



مدیریت آب و آبیاری

دوره ۱۲ ■ شماره ۱ ■ بهار ۱۴۰۱

صفحه‌های ۳۱-۱۵

DOI: 10.22059/jwim.2022.335702.949

مقاله پژوهشی:

پیش‌بینی جریان ماهانه رودخانه با استفاده از رویکردهای آنتروپی شانون و موجک (مطالعه موردی: رودخانه مارون)

محمد امین نکوتیان^۱، فریدون رادمنش^۲، فرشاد احمدی^{۳*}

۱. دانشجوی کارشناسی ارشد، گروه هیدرولوژی و منابع آب، دانشکده مهندسی آب و محیط زیست، دانشگاه شهید چمران اهواز، اهواز، ایران.

۲. دانشیار، گروه هیدرولوژی و منابع آب، دانشکده مهندسی آب و محیط زیست، دانشگاه شهید چمران اهواز، اهواز، ایران.

۳. استادیار، گروه هیدرولوژی و منابع آب، دانشکده مهندسی آب و محیط زیست، دانشگاه شهید چمران اهواز، اهواز، ایران.

تاریخ دریافت مقاله: ۱۴۰۱/۹/۲۵

تاریخ پذیرش مقاله: ۱۴۰۱/۱/۷

چکیده

جریان رودخانه از مهم‌ترین اجزای چرخه هیدرولوژی است که به عوامل اقلیمی متعددی وابسته بوده و برآورد دقیق آن در زمینه‌های مختلف مدیریت منابع آب کاربرد دارد. در مطالعه حاضر از مدل‌های جنگل‌های تصادفی (RF) و ماشین بردار پشتیبان (SVM) برای پیش‌بینی جریان ماهانه رودخانه مارون در دوره آماری ۱۳۶۰ تا ۱۳۹۶ استفاده شد. یکی از مراحل مهم در کاربرد مدل‌های هوش مصنوعی تعریف الگوهای ورودی و شناسایی پارامترهای مؤثر در فرایند مدل‌سازی است. برای انتخاب بهترین ورودی‌ها از بین بارش، تبخیر و دماهای کمینه، بیشینه و متوسط روش آنتروپی شانون استفاده شد. نتایج نشان داد که وزن بارش و تبخیر در مجموع بیش از ۸۵ درصد بود. در گام بعد، سه ساختار متفاوت برای ورودی مدل‌ها توسعه داده شد. در حالت اول الگوهای اقلیم‌پایه تعریف شدند که از داده‌های هواشناسی به‌عنوان ورودی استفاده می‌کردند. در حالت دوم خاصیت تناوبی غیرخطی به الگوهای اقلیم‌پایه افزوده شد و در حالت سوم داده‌های ورودی اقلیم پایه با استفاده از پنج تابع موجک مادر تجزیه شده و مدل‌های هیبریدی W-RF و W-SVM ایجاد شدند. عملکرد مدل‌های منفرد RF و SVM نشان داد که با افزودن ترم پریودیک، دقت در مقایسه با ورودی‌های اقلیم پایه تا حدودی افزایش می‌یابد، اما تجزیه داده‌ها با تئوری موجک به‌طور قابل‌ملاحظه‌ای خطای مدل‌سازی را کاهش داد. در این بین عملکرد دو مدل W-RF و W-SVM بسیار نزدیک به یکدیگر بود، اما با توجه به نمودار ویلونی، مدل W-SVM به‌عنوان مناسب‌ترین گزینه برای پیش‌بینی جریان ماهانه رودخانه مارون پیشنهاد می‌شود.

کلیدواژه‌ها: الگوهای اقلیم پایه، ترم پریودیک، سطح تجزیه، وزن‌دهی.

Prediction of Monthly Streamflow Using Shannon Entropy and Wavelet Theory Approaches (Case study: Maroon River)

Mohammad Amin Nekoeeyan¹, Feridon Radmanesh², Farshad Ahmadi^{3*}

1. M.Sc. Student, Department of Hydrology and Water Resources, Faculty of Water and Environmental Engineering, Shahid Chamran

University of Ahvaz, Ahvaz, Iran.

2. Associate Professor, Department of Hydrology and Water Resources, Faculty of Water and Environmental Engineering, Shahid Chamran

University of Ahvaz, Ahvaz, Iran.

3. Assistant Professor, Department of Hydrology and Water Resources, Faculty of Water and Environmental Engineering, Shahid Chamran

University of Ahvaz, Ahvaz, Iran.

Received: December 16, 2021

Accepted: March 27, 2022

Abstract

River flow is one of the most important components of the hydrological cycle, which depends on several climatic factors and its accurate estimation is used in various fields of water resources management. Therefore, in the present study, random forest (RF) and support vector machine (SVM) models were used to predict the monthly streamflow of the Maroon River in the period of 1981- 2017. One of the important steps in the application of artificial intelligence models is the definition of input patterns and determining the effective variables in the modeling process. The Shannon entropy method was used to select the most efficient inputs among precipitation, evaporation, and minimum, maximum, and average temperatures. The results showed that the total weight of precipitation and evaporation was more than 85 percent. In the next step, three different structures were developed for modeling. In the first case, climate-based patterns were defined that used meteorological data as input. In the second case, nonlinear periodicity was added to the climate-based patterns, and in the third case, the climate-based input data were decomposed using five mother wavelet functions, and W-RF and W-SVM hybrid models were created. The performance evaluation of the standalone RF and SVM models showed that by considering the periodic term, the accuracy is somewhat increased compared to the climate-based inputs, but the analysis of the data with wavelet theory significantly reduced the modeling error. In the meantime, the performance of the two models W-RF and W-SVM was very close to each other, but according to the violin plot, the W-SVM model is suggested as the most suitable option for predicting the monthly streamflow of the Maroon River.

Keywords: Climate based patterns, Decomposition level, Periodic term, Weighting.

مقدمه

تغییرات در جریان رودخانه یک فرایند هیدرولوژیکی است که به طور مستقیم بر ویژگی‌های محیطی، اکولوژیکی، بیولوژیکی و اقلیمی یک منطقه و در نهایت زندگی انسان تأثیر می‌گذارد. تقاضا برای آب در بخش‌های مختلف هر سال در حال افزایش است و به سرعت از تغییرات اقلیمی و انسانی تأثیر می‌پذیرد. دستیابی به روش‌های مناسب پیش‌بینی دقیق جریان رودخانه‌ها در برنامه‌ریزی و مدیریت بهینه منابع آب و بهبود عملکرد سیستم‌های ذخیره نظیر سدها بسیار مهم و حیاتی است (Khodakhah *et al.*, 2021). در سال‌های اخیر روش‌های متعددی برای این منظور توسعه یافته است که می‌توان آن‌ها را در دو دسته کلی مدل‌های مفهومی و مدل‌های داده محور تقسیم نمود (Nayak *et al.*, 2004; Aghelpour *et al.*, 2021).

مدل‌های مفهومی نیز به داده‌های کامل و جامع نیاز داشته و لازم است از فرایندهای فیزیکی مؤثر در پدیده موردبررسی آگاهی کامل حاصل شود. علاوه بر این مهارت و دید هیدرولوژیکی کارشناس بهره‌بردار از این مدل‌ها نیز بسیار تأثیرگذار و عاملی مهم در عملکرد مناسب مدل‌های مفهومی می‌باشد. با توجه به موارد مذکور مشاهده می‌شود که کاربست این نوع از مدل‌ها عملاً دشوار بوده و به دلیل فراهم نبودن کلیه اطلاعات موردنیاز و استفاده از حالت‌های پیش‌فرض، خطا در شبیه‌سازی‌ها نیز قابل توجه است (Khodakhah *et al.*, 2021). در نقطه مقابل مدل‌های داده محور می‌توانند با استفاده از داده‌های عددی ثبت شده متغیرهای آب و هواشناسی و برقراری رابطه بین آن‌ها، شبیه‌سازی را با دقت قابل‌قبولی انجام داده و در این بین نیازی به بررسی و تحلیل فرایندهای فیزیکی حاکم بر پدیده موردنظر ندارند (Mehdizadeh *et al.*, 2020; Hamed *et al.*, 2021; Ahmadi *et al.*, 2021).

تاکنون مدل‌های داده‌محور متعددی هم‌چون شبکه

عصبی مصنوعی، سیستم استنتاج عصبی فازی تطبیقی، روش نزدیکترین همسایگی، برنامه‌ریزی بیان ژن، جنگل‌های تصادفی (RF)^۱ و ماشین بردار پشتیبان (SVM)^۲ برای حل مسائل مختلف هیدرولوژیکی معرفی شده‌اند.

روش ماشین بردار پشتیبان یکی از روش‌های یادگیری تحت نظارت است که هم برای دسته‌بندی و هم رگرسیون قابل استفاده است. این روش توسط Vapnik (1998) بر پایه تئوری یادگیری آماری بنا نهاده شده است. ماشین بردار پشتیبان روشی برای طبقه‌بندی دوتایی در فضای ویژگی‌های دلخواه است و از این روش مناسب برای مسائل پیش‌بینی به‌شمار می‌رود (Pai & Hong, 2007). Choubey *et al.* (2014) از مدل SVM برای پیش‌بینی و داده‌کاوی جریان ورودی به سد نارامدا^۳ واقع در ایالت مادهاپارادش^۴ هند استفاده کرده و نشان دادند که این روش از توانایی بسیار خوبی در مدل‌سازی و پیش‌بینی جریان میان‌مدت ماهانه برخوردار است. Baydaroglu *et al.* (2018) به‌منظور پیش‌بینی جریان ماهانه رودخانه قزل ایرماق ترکیه در دوره آماری ۱۹۴۰ تا ۲۰۱۲ از مدل‌های ماشین بردار پشتیبان، تحلیل طیفی و تئوری آشوب استفاده نمودند. نتایج نشان داد که روش SVM نسبت به مابقی روش‌های مورد استفاده دقت بیشتری در پیش‌بینی جریان داشته است. Christian *et al.* (2021) از روش ماشین بردار پشتیبان برای پیش‌بینی دبی ماهانه در حوضه رودخانه پمالی، اندونزی بهره بردند. ایشان بدین منظور از داده‌های تأخیر یافته جریان ماهانه به‌عنوان ورودی مدل استفاده نمودند. نتایج نشان داد که داده‌های ورودی با همبستگی خوب می‌توانند دقت پیش‌بینی مناسبی را ارائه دهند. هم‌چنین درگیر شدن سری داده‌های ورودی با همبستگی ضعیف ممکن است عملکرد مدل را کاهش دهد. به‌طورکلی نتایج این پژوهش نشان‌دهنده دقت بسیار مناسب مدل SVM در برآورد جریان ماهانه رودخانه بود.

یکی دیگر از مشخصه‌های مهمی که در سری‌های زمانی ثبت شده هیدرولوژیک وجود دارد و می‌تواند کمک شایانی در بهبود عملکرد مدل‌ها داشته باشد، خاصیت تناوبی است. *Mirabbasi et al.* (2019) و *Lohani et al.* (2012) با استفاده از رویکردهای متفاوت خاصیت تناوبی اقدام به مدل‌سازی بارش و جریان رودخانه در مقیاس ماهانه نموده و تأثیر مثبت آن را در بهبود عملکرد مدل‌ها گزارش نمودند. اما هیچ‌یک از دو روش مذکور نمی‌توانند شرایطی را فراهم نمایند تا تأثیر نوفه‌ها در فرایند مدل‌سازی مورد توجه قرار گیرد. بنابراین در مطالعه حاضر برای حل این مشکل از روش تئوری موجک استفاده می‌شود. تجزیه کردن سری‌های زمانی با استفاده از توابع موجک دو نوع زیرسری را به وجود می‌آورد؛ بردارهای تقریبی یا باقی‌مانده و بردارهای جزئیات (Roushangar & Ghasempour, 2021). تابع تبدیل موجک قابلیت تجزیه سری زمانی به چندین زیرسری زمانی با مقیاس‌های مختلف را دارد و با بررسی زیرسری‌های زمانی متوجه از سری زمانی کلی، رفتار کوچک‌مقیاس و بزرگ‌مقیاس یک فرایند هیدرولوژیک را آنالیز می‌کند. این امر کمک می‌کند تا اطلاعات به صورت مفیدتری در اختیار مدل‌های هوشمند قرار گرفته و در نتیجه عملکرد آنها بهبود می‌یابد. در مطالعات متعددی هم‌چون *Saraiva et al.* (2021)، *Wang et al.* (2021) و *Drisya et al.* (2021) توابع موجک برای تجزیه داده‌های جریان رودخانه و پیش‌بینی آنها با استفاده از مدل‌های هوش مصنوعی بهره برده شده است. در این مطالعات بهبود عملکرد مدل‌های تلفیقی موجک پایه توسعه یافته گزارش شده است.

با عنایت به موارد فوق می‌توان دریافت که تخمین متغیرهای هیدرولوژیک هم‌چون جریان رودخانه‌ها همواره مورد توجه پژوهش‌گران امر بوده و بدین منظور

روش جنگل تصادفی یکی از روش‌های هوش مصنوعی است که از الگوریتم‌های چندگانه تکراری برای انجام فرایند مدل‌سازی بهره می‌برد. این شیوه از پتانسیل لازم به عنوان یک ابزار برای ارزیابی مسائل هیدرولوژیک برخوردار بوده و می‌تواند الگوهای پیچیده را یاد گرفته و ارتباط غیرخطی بین متغیرهای مستقل و وابسته را در نظر بگیرد (Booker & Snelder, 2012). *Ali et al.* (2020) در مطالعه خود از روش‌های رگرسیون رنج و RF برای پیش‌بینی جریان ماهانه رودخانه بهره برده و گزارش نمودند که روش جنگل‌های تصادفی در پیش‌بینی جریان میان مدت با روش‌های پیش‌پردازش داده‌ها عملکرد بهتری دارد. در پژوهش‌های دیگری هم‌چون *Ghorbani et al.* (2020)، *Tang et al.* (2020) و *Pham et al.* (2020) از روش RF برای پیش‌بینی جریان رودخانه استفاده شده و کارایی آن مورد تأیید قرار گرفته است.

در کاربرد روش‌های هوش مصنوعی مانند RF و SVM همواره تعیین داده‌های ورودی بهینه نقش بسیار مهمی در عملکرد نهایی دارند. به عبارت دیگر، ارائه بیش‌ترین تعداد ورودی الزاماً به مفهوم دستیابی به بیش‌ترین دقت نخواهد بود. یکی از راه‌کارهایی که برای انتخاب ورودی‌های بهینه در سال‌های اخیر پیشنهاد شده است، استفاده از روش آنتروپی شانون می‌باشد (Ahmadi et al., 2021). این شاخص نشان می‌دهد که وقایع با احتمال وقوع زیاد اطلاعات کم‌تری در اختیار می‌گذارند و برعکس هرچه احتمال وقوع یک رخداد کم‌تر باشد، اطلاعات حاصل از آن بیش‌تر است (Saray et al., 2020). با به دست آوردن اطلاعات جدید، در واقع عدم قطعیت‌ها کاهش یافته و ارزش اطلاعات جدید برابر با مقداری است که از عدم قطعیت کاسته شده است. بنابراین می‌توان با محاسبه آنتروپی و وزن‌دهی به هر یک از ورودی‌ها، مؤثرترین آنها را انتخاب و وارد فرایند مدل‌سازی نمود.

دشت بهبهان رسیده و پس از مشروب نمودن آن وارد مسیر کوهستانی با دامنه‌های کم ارتفاع می‌شود. رودخانه مارون توسط سرشاخه‌های سقاوه، لوراب، شور و چاروساق از ارتفاعات زاگرس سرچشمه گرفته و با مساحتی حدود ۲۷۵۰ کیلومترمربع در ایستگاه هیدرومتری ایدنک اندازه‌گیری می‌شود.

در این مطالعه، برای پیش‌بینی جریان رودخانه مارون در دوره آماری ۱۳۶۰ تا ۱۳۹۶، در محل ایستگاه ایدنک، کلیه ایستگاه‌های موجود در منطقه بررسی و مشخص شد که در کل محدوده مورد مطالعه در مجموع ۲۳ ایستگاه باران‌سنجی و تبخیرسنجی وجود دارد. در این بین تعداد پنج ایستگاه به دلیل دارا بودن آمار مناسب برای محاسبه بارش‌های منطقه‌ای با استفاده از روش چندضلعی تیسن مورد بهره‌برداری قرار گرفتند. در داخل حوضه آبریز نیز دو ایستگاه تبخیرسنجی شامل دهنو و ایدنک واقع شده‌اند. ایستگاه دهنو از کیفیت مناسبی در ثبت داده‌های اقلیمی شامل دما و تبخیر برخوردار نبود. بنابراین در پژوهش حاضر داده‌های تبخیر، دمای کمینه، دمای بیشینه، و دمای متوسط فقط از ایستگاه تبخیرسنجی ایدنک استخراج شد. مشخصات ایستگاه‌های مورد استفاده در جدول (۱) ارائه شده و شکل (۱) موقعیت جغرافیایی منطقه مورد مطالعه نشان می‌دهد.

ماشین بردار پشتیبان

روش ماشین بردار پشتیبان یک الگوریتم کامپیوتری است که با استفاده از فرایند آموزش و مثال بهترین تابع را به‌عنوان طبقه‌بندی‌کننده یا ابرصفحه برای حل مسائل مختلف پیچیده و غیرخطی برازش می‌دهد (Ghorbani et al., 2016). روش SVM می‌تواند داده‌های خطی و غیرخطی مورد بررسی و تحلیل قرار دهد. مثالی از نحوه حل مسأله در روش SVM برای داده‌های خطی در شکل (۲) ارائه شده است (Ghorbani et al., 2016).

روش‌های متعددی از جمله مدل‌های تجربی - نیمه تجربی، سری‌های زمانی و مدل‌های هوشمند توسعه یافته‌اند. در این میان مدل‌های هوشمند قادر به تخمین متغیرهای هیدرولوژیک با دقت قابل قبول بوده و به اطلاعات و فرضیات کم‌تری نسبت به سایر روش‌ها نیاز دارند. از این‌رو، هدف از پژوهش حاضر پیش‌بینی جریان ماهانه حوضه آبریز مارون در محل ایستگاه هیدرومتری ایدنک با استفاده از مدل‌های RF و SVM بوده و بدین منظور مؤثرترین اطلاعات ورودی با توجه به داده‌های اقلیمی، دخالت دادن خاصیت تناوبی و پیش‌پردازش داده‌ها با استفاده از تئوری موجک مورد توجه قرار گرفته و رویکردهای مستقل و هیبریدی از نظر مدل‌سازی بررسی شدند.

مواد و روش‌ها

داده‌ها و منطقه مورد مطالعه

حوضه آبخیز سد مارون یکی از زیرحوضه‌های رودخانه مارون - جراحی است که در جنوب غرب ایران قرار دارد. طول رودخانه مارون از سرچشمه تا سد مارون حدود ۱۵۰ کیلومتر و از سد مارون تا ابتدای رودخانه جراحی حدود ۱۳۰ کیلومتر می‌باشد. سد مخزنی مارون دومین سد سنگریزه‌ای مرتفع ایران بوده و در فاصله ۱۹ کیلومتری شمال شرقی شهر بهبهان احداث شده است. رودخانه مارون از ارتفاعات زاگرس و از چشمه‌سارهای کوه‌های سادات و نیل (نیر) سرچشمه گرفته و حوضه آبریز آن در مناطق غربی بویراحمد سفلی و قسمت اعظم طبیعی و دشمن زیاری است که عمود بر جهت کلی سلسله جبال زاگرس به طرف جنوب غربی جریان می‌یابد. این رودخانه پس از طی مسیر پرشیب کوهستانی از ارتفاعات حدود ۳۱۰۰ متری عبور نموده و بعد از الحاق چندین رودخانه فرعی دیگر در ارتفاع حدود ۳۵۰ متری از سطح دریا به

پیش‌بینی جریان ماهانه رودخانه با استفاده از رویکردهای آنتروپی شانون و موجک (مطالعه موردی: رودخانه مارون)

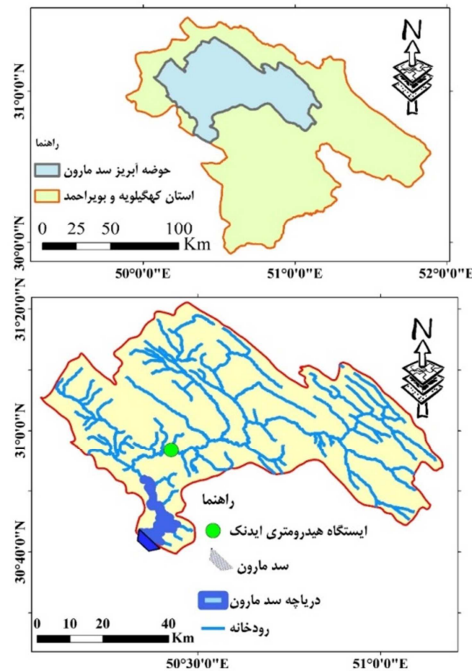


Figure 1. Location of the study area along Maroon river basin

Table 1. Spatial and statistical characteristics of selected stations (1981 to 2017)

Row	Variable	Station	Type of station	Coordinates (m)		Annual average
				Longitude	Latitude	
1	Streamflow	Idanak	Hydrometric	445324	3423876	48.39 (m ³ /s)
2	Rainfall	Pa ghale	Rain gauge	442670	3420521	599.20 (mm)
3		Dehno	Climatology	485677	3427580	939.30 (mm)
4		Dehdasht	Rain gauge	458551	3409181	499.50 (mm)
5		Tasoj	Rain gauge	507981	3394328	833.10 (mm)
6		Idanak	Climatology	444282	3424207	614.16 (mm)
7	Evaporation					2780.50 (mm)
8	Tmin	Idanak	Climatology	444282	3424207	14.16 (°C)
9	Tmax					32.03 (°C)
10	Tave					23.15 (°C)

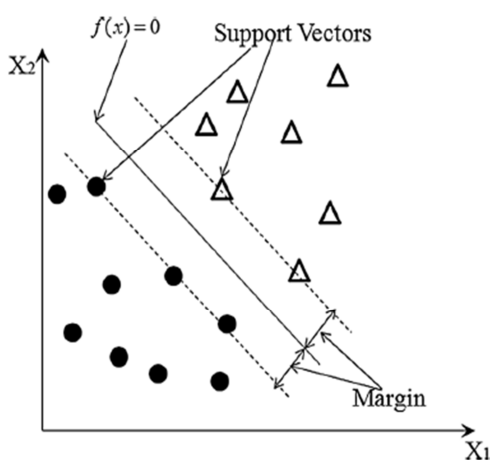


Figure 2. A schematic of linearly isolated data (Ghorbani et al., 2016)

بهترین ابرصفحه بین داده‌ها با استفاده از اندازه‌گیری حاشیه هر صفحه از نقاط حداکثر برازش داده می‌شود. حاشیه‌ها به‌عنوان فاصله بین ابرصفحه منتخب و نزدیک‌ترین حالت به هر کلاس داده در نظر گرفته شده و بردار پشتیبان نامیده می‌شوند (Vapnik, 1998).

فرض کنید داده‌های آموزشی با تعداد N نمونه و به‌صورت $(x_1, y_1), (x_2, y_2), \dots, (x_N, y_N)$ بوده و در آنها x ورودی‌ها و y خروجی وابسته به x باشد. در این صورت تابع برآوردکننده SVM به شرح زیر بیان می‌شود:

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

تابع کرنلی را انتخاب نمود که توانایی آموزش برای ورودی‌های مسأله را داشته باشد. در نهایت، تابع تصمیم رگرسیون بردار پشتیبان غیرخطی به صورت معادله زیر خواهد بود (Ghorbani et al., 2016):

$$f(x, a_i, a_i^*) = \sum_{i=1}^N (a_i + a_i^*) \cdot k(x, x_i) + b \quad (3)$$

در رابطه فوق α_i, α_i^* ضرایب تابع لاگرانژ و $k(x, x_i)$ تابع کرنل مورد استفاده می‌باشد. در عمل چهار نوع کرنل خطی، کرنل چندجمله‌ای، کرنل تانژانت هیپربولیک و کرنل شعاعی (RBF) به کار گرفته می‌شوند. در جدول (۲) معادلات کرنل‌های رایج ارائه شده‌اند.

Table 2. Common kernel functions in support vector machines (Hamel, 2011)

Kernel functions	Equation
Linear	$k(x_i, x_j) = x_i^T \cdot x_j$
Polynomial	$k(x_i, x_j) = (\gamma x_i^T \cdot x_j + C)^d$
Hyperbolic tangent	$k(x_i, x_j) = \tanh(\gamma x_i^T \cdot x_j + C)$
Radial Basis Function (RBF)	$k(x_i, x_j) = \exp(-\gamma x_i - x_j ^2)$

الگوریتم جنگل تصادفی

الگوریتم جنگل تصادفی توسط Breiman (2001) به عنوان یک روش یادگیری تجمعی برای مسائل مبتنی بر رگرسیون و خوشه‌بندی براساس توسعه درخت تصمیم، ارائه شده است. یک جنگل تصادفی، مجموعه‌ای از درخت‌های هرس نشده است که هر درخت با الگوریتم جزءبندی بازگشتی حاصل می‌شود. به عبارتی، جنگل تصادفی ترکیبی از چندین درخت تصمیم است که در ساخت آن چندین نمونه خودسازمانده از داده‌ها شرکت دارند (Friedman et al., 2001).

به منظور ایجاد درخت رگرسیونی از جزءبندی بازگشتی و رگرسیون‌های چندگانه استفاده می‌شود. فرایند تصمیم در هر گره داخلی از گره ریشه، طبق قانون درختی تکرار می‌شود تا زمانی که شرط توقف قبلی تعیین شده به دست آید (Breiman, 2001).

در رابطه فوق، w وزن بردار، b عرض از مبدا و ϕ تابع نگاشت غیرخطی است. مقادیر کوچک‌تر w نشان‌دهنده تابع تقریب ساده‌تر و با پیچیدگی کم‌تر است. تابع تخمین‌گر به صورتی طراحی می‌شود که کم‌ترین انحراف از مشاهدات واقعی در مجموعه داده‌های آموزش را داشته باشد. بدین منظور Vapnik (1998) یک روش بهینه‌سازی محدب را برای کمینه‌کردن انحرافات به صورت رابطه (۲) توسعه داد:

$$\begin{aligned} \text{Minimize: } & \frac{1}{2} w^2 + C \sum_{k=1}^N (\xi_k^+ + \xi_k^-) \quad (2) \\ \text{subject to: } & \begin{cases} y_k - (w \cdot \phi(x_k) + b) \leq \varepsilon + \xi_k^+ \\ w \cdot \phi(x_k) + b - y_k \leq \varepsilon + \xi_k^- \\ \xi_k^+, \xi_k^- \geq 0, k = 1, 2, 3, \dots, N \end{cases} \\ \text{Minimize: } & \frac{1}{2} w^2 + C \sum_{k=1}^N (\xi_k^+ + \xi_k^-) \end{aligned}$$

که در آن، C ثابتی است که توسط کاربر تعیین می‌شود و درجه اهمیت (سبک- سنگینی) بین همواری $f(x)$ و مقدار مجاز انحراف را نشان می‌دهد. هم‌چنین پارامترهای ξ_k^+, ξ_k^- متغیرهای انقطاع هستند که برای نشان دادن مقادیر پرت در داده‌های آموزش استفاده می‌شود. حداقل‌سازی ترم اول معادل با حداقل‌سازی فاصله اطمینان از ماشین یادگیری است و حداقل‌سازی ترم دوم مطابق با حداقل‌سازی ریسک تجربی می‌باشد. برای حل این مسأله از روش ضرایب لاگرانژ با در نظر گرفتن شرایط کاروش-کان-تاکر^۱ استفاده می‌شود (Vapnik, 1998).

در بیشتر مسائل جهان واقعی، تقریب تابع خطی استفاده‌های عملی محدودی دارد. Vapnik (1998) با استفاده از مفهوم حاصل ضرب داخلی نشان داد که ابتدا می‌توان بردار ورودی x را با یک تبدیل غیرخطی یا یک تابع کرنل به فضایی با بعد زیاد انتقال و محاسبات مورد نیاز را انجام داد. ماشین‌های بردار پشتیبان برای حل مسائل غیرخطی، ابعاد مسأله را از طریق توابع کرنل تغییر می‌دهند. انتخاب کرنل برای SVM به حجم داده‌های آموزشی و ابعاد بردار ویژگی بستگی دارد. به عبارت دیگر، باید با توجه به این پارامترها

سری‌های زمانی را در فضای زمان-فرکانس از هم جدا می‌کند. تبدیل موجک همانند تبدیل فوریه، سری زمانی را به صورت ترکیب خطی از چند تابع اصلی در نظر می‌گیرد. یکی از مشخصات مهم تبدیل موجک توانایی به دست آوردن اطلاعات در زمان، تناوب و موقعیت به طور همزمان است (Misiti et al., 1996). تبدیل موجک پیوسته توانایی عملکرد در هر مقیاس را داراست، هرچند که محاسبه ضرایب موجک در هر مقیاس بسیار مشکل است و نیازمند زمان محاسباتی بالا و تولید حجم زیادی داده می‌باشد. برای رفع این مشکل می‌توان از روش تبدیل موجک گسسته (DWT) استفاده نمود (Chen et al., 1999).

برای اجرای روش DWT الگوریتم مالات یا روش تحلیل چند تفکیکی (MAR)^۸ ارائه شد (Mallat, 2009). در این روش سیگنالی که تجزیه می‌شود از میان فیلترهای پایین‌گذر^۹ و بالا‌گذر^{۱۰} عبور داده می‌شود. در تحلیل گسسته موجک محتوای فرکانس پایین سیگنال تقریب^{۱۱} و محتوای فرکانس بالای سیگنال جزئیات^{۱۲} نامگذاری شده است. از این تکنیک فیلترکردن می‌توان برای به دست آوردن نمایش زمان-مقیاس یک سیگنال استفاده نمود (Polikar, 1999). در تبدیل گسسته موجک می‌توان با استفاده از برهم‌گذاری^{۱۳} ضرایب موجک، سیگنال اولیه را بازسازی نمود. این عمل از آخرین سطح تجزیه آغاز شده و با تجمیع سری‌های تقریب و جزئیات می‌توان سیگنال اولیه را بازسازی نمود (Polikar, 1996).

یکی از مراحل مهم در کاربرد توابع موجک انتخاب سطح تجزیه مناسب برای تحلیل سیگنال مورد نظر می‌باشد. بدین منظور در مطالعه حاضر از رابطه پیشنهادی Wang & Ding (2003) به شرح زیر استفاده شد:

$$L = \text{Int}[\text{Log}(N)] \quad (۸)$$

در رابطه فوق، L تعداد سطح تجزیه، N طول سری داده‌های اولیه و Int عملگر صحیح می‌باشد. با تجزیه هر

در روش RF، بردار تصادفی X_n که مستقل از بردارهای تصادفی X_1, X_2, \dots, X_{n-1} بوده، برای درخت n ام، تولید می‌شود. همچنین همه بردارها از توزیع مشابهی تبعیت می‌کنند. رگرسیون درختی با استفاده از مجموعه داده‌های آموزش و X_n محاسبه شده، مجموعه درخت‌هایی برابر با n را به شرح زیر تولید می‌نماید (Breiman, 2001):

$$X_n = \{h_1(x), h_2(x), \dots, h_n(x)\} \quad (۴)$$

$$h_n = h(x, X_n), x = \{x_1, x_2, \dots, x_p\} \quad (۵)$$

بردار p بعدی فوق یک جنگل را تشکیل داده و خروجی‌ها برای هر درخت به صورت زیر ارائه می‌شود:

$$\hat{y}_1 = h_1(x), \hat{y}_2 = h_2(x), \dots, \hat{y}_n = h_n(x) \quad (۶)$$

که در رابطه فوق، \hat{y}_n خروجی درخت n ام می‌باشد. برای به دست آوردن خروجی نهایی، متوسط همه پیش‌بینی‌های درخت‌ها محاسبه می‌شود (Breiman, 2001).

خطای پیش‌بینی نیز طبق رابطه (۷) محاسبه می‌شود.

$$MSE = \frac{\sum_{i=1}^n [y(x_i) - \hat{y}_i]^2}{n} \quad (۷)$$

در رابطه فوق، $\hat{y}(x_i)$ نتایج محاسباتی، y_i نتایج مشاهداتی و n تعداد کل مشاهدات است و MSE میزان خطای بین مقادیر مشاهداتی و محاسباتی را نشان می‌دهد. با اجرای مدل و تفسیر میانگین مربعات خطا در داده‌های آموزش و آزمون، جایی که با افزایش تعداد درختان تصمیم، تغییری در کاهش میانگین مربعات خطا مشاهده نشود، به عنوان تعداد درخت بهینه انتخاب می‌شود (Shataee et al., 2012).

تبدیل موجک

موجک دسته‌ای از توابع ریاضی است که برای تجزیه سیگنال پیوسته به مؤلفه‌های فرکانسی آن به کار می‌رود. این روش یک تحلیل طیفی مستقل از زمان است که به منظور توصیف مقیاس زمانی فرایندها و روابط آن‌ها،

تصمیم‌گیری زیر m گزینه و n شاخص (معیار) مدنظر می‌باشند (Shannon, 2001).

$$D = \begin{bmatrix} a_{11} & \dots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \dots & a_{mn} \end{bmatrix} \quad (11)$$

نتایج ماتریس بالا برای شاخص j (P_{ij}) به شرح زیر می‌باشد:

$$P_{ij} = \frac{a_{ij}}{\sum_{i=1}^m r_{ij}}; i \text{ and } j = 1, \dots, n; \forall ij \quad (12)$$

آن‌تروپی E_j به شرح زیر محاسبه می‌شود:

$$E_j = -K \sum_{i=1}^m P_{ij} \ln P_{ij}; \forall ij \quad (13)$$

و K به‌عنوان مقدار ثابت به‌صورت زیر محاسبه می‌شود:

$$K = \frac{1}{\ln m} \quad (14)$$

در ادامه مقدار d_j (درجه انحراف) محاسبه می‌شود که بیان می‌کند شاخص مربوطه (j) چه میزان اطلاعات مفید برای تصمیم‌گیری در اختیار تصمیم‌گیرنده قرار می‌دهد (Shannon, 2001).

$$d_j = 1 - E_j; \forall ij \quad (15)$$

در نهایت مقدار وزن W_j محاسبه می‌شود که در آن بزرگ‌ترین وزن نشان‌دهنده اهمیت پارامتر موردنظر می‌باشد (Shannon, 2001):

$$W_j = \frac{d_j}{\sum_{i=1}^n d_j} \quad (16)$$

توسعه مدل‌ها

برای مدل‌سازی پدیده‌های هیدرولوژیک، انتخاب مناسب‌ترین داده‌های ورودی می‌تواند تأثیر به‌سزایی در عملکرد روش‌های مورد استفاده داشته باشد. به‌طوری‌که در مرحله آموزش افزایش تعداد ورودی‌ها نه تنها سبب پیچیدگی الگو و افزایش حافظه درگیر خواهد شد، بلکه سبب کاهش دقت نیز می‌شود (Ahmadi et al., 2021). بنابراین در مطالعه حاضر سعی شد الگوهای ورودی با حالت‌های مختلف و براساس انتخاب از طریق تئوری آن‌تروپی شانون توسعه یابد. بدین منظور سه ساختار ورودی تعریف شد. در ساختار اول،

یک از سری داده‌ها، واریانس سری‌های زمانی اولیه به زیرسری‌های تقریب و جزئیات منتقل شده و هیچ‌گونه اطلاعاتی از سری زمانی اولیه از بین نخواهد رفت.

آن‌تروپی شانون

همان‌گونه که ذکر شد در این مطالعه به‌منظور انتخاب مناسب‌ترین متغیرهای اقلیمی در فرایند پیش‌بینی جریان رودخانه در محل ایستگاه ایدنک از آن‌تروپی شانون استفاده می‌شود. در مطالعات مختلف اغلب از روش آزمون و خطا یا خودهمبستگی در انتخاب ورودی‌ها مدنظر قرار گرفته و کم‌تر به اهمیت و وزن داده‌های در دسترس پرداخته می‌شود. مزیت روش حاضر نسبت به سایر روش‌های استخراج وزن، این است که یک روش کاملاً عینی بوده و نظرهای خبرگان در آن وجود ندارد؛ بنابراین اگر شرایط به گونه‌ای باشد که احتمال خطا در قضاوت خبرگان وجود داشته باشد، استفاده از این روش می‌تواند جایگزین خوب و قابل‌قبولی باشد (Bednarik et al., 2010).

مفهوم آن‌تروپی توسط Boltzmann در علم ترمودینامیک مطرح و به‌عنوان ابزاری برای بیان میزان عدم اطمینان به‌کار گرفته شد (Shannon, 2001). آن‌تروپی در نظریه اطلاعات به‌وسیله توزیع احتمال مشخص P_i بیان می‌شود. اندازه‌گیری این عدم اطمینان به‌وسیله شانون به‌صورت زیر قابل ارائه می‌باشد:

$$E_j = S(P_1, P_2, \dots, P_n) = -K \sum_{i=1}^m P_i \ln P_i \quad (9)$$

$$i = 1, \dots, m$$

در این رابطه، K یک مقدار ثابت است. زمانی که P_i مساوی با یکدیگر باشند (برای مقادیر j و i داده‌شده) در این صورت (Mukhopadhyay & Khan, 2015):

$$P_i = \frac{1}{n} \quad (10)$$

در یک ماتریس تصمیم‌گیری P_{ij} می‌تواند برای ارزیابی گزینه‌های مختلف به‌کار رود. در ماتریس

می‌شود که کم‌ترین (بیش‌ترین) مقدار RMSE و MAE (KGE) را به‌خود اختصاص دهد.

نتایج و بحث

آماده‌سازی الگوهای ورودی و نتایج تئوری آنتروپی شانون

با توجه به نکات ارائه‌شده، در مطالعه حاضر انتخاب ورودی‌های بهینه از بین متغیرهای اقلیمی شامل بارش، تبخیر، دمای کمینه، دمای بیشینه و دمای متوسط براساس وزن‌دهی با استفاده از روش آنتروپی شانون تعیین شد. در شکل (۳) نتایج حاصل از روش آنتروپی برای پارامترهای فوق ارائه شده است. با توجه به این شکل مشاهده می‌شود که متغیر بارش بیش‌ترین وزن را به‌خود اختصاص داده و در رتبه بعدی متغیر تبخیر قرار می‌گیرد. نکته مهم دیگر، تأثیر اندک متغیرهای دمایی می‌باشد به‌طوری‌که در مجموع کم‌تر از ۱۵ درصد را از نظر وزن و اهمیت به‌خود اختصاص داده‌اند. پس از تعیین پارامترهای مؤثر، خاصیت تناوبی غیرخطی به آن‌ها افزوده شده و الگوهای غیر خطی تناوبی ایجاد شدند. در حالت سوم نیز تجزیه داده‌ها با تئوری موجک انجام و الگوهای موجک پایه ایجاد شدند. بدین منظور باید سطح تجزیه مناسب انتخاب شود. همان‌گونه که ذکر شد، در مطالعه حاضر از رابطه Wang & Ding (2003) استفاده شده و بر این اساس سطح تجزیه مناسب برای ۴۳۲ داده موردبررسی برابر با عدد دو به‌دست آمد. در جدول (۳) ساختار الگوهای مختلف ورودی نیز ارائه شده است.

مدل‌سازی جریان ماهانه با استفاده از مدل‌ها منفرد

در این پژوهش برای مدل‌سازی جریان ماهانه ۷۵ درصد داده‌ها برای آموزش (از مهرماه ۱۳۶۰ تا شهریورماه ۱۳۸۶ به‌مدت ۳۱۲ ماه) و ۲۵ درصد برای صحت‌سنجی (از مهرماه ۱۳۸۶ تا شهریورماه ۱۳۹۶ به‌مدت ۱۲۰ ماه) در نظر گرفته شد.

داده‌های اقلیمی بارش، تبخیر، دمای کمینه، دمای بیشینه و دمای متوسط به‌عنوان اطلاعات در دسترس با استفاده از روش آنتروپی شانون وزن‌دهی شده و پس از انتخاب متغیرهای تأثیرگذار الگوهای اقلیم پایه (CB)^{۱۴} ایجاد شدند. در ساختار دوم، علاوه بر متغیرهای منتخب، از خاصیت فصلی غیرخطی بهره گرفته شد. بدین منظور از عبارت‌های $\cos[2\pi.i/12]$ و $\sin[2\pi.i/12]$ ($i = 1, \dots, 12$) برای بسط اثر خاصیت پریودیک غیرخطی استفاده شده و به الگوهای ورودی افزوده شد. در ساختار سوم، پیش‌پردازش داده‌ها به‌منظور حذف اثر نوفه‌ها و نیز انتقال بهتر اطلاعات به مدل‌های منفرد، از پنج تابع موجک مادر شامل تابع موجک هار، دابچیزه، سیملت، کویفلت و موجک فجرکروکین استفاده شد. به‌عبارت دیگر، در ابتدا الگوهای ورودی CB تجزیه شده و سپس به مدل‌ها معرفی شده و روش‌های تلفیقی W-SVM و W-RF توسعه داده شدند.

ارزیابی مدل‌ها

در این پژوهش، برای ارزیابی مدل‌های به‌کار گرفته شده در پیش‌بینی جریان از معیارهای جذر میانگین مربعات خطا (RMSE)، میانگین خطای مطلق (MAE) و کلینگ گوپتا^{۱۵} (KGE) استفاده می‌شود:

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (O_i - P_i)^2}{n}} \quad (17)$$

$$MAE = \left| \frac{O_i - P_i}{n} \right| \quad (18)$$

$$KGE = 1 - \sqrt{(CC - 1)^2 + (\alpha - 1)^2 + (\beta - 1)^2} \quad (19)$$

که در روابط فوق، O_i مقادیر مشاهداتی، P_i مقادیر پیش‌بینی‌شده، \bar{O} میانگین جریان مشاهداتی، CC ضریب همبستگی بین داده‌های مشاهداتی و محاسباتی، α نسبت انحراف معیار O_i و P_i ، β نسبت میانگین O_i و P_i و n تعداد داده‌ها می‌باشد. مدلی به‌عنوان مناسب‌ترین گزینه انتخاب

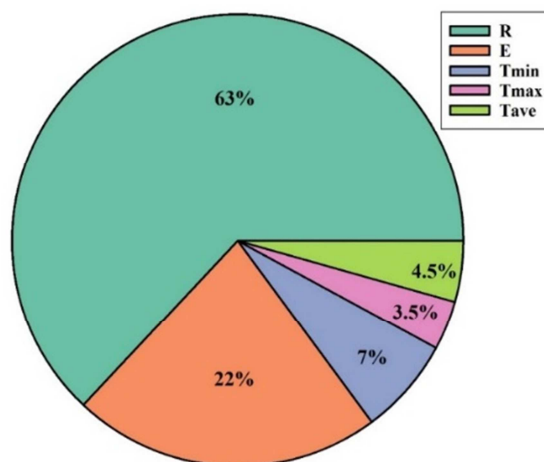


Figure 3. Shannon entropy results in determining the weight of climatic variables

Table 3. Input and output variables used to develop the climatic based and wavelet based models

Row	Inputs structure	Pattern	Inputs	Output
1	Climatic based patterns (CB)	CB1	<i>R</i>	Q_i
2		CB2	<i>E</i>	Q_i
3		CB3	<i>Tave</i>	Q_i
4		CB4	<i>Tmax</i>	Q_i
5		CB5	<i>Tmin</i>	Q_i
6		CB6	<i>R, E</i>	Q_i
7	Periodic climatic based patterns (PCB)	PCB1	<i>R, PER</i>	Q_i
8		PCB2	<i>E, PER</i>	Q_i
9		PCB3	<i>Tave, PER</i>	Q_i
10		PCB4	<i>Tmax, PER</i>	Q_i
11		PCB5	<i>Tmin, PER</i>	Q_i
12		PCB6	<i>R, E, PER</i>	Q_i
13	Wavelet based patterns in decomposition levels one (W1) and tow (W2)	W(1)CB1	<i>decomposed inputs level 1</i>	Q_i
14		W(2)CB1	<i>decomposed inputs level 2</i>	Q_i
15		W(1)CB2	<i>decomposed inputs level 1</i>	Q_i
16		W(2)CB2	<i>decomposed inputs level 2</i>	Q_i
17		W(1)CB3	<i>decomposed inputs level 1</i>	Q_i
18		W(2)CB3	<i>decomposed inputs level 2</i>	Q_i
19		W(1)CB4	<i>decomposed inputs level 1</i>	Q_i
20		W(2)CB4	<i>decomposed inputs level 2</i>	Q_i
21		W(1)CB5	<i>decomposed inputs level 1</i>	Q_i
22		W(2)CB5	<i>decomposed inputs level 2</i>	Q_i
23		W(1)CB6	<i>decomposed inputs level 1</i>	Q_i
24		W(2)CB6	<i>decomposed inputs level 2</i>	Q_i

دیگر نزدیک بودن نتایج ارزیابی الگوهای CB6 و CB1 در مدل سازی جریان ماهانه می باشد. با این وجود مدل های RF و SVM هرکدام رفتا متفاوتی را با الگوی ورودی یکسان داشته اند، به طوری که مدل SVM با الگوی CB1 به بیشترین دقت دست یافته و افزودن متغیر تبخیر موجب کاهش دقت آن شده است، اما در مدل RF وضعیت کاملاً برعکس بوده و بهترین دقت با الگوی CB6 به دست آمده

نتایج حاصل از اجرای مدل های مستقل برای الگوهای اقلیم پایه پریودیک و غیر پریودیک در جدول (۴) ارائه شده است. با توجه به این جدول مشاهده می شود که استفاده از متغیر بارش به عنوان ورودی، تأثیر بسزایی در افزایش دقت مدل های RF و SVM داشته است. این امر نشان می دهد که آنتروپی شانون توانسته است در تعیین متغیر مؤثر عملکرد بسیار خوبی را ارائه نماید. نکته مهم

را کاهش داده است. (Montaseri & Ghavidel (2014). عملکرد مدل‌ها را با افزودن ترم پریودیک گزارش نموده‌اند، اما در هیچ یک از مطالعات مذکور از روش RF برای مدل‌سازی استفاده نشده بود.

نکته مهم دیگری که می‌توان از نتایج جدول (۴) استنباط نمود، نزدیک‌شدن عملکرد مدل‌های RF و SVM با الگوهای پریودیک PCB1 و PCB6 می‌باشد. همان‌گونه که مشاهده می‌شود با در نظر گرفتن خاصیت تناوبی، اختلاف موجود در بین الگوهای ورودی CB1 و CB6 برای مدل‌های منفرد از بین رفته است. این امر می‌تواند در کاهش تعداد ورودی‌ها بسیار مؤثر باشد، چراکه افزودن یک متغیر جدید به الگوهای ورودی هزینه‌های زیادی داشته و دسترسی به داده‌ها در بسیاری از موارد ممکن نیست، اما اضافه‌کردن یک عبارت ساده ریاضی به الگوها بدون تحمیل هرگونه هزینه مازاد بوده و مشاهده می‌شود که تقریباً نتایج یکسانی را نیز به دست داده است.

است. (Ahmadi et al. (2021) و Saray et al. (2020) نیز عملکرد مناسب آنتروپی شانون را در انتخاب متغیرهای مؤثر گزارش نموده‌اند که با نتایج حاصل از پژوهش حاضر مطابقت دارد. در جدول (۴) نتایج حاصل از تأثیر دادن خاصیت تناوبی در عملکرد مدل‌های RF و SVM نیز ارائه شده است. در این جدول می‌توان مشاهده نمود که با در نظر گرفتن ترم تناوبی غیرخطی دقت در فرایند مدل‌سازی با الگوهای مختلف ورودی تا حدودی بهبود یافته، اما تأثیر یکسانی در مدل‌های منفرد مورد بررسی نداشته است. به عبارت دیگر، در مدل SVM کاربرد ترم تناوبی در کلیت فرایند مدل‌سازی مثبت بوده و توانسته مقدار شاخص RMSE را در بهترین حالت از ۳۰/۴۸۵ مترمکعب بر ثانیه کاهش دهد، اما برای مدل RF چنین چیزی مشاهده نمی‌شود و کاربرد خاصیت تناوبی موجب افزایش اندک خطا در بهترین حالت ورودی شده (الگوی CB6 در مقایسه با الگوی PCB6) اما در سایر الگوها به‌طور چشم‌گیری خطا

Table 4. Values of RMSE (m^3/s), MAE (m^3/s), and KGW statistics obtained from the periodic and non-periodic standalone models

Model	Pattern	Train			Test		
		RMSE	MAE	KGW	RMSE	MAE	KE
SVM	CB1	43.171	30.085	0.492	30.485	23.507	0.509
	CB2	54.490	38.160	0.191	34.971	28.138	0.315
	CB3	51.419	34.201	0.280	35.940	27.091	0.395
	CB4	52.952	36.485	0.236	32.058	23.880	0.413
	CB5	50.928	35.050	0.293	38.150	27.306	0.226
	CB6	39.130	27.669	0.583	34.858	25.042	0.426
	PCB1	30.467	18.803	0.747	28.941	19.593	0.545
	PCB2	42.156	25.415	0.516	37.820	26.736	0.412
	PCB3	44.478	26.378	0.461	35.242	25.985	0.450
	PCB4	41.930	24.424	0.521	31.706	19.019	0.508
	PCB5	42.515	25.820	0.508	37.775	26.600	0.351
	PCB6	33.053	20.410	0.702	28.514	19.570	0.553
RF	CB1	20.860	15.540	0.829	36.019	25.390	0.412
	CB2	27.431	17.612	0.598	53.673	36.146	0.145
	CB3	33.396	19.243	0.607	47.900	32.807	0.253
	CB4	35.709	21.391	0.580	48.668	33.084	0.110
	CB5	38.389	24.411	0.545	45.319	30.329	0.116
	CB6	17.548	12.025	0.827	34.882	24.010	0.421
	PCB1	15.333	10.071	0.875	35.206	22.713	0.452
	PCB2	20.952	11.783	0.807	41.891	28.228	0.357
	PCB3	26.660	14.484	0.742	46.730	29.350	0.279
	PCB4	22.726	12.208	0.807	32.470	20.335	0.497
	PCB5	73.366	51.967	0.025	35.847	23.928	0.393
	PCB6	13.841	8.397	0.881	35.209	24.664	0.447

مدل سازی جریان ماهانه با استفاده از مدل های تلفیقی

در این پژوهش، به منظور توسعه مدل های W-SVM و W-RF از پنج تابع موجک مادر شامل هار، دابچیز، سیملت، کویفلت و فجر-کروکین^{۱۶} (FK) استفاده شد. تعیین سطح تجزیه مناسب یکی از مراحل مهم در کاربرد توابع موجک است که در مطالعه حاضر با توجه به تعداد داده ها (۴۱۲ ماه) و رابطه (۸) برابر با عدد دو به دست آمد.

در جدول های (۵) و (۶) نتایج حاصل از روش های تلفیقی W-SVM و W-RF در مرحله آزمون ارائه شده است. با توجه این جدول می توان مشاهده نمود که استفاده از توابع موجک و تجزیه داده ها تأثیر بسیار مثبتی در بهبود عملکرد مدل ها داشته است. به عنوان مثال، در مدل SVM با الگوی CB1 مقدار RMSE از ۳۰/۴۸۵ مترمکعب در ثانیه بوده که به ۲۴/۹۰۶ مترمکعب در ثانیه کاهش یافته است.

با بررسی جدول (۵) می توان نتیجه گرفت بهترین عملکرد مدل W-SVM با الگوی W(db4)(2)CB6 به دست آمده که از داده های تجزیه شده بارش و تبخیر در سطح دو با استفاده از تابع موجک دابچیز چهار حاصل شده است. در همه اجراهای انجام شده (به جز W(coif)(1)CB4) سطح دوم تجزیه بهترین عملکرد را به خود اختصاص داده است. نکته قابل توجه، تأثیر انتخاب نوع تابع موجک مورد استفاده در عملکرد مدل SVM می باشد که یکی نقاط ضعف کاربرد تئوری موجک به شمار آمده و برای دسترسی به بهترین انتخاب باید سعی نمود بیشترین تعداد تابع موجک مادر را به کار گرفت که متعاقباً زمان مدل سازی را افزایش می دهد.

نتایج مدل W-RF در جدول (۶) نیز مانند روش W-SVM بوده و بهبود دقت مدل RF با ورودی های تجزیه شده کاملاً مشهود می باشد. در این روش نیز بهترین عملکرد با تابع موجک سیملت در سطح دو و با ورودی های بارش و تبخیر (W(sym)(2)CB6)

حاصل شده است. با مقایسه آماره های RMSE و MAE مدل W-RF اندکی بهتر از W-SVM بوده، اما آماره KGE که اصلاح شده شاخص نش- ساتکلیف بوده و نشان دهنده کارایی مدل می باشد، روش W-SVM را به عنوان گزینه برتر پیشنهاد می دهد. در این شرایط یکی از راه های مناسب ارزیابی، استفاده از روش های گرافیکی می باشد. به منظور مقایسه هرچه بهتر عملکرد مدل های منفرد (با بهترین ورودی ها) و مدل های هیبریدی، از نمودارهای تلفیقی گرافیکی و ویولونی استفاده شد.

در شکل (۴) نمودارهای گرافیکی و ویولونی مدل های پرودییک، غیرپرودییک و هیبریدی موجک پایه برای بهترین عملکرد ارائه شده است. همان گونه که ذکر شد نتایج مدل های هیبریدی موجک پایه در مدل سازی جریان ماهانه، براساس معیارهای ارزیابی بسیار بهبود یافته و این امر در معیارهای ارزیابی عددی ارائه شده انعکاس یافته است. با این وجود روش های ارزیابی مذکور هیچ گونه اطلاعاتی درخصوص نحوه توزیع داده ها در اختیار تحلیل گران قرار نمی دهند. به عبارت دیگر، ممکن است مدلی فقط در برآورد دامنه مشخصی از داده ها که بیشترین فراوانی را دارند، عملکرد بهتری داشته باشد و نتواند مقادیر حداکثر یا حداقل را به خوبی برآورد نماید. برای بررسی هرچه بهتر این موضوع می باید از روش هایی استفاده شود که توانایی ارزیابی توزیع داده های برآورد شده را با توزیع داده های مشاهداتی داشته باشند. بدین منظور استفاده از نمودار ویولونی می تواند بسیار سودمند باشد. نمودار ویولونی حالت دیگری از نمودار جعبه ای می باشد. در نمودار جعبه ای صرفاً کمینه، بیشینه، میانگین و چارک های داده ها ارائه می شود، اما از نمودار ویولونی برای بصری سازی توزیع داده ها و تراکم احتمالی آن استفاده می شود (Ahmadi et al., 2021).

پیش‌بینی جریان ماهانه رودخانه با استفاده از رویکردهای آنتروپی شانون و موجک (مطالعه موردی: رودخانه مارون)

Table 5. Values of RMSE (m^3/s), MAE (m^3/s), and KGE statistics obtained from the W-SVM hybrid model

Row	Pattern*	RMSE	MAE	KGE	Row	Pattern	RMSE	MAE	KGE
1	W(haar)(1)CB1	32.668	24.720	0.466	31	W(haar)(1)CB4	32.056	22.756	0.394
2	W(haar)(2)CB1	28.291	20.209	0.519	32	W(haar)(2)CB4	33.289	26.430	0.148
3	W(db4)(1)CB1	28.475	22.634	0.521	33	W(db4)(1)CB4	31.246	23.247	0.438
4	W(db4)(2)CB1	28.286	22.306	0.523	34	W(db4)(2)CB4	32.809	24.786	0.463
5	W(sym)(1)CB1	28.485	22.669	0.528	35	W(sym)(1)CB4	45.734	27.843	0.239
6	W(sym)(2)CB1	24.906	20.154	0.545	36	W(sym)(2)CB4	35.803	24.101	0.414
7	W(coif)(1)CB1	29.863	22.678	0.513	37	W(coif)(1)CB4	30.335	23.049	0.441
8	W(coif)(2)CB1	28.331	22.066	0.527	38	W(coif)(2)CB4	33.910	26.225	0.437
9	W(FK)(1)CB1	32.008	24.394	0.481	39	W(FK)(1)CB4	34.166	23.951	0.400
10	W(FK)(2)CB1	33.659	25.087	0.449	40	W(FK)(2)CB4	32.737	22.678	0.376
11	W(haar)(1)CB2	37.977	31.163	0.324	41	W(haar)(1)CB5	51.315	39.815	0.123
12	W(haar)(2)CB2	37.352	30.728	0.321	42	W(haar)(2)CB5	74.287	56.747	-0.129
13	W(db4)(1)CB2	35.565	26.963	0.393	43	W(db4)(1)CB5	39.345	30.224	0.232
14	W(db4)(2)CB2	34.775	26.963	0.461	44	W(db4)(2)CB5	37.453	28.649	0.281
15	W(sym)(1)CB2	38.555	27.781	0.367	45	W(sym)(1)CB5	40.096	29.218	0.211
16	W(sym)(2)CB2	33.799	24.885	0.470	46	W(sym)(2)CB5	42.619	30.309	0.241
17	W(coif)(1)CB2	38.145	28.806	0.394	47	W(coif)(1)CB5	42.224	34.653	0.278
18	W(coif)(2)CB2	34.732	27.487	0.460	48	W(coif)(2)CB5	38.049	29.947	0.316
19	W(FK)(1)CB2	35.784	29.414	0.318	49	W(FK)(1)CB5	39.371	30.756	0.315
20	W(FK)(2)CB2	33.932	25.815	0.459	50	W(FK)(2)CB5	37.244	28.966	0.348
21	W(haar)(1)CB3	34.860	25.579	0.417	51	W(haar)(1)CB6	33.595	25.931	0.435
22	W(haar)(2)CB3	33.596	25.297	0.446	52	W(haar)(2)CB6	30.163	22.688	0.492
23	W(db4)(1)CB3	34.730	26.320	0.375	53	W(db4)(1)CB6	33.673	24.153	0.439
24	W(db4)(2)CB3	32.421	25.195	0.459	54	W(db4)(2)CB6	22.183	16.903	0.667
25	W(sym)(1)CB3	32.557	25.693	0.408	55	W(sym)(1)CB6	37.095	24.919	0.415
26	W(sym)(2)CB3	33.113	25.035	0.481	56	W(sym)(2)CB6	22.444	16.705	0.419
27	W(coif)(1)CB3	35.217	26.938	0.375	57	W(coif)(1)CB6	34.580	24.532	0.440
28	W(coif)(2)CB3	33.783	26.388	0.438	58	W(coif)(2)CB6	24.363	18.563	0.516
29	W(FK)(1)CB3	34.358	26.038	0.437	59	W(FK)(1)CB6	33.799	25.149	0.436
30	W(FK)(2)CB3	36.153	26.106	0.439	60	W(FK)(2)CB6	30.953	22.761	0.514

* در این جدول منظور از W(haar)(1)CB1 داده‌های تجزیه‌شده ورودی‌های الگوی CB1 با سطح تجزیه یک از طریق موجک مادر هر می‌باشد.

Table 6. Values of RMSE (m^3/s), MAE (m^3/s), and KGE statistics obtained from the W-RF hybrid model

Row	Pattern*	RMSE	MAE	KGE	Row	Pattern	RMSE	MAE	KGE
1	W(haar)(1)CB1	36.885	26.742	0.353	31	W(haar)(1)CB4	32.896	23.787	0.333
2	W(haar)(2)CB1	35.077	24.080	0.408	32	W(haar)(2)CB4	27.495	18.804	0.354
3	W(db4)(1)CB1	33.791	25.142	0.436	33	W(db4)(1)CB4	36.772	24.429	0.327
4	W(db4)(2)CB1	36.593	26.479	0.413	34	W(db4)(2)CB4	38.152	25.970	0.380
5	W(sym)(1)CB1	35.959	23.053	0.418	35	W(sym)(1)CB4	43.585	29.030	0.251
6	W(sym)(2)CB1	25.821	19.038	0.453	36	W(sym)(2)CB4	40.268	25.038	0.323
7	W(coif)(1)CB1	35.115	25.575	0.393	37	W(coif)(1)CB4	41.297	28.929	0.285
8	W(coif)(2)CB1	34.087	24.224	0.449	38	W(coif)(2)CB4	37.930	26.800	0.367
9	W(FK)(1)CB1	41.440	27.492	0.309	39	W(FK)(1)CB4	34.060	22.703	0.379
10	W(FK)(2)CB1	36.273	25.611	0.379	40	W(FK)(2)CB4	30.784	20.913	0.454
11	W(haar)(1)CB2	38.227	29.962	0.391	41	W(haar)(1)CB5	69.025	48.640	0.020
12	W(haar)(2)CB2	41.260	32.900	0.354	42	W(haar)(2)CB5	37.308	28.884	0.412
13	W(db4)(1)CB2	42.403	30.348	0.264	43	W(db4)(1)CB5	46.205	31.474	0.192
14	W(db4)(2)CB2	40.274	29.390	0.374	44	W(db4)(2)CB5	48.171	34.873	0.149
15	W(sym)(1)CB2	48.433	34.721	0.241	45	W(sym)(1)CB5	44.610	32.004	0.101
16	W(sym)(2)CB2	40.489	29.196	0.362	46	W(sym)(2)CB5	48.957	36.350	0.177
17	W(coif)(1)CB2	36.730	26.087	0.403	47	W(coif)(1)CB5	44.892	30.148	0.171
18	W(coif)(2)CB2	39.269	29.333	0.392	48	W(coif)(2)CB5	45.919	32.402	0.196
19	W(FK)(1)CB2	41.492	30.128	0.327	49	W(FK)(1)CB5	52.064	38.308	0.156
20	W(FK)(2)CB2	34.979	26.632	0.427	50	W(FK)(2)CB5	49.339	41.915	0.221
21	W(haar)(1)CB3	35.478	25.552	0.425	51	W(haar)(1)CB6	35.437	26.394	0.379
22	W(haar)(2)CB3	39.204	26.594	0.311	52	W(haar)(2)CB6	31.307	23.146	0.484
23	W(db4)(1)CB3	40.873	28.843	0.316	53	W(db4)(1)CB6	33.330	23.591	0.432
24	W(db4)(2)CB3	40.783	28.718	0.351	54	W(db4)(2)CB6	25.277	19.218	0.557
25	W(sym)(1)CB3	38.440	26.788	0.359	55	W(sym)(1)CB6	30.215	22.579	0.530
26	W(sym)(2)CB3	41.756	29.479	0.359	56	W(sym)(2)CB6	21.718	15.895	0.608
27	W(coif)(1)CB3	44.951	33.062	0.237	57	W(coif)(1)CB6	32.280	22.887	0.476
28	W(coif)(2)CB3	39.321	27.260	0.387	58	W(coif)(2)CB6	25.093	19.801	0.582
29	W(FK)(1)CB3	40.753	28.746	0.296	59	W(FK)(1)CB6	36.577	26.327	0.374
30	W(FK)(2)CB3	37.966	28.499	0.414	60	W(FK)(2)CB6	29.100	20.930	0.524

* در این جدول منظور از W(haar)(1)CB1 داده‌های تجزیه‌شده ورودی‌های الگوی CB1 با سطح تجزیه یک از طریق موجک مادر هر می‌باشد.

SVM(PCB6) کم تر رخ داده اما مقادیر متوسط و کمینه در مدل های مذکور دچار بیش برآورد شده است. اما کاربرد مدل های هیبریدی توانسته تا حد بسیار زیادی مشکل بیش برآورد داده های کمینه و متوسط را برطرف نموده و در نتیجه مطابق با معیارهای ارزیابی خطای پیش بینی را کاهش دهد، اما همچنان برآورد مقادیر بیشینه دچار اشکال شده و کم تخمینی رخ داده است. به طور کلی با بررسی معیارهای ارزیابی عددی و نموداری می توان نتیجه گرفت که هیبرید روش موجک و مدل SVM بهترین عملکرد را در برآورد جریان ماهانه در حوضه مارون داشته است.

با توجه به شکل (۴) می توان نتیجه گرفت که مدل های منفرد با ورودی های تناوبی و غیرتناوبی جریان ماهانه را بزرگ تر از مقادیر مشاهداتی برآورد نموده و در نتیجه میانگین داده ها در مقایسه با مقادیر مشاهداتی افزایش یافته است. اما با استفاده از تبدیل موجکی این امر تا حد زیادی برطرف شده و میانگین مقادیر مشاهداتی و محاسباتی به یکدیگر بسیار نزدیک تر شده اند. با ارزیابی همزمان نمودار ویلونی و گرافیکی مشاهده می شود که مدل های RF(PCB6) و RF(CB6) مقادیر حدی را بسیار بیش تر از حالت واقعی و در زمان های نامناسب برآورد نموده اند. این اتفاق در مدل های SVM(CB1) و

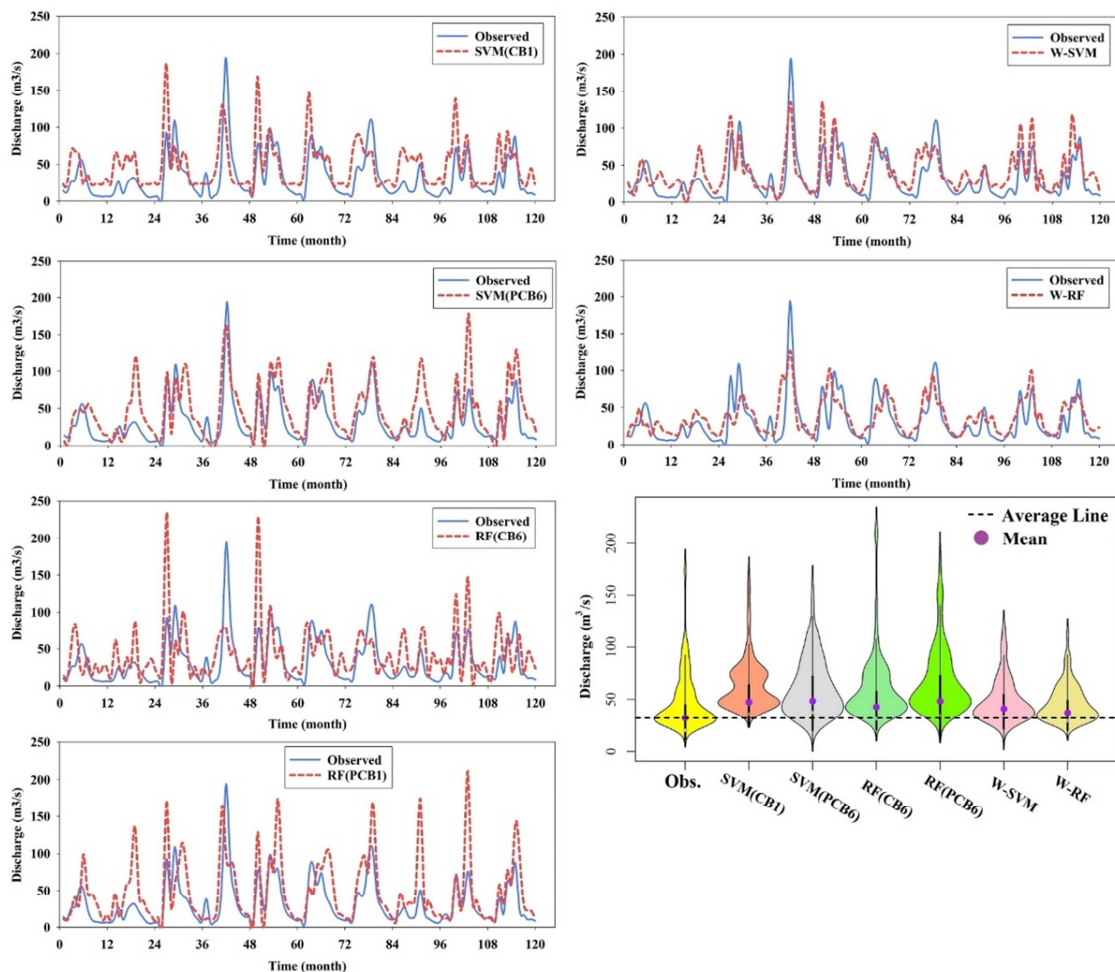


Figure 4. Graphical and violin plots of standalone and hybrid models with optimal input patterns

نتیجه‌گیری

مدل‌سازی نمایند. با توجه به نتایج به‌دست‌آمده مشاهده می‌شود که استفاده از داده‌های تجزیه‌شده می‌تواند تأثیر مثبتی در بهبود عملکرد مدل‌های هوشمند داشته باشد، از این‌رو پیشنهاد می‌شود که سایر روش‌های تجزیه سیگنال از جمله روش تجزیه مد یکپارچه کامل نیز برای توسعه مدل‌های تلفیقی موردبررسی قرار گیرد.

در این مطالعه، از مدل‌های RF و SVM و تلفیق آن‌ها با تئوری موجک برای پیش‌بینی جریان ماهانه رودخانه مارون در ایستگاه ایدنک استفاده‌شده و دقت آن‌ها موردارزیابی قرار گرفت. بدین منظور الگوهای ورودی براساس اطلاعات اقلیمی توسعه داده شده و مؤثرترین متغیرها با استفاده از تئوری آنتروپی شانون انتخاب و تأثیر خاصیت تناوبی نیز موردارزیابی قرار گرفت.

پی‌نوشت‌ها

1. Random Forests (RF)
2. Support Vector Machine (SVM)
3. Narmada
4. Madhya Pradesh
5. Noise
6. Karush-Khuan-Tucker
7. Discrete Wavelet Transform
8. Multi Resolution Analysis
9. Low-Pass Filters
10. High Pass Filters
11. Approximation
12. Detail
13. Synthesizing
14. Climatic Based
15. Kling-Gupta Efficiency
16. Fejer-Korovkin

نتایج حاصل از پژوهش حاضر نشان داد که مدل‌های منفرد RF و SVM در برآورد جریان ماهانه رودخانه با ورودی‌های پریرودیک عملکرد بهتری نسبت به حالت غیرتناوبی داشته‌اند. به عبارت دیگر، لحاظ کردن خاصیت تناوبی اندکی خطاهای پیش‌بینی را کاهش داده است. ترم پریرودیک صرفاً براساس یک رابطه ریاضی ساده محاسبه شده و هیچ‌گونه زمان و هزینه مازادی را از لحاظ جمع‌آوری داده‌ها به پژوهش‌گران تحمیل نمی‌کند. بنابراین پیشنهاد می‌شود که در مطالعات آتی نیز تأثیر کاربرد ترم پریرودیک با مدل‌ها و متغیرهای مختلف در بهبود نتایج مدل‌سازی مورد ارزیابی قرار گیرد. براساس نتایج به‌دست‌آمده مدل SVM در مقایسه با RF عملکرد بهتری داشته و به‌عنوان مدل منفرد مناسب انتخاب می‌شود.

تعارض منافع

هیچ‌گونه تعارض منافع توسط نویسندگان وجود ندارد.

منابع

1. Aghelpour, P., Bahrami-Pichaghchi, H., & Varshavian, V. (2021). Hydrological drought forecasting using multi-scalar streamflow drought index, stochastic models and machine learning approaches, in northern Iran. *Stochastic Environmental Research and Risk Assessment*, 1-21.
2. Ahmadi, F. (2021). Evaluation of the Efficiency of Data Preprocessing Methods on Improving the Performance of Gene Expression Programming Model (Case Study: Ab Zal River). *Water and Soil*, 35(2), 153-165. (In Persian).
3. Ahmadi, F., Mehdizadeh, S., Mohammadi, B., Pham, Q. B., Doan, T. N. C., & Vo, N. D. (2021). Application of an artificial intelligence technique enhanced with intelligent water drops for monthly reference evapotranspiration estimation. *Agricultural Water Management*, 244, 106622.

در این مطالعه علاوه بر مدل‌های منفرد، از یک روش ریاضی پیچیده مبتنی بر تجزیه سیگنال تحت عنوان تئوری موجک برای توسعه مدل‌های هیبریدی W-SVM و W-RF استفاده شد. نتایج نشان داد که استفاده از این روش دقت برآوردها را به مراتب بهبود بخشیده و شاخص RMSE با بهترین الگوی ورودی برای مدل‌های SVM و W-SVM به ترتیب از ۲۸/۵۱۴ مترمکعب بر ثانیه به ۲۱/۷۱۸ مترمکعب بر ثانیه کاهش یافته است. این موضوع برای مدل‌های RF و W-RF نیز صادق بوده و باتوجه به نمودار ویلونی روش‌های هیبریدی می‌توانند میانگین و واریانس داده‌ها را به‌خوبی

4. Ali, M., Prasad, R., Xiang, Y., & Yaseen, Z. M. (2020). Complete ensemble empirical mode decomposition hybridized with random forest and kernel ridge regression model for monthly rainfall forecasts. *Journal of Hydrology*, 584, 124647.
5. Baydaroglu, Ö., Koçak, K., & Duran, K. (2018). River flow prediction using hybrid models of support vector regression with the wavelet transform, singular spectrum analysis and chaotic approach. *Meteorology and Atmospheric Physics*, 130(3), 349-359.
6. Bednarik, M., Magulová, B., Matys, M., & Marschalko, M. (2010). Landslide susceptibility assessment of the Kral'ovany–Liptovský Mikuláš railway case study. *Physics and Chemistry of the Earth, Parts A/B/C*, 35(3-5), 162-171.
7. Booker, D. J., & Snelder, T. H. (2012). Comparing methods for estimating flow duration curves at ungauged sites. *Journal of Hydrology*, 434, 78-94.
8. Breiman, L. (2001). Random Forests. *Machine Learning*, 45(1), 5-32.
9. Chen, B. H., Wang, X. Z., Yang, S. H., & McGreavy, C. (1999). Application of wavelets and neural networks to diagnostic system development, 1, feature extraction. *Computers & Chemical Engineering*, 23(7), 899-906.
10. Choubey, V., Mishra, S., & Pandey, S. K. (2014). Time series data mining in real time surface runoff forecasting through Support Vector Machine. *International Journal of Computer Applications*, 98(3), 23-30.
11. Christian, K., Roy, A. F., Yudianto, D., & Zhang, D. (2021). Application of optimized Support Vector Machine in monthly streamflow forecasting: using Autocorrelation Function for input variables estimation. *Sustainable Water Resources Management*, 7(3), 1-14.
12. Drisya, J., Kumar, D. S., & Roshni, T. (2021). Hydrological drought assessment through streamflow forecasting using wavelet enabled artificial neural networks. *Environment, Development and Sustainability*, 23(3), 3653-3672.
13. Friedman, J., Hastie, T., & Tibshirani, R. (2001). The elements of statistical learning (Vol. 1, No. 10). *New York: Springer series in statistics*.
14. Ghorbani, M. A., Deo, R. C., Kim, S., Kashani, M. H., Karimi, V., & Izadkhah, M. (2020). Development and evaluation of the cascade correlation neural network and the random forest models for river stage and river flow prediction in Australia. *Soft Computing*, 1-12.
15. Ghorbani, M. A., Zadeh, H. A., Isazadeh, M., & Terzi, O. (2016). A comparative study of artificial neural network (MLP, RBF) and support vector machine models for river flow prediction. *Environmental Earth Sciences*, 75(6), 476.
16. Hammad, M., Shoaib, M., Salahudin, H., Baig, M. A. I., Khan, M. M., & Ullah, M. K. (2021). Rainfall forecasting in upper Indus basin using various artificial intelligence techniques. *Stochastic Environmental Research and Risk Assessment*, 1-23.
17. Khodakhah, H., Aghelpour, P., & Hamed, Z. (2021). Comparing linear and non-linear data-driven approaches in monthly river flow prediction, based on the models SARIMA, LSSVM, ANFIS, and GMDH. *Environmental Science and Pollution Research*, 1-20.
18. Lohani, A. K., Kumar, R., & Singh, R. D. (2012). Hydrological time series modeling: A comparison between adaptive neuro-fuzzy, neural network and autoregressive techniques. *Journal of Hydrology*, 442, 23-35.
19. Mallat, S. (1999). A wavelet tour of signal processing. *Elsevier*.
20. Mallat, S. G. (2009). A theory for multiresolution signal decomposition: the wavelet representation. In *Fundamental Papers in Wavelet Theory* (pp. 494-513). *Princeton University Press*.
21. Mehdizadeh, S., Ahmadi, F., Mehr, A. D., & Safari, M. J. S. (2020). Drought modeling using classic time series and hybrid wavelet-gene expression programming models. *Journal of Hydrology*, 587, 125017.
22. Mirabbasi, R., Kisi, O., Sanikhani, H., & Meshram, S. G. (2019). Monthly long-term rainfall estimation in Central India using M5Tree, MARS, LSSVR, ANN and GEP models. *Neural Computing and Applications*, 31(10), 6843-6862.
23. Misiti, M., Misiti, Y., Oppenheim, G., & Poggi, J. M. (1996). Wavelet Toolbox for Use with Matlab. *The Mathworks, Inc.: Natick, Massachusetts, USA*.
24. Montaseri, M., & Ghavidel, S. (2014). River Flow Forecasting by Using Soft computing. *Water and Soil*, 28(2), 394-405. (In Persian).
25. Mukhopadhyay, B., & Khan, A. (2015). Boltzmann–Shannon entropy and river flow stability within Upper Indus Basin in a changing climate. *International Journal of River Basin Management*, 13(1), 87-95.

26. Nayak, P. C., Sudheer, K. P., Rangan, D. M., & Ramasastri, K. S. (2004). A neuro-fuzzy computing technique for modeling hydrological time series. *Journal of Hydrology*, 291(1-2), 52-66.
27. Pai, P.F., & Hong, W.C. (2007). A recurrent support vector regression model in rainfall forecasting. *Hydrological Process*, 21, 819-827.
28. Pham, L. T., Luo, L., & Finley, A. O. (2020). Evaluation of Random Forest for short-term daily streamflow forecast in rainfall and snowmelt driven watersheds. *Hydrology and Earth System Sciences Discussions*, 1-33.
29. Phillis, G.D. (1996). Wavelets: a new alternative to Fourier transforms. *Computers in Physics*, 10(3), 247-252.
30. Polikar, R. (1996). Fundamental concepts & an overview of the wavelet theory. The Wavelet Tutorial Part I, *Rowan University, College of Engineering Web Servers*, 15.
31. Polikar, R. (1999). Fundamental concepts and overview of the wavelet theory: the wavelet tutorial—part I. *Rowan University, College of Engineering Web Servers*, 22.
32. Quinlan, J. R. (2014). C4. 5: programs for machine learning. Elsevier. *San Mateo, CA: Morgan Kaufmann*. 303 pp.
33. Roushangar, K., & Ghasempour, R. (2020). Monthly precipitation prediction improving using the integrated model based on kernel-wavelet and complementary ensemble empirical mode decomposition. *Amirkabir Journal of Civil Engineering*, 52(10), 2649-2660.
34. Saraiva, S. V., de Oliveira Carvalho, F., Santos, C. A. G., Barreto, L. C., & Freire, P. K. D. M. M. (2021). Daily streamflow forecasting in Sobradinho Reservoir using machine learning models coupled with wavelet transform and bootstrapping. *Applied Soft Computing*, 102, 107-121.
35. Saray, M. H., Eslamian, S. S., Klöve, B., & Gohari, A. (2020). Regionalization of potential evapotranspiration using a modified region of influence. *Theoretical and Applied Climatology*, 140(1), 115-127.
36. Shannon, C. E. (2001). A mathematical theory of communication. *ACM SIGMOBILE Mobile Computing and Communications Review*, 5(1), 3-55.
37. Shataee, S., Kalbi, S., Fallah, A., & Pelz, D. (2012). Forest attribute imputation using machine-learning methods and ASTER data: comparison of k-NN, SVR and random forest regression algorithms. *International Journal of Remote Sensing*, 33(19), 6254-6280.
38. Tang, T., Liang, Z., Hu, Y., Li, B., & Wang, J. (2020). Research on flood forecasting based on flood hydrograph generalization and random forest in Qiushui River basin, China. *Journal of Hydroinformatics*, 22(6), 1588-1602.
39. Vapnik, V.N. (1998). *Statistical Learning Theory*. Wiley, New York.
40. Wang, J., Bao, W., Gao, Q., Si, W., & Sun, Y. (2021). Coupling the Xinanjiang model and wavelet-based random forests method for improved daily streamflow simulation. *Journal of Hydroinformatics*, 23(3), 589-604.
41. Wang, W., & Ding, J. (2003). Wavelet network model and its application to the prediction of hydrology. *Nature and Science*, 1(1), 67-71.