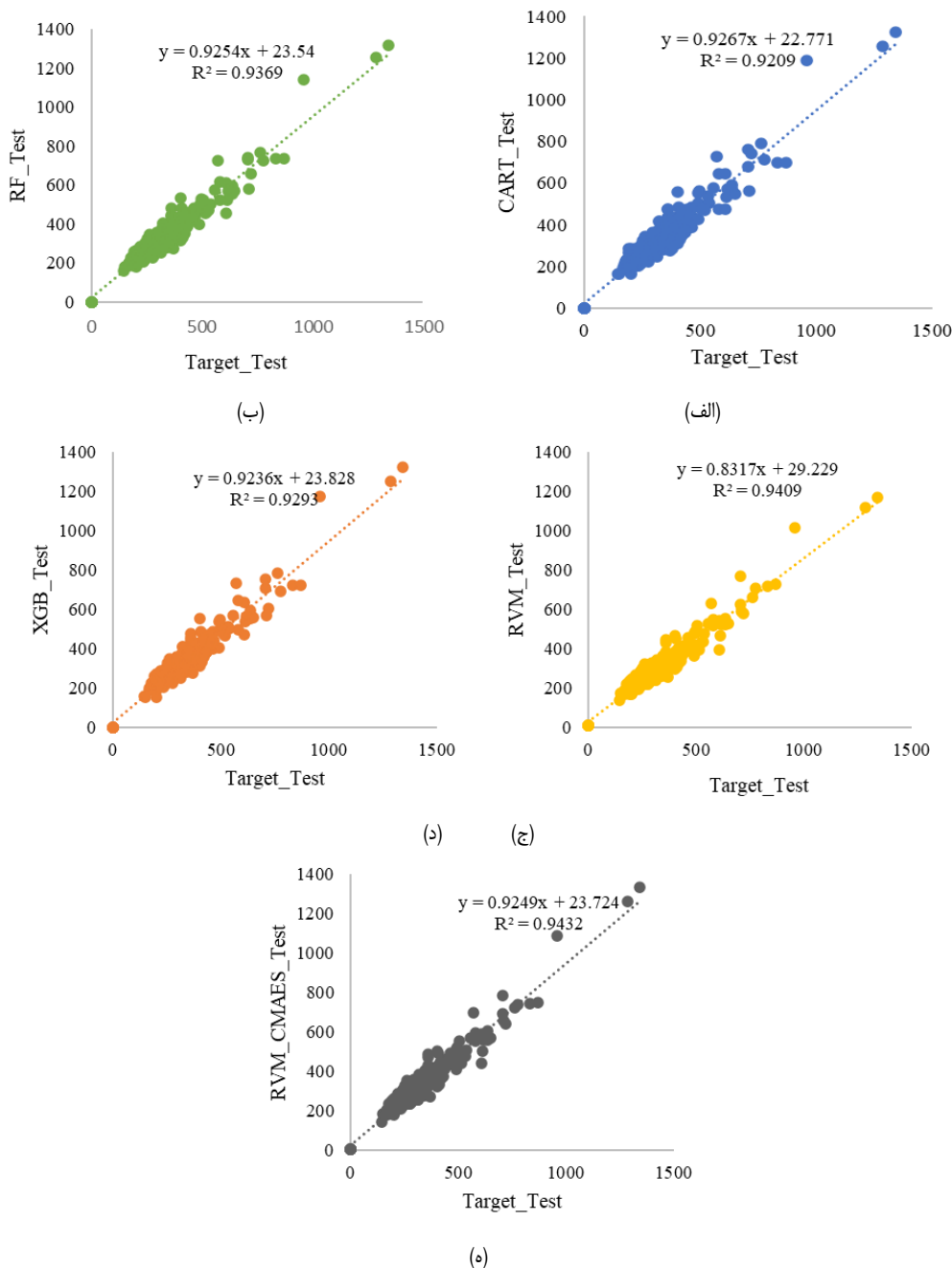


(ج) (د)

شکل ۴- نمودار سری- زمانی روش‌های استفاده شده برای پیش‌بینی جریان رودخانه

الگوریتم‌ها می‌باشند. الگوریتم نزدیک‌تر به داده‌های مشاهداتی دارای بهترین عملکرد است. با مقایسه موقعیت نقاط می‌توان فهمید که کدام مدل در فاز آزمون دقیق‌تر عمل کرده است. R^2 برای RF برابر با ۰/۹۳۶۹ و برای RVM-CMAES برابر با ۰/۹۴۳۲ است. همان‌طور که مشاهده می‌شود، RVM-CMAES دارای نزدیک‌ترین فاصله با نقطه مربوط به داده‌های مشاهداتی و کمترین انحراف استاندارد را در فاز آزمون دارد، که نشان‌دهنده دقت و قدرت پیش‌بینی بالای آن مدل و در نتیجه بهترین عملکرد است. بنابراین، RVM-CMAES از نظر عملکرد و دقت، روشی مناسب در مدل‌سازی هیدروگراف جریان رودخانه در بازه‌های ذکر شده بوده و روندیابی را با دقت مناسبی انجام داده است.

در شکل ۶ نتایج مدل‌سازی پیش‌بینی هیدروگراف جریان رودخانه بر اساس الگوریتم‌های ذکر شده با به کار گرفتن نمودار تیلور نمایش مقایسه شده‌اند. مدل‌های مختلف با رنگ‌های متفاوت نشان داده شده‌اند و اگر یک مدل به نقطه مرکزی نزدیک‌تر باشد، نشان‌دهنده عملکرد بهتر آن مدل می‌باشد. این نمودار به صورت بصری امکان انتخاب بهترین مدل را فراهم می‌کند. در دیاگرام تیلور، هر نقطه نشان‌دهنده انطباق بین پیش‌بینی‌های مدل و داده‌های مشاهده شده است. در این نمودار محور افقی و عمودی انحراف معیار، اعداد روی کمان دایره مقدار ضریب همبستگی پیرسون و دایره‌های سبزرنگ مقدار RMSE را نمایش می‌دهند. مربع سیاه‌رنگ نیز نشان‌دهنده داده‌های مشاهداتی و دیگر نقاط مشخص شده بیان‌گر

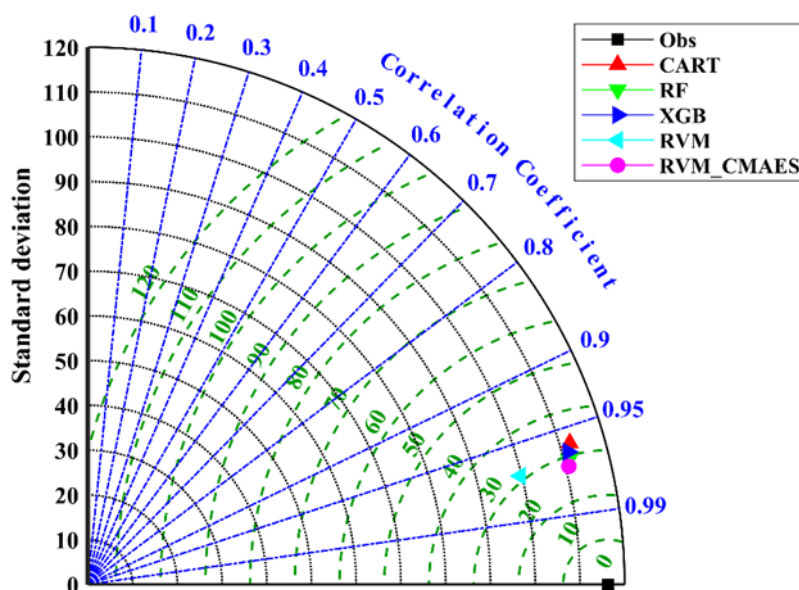


شکل ۵- نمودار R^2 روش‌های مدل‌سازی در دوره آزمون

نتیجه‌گیری

سیل یکی از مخرب‌ترین بلاهای طبیعی به شمار می‌رود که خسارات جانی و مالی سنگینی به بار می‌آورد. بر این اساس اقدامات زیادی قبل، حین و بعد از سیل مؤثر هستند. با توجه به اهمیت تخمین دقیق هیدروگراف سیلاب خروجی از بازه‌های پایین‌دست در مدیریت و کنترل سیل، ارائه روش‌های بهینه در این زمینه دارای اهمیت

می‌باشد. در این مقاله یک روش هیبریدی جدید برای مدل‌سازی هیدروگراف جریان رودخانه ارائه شد. در این مطالعه، الگوریتم RVM با الگوریتم بهینه‌ساز CMAES برای پیش‌بینی سیل در ملاتانی اهواز باهم ترکیب شدند. روش پیشنهادی یک روش کلی است و برای یک رودخانه خاص ارائه نشده است. همچنین به‌منظور ارزیابی عملکرد الگوریتم پیشنهادی از رودخانه کارون به‌عنوان مطالعه موردی استفاده شده است.



شکل ۶- نمودار تیلور

Karami, H., Nourani, V. and Shiri, J. 2019. Comparative evaluation of intelligent algorithms to improve adaptive neuro-fuzzy inference system performance in precipitation modelling. *Journal of hydrology*. 571: 214-224

Breiman, L. 2001. Random forests. *Machine learning*. 45: 5-32

Breiman, L. 2017. *Classification and regression trees*. Routledge.

Bui, D. T., Ngo, P. T. T., Pham, T. D., Jaafari, A., Minh, N. Q., Hoa, P. V. and Samui, P. 2019. A novel hybrid approach based on a swarm intelligence optimized extreme learning machine for flash flood susceptibility mapping. *Catena*. 179: 184-196

Byeon, H. 2015. Development of prediction model for endocrine disorders in the Korean elderly using CART algorithm. *Development*. 6(9)

Chen, T. and Guestrin, C. 2016. Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*. 785-794

Chen, Y., Zhang, W., Yang, J., Xu, Y., Cheng, W. and Peng, L. 2023. Extraction Methods for Small-scale Features on a Large Scale: Investigating Object-oriented Cart Decision Tree for Gravel Information Extraction. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*.

Ehteram, M., Mousavi, S. F., Karami, H., Farzin, S., Singh, V. P., Chau, K. W. and El-Shafie, A. 2018. Reservoir operation based on evolutionary

این روش‌ها را در رودخانه‌ها با شرایط مختلف می‌توان استفاده کرد و در صورت با کیفیت بودن و کافی بودن داده‌های ورودی و خروجی هدف، انتظار می‌رود که عملکرد مطلوبی داشته باشند.

مهم‌ترین نتایج به‌دست‌آمده از تحقیق به شرح زیر می‌باشد:

- در دوره آموزش، عملکرد الگوریتم‌های بررسی‌شده مناسب بود. اما در دوره آزمون دقت برخی الگوریتم‌ها کاهش یافت.
- الگوریتم هیبریدی RVM - CMAES نسبت به سایر الگوریتم‌ها دارای برتری بود.
- الگوریتم CMAES پارامترهای RVM را با دقت خوبی تخمین زد و سبب افزایش دقت RVM شد.
- روندیابی سیل توسط الگوریتم پیشنهادی با دقت خوبی انجام گرفت.

با وجود دقت قابل قبول الگوریتم پیشنهادی، برای آموزش آن نیاز به داده‌های با کیفیت و طول دوره آماری کافی است. همچنین، اجرای آن نیاز به استفاده از زبان برنامه‌نویسی پایتون و بسته‌های نرم‌افزاری مربوطه دارد. با توجه به دقت بالای الگوریتم RVM-CMAES استفاده از آن جهت روندیابی جریان رودخانه در دیگر ایستگاه‌ها و همچنین توسعه سیستم‌های هشدار سیل پیشنهاد می‌گردد. پیشنهاد می‌شود در مطالعات آینده با تعریف کردن چند مدل با پارامترهای ورودی متفاوت، تأثیر پارامترها را نیز در مدل‌سازی بررسی شود.

منابع

Azad, A., Manoochehri, M., Kashi, H., Farzin, S.,

- evolutionary computation. 1177-1184
- Norouzi, H. and Bazargan, J. 2019. Using the linear muskingum method and the particle swarm optimization (PSO) algorithm for calculating the depth of the rivers flood. *Iran-Water Resources Research*. 15(3): 344-347
- Pal, S. C., Chowdhuri, I., Das, B., Chakraborty, R., Roy, P., Saha, A. and Shit, M. 2022. Threats of climate change and land use patterns enhance the susceptibility of future floods in India. *Journal of environmental management*. 305: 114317
- Safavi, H.R. 2011. *Engineering Hydrology*. Arkan Danesh. Isfahan.724
- Sarıgöl, M. and Katipoğlu, O. M. 2023. Estimation of hourly flood hydrograph from daily flows using machine learning techniques in the Büyük Menderes River. *Natural Hazards*. 119(3): 1461-1477
- Sharifi, A. 2020. Flood mapping using relevance vector machine and SAR data: A case study from Aqqala, Iran. *Journal of the Indian Society of Remote Sensing*. 48(9), 1289-1296
- Sultana, Z., Sieg, T., Kellermann, P., Müller, M. and Kreibich, H. 2018. Assessment of business interruption of flood-affected companies using random forests. *Water*. 10(8): 1049
- Tayfur, G. 2017. Modern optimization methods in water resources planning, engineering and management. *Water Resources Management*. 31: 3205-3233
- Tawfik, A. M. 2023. River flood routing using artificial neural networks. *Ain Shams engineering journal*. 14(3): 101904.
- Tayfur, G., Moramarco, T. and Singh, V. P. 2007. Predicting and forecasting flow discharge at sites receiving significant lateral inflow. *Hydrological Processes: An International Journal*. 21 (14): 1848-1859.
- Tipping, M. 1999. The relevance vector machine. *Advances in neural information processing systems*. 12
- Valikhan Anaraki, M., Mousavi, S. F., Farzin, S., Karami, H., Ehteram, M., Kisi, O., Fai, CM., Hossain, MS., Hayder, G., Ahmed, AN. and El-Shafie, A. 2019. Development of a novel hybrid optimization algorithm for minimizing irrigation deficiencies. *Sustainability*. 11(8): 2337
- Valikhan Anaraki, M., Farzin, S., Ahmadianfar, I. and Shams, A. 2024. Development a novel discharge routing method based on the large discharge dataset, Muskingum model, optimization methods, and multi-criteria decision making. *Journal of Soft Computing in Civil Engineering*. 8(4).
- Wang, Z., Chen, H., Zhu, J. and Ding, Z. 2022. Daily algorithms and multi-criteria decision-making under climate change and uncertainty. *Journal of Hydroinformatics*. 20(2): 332-355
- El-Haddad, B. A., Youssef, A. M., Pourghasemi, H. R., Pradhan, B., El-Shater, A. H. and El-Khashab, M. H. 2021. Flood susceptibility prediction using four machine learning techniques and comparison of their performance at Wadi Qena Basin, Egypt. *Natural Hazards*, 105: 83-114.
- Elsafi, S. H. 2014. Artificial neural networks (ANNs) for flood forecasting at Dongola Station in the River Nile, Sudan. *Alexandria Engineering Journal*. 53 (3), 655-662
- Gayathri, B. M. and Sumathi, C. P. 2016. Comparative study of relevance vector machine with various machine learning techniques used for detecting breast cancer. In 2016 IEEE International Conference on Computational Intelligence and Computing Research (ICIC). 1-5
- Gelbart, M. A., Snoek, J. and Adams, R. P. 2014. Bayesian optimization with unknown constraints. 1403.5607
- Gholizadeh, M., Jamei, M., Ahmadianfar, I. and Pourrajab, R. 2020. Prediction of nanofluids viscosity using random forest (RF) approach. *Chemometrics and Intelligent Laboratory Systems*. 201: 104010
- Han, Y., Tang, R., Liao, Z., Zhai, B. and Fan, J. 2022. A novel hybrid GOA-XGB model for estimating wheat aboveground biomass using UAV-based multispectral vegetation indices. *Remote Sensing*. 14(14): 3506
- Hansen, N. and Ostermeier, A. 2001. Completely derandomized self-adaptation in evolution strategies. *Evolutionary computation*. 9(2): 159-195
- Hansen, N., Müller, S. D. and Koumoutsakos, P. 2003. Reducing the time complexity of the derandomized evolution strategy with covariance matrix adaptation (CMA-ES). *Evolutionary computation*. 11(1): 1-18
- Katipoğlu, O. M. and Sarıgöl, M. 2023. Prediction of flood routing results in the Central Anatolian region of Türkiye with various machine learning models. *Stochastic Environmental Research and Risk Assessment*. 37(6): 2205-2224
- Lam, F. M., Leung, J. C. and Kwok, T. C. 2019. The clinical potential of frailty indicators on identifying recurrent fallers in the community: The Mr. Os and Ms. OS cohort study in Hong Kong. *Journal of the American Medical Directors Association*. 20(12): 1605-1610
- Loshchilov, I., Schoenauer, M. and Sebag, M. 2013. Bi-population CMA-ES algorithms with surrogate models and line searches. In Proceedings of the 15th annual conference companion on Genetic and

Zare, M. and Koch, M. 2014. An analysis of MLR and NLP for use in river flood routing and comparison with the Muskingum method. In 11th International Conference on Hydroscience & Engineering (ICHE).

Zhou, L. and Kang, L. 2023. A comparative analysis of multiple machine learning methods for flood routing in the Yangtze River. *Water*. 15(8): 1556

PM2. 5 and PM10 forecasting using linear and nonlinear modeling framework based on robust local mean decomposition and moving window ensemble strategy. *Applied Soft Computing Journal*. 114: 108110

Yuan, X., Zhang, X. and Tian, F. 2020. Research and application of an intelligent networking model for flood forecasting in the arid mountainous basins. *Journal of Flood Risk Management* .13: 12638.

A New Approach in Flood Routing Based on the Integration of Bayes Theory, Support Vector Machine and Meta-Heuristic Optimization Algorithm

Z. Khoramipoor¹, M. Valikhan Anaraki², S. Farzin^{3,*}

Received: Jan.08, 2024

Accepted: Jun.27, 2024

Abstract:

Flood control through rivers is one of the important measures to control and prevent flood damage. Flood hydrograph prediction is generally done using two hydrological methods (Manning and numerical) or hydraulic methods (solving Saint-Venant equations). Numerical methods are widely considered due to the fewer parameters, the low cost of calculations, and finally the appropriate accuracy. In this study, the use of artificial intelligence has been investigated for flow routing of the river. For this purpose, the combined algorithm, covariance matrix adaptation evolutionary strategy (CMAES), and Bayesian vector machine (RVM) have been used. Also, its results have been compared with the outputs of decision tree classification algorithms (CART), double gradient boosting algorithm XGBoost, simple Bayesian vector machine (RVM), and random forest (RF). Finally, to evaluate the performance and efficiency of the algorithms, the evaluation criteria of mean absolute error (MAE), root mean square error (RMSE), and coefficient of determination (R^2) have been used. Calculations of flood routing in the Mollasani area of Ahvaz have been performed for the study period 2011-2017. The obtained results indicate the appropriate success of the combined RVM-CMAES algorithm as a powerful tool in hydrograph forecasting for flood management, with evaluation criteria of MAE=19.69, RMSE=29.37, and $R^2=0.93$ in the training period, and MAE=19, RMSE=27.83, and $R^2=0.94$ in the test period. Furthermore, the Taylor diagrams and the routed hydrograph time series demonstrate that the combined algorithm is superior to other algorithms. The proposed algorithm has high potential for use in flood warning systems and other issues related to water resource management.

Keywords: Flood Forecasting, Flood routing, Hydrograph, Machine Learning Methods

1- Masters Student, Department of Water Engineering and Hydraulic Structures, Faculty of Civil Engineering, Semnan University, Semnan, Iran

2- Ph.D. Graduated, Department of Water Engineering and Hydraulic Structures, Faculty of Civil Engineering, Semnan University, Semnan, Iran

3- Associate Professor, Department of Water Engineering and Hydraulic Structures, Faculty of Civil Engineering, Semnan University, Semnan, Iran

(* Corresponding Author Email: saeed.farzin@semnan.ac.ir)