



Available Online: <http://jmst.kmsu.ac.ir>

Original Article



Application of soft computing models in predicting monthly river flow (Case study: Volgograd hydrometric station)

Hafez Koohpari ¹, Khosrow Hosseini ^{*1}, Saeed Khorram ²

1. Department of Water Engineering and Hydraulic Structures, Faculty of Civil Engineering, Semnan University, Semnan, Iran.

2. Department of Civil Engineering, Marvdasht Branch, Islamic Azad University, Marvdasht, Iran.

*** Corresponding Author Email:** khoseini@semnan.ac.ir

Received: 20 January 2023

Revise Date: 22 February 2023

Accepted: 11 March 2023

DOI: 10.22113/jmst.2023.382026.2511

ABSTRACT

Today, river flow forecasting is one of the important topics in hydrology and water resources, water structures management, and flood forecasting. In this research, the performance of a combined model consisting of the slime mold optimizer algorithm and the support vector regression simulator algorithm (SVR-SMA) was evaluated in forecasting the average monthly discharge of the Volga River at the Volgograd hydrometric station during the statistical period of 1980–2020. In this research, various combinations of climatic observation data, including minimum temperature, maximum temperature, average rainfall, and sunshine hours, were used as input patterns to predict flow rate. The results obtained from this research indicate the better performance of the combined slime mold–support vector regression model, with a coefficient of determination equal to 0.80 ($R^2 = 0.80$), a root mean square error equal to 7.19 (RMSE = 7.19), and a mean absolute error equal to 0.56 (MAE = 0.56) in the validation stage for forecasting the monthly discharge of the Volga River at the Volgograd station, in comparison with genetic algorithms, artificial neural networks, and support vector regression alone. Therefore, the findings of this research indicate that the use of the slime mold algorithm together with the support vector regression algorithm has been able to improve its speed and accuracy.

Keywords: Caspian Sea, Volga River, flow estimation, slime mold algorithm, support vector regression

1. INTRODUCTION

The Volga River is the main supplier of the Caspian Sea. It provides 90% of the water entering the Caspian Sea (Kalugin, 2022), which is of great importance for water resources and environmental studies. Therefore, predicting the discharge of the Volga River is one of the important topics of the hydrology of this basin, which needs more research and studies. The purpose of this research is to investigate the effectiveness of new solutions by combining the slime mold algorithm and support vector regression as a new model in predicting the monthly river discharge and finally comparing the results of other models and providing a suitable and accurate method.

Copyrights:

Copyright for this article is retained by the author(s), with publication rights granted Journal of Marine Science and Technology. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



2. MATERIALS AND METHODS

The Slime Mold Optimizer (SMA) is a new population-based meta-heuristic algorithm inspired by the intelligent behavior of a type of mold called slime mold. The mold has intelligent behavior and can perform very fast, without error. The Support Vector Machine (SVM) is an efficient learning system based on the theory of constrained optimization that uses the inductive principle of structural error minimization and leads to an overall optimal solution (Vapnik et al., 1996). Support vector machines include two categories: the support vector classifier and support vector regression. A support vector machine algorithm seeks a decision plane that separates data of different classes. In some cases, a complex and non-linear structure is needed for data separation, in which the support vector machine maps and recovers the original data in a new space by using a set of mathematical functions called kernel functions. In this research, the optimal values of hyper-parameters required by the model were obtained by trial and error. Often, data models are evaluated based on the correlation coefficient (Coefficient of Determination, R^2), Root Mean Squared Error (RMSE), and Mean Absolute Error. In this study, the average daily rainfall data for 41 years from the Volgograd hydrometric station and the climate parameters of the Volgograd synoptic station were used between 1980 and 2020. From a total of 492 data points from the monthly hydrometric station and climatic parameters of the synoptic station, 369 data points were used for training and the remaining 123 data points were used for testing.

3. RESULTS

In this research, eight different models for predicting river discharge at the Volgograd hydrometric station were investigated. The evaluated models are based on the combination of parameters affecting the hydrometric station of Volgograd. The meteorological parameters used for modeling are: average monthly maximum temperature (T_{max}), average monthly minimum temperature (T_{min}), average monthly temperature (T_{mean}), average total monthly precipitation (P), and average total monthly sunshine hours (Rad). In this research, the slime mold algorithm was used as an optimization algorithm and the support vector regression algorithm was used as a simulation algorithm to predict the Volga River discharge.

4. DISCUSSION AND CONCLUSION

In this research, the performance of the combination of the slime mold optimizer and the support vector regression simulator was investigated in forecasting the monthly flow of the Volga River in the Caspian Sea catchment. In this regard, the average monthly discharge data of the Volgograd hydrometric station between 1980 and 2020 have been used. Different rainfall forecasting patterns with the influence of climatic variables were used for modeling. In the models used in this research, the best performance was related to the scenario of total rainfall, maximum and minimum temperatures, and monthly solar radiation using the RBF kernel. Also, the results of this research show that the use of the slime mold algorithm has been able to improve the speed and accuracy of the results of the support vector regression algorithm in predicting the monthly flow of the Volga River. The results of this research showed that the IHACRES model has good abilities in simulating the monthly discharge of the river, but the accuracy of data mining models was better than that of this model.

REFERENCES

- Kalugin, A., 2022. Hydrological and Meteorological Variability in the Volga River Basin under Global Warming by 1.5 and 2 Degrees. *Climate*, 107(10), pp.1–23. Doi: 10.3390/cli10070107
- Vapnik, V., Golwich, E.S. and Smola, A., 1996. Support vector method for function approximation, regression estimation and signal processing, *Journal of Advances in Neural Information Processing Systems*, 9, pp.281–287.



مقاله پژوهشی

Available Online: <http://jmst.kmsu.ac.ir>



کاربرد مدل های محاسبات نرم در پیش بینی دبی ماهانه رودخانه (مطالعه موردی: ایستگاه هیدرومتری ولگوگراد)

حافظ کوه پری^۱، خسرو حسینی^{۱*}، سعید خرم^۲

۱. گروه مهندسی آب و سازه های هیدرولیکی، دانشکده مهندسی عمران، دانشگاه سمنان، سمنان، ایران.

۲. گروه مهندسی عمران، دانشکده مهندسی عمران، واحد مرودشت، دانشگاه آزاد اسلامی، مرودشت، ایران.

پست الکترونیک نویسنده مسئول: khoseini@semnan.ac.ir

تاریخ پذیرش: ۱۴۰۱/۱۲/۲۰

تاریخ بازنگری: ۱۴۰۱/۱۲/۰۳

تاریخ ارسال: ۱۴۰۱/۱۰/۳۰

شناسه دیجیتال (DOI): 10.22113/jmst.2023.382026.2511

چکیده

امروزه، پیش بینی جریان رودخانه ها از مباحث مهم در هیدرولوژی و منابع آب است که می توان از نتایج الگوبندی جریان رودخانه در مدیریت منابع آب، مدیریت سازه های آبی و پیش بینی سیل استفاده کرد. در این تحقیق، عملکرد مدل ترکیبی الگوریتم بهینه ساز کپک مخاطی و الگوریتم شبیه ساز رگرسیون بردار پشتیبان (SVR-SMA) در پیش بینی متوسط دبی ماهانه رودخانه ولگا در محل ایستگاه هیدرومتری ولگوگراد طی دوره آماری ۱۹۸۰-۲۰۲۰ مورد ارزیابی قرار گرفته است. در این تحقیق ترکیب های مختلفی از داده های مشاهداتی اقلیمی شامل دمای حداقل، دمای حداکثر، میانگین بارش و ساعت آفتابی به عنوان الگوی ورودی جهت پیش بینی دبی جریان استفاده شد. نتایج به دست آمده از این پژوهش حاکی از عملکرد بهتر راه حل ترکیبی کپک مخاطی - رگرسیون بردار پشتیبان با ضریب تبیین برابر 0.80 ($R^2=0.80$)، جذر میانگین مربعات خطا برابر 7.19 ($RMSE=7.19$) و میانگین خطای مطلق برابر 0.56 ($MAE=0.56$) در مرحله صحت سنجی برای پیش بینی دبی ماهانه رودخانه ولگا در ایستگاه ولگوگراد در مقایسه با الگوریتم های ژنتیک، شبکه های عصبی مصنوعی و نیز رگرسیون بردار پشتیبان به تنهایی بوده است. بنابراین یافته های این تحقیق بیان گر آن است که استفاده از الگوریتم کپک مخاطی همراه با الگوریتم رگرسیون بردار پشتیبان توانسته است سرعت و دقت آن را ارتقاء بخشد.

واژگان کلیدی: دریای خزر، رودخانه ولگا، برآورد جریان، الگوریتم کپک مخاطی، رگرسیون بردار پشتیبان.

Copyrights:

Copyright for this article is retained by the author(s), with publication rights granted Journal of Marine Science and Technology. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



۱. مقدمه

با توجه به محدودیت و توزیع زمانی و مکانی غیریکنواخت منابع آبی در طبیعت، افزایش آلودگی‌ها، تخریب منابع طبیعی، رشد و توسعه جوامع شهری و فعالیت‌های کشاورزی و صنعتی، ضرورت برنامه‌ریزی و مدیریت صحیح منابع آبی اجتناب‌ناپذیر است. در این میان آگاهی از میزان آورد رودخانه یکی از مسائل مهم و کاربردی در مدیریت منابع آب و از اطلاعات ضروری به‌منظور اجرای پروژه‌های آبی است (Kemp et al., 2011). جهت پیش‌بینی جریان رودخانه در طی سالیان گذشته، روش‌های مختلفی ابداع شده است که می‌توانند به دو دسته مدل‌های مفهومی و مدل‌های مبتنی بر داده طبقه‌بندی شوند. مدل‌های مبتنی بر داده، کاربرد بیشتری از مدل‌های مفهومی دارند، زیرا درک ساختار داخلی آن‌ها، از فرآیندهای فیزیکی مدل‌سازی، مستقل است (Saghebzan, 2020).

علی‌رغم این‌که مدل‌های مختلفی برای پیش‌بینی آبدی رودخانه پیشنهاد شده است، لیکن با توجه به ماهیت غیرخطی و ویژگی‌های متغیر زمانی و مکانی در سیستم گردش آبی، هیچ یک از مدل‌های مفهومی ارائه شده به منظور الگوسازی دقیق دبی جریان نتوانسته‌اند به صورت یک مدل کارا معرفی گردند و عدم قطعیت‌های موجود در پیش‌بینی‌ها از اعتبار برآوردها از طریق مدل‌های گوناگون می‌کاهد. از سوی دیگر در سال‌های اخیر محققین با پیشرفت روش‌های هوش مصنوعی در جستجوی راه‌حلی برای پیش‌بینی پارامترهای مهم هیدرولیکی و هیدرولوژیکی حوضه‌های آبخیز و رودخانه‌ها هستند. در سال‌های اخیر روش‌های داده مینا برای شبیه‌سازی رفتار سیستم‌ها در مهندسی آب مورد استفاده قرار گرفته‌اند. این روش‌ها اکثراً الهام گرفته از پدیده‌های موجود در طبیعت هستند و به عنوان ابزاری قدرتمند در حل مسائل پیچیده به شمار می‌روند (Saghebzan, 2020).

Cheng et al. (2015) با استفاده از سه مدل شبکه عصبی مصنوعی (Artificial neural network (ANN)، ماشین بردار پشتیبان (SVM)) و مدل ترکیبی اقدام به پیش‌بینی دبی ماهانه سد خین فن جیانگ (Xinfengjiang) در چین نمودند. یافته‌های این تحقیق بیان می‌کند که تمامی مدل‌های مورد استفاده عملکرد قابل قبولی داشته‌اند. همچنین، مطابق معیارهای قضاوت آماری، در شرایطی که مدل‌ها به‌صورت ترکیبی به کار رفته‌اند، عملکرد بهتری نسبت به حالتی که مدل‌ها به صورت منفرد به کار می‌روند داشته‌اند.

Ghorbani et al. (2016) اقدام به ارزیابی روش‌های هیدرولوژیکی و داده‌کاوی در شبیه‌سازی و پیش‌بینی جریان ماهانه در ایستگاه هیدرومتری اراز کوسه کردند. نتایج این تحقیق نشان داد که کاربرد

مدل‌های داده‌کاوی منجر به بهبود نتایج شبیه‌سازی در مقایسه با مدل‌های هیدرولوژیکی می‌شود.

Young et al. (2017) از یک راه‌حل ترکیبی برای پیش‌بینی دبی در حوضه چیشان کریک (Chishan Creek) در جنوب تایوان استفاده کردند. یافته‌های آنان نشان می‌دهد که مدل ترکیبی نسبت به سه مدل منفرد هیدرولوژیکی، شبکه عصبی مصنوعی و ماشین بردار پشتیبان نتایج بهتری ارائه می‌دهد.

Wen and Zhong (2020) با استفاده از روش ترکیبی شبکه‌های عصبی مصنوعی و الگوریتم جستجوی همکاری (Cooperation (CSA) Search Algorithm) پیش‌بینی سری زمانی آبدی رودخانه در مناطق مرطوب و نیمه مرطوب را مورد بررسی قرار دادند. نتایج این تحقیق نشان داد که راه‌کار جدید توانایی بالایی در پیش‌بینی سری زمانی آبدی رودخانه در مناطق مورد مطالعه داشته است. Zounemat-Kermani et al. (2021) به مقایسه روش‌های شبکه‌های عصبی مدرن و مساحی عرفی در پیش‌بینی آبدی رودخانه‌ها پرداختند. این مطالعه بر روی پیش‌بینی آبدی روزانه رودخانه تایمز (Thames) در کشور انگلستان انجام پذیرفت که نتایج آن نشان دهنده توانایی بالای شبکه‌های عصبی مدرن در پیش‌بینی سری زمانی آبدی رودخانه مورد مطالعه بوده است.

Saghebzan (2020) از شیوه‌های پیش‌پردازش و پس‌پردازش سری‌های زمانی به همراه روش‌های مبتنی بر کرنل ماشین بردار پشتیبان و رگرسیون فرآیند گاوسی (Gaussian Process Regression (GPR)) جهت تخمین دبی جریان رودخانه‌ها برای دو رودخانه هوستونیک (Houstonic) و می‌سی‌سی‌پی (Mississippi) در ایالات متحده آمریکا استفاده نمودند. نتایج حاصل از تحلیل مدل‌های تعریف شده در این پژوهش، دقت بالای روش‌های تلفیقی به کار رفته در تحقیق را در تخمین دبی جریان به خوبی نشان می‌دهد.

Alinezhadi et al. (2021) جهت پیش‌بینی میانگین دبی روزانه ایستگاه هیدرومتری ملاثانی بر روی رودخانه کارون، روش‌های مختلف شامل مدل برنامه ریزی بیان ژن (GEP)، رگرسیون پارامتریک خطی (LR)، رگرسیون پارامتریک غیر خطی (NLR) و روش ناپارامتریک K - نزدیک‌ترین همسایگی (K-NN) را با الگوهای ورودی مختلف مورد بررسی قرار دادند که برای دوره زمانی ۵۰ ساله این ایستگاه، مدل برنامه ریزی بیان ژن بهترین عملکرد را داشت.

از آنجائی که سهم رودخانه ولگا از دبی ورودی جریان آب به دریای خزر حدود ۹۰ درصد است (Kalugin, 2022)، به عنوان مهم‌ترین منبع تأمین آب دریای خزر دارای اهمیت بالایی برای مطالعات منابع آب و محیط زیستی است. بنابراین پیش‌بینی آبدی این رودخانه به دریای خزر

(میانگین دبی سالانه ۸۰۰۰ مترمکعب بر ثانیه) پرآب‌ترین رودخانه اروپاست. رودخانه ولگا پس از عبور از شهرهای مختلف روسیه با یک دلتای بزرگ در ۲۸ متری زیر تراز آب‌های آزاد به دریای خزر می‌ریزد که حدود ۹۰ درصد از آب ورودی به دریای خزر را تأمین می‌کند که تأثیر زیادی در نوسانات تراز آب دریای خزر دارد (Kalugin, 2022). رودخانه‌های کاما (Kama)، اوکا (Oka)، ویاتکا (Vyatka) و بلایا (Belaya)، شاخه‌های اصلی رودخانه ولگا هستند که هر کدام بیش از ۱۰۰۰ کیلومتر طول دارند (Middelkoop, 2005). در پایین‌دست شهر ولگوگراد تا خروجی دلتا هیچ شاخه مهمی وارد جریان اصلی نمی‌شود (Leummens, 2016).

رژیم طبیعی آب رودخانه ولگا با سیلاب‌های بهاری، دوره‌های تابستان و زمستان کم آب و سیلاب‌های پاییزی مشخص می‌شود. میانگین دمای هوای سالانه در حوضه ولگا ۴ درجه سانتی‌گراد است که میزان بارش برف و باران به ترتیب ۳۰ و ۷۰ درصد است. میانگین بارندگی سالانه برای کل حوضه ۵۸۵ میلی‌متر است (Kalugin, 2022).

موقعیت حوضه آبریز و مسیر رودخانه ولگا در شکل (۱) نشان داده شده است. نمودار دبی ماهانه خروجی رودخانه ولگا در ایستگاه هیدرومتری ولگوگراد به دریای خزر در شکل (۲) نشان داده شده است.

از مباحث مهم هیدرولوژی این حوضه به شمار می‌رود که تحقیقات و مطالعات بیشتری را در این مورد می‌طلبد. هدف از انجام این پژوهش بررسی کارایی راه حل‌های نو برای نخستین بار با ترکیب الگوریتم کپک مخاطی و رگرسیون بردار پشتیبان به عنوان یک مدل جدید در پیش‌بینی میزان دبی ماهانه رودخانه و در نهایت مقایسه نتایج مدل‌های مختلف دیگر و ارائه روش مناسب و با دقت کافی برای پیش‌بینی آورد ماهانه این رودخانه است.

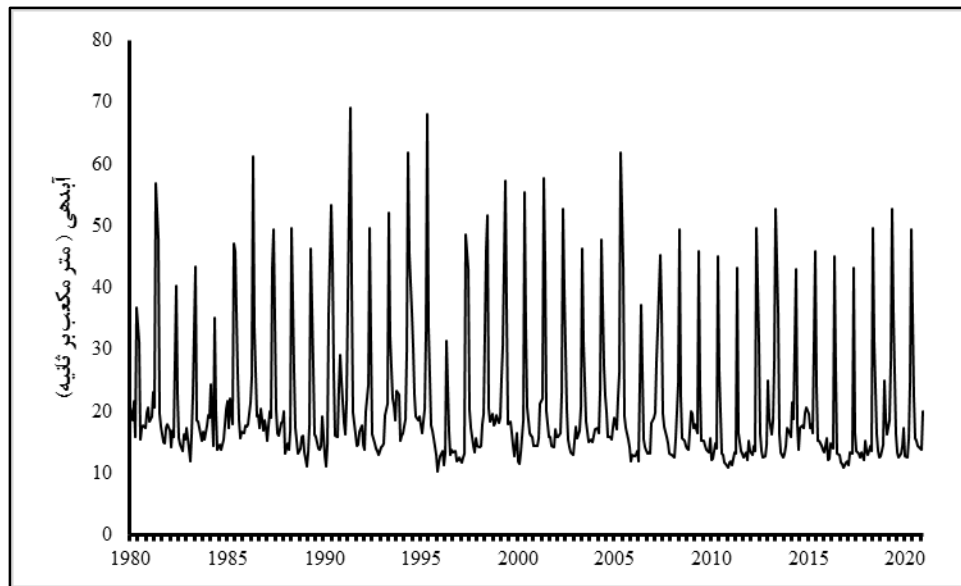
۲. مواد و روش‌ها

رودخانه ولگا بخشی از حوضه زهکشی بسته دریای خزر است که از نواحی شمالی منطقه سیبری در تپه‌های والداج در شمال مسکو در ارتفاع ۲۲۸ متری از سطح آب‌های آزاد سرچشمه می‌گیرد (Kalugin, 2022). حوضه آبریز این رودخانه تقریباً به طور کامل در کشور روسیه واقع شده است که حدود یک سوم از کل مساحت این کشور را در برمی‌گیرد. مساحت حوضه آبریز این رودخانه ۱۳۶۰۰۰۰ کیلومترمربع است که بخش عمده آن را دشت‌های نسبتاً هموار است و کمتر از ۵ درصد آن را کوه‌ها تشکیل می‌دهند (Middelkoop, 2005). این رودخانه به طول حدود ۳۵۳۰ کیلومتر طولانی‌ترین و با جریان حدود ۲۵۰ کیلومترمکعب در سال



شکل ۱- موقعیت حوضه آبریز رودخانه ولگا و شهرهای مهم آن

Fig. 1- Volga River basin and its important cities



شکل ۲- آبدهی ماهانه خروجی رودخانه ولگا به دریای خزر در دوره مشاهداتی (۱۹۸۰-۲۰۲۰)
Fig. 2- Volga River monthly discharge to the Caspian Sea (1980-2020)

به‌منظور ارزیابی از قرار داشتن وقایع حدی در یک سری زمانی و نیز کافی بودن دوره آماری سری زمانی مورد استفاده در مدل‌سازی از آزمون کفایت داده‌ها استفاده می‌شود. یکی از روش‌های آزمون کفایت طول داده‌های سری‌های زمانی، روش پیشنهادی هرست (Hurst Coefficient) است. هرست پارامتری را به نام ضریب هرست پیشنهاد نمود که توسط رابطه (۴) معرفی می‌شود. در رابطه (۴) پارامتر N تعداد داده‌های سری زمانی و σ انحراف معیار آن سری است. مقدار R از رابطه (۵) برابر است با تفاوت بین مقادیر مثبت و منفی انحراف از میانگین سری‌های زمانی که به صورت تجمعی محاسبه شده باشد. هرگاه مقدار ضریب هرست برای یک سری زمانی برابر با $0/5$ باشد بدان معناست که سری زمانی مورد استفاده یک سری مستقل نرمال است. هرچه مقدار این ضریب از عدد $0/5$ بیشتر باشد نشان‌دهنده وجود حافظه بلندمدت در سری زمانی است. به عبارت دیگر سری زمانی مورد استفاده نیازی به افزایش طول آماری آن نیست (Karamouz and Araghinejad, 2005).

آزمون ترسیمی روند من- کندال (Mann-Kendall) یک آزمون ناپارامتریک رتبه مبنا جهت بررسی روند داده‌های هیدرولوژیک است که در سال ۱۹۴۵ توسط من و کندال ارائه شده است. روابط استفاده شده در آزمون من- کندال به‌صورت رابطه (۱) است. واریانس فرایند توسط رابطه (۲) تعیین می‌شود که m تعداد دسته‌ها با داده‌های گره‌دار و t_i تعداد داده‌ها گره‌دار در هر دسته m است. مقادیر Z به دست آمده از رابطه (۳) با مقادیر Z به دست آمده از جدول توزیع نرمال استاندارد با سطح اطمینان مورد نظر مقایسه می‌گردد. چنانچه Z محاسباتی بزرگ‌تر از Z جدول باشد فرضیه صفر مبنی بر وجود روند با سطح اطمینان مورد نظر رد می‌شود و داده‌ها در این حالت فاقد روند خواهند بود. چنانچه فرضیه صفر مبنی بر وجود روند با سطح اطمینان مورد نظر تأیید شود، داده‌ها دارای روند می‌شوند. در صورت وجود روند، مقادیر Z مثبت نشان‌دهنده روند مثبت و مقادیر منفی Z نشان‌دهنده روند منفی هستند (Bagherpoor et al., 2017).

$$s = \sum_{i=1}^{N-1} \sum_{j=i+1}^N \text{sgn}(X_j - X_i), \text{ Where } \text{sgn}(X) = \begin{cases} +1, & X > 0 \\ 0, & X = 0 \\ -1, & X < 0 \end{cases} \quad \text{رابطه (۱)}$$

$$\sigma_s^2 = \frac{1}{18} \left[N(N-1)(2N+5) - \sum_{i=1}^m t_i(t_i-1)(2t+5) \right] \quad \text{رابطه (۲)}$$

$$z = \begin{cases} \frac{s-1}{\sigma_s}, & \text{if } s > 0 \\ 0, & \text{if } s = 0 \\ \frac{s+1}{\sigma_s}, & \text{if } s < 0 \end{cases} \quad \text{رابطه (۳)}$$

$$k = \frac{\log \frac{R}{\sigma}}{\log \frac{N}{2}} \quad \text{رابطه (۴)}$$

$$R = S^+ + S^- \quad \text{رابطه (۵)}$$

محیط است، $\overrightarrow{X(t)}$ نشان دهنده موقعیت کپک مخاطی در تکرار t ام است، $\overrightarrow{X(t+1)}$ موقعیت بعدی کپک مخاطی در تکرار t است، $\overrightarrow{X_B(t)}$ و $\overrightarrow{X_A(t)}$ نشان دهنده دو موقعیت تصادفی انتخاب شده از کپک مخاطی است و $\overrightarrow{W}$ نشان دهنده وزن کپک مخاطی است. مقدار p با استفاده از رابطه (۸) به دست می‌آید که در آن $i = 1, 2, \dots, n$ است و n نشان دهنده تعداد سلول‌های موجود در کپک مخاطی است، $S(i)$ نشان دهنده مقدار برازندگی $\overrightarrow{X(t)}$ است و DF نشان دهنده بهترین برازندگی به دست آمده طی همه تکرارها است. $\overrightarrow{vb}$ از رابطه (۹) به دست می‌آید. همچنین $\overrightarrow{W}$ با استفاده از رابطه (۱۲) به دست می‌آید. در رابطه (۱۲)، $condition$ برابر این است که $S(i)$ در نیمه اول جمعیت قرار دارد، r نشان دهنده یک مقدار تصادفی در بازه $[0, 1]$ است، bF نشان دهنده برازندگی بهینه به دست آمده در تکرار فعلی است، wF نشان دهنده بدترین مقدار برازندگی به دست آمده در تکرار فعلی است و $SmellIndex$ نشان دهنده توالی مقادیر برازندگی مرتب شده است (به صورت صعودی). همچنین موقعیت کپک‌های مخاطی با استفاده از رابطه (۱۳) به روز رسانی می‌شود که در آن UB و LB به ترتیب حد پایین و حد بالای بازه جستجو است، $rand$ و r به طور تصادفی در بازه $[0, 1]$ تعیین می‌شوند و مقدار Z در آزمایش تنظیم پارامتر مورد بحث قرار خواهد گرفت. بنابراین این روند تکرار می‌شود تا زمانی که شرط توقف برقرار باشد. سپس خروجی $\overrightarrow{X^*}$ که نشان دهنده موقعیت ویژگی‌های بهینه است به دست می‌آید (Li et al., 2020).

$$\overrightarrow{W(SmellIndex(i))} = \begin{cases} 1 + r \cdot \log \left(\frac{bF - S(i)}{bF - wF} \right) + 1, & condition \\ 1 - r \cdot \log \left(\frac{bF - S(i)}{bF - wF} \right) + 1, & others \end{cases} \quad \text{رابطه (۱۱)}$$

$$SmellIndex = sort(S) \quad \text{رابطه (۱۲)}$$

$$\overrightarrow{X^*} = \begin{cases} rand.(UB - LB) + LB, & rand < z \\ \overrightarrow{X_b(t)} + \overrightarrow{vb} \cdot (W \cdot \overrightarrow{X_A(t)} - \overrightarrow{X_B(t)}), & r < p \\ \overrightarrow{vc} \cdot \overrightarrow{X(t)}, & r \geq p \end{cases} \quad \text{رابطه (۱۳)}$$

الگوریتم بهینه‌ساز کپک مخاطی (SMA) یک الگوریتم فراابتکاری جدید مبتنی بر جمعیت است که با الهام از رفتار هوشمندانه نوعی کپک به نام کپک مخاطی ارائه شده است. کپک رفتاری هوشمندانه دارد و می‌تواند خیلی سریع، بدون خطا مسیریابی انجام دهد. این نوع کپک، موجودات تک سلولی هستند که گرد هم جمع می‌شوند و ساختاری چند سلولی و تناسلی را ایجاد می‌کنند. کپک‌های مخاطی مغز ندارند ولی به طرزی عجیب می‌توانند کارهای هوشمندانه‌ای انجام داده و در مسیرهای پیچ و خم دار راه خود را پیدا کنند. آن‌ها به دقت در تغذیه خود تعادل برقرار کرده و کیفیت غذا و ریسک آن را می‌سنجند. به منظور تعیین موقعیت اولیه تصادفی کپک‌ها یا ویژگی‌های یعنی $X_{i,k}$ از رابطه (۶) استفاده می‌شود. که در آن، ub و lb به ترتیب حد بالا و حد پایین برای هر راه‌حل یا هر ویژگی منتخب است. حال مقدار تابع برازندگی برای تمام کپک‌ها محاسبه می‌شود و آن کپکی که تابع برازندگی به دست آمده برای آن بیشترین مقدار ممکن باشد به عنوان معیار قرار می‌گیرد و موقعیت مربوط به آن که به صورت $\overrightarrow{X^*}$ است به عنوان ویژگی مورد نظر تعیین می‌شود. کپک مخاطی برای نزدیک شدن

به طعمه و مسیریابی آن از بوی طعمه که در هوا پخش می‌شود استفاده می‌کند. رابطه (۷) رفتار مسیریابی کپک مخاطی را بر اساس بوی طعمه توصیف می‌کند، که در آن $\overrightarrow{vb}$ یک پارامتر است که در بازه $[-a, a]$ قرار دارد، $\overrightarrow{vc}$ یک پارامتر است که به طور خطی از یک به صفر کاهش می‌یابد، t نشان دهنده تکرار فعلی است، $\overrightarrow{X_b(t)}$ نشان دهنده موقعیت کپک مخاطی در تکرار t ام با بیشترین غلظت بوی موجود در

$$X_{i,k} = lb + rand(0,1) \times (ub - lb) \quad \text{رابطه (۶)}$$

$$\overrightarrow{X(t+1)} = \begin{cases} (\overrightarrow{X_b(t)}) + (\overrightarrow{vb}) \cdot (W \cdot (\overrightarrow{X_A(t)}) - (\overrightarrow{X_B(t)})), & r < p \\ \overrightarrow{vc} \cdot \overrightarrow{X(t)}, & r \geq p \end{cases} \quad \text{رابطه (۷)}$$

$$p = \tanh(|S(i) - DF|) \quad \text{رابطه (۸)}$$

$$\overrightarrow{vb} = [-a, a] \quad \text{رابطه (۹)}$$

$$a = \arctanh\left(-\left(\frac{t}{\max t}\right) + 1\right) \quad \text{رابطه (۱۰)}$$

یکی از روش‌های متداول برای حل مسائل غیرخطی، استفاده از توابع کرنل است. این توابع بر اساس ضرب داخلی داده‌های مفروض تعریف می‌شوند. طراحی روش‌های رگرسیون مبتنی بر فرایند گوسی نیز شامل استفاده از مفهوم تابع کرنل است. در واقع با یک تبدیل غیرخطی از فضای ورودی به فضای خصیصه با ابعاد بیشتر می‌توان مسائل را به صورت خطی تفکیک‌پذیر ساخت. با تبدیل نمونه‌ها از فضای ورودی به فضای ویژگی، تفکیک‌کننده غیرخطی به حالت خطی تبدیل خواهد شد. از مهم‌ترین توابع کرنل می‌توان به چندجمله‌ای ساده (رابطه ۲۱)، چندجمله‌ای نرمال شده (رابطه ۲۲)، کرنل تابع شعاع محور (رابطه ۲۳) اشاره کرد. هر یک از این توابع کرنل در ساختار خود دارای پارامترهایی هستند که به نام هاپر پارامتر (Hyper Parameter) شناخته می‌شوند. در کاربرد روش‌های مبتنی بر کرنل، علاوه بر انتخاب تابع کرنل بهینه، تعیین مقدار بهینه هاپر پارامترهای مربوط به هر تابع بسیار حائز اهمیت است زیرا دقت ین روش رگرسیونی وابستگی زیادی به انتخاب این پارامترها دارد. برای انتخاب پارامترهای انتخاب شده توسط کاربر $(C, \varepsilon, \gamma, \sigma)$ روش‌های متعددی مانند روش سعی و خطا، روش جستجوی شبکه، روش الگوریتم ژنتیک و برخی از روش‌های دیگر ارائه شده است (Pal and Deswal, 2010).

ماشین بردار پشتیبان (SVM) یک سیستم یادگیری کارآمد بر مبنای تئوری بهینه‌سازی مقید است که از اصل استقرای کمینه‌سازی خطای ساختاری استفاده می‌کند و منجر به یک جواب بهینه کلی می‌گردد (Vapnik et al., 1996). ماشین‌های بردار پشتیبان شامل دو دسته طبقه‌کننده بردار پشتیبان و رگرسیون بردار پشتیبان هستند. یک الگوریتم ماشین بردار پشتیبان به دنبال صفحه تصمیم است که داده‌های با کلاس‌های مختلف را از هم تفکیک کند. در برخی موارد، به ساختار پیچیده و غیرخطی برای جداسازی داده‌ها نیاز است که در این مواقع ماشین بردار پشتیبان داده‌های اصلی را با به کارگیری مجموعه‌ای از توابع ریاضی که کرنل (Kernel Functions) نام دارند در فضای جدیدی نگاشت و بازایی می‌کند. اگر داده‌های آموزش به صورت k سری به صورت $(x_k, y_k), \dots, (x_i, y_i)$ باشند تابع خطی سازی مربوطه به صورت روابط (۱۴) و (۱۵) قابل بیان است که x بردار ورودی، w وزن بردار b میزان اختلال است. تابع $f(x)$ همواره می‌بایست به گونه‌ای تعیین شود که به طور هم‌زمان میزان کمترین انحراف (ε) تعیین شده و همچنین تابع، مقدار مناسب w را اختیار نماید. ایجاد چنین شرایطی با حل معادله بهینه‌سازی زیر قابل دستیابی است که w, b, x_i و y_i همان پارامترهای رابطه (۱۵) بوده و ε نیز میزان انحراف اعمال شده است (Shahinejad and Dehghani, 2017).

$$f(x) = (w.x) + b \quad \text{رابطه (۱۴)}$$

$$\text{with } (w.x) \in R^N, b \in R \quad \text{رابطه (۱۵)}$$

$$\text{Minimize } \frac{1}{2} \|w\|^2 + C \sum_{i=1}^m (\xi_i^- + \xi_i^+) \quad \text{رابطه (۱۶)}$$

$$(w_i x_i + b) - y_i \leq \varepsilon + \xi_i^+, i = 1, 2, 3, \dots, m, \xi_i^+ \geq 0 \quad \text{رابطه (۱۷)}$$

$$y_i - (w_i x_i + b) \leq \varepsilon + \xi_i^-, i = 1, 2, 3, \dots, m, \xi_i^- \geq 0 \quad \text{رابطه (۱۸)}$$

$$f(x) = (w^{Tr} . K(x_i, x_j)) + b \quad \text{رابطه (۱۹)}$$

$$K(x, x_i) = \exp\left(-\frac{|x - x_i|}{2\gamma^2}\right), i = 1, 2, 3, \dots, m \quad \text{رابطه (۲۰)}$$

$$K(x_i, x_j) = (x_i^T x_j + 1) \quad \text{رابطه (۲۱)}$$

$$K(x_i, x_j) = \frac{(x_i^T x_j + 1)^P}{\sqrt{(x_i^T x_i)(x_j^T x_j)}} \quad \text{رابطه (۲۲)}$$

$$K(x_i, x_j) = \exp(-\gamma |x_i - x_j|^2) \quad \text{رابطه (۲۳)}$$

$$R^2 = \frac{\sum_{i=1}^n (X_O - \bar{X}_O)(X_P - \bar{X}_P)}{\sqrt{\sum_{i=1}^n (X_O - \bar{X}_O)^2 \sum_{i=1}^n (X_P - \bar{X}_P)^2}} \quad \text{رابطه (۲۴)}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (X_O - X_P)^2}{n}} \quad \text{رابطه (۲۵)}$$

$$MAE = \frac{\sum_{i=1}^n |X_O - X_P|}{n} \quad \text{رابطه (۲۶)}$$

همان گونه که قبلاً بیان شد ارزیابی روند داده‌های اقلیمی ماهانه (دمای حداقل، دمای حداکثر، میانگین بارش و ساعت آفتابی) ایستگاه‌های سینوپتیک هواشناسی ولگوگراد و نیز آبدهی ماهانه ایستگاه ولگوگراد در بازه زمانی ۱۹۸۰-۲۰۲۰ با بهره‌گیری از آزمون ترسیمی من-کندال انجام گردید. همان گونه که نتایج شکل‌های (۳) و (۴) نشان می‌دهند مقادیر Z داده‌های مورد استفاده در این تحقیق، در بازه ۱/۹۶+ تا ۱/۹۶- قرار داشته، بنابراین در سطح اعتماد ۹۵ درصد فاقد روند بوده و قابلیت اعتماد دارند. مقادیر ضریب هرست برای پارامترهای اقلیمی ایستگاه‌های هواشناسی مورد استفاده در این تحقیق در جدول (۲) ارائه شده است. نتایج این جدول نشان‌دهنده مناسب بودن حافظه بلندمدت طول دوره آماری سری زمانی این ایستگاه‌ها است. نتایج این جدول نشان‌دهنده مناسب بودن حافظه بلندمدت طول دوره آماری سری زمانی ایستگاه‌های مورد استفاده است.

در این تحقیق هشت مدل مختلف جهت پیش‌بینی آبدهی در ایستگاه هیدرومتری ولگوگراد مورد بررسی قرار گرفت. مدل‌های ارزیابی شده بر اساس ترکیب پارامترهای مؤثر بر آبدهی ایستگاه هیدرومتری ولگوگراد در نظر گرفته شده است که در جدول (۳) مشخصات هر مدل ارائه شده است. پارامترهای ایستگاه‌های هواشناسی استفاده شده در مدل‌سازی عبارت‌اند از: متوسط دمای بیشینه ماهیانه (T_{max})، متوسط دمای کمینه ماهیانه (T_{min})، متوسط دمای ماهیانه (T_{mean})، متوسط مجموع بارش ماهیانه (P) و متوسط مجموع ساعات آفتابی ماهیانه (Rad). در این تحقیق الگوریتم کپک مخاطی به عنوان الگوریتم بهینه‌یاب و الگوریتم رگرسیون بردار پشتیبان به عنوان الگوریتم شبیه‌ساز جهت پیش‌بینی آبدهی رودخانه ولگا استفاده گردید. دامنه پارامتر مؤثر و نیز مقادیر بهینه مورد استفاده در الگوریتم‌های کپک مخاطی و رگرسیون بردار پشتیبان در جداول (۴) و (۵) ارائه شده است.

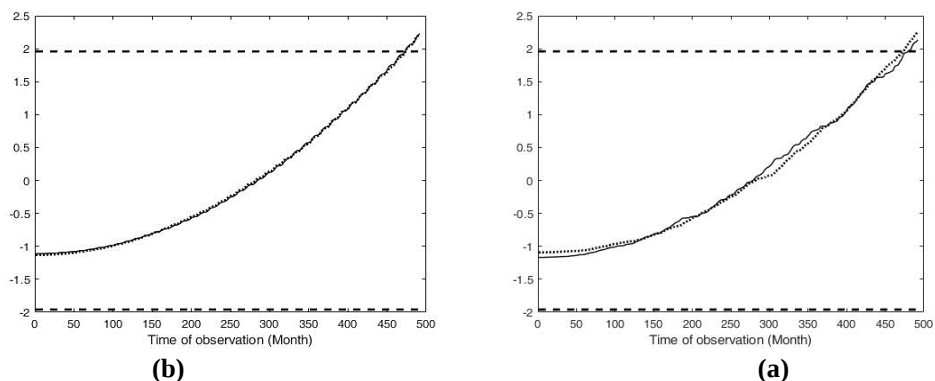
در این تحقیق، مقادیر بهینه‌های پارامترهای مورد نیاز مدل به روش سعی و خطا به دست آمده‌اند. غالباً مدل‌های داده‌منا بر اساس ضریب همبستگی (R^2 ، Coefficient of Determination)، جذر میانگین مربعات خطا (Root Mean Squared Error, RMSE) و میانگین مطلق خطا (Mean Absolute Error) مورد ارزیابی قرار می‌گیرند. روابط (۲۴) تا (۲۶) معیارهای ارزیابی مورد استفاده در این تحقیق را نشان می‌دهند. X_O مقادیر داده‌مشاهدیر داده پیش‌بینی شده در همان زمان، $\bar{X}_O$ میانگین مقادیر داده‌های مشاهداتی، $\bar{X}_P$ میانگین مقادیر داده‌های پیش‌بینی شده و n تعداد داده‌های مورد بررسی می‌باشد (Karamouz and Araghinejad, 2005).

۳. نتایج

در این مطالعه، از داده‌های میانگین آبدهی روزانه ۴۱ ساله ایستگاه هیدرومتری ولگوگراد و پارامترهای اقلیمی ایستگاه سینوپتیک ولگوگراد در فاصله سال‌های ۱۹۸۰-۲۰۲۰ استفاده شده است. از مجموع ۴۹۲ پارامتر آبدهی ماهانه ایستگاه هیدرومتری و پارامترهای اقلیمی ایستگاه سینوپتیک مورد بررسی، تعداد ۳۶۹ داده به منظور آموزش و تعداد ۱۲۳ داده باقیمانده برای آزمون مدل استفاده گردید. به طور کلی ۷۵ درصد داده‌ها برای آموزش و ۲۵ درصد برای آزمون در نظر گرفته شده است. پارامترهای آماری دبی ماهیانه ایستگاه ولگوگراد در جدول (۱) ارائه گردیده‌اند. مطابق نتایج این جدول حداکثر دبی در دوره آموزش بیشتر از دوره آزمون است، ولی حداقل دبی در دوره آموزش کمتر از دوره آزمون بوده است. این موضوع جهت برون‌یابی داده‌های مورد آزمون به‌منظور ارزیابی بهتر مدل‌های مورد استفاده در این تحقیق دارای اهمیت است (Kisi., 2007).

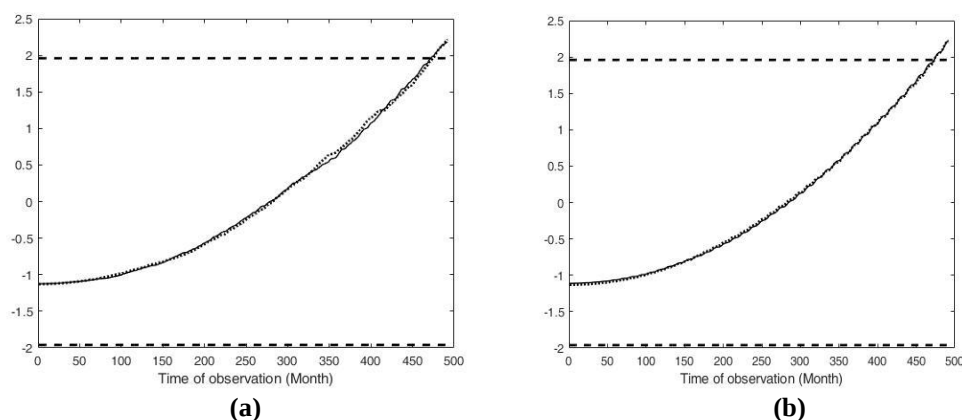
جدول ۱- مقادیر پارامترهای آماری داده‌های ماهیانه جریان رودخانه ولگا در ایستگاه ولگوگراد (۱۹۸۰-۲۰۲۰)
Table 1- Monthly statistical data of Volga River flow at Volgograd station (1980-2020)

Stage name	Number of statistical data	Mean (m3/s)	Maximum (m3/s)	Minimum (m3/s)	Standard deviation (m3/s)	Coefficient of variation (%)	Skewness (m3/s)
Test	369	21.68	69.1	10.2	11.47	0.53	2.04
Train	123	19.07	52.8	10.8	9.86	0.52	1.95



شکل ۳- نتایج آزمون من-کندال داده‌های (a) دبی ماهیانه ایستگاه هیدرومتری ولگوگراد، (b): متوسط حداکثر دمای ماهیانه ایستگاه سینوپتیک ولگوگراد

Fig. 3- Mann-Kendall Test results: (a) Volgograd Hydrometric Station monthly discharge, (b) Volgograd Synoptic Station average maximum monthly temperature



شکل ۴- نتایج آزمون من-کندال داده‌های (a): متوسط حداقل دمای ماهیانه ایستگاه سینوپتیک ولگوگراد، (b): متوسط بارش ماهیانه ایستگاه سینوپتیک ولگوگراد

Fig. 4- Mann-Kendall Test results for Volgograd Synoptic Station: a) average minimum monthly temperature, b) average monthly precipitation

جدول ۲- ضریب هرست پارامترهای اقلیمی ایستگاه‌های سینوپتیک مورد استفاده

Table 2- Hurst coefficient of the climatic parameters of the used synoptic stations

Sstation name	Monthly precipitation	Monthly sun time	Monthly maximum temperature (m3/s)	Monthly minimum temperature (m3/s)	Monthly discharge (m3/s)
Volgograd synoptic	0.723	0.693	0.615	0.672	-
Volgograd hydrometry	-	-	-	-	0.639

جدول ۳- الگوهای مختلف مورد بررسی برای پیش‌بینی آبدهی ماهیانه ایستگاه هیدرومتری ولگوگراد

Table 3- Different models studied to predict the Volgograd Hydrometric Station monthly discharge

No.	Model structure
1	$Q(t) = f \{P(t), T_{max}(t), T_{min}(t), Rad(t)\}$
2	$Q(t) = f \{P(t-1), T_{max}(t-1), T_{min}(t), T_{min}(t-1), Rad(t-1)\}$
3	$Q(t) = f \{P(t), P(t-1), T_{max}(t), T_{max}(t-1), T_{min}(t), T_{min}(t-1), Rad(t), Rad(t-1)\}$
4	$Q(t) = f \{P(t), P(t-1), T_{max}(t), T_{max}(t-1), T_{min}(t), T_{min}(t-1)\}$
5	$Q(t) = f \{P(t), P(t-1), T_{max}(t), T_{max}(t-1), T_{min}(t), T_{min}(t-1), Rad(t)\}$
6	$Q(t) = f \{P(t), P(t-1), T_{max}(t), T_{max}(t-1), T_{min}(t), T_{min}(t-1), Rad(t-1)\}$
7	$Q(t) = f \{P(t), T_{max}(t), T_{max}(t-1), T_{min}(t), T_{min}(t-1), Rad(t), Rad(t-1)\}$
8	$Q(t) = f \{P(t), T_{max}(t), T_{max}(t-1), T_{min}(t), T_{min}(t-1), Rad(t-1)\}$

جدول ۴- مقادیر پارامترهای بهینه الگوریتم کپک مخاطی

Table 4- Optimal parameter values of the slime mold algorithm

No.	Parameter	variation range	Optimal values
1	Z Parameter	0-0.01	0.008
2	Initial society size	10-500	100
3	maximum algorithm iterations	1000	1000

جدول ۵- دامنه و مقادیر بهینه هایپر پارامترهای مؤثر در الگوریتم رگرسیون بردار پشتیبان برای مدل‌سازی پیش‌بینی آبدهی رودخانه ولگا در ایستگاه ولگوگراد

Table 5- The range and optimal values of the effective hyperparameters in the support vector regression algorithm for the modeling of the Volga River discharge prediction at Volgograd Station

No	SVR algorithm hyperparameters values			
	d	ϵ	γ	C
variation range	1-4	0-1.5	0.1-5	10-40
Optimal values	1.5	0.18	3.2	16

دارای بهترین عملکرد در میان سایر مدل‌ها و کرنل‌های مورد استفاده بوده و به‌عنوان بهترین مدل برای پیش‌بینی آبدهی ماهیانه ایستگاه هیدرومتری ولگوگراد انتخاب گردیده است. شکل شماره (۵) نمودار پراکندگی نقاط داده‌های مشاهداتی و بهترین مدل پیش‌بینی انتخابی را نشان می‌دهد. همچنین شکل شماره (۶) تأیید کننده این موضوع است که دقت مدل انتخابی با کاربرد راه حل ترکیبی الگوریتم کپک مخاطی و رگرسیون بردار

همان‌گونه که نتایج در جداول (۶) و (۷) نشان داده شده است، هشت سناریوی مختلف جهت پیش‌بینی آبدهی ایستگاه هیدرومتری ولگوگراد مورد بررسی قرار گرفته است که نتایج ارزیابی آن‌ها با استفاده از روش ترکیبی الگوریتم‌های کپک مخاطی و رگرسیون بردار پشتیبان نشان می‌دهد که مدل شماره ۵ در مراحل واسنجی با $R^2 = 0.79$ ، $MAE = 0.2$ ، $RMSE = 0.61$ و در مرحله صحت‌سنجی با $MAE = 0.80$ ، $RMSE = 0.79$ ، $R^2 = 0.56$ با بهره‌گیری از کرنل RBF

یافتن ساختار شبکه استفاده گردید. تعداد نرون‌ها در لایه ورودی مدل سه نرون و در لایه میانی شش نرون است. توابع تبدیل نرون‌های میانی در این مدل تانژانت سیگموئید و نرون‌های لایه خروجی نیز تابع تبدیل خطی بوده است. در مدل پیش‌بینی آینده مبتنی بر الگوریتم ژنتیک نیز جمعیت اولیه ۵۰۰، احتمال ترکیب ۰/۹، احتمال جهش ۰/۳ و انتخاب والدین بر اساس چرخه رولت در نظر گرفته شده است.

نتایج جدول (۸) نشان می‌دهد که استفاده از الگوریتم رگرسیون بردار پشتیبان به تنهایی در مقایسه با برخی روش‌های داده‌منا دارای دقت بالایی نمی‌باشد لیکن ترکیب آن با الگوریتم کپک مخاطی موجب ارتقا دقت و سرعت مدل شده است. بنابراین این راه حل ترکیبی در مقایسه با سایر روش‌های معمول مورد استفاده در پیش‌بینی آینده، دقت بالایی را نشان می‌دهد.

پشتیبان (SVR-SMA) و با استفاده از کرنل RBF در پیش‌بینی جریان ماهیانه ایستگاه ولگوگراد دارای دقت بالایی است.

میزان توانایی راه حل ترکیبی الگوریتم کپک مخاطی و رگرسیون بردار پشتیبان (SVR-SMA) در پیش‌بینی آینده ایستگاه هیدرومتری ولگوگراد با نتایج الگوریتم رگرسیون بردار پشتیبان به تنهایی (SVR)، الگوریتم ژنتیک (GA) و همچنین الگوریتم شبکه عصبی مصنوعی (ANN) مورد ارزیابی قرار گرفت که نتایج آن در جدول (۸) ارائه شده است.

در این بخش از تحقیق، هاپیر پارامترهای مورد استفاده در الگوریتم رگرسیون بردار پشتیبان در راه حل ترکیبی و نیز استفاده منفرد از رگرسیون بردار پشتیبان به صورت یکسان در نظر گرفته شده است. همچنین، در مدل پیش‌بینی جریان ساخته شده از الگوریتم شبکه‌های عصبی مصنوعی، از شبکه پرسپترون چند لایه و الگوریتم پس انتشار خطا برای

جدول ۶- معیارهای قضاوت آماری انتخاب بهترین مدل پیش‌بینی دبی ایستگاه هیدرومتری ولگوگراد با استفاده از الگوریتم ترکیبی کپک مخاطی و رگرسیون بردار پشتیبان (مرحله واسنجی)

Table 6- Statistical judgment criteria to select the best flow prediction model for Volgograd Hydrometric Station using the SVR-SMA model (calibration stage)

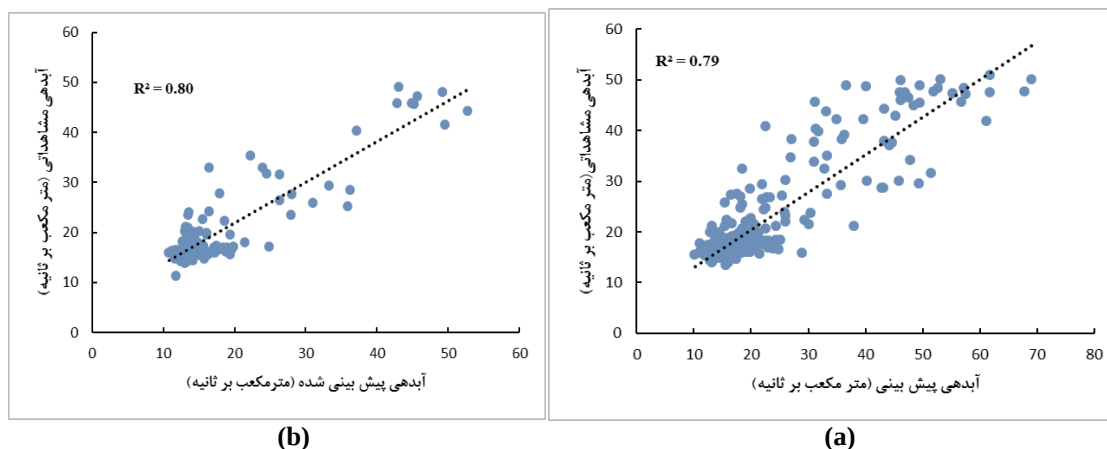
No.	Kernel type								
	RBF kernel			Poly kernel			Line kernel		
	MAE	R ²	RMSE	MAE	R ²	RMSE	MAE	R ²	RMSE
1	0.85	0.73	9.19	1.43	0.6	10.08	3.22	0.35	29.6
2	0.71	0.74	9.04	2.32	0.45	15.49	4.11	0.23	37.16
3	0.65	0.78	8.11	1.50	0.59	11.72	3.20	0.35	29.52
4	0.73	0.74	9.08	1.67	0.53	12.34	4.00	0.28	33.05
5	0.61	0.79	8.02	1.56	0.55	11.68	1.51	0.58	11.81
6	0.63	0.79	8.06	1.57	0.55	11.97	3.47	0.28	33.11
7	0.70	0.78	8.15	0.79	0.76	9.11	3.27	0.35	29.47
8	0.71	0.76	8.86	1.02	0.70	9.86	3.50	0.28	32.97

جدول ۷- معیارهای قضاوت آماری انتخاب بهترین مدل پیش‌بینی دبی ایستگاه هیدرومتری ولگوگراد با استفاده از الگوریتم ترکیبی کپک مخاطی و رگرسیون بردار پشتیبان (مرحله صحت‌سنجی)

Table 7- Statistical judgment criteria to select the best flow prediction model for Volgograd Hydrometric station using the SVR-SMA model (validation stage)

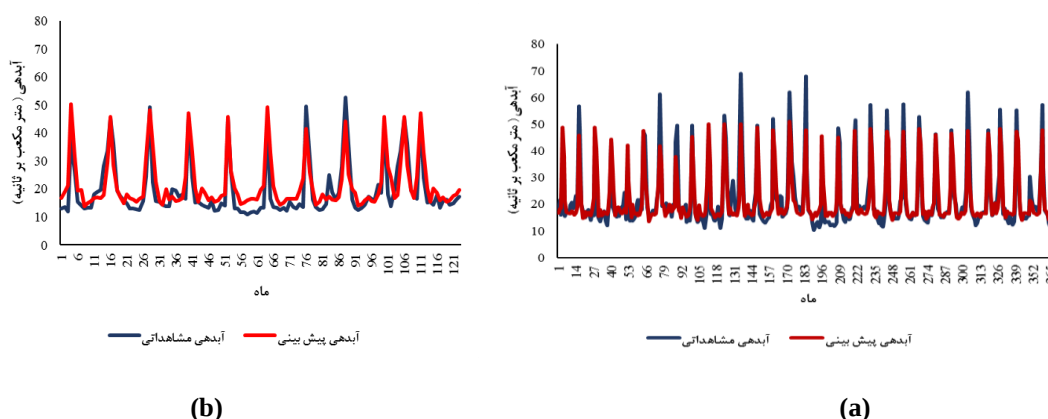
No.	Kernel type								
	RBF kernel			Poly kernel			Line kernel		
	MAE	R ²	RMSE	MAE	R ²	RMSE	MAE	R ²	RMSE
1	1.06	0.78	8.77	1.13	0.77	9.17	4.1	0.42	21.75

2	1.04	0.78	8.54	2.35	0.062	13.01	6.36	0.031	35.86
3	1.03	0.79	8.06	1.79	0.76	9.28	3.69	0.056	17.11
4	1.03	0.79	8	1.41	0.71	9.89	3.84	0.51	20.85
5	0.56	0.8	7.19	0.98	0.79	7.23	1.76	0.7	10.81
6	0.93	0.79	7.75	1.34	0.72	9.64	4.39	0.4	24.88
7	1.02	0.79	7.93	1.09	0.77	8.93	5.74	0.39	25.31
8	1.03	0.79	8.07	1.01	0.79	7.69	6.06	0.36	28.04



شکل ۵- نمودار پراکندگی نقاط داده‌های مشاهداتی و بهترین مدل پیش‌بینی در ایستگاه هیدرومتری ولگوگراد (a): مرحله واسنجی - (b) مرحله صحت‌سنجی

Fig.5- Scatter diagram of the observed and the best prediction model data points at Volgograd Hydrometric Station: a) calibration stage, b) validation stage



شکل ۶- نمودار شبیه‌سازی مقایسه‌ای جریان ماهیانه در ایستگاه هیدرومتری ولگوگراد (a): مرحله واسنجی - (b) مرحله صحت‌سنجی
Fig. 6- Comparative simulation diagram of the monthly flow in Volgograd Hydrometric Station: a) calibration stage, b) validation stage

جدول ۸. مقایسه عملکرد بهترین مدل ترکیبی کپک مخاطی و رگرسیون بردار پشتیبان با برخی دیگر از روش‌های داده‌مبنای پیش‌بینی آبدهی ایستگاه هیدرومتری ولگوگراد

Table 8- Comparison of the performance of the best SVR-SMA model with some other data-based discharge prediction methods of Volgograd Hydrometric Station

Model	calibration stage			validation stage		
	MAE	R ²	RMSE	MAE	R ²	RMSE
SVR-SMA	0.56	0.8	7.19	0.61	0.79	8.02
SVR	0.84	0.77	9.23	1.94	0.72	10.86
ANN	1.61	0.71	11.03	2.15	0.7	11.49
GA	0.71	0.78	8.37	1.85	0.73	10.15

دارد. اما دقت مدل‌های داده‌کاوی بهتر از این مدل بود که مقایسه نتایج نشان دهنده این بود که روش M5 دقت بیشتری نسبت به روش KNN بود. در مطالعه حاضر به بررسی روش‌های مختلف داده‌کاوی جهت پیش‌بینی دبی ماهانه رودخانه پرداخته شد که ترکیب الگوریتم‌های بهینه‌ساز کپک مخاطی و شبیه‌ساز رگرسیون بردار پشتیبان دارای دقت بیشتری نسبت به سایر مدل‌های استفاده شده بود.

Alinezhadi et al. (2021) جهت پیش‌بینی میانگین دبی روزانه ایستگاه هیدرومتری ملائانی بر روی رودخانه کارون، از روش‌های مختلف داده‌کاوی استفاده کردند که مدل برنامه‌ریزی بیان ژن (GEP)، نسبت به روش‌های دیگر بهترین عملکرد را داشت. نویسندگان تحقیق حاضر از ترکیب دو مدل داده‌کاوی جهت پیش‌بینی دبی ماهانه رودخانه استفاده نمودند که نتایج نشان دهنده دقت بالاتر مدل ترکیبی نسبت به سایر مدل‌ها بوده است.

پیشنهاد می‌شود در تحقیقات بعدی از مدل‌های هیدرولوژیکی مانند IHACRES جهت پیش‌بینی دبی ماهانه رودخانه ولگا استفاده و با نتایج این تحقیق مقایسه نمود. همچنین مدل برنامه‌ریزی بیان ژن (GEP) و روش‌های دیگر با عملکرد مدل ترکیبی استفاده شده در این تحقیق مورد ارزیابی قرار گیرند.

۴. بحث و نتیجه‌گیری

در این تحقیق، به بررسی عملکرد ترکیب الگوریتم‌های بهینه‌ساز کپک مخاطی و شبیه‌ساز رگرسیون بردار پشتیبان در پیش‌بینی جریان ماهانه رودخانه ولگا در حوضه آبریز دریای خزر پرداخته شد.

در این راستا، از داده‌های متوسط آبدهی ماهیانه ایستگاه هیدرومتری ولگوگراد در فاصله سال‌های ۱۹۸۰ تا ۲۰۲۰ استفاده گردیده است. بدین منظور، از الگوهای مختلف پیش‌بینی آبدهی با تأثیرپذیری از متغیرهای اقلیمی جهت مدل‌سازی استفاده گردید. در مدل‌های استفاده شده در این تحقیق، بهترین عملکرد مربوط به سناریوی تأثیر مجموع بارش، دماهای بیشینه و کمینه و نیز تابش خورشیدی ماهانه با بهره‌گیری از کرنل RBF در پیش‌بینی آبدهی جریان بوده است. همچنین نتایج این پژوهش نشان می‌دهد که استفاده از الگوریتم کپک مخاطی توانسته است سرعت و دقت نتایج الگوریتم رگرسیون بردار پشتیبان را در پیش‌بینی آبدهی ماهانه رودخانه ولگا ارتقاء دهد.

Ghorbani et al. (2016) در تحقیق خود به پیش‌بینی میزان جریان ماهانه رودخانه چهل‌چای گلستان در ایستگاه هیدرومتری ارازکوسه با استفاده از مدل هیدرولوژیکی IHACRES و روش‌های داده‌کاوی M5 و KNN استفاده کردند. نتایج این تحقیق نشان داد که مدل IHACRES توانایی خوبی جهت شبیه‌سازی جریان ماهانه رودخانه را

References:

Alinezhadi, M., Mousavi, S.F. and Hosseini, Kh., 2021. Comparison of Gene Expression Programming (GEP) and Parametric and Non-parametric Regression Methods in the Prediction of the Mean Daily Discharge of Karun River (A case Study: Mollasani Hydrometric Station). *Journal of Water and Soil Science*. 25(1), pp.43-62. Doi: 10.47176/jwss.25.1.1012. (In Persian).

Bagherpoor, M., Seyedian, S.M., Fathabadi, A.H. and Mohamadi, A., 2017. Study of Mann-Kendall Test Performance in Detecting the Series of Autocorrelation. *Iran-Watershed Management Science & Engineering*, 11(36), pp.11-22. Dor : 20.1001.1.20089554.1396.11.36.2.4. (In Persian).

Cheng, C.T., Feng, Z.K., Niu, W.J. and Liao, S.L., 2015. Heuristic methods for reservoir monthly inflow forecasting, a case study of Xinfengjiang

- Reservoir in Pearl River China. *Journal of Water*, 7(8), pp.4477-4495. Doi:10.3390/w7084477.
- Ghorbani, Kh., Sohrabian, E. and Salarijazi, M., 2016. Evaluation of hydrological and data mining models in monthly river discharge simulation and prediction (Case study: Araz-Kouseh watershed). *Journal of Water and Soil Conservation*, 23(1), pp.203-217. Doi: 10.22069/JWFST.2016.3027. (In Persian).
- Kalugin, A., 2022. Hydrological and Meteorological Variability in the Volga River Basin under Global Warming by 1.5 and 2 Degrees. *Climate*, 107(10), pp.1-23. Doi: 10.3390/cli10070107
- Karamouz, M. and Araghinejad, Sh., 2005. *Advanced Hydrology*. Amirkabir University of Technology, Tehran. (In Persian).
- Kemp, P., Sear, D., Collins, A., Naden, P. and Jones, I., 2011. The impacts of fine sediment on riverine fish. *Hydrological Processes*, 25, pp.1800–1821.
- Kisi, O., 2007. Streamflow forecasting using different artificial neural network algorithms. *Journal of Hydrologic Engineering*, 12(5), pp.532-539. [https://doi.org/10.1061/\(ASCE\)1084-0699\(2007\)12:5\(532](https://doi.org/10.1061/(ASCE)1084-0699(2007)12:5(532)
- Li, Sh., Chen, H., Wang, M., Heidari, A.A. and Mirjalili, S.A., 2020. Slime Mould Algorithm: A New Method For Stochastic Optimization. *Future Generation Computer Systems*, 111, pp.300-323. <https://doi.org/10.1016/j.future.2020.03.055>
- Middelkoop, H., 2005. Large European river system responses to global change and human activities-The Volga and Rhine rivers. Final report of the NWO-RFBR project nr. 047.014. 010.
- Pal, M. and Deswal, S., 2010. Modelling pile capacity using Gaussian process regression. *Journal of Computers and Geotechnics*, 37, pp.942-947. Doi:10.1016/j.compgeo.2010.07.012
- Saghebzan, S.M., 2020. Temporal and Spatial Flow Discharge Prediction using Integrated Artificial Intelligence and Pre and Post-processing Time Series Method. *Iranian Journal of Irrigation and Drainage*, 14(4), pp. 1137-1151. (In Persian).
- Shahinejad, B. and Dehghani, R., 2017. Evaluation and Performance of Support Vector Machine Model in Estimation of Suspended Sediment. *Irrigation and water engineering scientific research*, 29(8), pp.30-42.
- https://www.waterjournal.ir/article_74207_en.html. (In Persian).
- Vapnik, V., Golwich, E.S. and Smola, A., 1996. Support vector method for function approximation, regression estimation and signal processing, *Journal of Advances in Neural Information Processing Systems*, 9, pp.281–287.
- Young, C.C., Liu, W.C. and Wu, M.C., 2017. A physically based and machine learning hybrid approach for accurate rainfall-runoff modeling during extreme typhoon events. *Applied Soft Computing*, 53, pp.205-216. <https://doi.org/10.1016/j.asoc.2016.12.052>
- Zhong, K. F., Wen, J. N., Zheng, Y. T. and Zhi. Q. J., 2020. Monthly runoff time series prediction by variational mode decomposition and support vector machine based on quantum-behaved particle swarm optimization, *Journal of Hydrology*, 583, pp.1-12. Doi: 10.1016/j.jhydrol.2020.124627.
- Zounemat-Kermani, M., Mahdavi-Meymand, A. and Hinkelmann, R., 2021. A comprehensive survey on conventional and modern neural networks: application to river flow forecasting. *Earth Science Informatics*, 14, 893–911. <https://doi.org/10.1007/s12145-021-00599-1>