

Research Paper



Improving the hybrid IHACRES–XGBoost model for simulating daily runoff in the Ghareesu Basin



Vahid Kakapor^{1*}, Maryam Bayatikhatibi², Saeed Samadianfar³

1. Ph.D. Student in Remote Sensing, Department of Remote Sensing, Faculty of Environmental Planning and Sciences, University of TabrizIran. v.kakpor@gmail.com
 2-Professor of Geomorphology, Department of Remote Sensing, Faculty of Environmental Planning and Sciences, University of Tabriz Iran. m_bayati@tabrizu.ac.ir
 3-Associate Professor, Department of Water Engineering, Tabriz University, Tabriz, Iran. s.samadianfard@tabrizu.ac.ir

Keywords

IHACRES, Genetic Algorithm, XGBoost, Runoff Simulation, Moving Average, Machine Learning, Nash–Sutcliffe Index Ghareesu Basin Western Iran.

ABSTRACT

Introduction

Runoff is influenced by the interaction of multiple variables, including precipitation, temperature, evaporation, soil moisture, snow cover, and physical characteristics of the basin such as topography and land use. The complex and nonlinear behavior of these systems has led to the development of various models for simulation and prediction of runoff. In general, the models used in this field are divided into two main categories: conceptual models (white-box) and data-driven or machine learning-based models (black-box). Conceptual models such as IHACRES and SWAT are developed based on physical principles and hydrological concepts and often require data such as basin physiography, soil type, and evapotranspiration parameters (Fatiyan et al. 2019). However, each of these approaches has limitations. Despite their high interpretability, conceptual models usually require extensive input data and may not have the desired accuracy in predicting runoff. On the other hand, data-driven models, although capable of identifying complex and nonlinear relationships between variables and having higher prediction performance, lack physical interpretability. In response to these limitations, researchers have tended to develop hybrid models that combine the advantages of the two approaches (Fattahi et al., 2022, Niromandfard, 2018). In the meantime, the IHACRES model has attracted the attention of many researchers due to its simple structure, less need for input data, and the ability to be implemented in basins without long-term statistics. This model was first introduced by Jakeman et al. (1990) and includes two main components: a nonlinear module to determine effective precipitation and a linear module to simulate the flow response. An updated version of this model was also developed by Croke and Jakeman (2004) with the aim of reducing collinearity between parameters and improving performance under data constraints (Croke & Jakeman 2008).

Received: 2024/07/15

Accepted: 2024/09/20

Published: 2026/01/22

*Correspondin Author: Vahid Kakapor: v.kakpor@gmail.com

How to cite this article: Kakapor,Vahid; Bayatikhatibi, Maryam; Samadianfard, Saeed. (2026). Improving the hybrid IHACRES–XGBoost model for simulating daily runoff in the Ghareesu Basin. rogeomorphology, 12(45): 100 – 122.

DOI: 10.22034/hyd.2025.68164.1800



Copyright: © by the authors

Publisher: University of Tabriz

Methodology

The present study was designed and implemented in two main stages. In the first step, the semi-distributed IHACRES model was used to simulate the precipitation-runoff process and the model parameters were optimized using a genetic algorithm. Next, using the model outputs and its derived variables, an XGBoost machine learning model has been developed to increase the accuracy of daily discharge prediction. The data used in this study include daily precipitation values, minimum, maximum, and average temperatures, as well as daily river discharge during the period 1995 to 2023. These data were collected from the hydrometric stations of Pol-e-Kohne and Qorbaghestan, as well as from the Kermanshah Synoptic Meteorological Station, which are located in the Qarasu sub-basin in Kermanshah province. After reviewing and cleaning the data, the final set was divided into two parts: 70% of the data was used for training the model and the other 30% was used to evaluate its performance (test). These refined data formed the basis of all modeling steps in this study. In the first step, the IHACRES model was run. This model consists of two main parts: the first part includes the estimation of water storage using effective precipitation and approximate evaporation, which is done by combining daily precipitation and average temperature (derived from the average of minimum and maximum temperatures); and the second part includes modeling the delay of fast and slow runoff and converting it into the final river flow (Chen, T.Q. and Guestrin, 2016). To increase the accuracy of the simulation of the IHACRES model, the optimization process of its parameters was carried out with a genetic algorithm. In this algorithm, an appropriate range for each parameter was first determined based on previous experiences and trial and error. Then, by defining the objective function based on minimizing the RMSE error, an initial population of responses was randomly generated and by implementing the selection (tournament method), crossover (Blend) and mutation (polynomial mutation with bounds) operators, the evolutionary process of the algorithm was run in 100 generations to achieve the best combination of parameters. In the final stage, the XGBoost model was developed to predict daily discharge (dependent variable). In order to fully exploit the available information, a set of derived features was generated as input to the machine learning model. These features included lagged variables of one- to four-day discharges, sum and moving average of precipitation and temperature in three-day and seven-day time windows, temporal indices such as sine and cosine components of the day of the year DOY_sin and DOY_cos, season, seasonal precipitation, and the Standard Drought Index (SPI). Also, in order to create a hybrid structure, the output of the IHACRES model was also considered as one of the inputs of the XGBoost model. All of these features were normalized and used to train the XGBoost model. The XGBoost (Extreme Gradient Boosting) algorithm is one of the powerful algorithms in the field of machine learning that operates based on a set of reinforcing and sequential decision trees and improves overall performance by reducing the residual error from previous models. This algorithm, with features such as internal regularization, tree pruning, parallel processing, and missing data handling, has high accuracy in regression and classification problems and is resistant to overfitting (Chen, T.Q. and Guestrin, 2016). The training data was evaluated using 10-layer cross-validation, and the final model performance was measured with indicators such as root mean square error (RMSE) and Nash–Sutcliffe efficiency (NSE).

Results and Discussion

In this section, the performance of four models and techniques including basic IHACRES, basic IHACRES with 3-day moving average (smooth_ihacres), and IHACRES optimized with genetic algorithm (IHACRES-GA) and the hybrid IHACRES–XGBoost model were compared and evaluated. The statistical indicators used to measure the performance of the models included Nash–Sutcliffe coefficient (NSE) and root mean square error (RMSE) in the calibration and validation periods. The NSE coefficient is usually used to evaluate the performance of rainfall–runoff models. The results showed that the NSE value in the training period for the Pol-e-Kohne and Qorbaghestan hydrometric stations was 0.39 and 0.4, respectively, and in the testing period it was 0.44 and 0.47. Also, the RMSE value was reported for the Pol-e-Kohne and Qorbaghestan hydrometric stations in the training period as 29.5 and 27.1, and for the testing period as 23.11 and 22.4, respectively. These results indicate that the performance of the IHACRES model in this study, although acceptable, was generally weaker than expected, which was probably due to the low quality of the input data (rainfall and temperature) and especially the low accuracy of the flow data recorded at the measurement stations, along with short-term fluctuations and sudden peaks on a daily scale, which led to a decrease in the accuracy of the model in reconstructing runoff. Which is consistent with the study of Khatibi et al. (2025) in the Qara Su basin. To reduce the effect of short-term fluctuations caused by unsystematic factors or measurement errors, a three-day moving average filter was used as a preprocessing technique on the output of the basic IHACRES model. In the implementation of the Genetic Algorithm, effective settings including population size, number of generations, crossover rate, and mutation rate were carefully selected to establish a proper balance between convergence and population diversity. Then, the value of the objective function, which was a combination of the statistical indicators NSE (Nash-Sutcliffe efficiency

coefficient) and RMSE (root mean square error), was evaluated in each generation. The results of the implementation of the GA-IHACRES model showed that the NSE value increased significantly in both training and testing periods and at the same time the RMSE value decreased, indicating an improvement in the model's efficiency compared to the baseline (without optimization). The use of genetic algorithm to optimize IHACRES parameters has led to a 25 to 34 percent improvement in model performance at both stations and in both training and testing stages. In the final stage of this research, in order to improve the performance of the basic IHACRES model and overcome its limitations in accurately simulating daily runoff, a hybrid modeling framework based on the XGBoost algorithm was used. In this approach, the output of the IHACRES model along with a set of effective variables including precipitation data, temperature, drought index, time components, and especially lagged flow were defined as the input of the XGBoost model to increase the accuracy of runoff simulation by utilizing the power of this algorithm in discovering nonlinear and complex relationships. In the design of this hybrid model, a targeted set of features was selected based on hydrological principles and statistical analyses. Among these features, the variables Q_lag1 to Q_lag4 , which represent the normalized runoff of the past four days, reflect the role of the watershed's hydrological memory in runoff generation. Also, data such as discharge simulated by IHACRES ($Q_sim_IHACRES$), SPI index, multi-day average temperatures, and time components DOY_sin , DOY_cos , season were also included in the model to consider climatic and seasonal aspects in flow behavior. The analysis of the importance of features using Gradient Boosting Feature Importance in the XGBoost model clearly showed that Q_lag1 has the highest impact on improving runoff simulation by a large margin compared to other variables. This finding is fully consistent with the physical logic of the hydrological system, since the water conditions of the previous day usually determine the flow trend of the next day.

Conclusions

In this study, four different approaches including: the basic IHACRES model, IHACRES with three-day moving average, IHACRES optimized with genetic algorithm, and the combined IHACRES–XGBoost model were investigated and compared to simulate daily runoff in the Gharezu watershed. Despite its simple and understandable structure, the conceptual IHACRES model could not reconstruct peak discharges with sufficient accuracy. One of the reasons for this was the neglect of processes such as snowmelt runoff, which has also been discussed in previous studies by Khatibi et al. Next, in order to reduce short-term fluctuations and increase the stability of the outputs, a three-day moving average filter was used, which led to a relative improvement in the results. Following this, optimization of the model parameters was carried out using genetic algorithm, which led to a 25 to 34 percent improvement in the model performance in indicators such as NSE and RMSE. This optimization played an effective role in reducing the uncertainty caused by manual adjustments. Finally, by combining the IHACRES conceptual model and the powerful XGBoost algorithm, a hybrid modeling framework was presented. In this structure, the output of the conceptual model along with a set of hydrological and climatic variables such as delayed runoffs Q_lag1 to Q_lag4 , cumulative precipitation, drought index SPI, and time components were defined as the input of the XGBoost model. This hybrid model was able to significantly increase the accuracy of runoff modeling by utilizing the capacity of machine learning algorithms in analyzing complex and nonlinear relationships. The analysis of the importance of features showed that variables such as Q_lag1 and Q_lag2 play a key role in the reconstruction of runoff, which indicates the existence of hydrological memory in the flow behavior. The results of the hybrid model at Qorbaghestan and Pol-e-Kohne stations showed that the RMSE decreased by more than 70% and the NSE value increased from 0.40 to 0.97, confirming the high accuracy, stability, and efficiency of the proposed framework. Overall, this study showed that the combination of conceptual and data-driven approaches can compensate for the limitations of each and provide a powerful tool for runoff simulation in variable climate conditions and incomplete data. The findings are consistent with the studies of Fattahi et al. (2022) and Mohammadi et al. (2022) who emphasize the advantage of using hybrid models. Therefore, the IHACRES–XGBoost modeling framework can be proposed as an effective, accurate, and reliable solution for runoff prediction, water resources management, and flood warning in semi-arid watersheds.



توسعه مدل ترکیبی IHACRES-XGBoost برای شبیه‌سازی رواناب روزانه در حوضه قره‌سو



وحید کاکاپور^۱، مریم بیاتی خطیبی^۲، سعید صمدیان فرد^۳

۱- دانشجوی دکتری سنجش از دور، گروه سنجش از دور، دانشکده برنامه‌ریزی و علوم محیطی، دانشگاه تبریز، ایران. v.kakpor@gmail.com

۲- استاد ژئومورفولوژی، گروه سنجش از دور، دانشکده برنامه‌ریزی و علوم محیطی، دانشگاه تبریز، ایران. m_bayati@tabrizu.ac.ir

۳- دانشیار گروه مهندسی آب، دانشگاه تبریز، تبریز، ایران. s.samadianfard@tabrizu.ac.ir

چکیده

مدیریت منابع آب در حوضه‌های نیمه‌خشک با چالش‌هایی چون بارش نامنظم، تبخیر بالا و دسترسی محدود به داده‌های هیدرولوژیکی مواجه است. دقت پایین در شبیه‌سازی رواناب، برنامه‌ریزی‌های آبیاری و پیش‌بینی سیلاب را با مشکل مواجه می‌کند. از این رو، توسعه روش‌هایی که بتوانند فرآیند بارش-رواناب را با دقت بالاتر مدل‌سازی کنند، به‌ویژه در حوضه‌های با داده نامحدود و شرایط اقلیمی متغیر، ضرورت دارد. در این پژوهش، با هدف بهبود دقت شبیه‌سازی رواناب روزانه در حوضه قره‌سو واقع در استان کرمانشاه، یک چارچوب مدل‌سازی ترکیبی مبتنی بر تلفیق مدل مفهومی IHACRES و الگوریتم‌های پیشرفته یادگیری ماشین توسعه داده شد. داده‌های مورد استفاده شامل بارش، دمای حداقل، حداکثر و میانگین و دبی رودخانه از ایستگاه‌های هیدرومتری پل کهنه و قورباغستان در بازه‌ی زمانی ۱۹۹۵ تا ۲۰۲۳ می‌باشند. در مرحله نخست، مدل نیمه‌توزیعی IHACRES جهت شبیه‌سازی فرآیند بارش-رواناب پیاده‌سازی گردید و با بهره‌گیری از الگوریتم ژنتیک، پارامترهای آن بهینه‌سازی شد. به‌منظور حذف نوسانات کوتاه‌مدت، یک فیلتر میانگین متحرک سه‌روزه بر خروجی مدل اعمال شد. سپس، با استفاده از خروجی IHACRES و مجموعه‌ای از متغیرهای مشتق‌شده شامل ویژگی‌های تأخیری، آماره‌های بارش و دما، شاخص‌های زمانی و خشکسالی، یک مدل یادگیری ماشین نوع XGBoost طراحی گردید. عملکرد مدل‌ها با شاخص‌های آماری RMSE و NSE در دو دوره آموزش و آزمون مورد ارزیابی قرار گرفت. نتایج نشان داد که مدل پایه‌ی IHACRES عملکردی در حد قابل قبول داشت ($NSE \approx 0.44$)، اما با اعمال فیلتر میانگین متحرک و بهینه‌سازی پارامترها، دقت آن به‌طور متوسط تا ۳۰٪ افزایش یافت. در نهایت، مدل ترکیبی IHACRES-XGBoost با دستیابی به مقدار NSE بیش از ۰/۹۷ و RMSE کمتر از ۱۰، بالاترین دقت را ارائه کرد. این یافته‌ها نشان‌دهنده‌ی کارایی بالا و پتانسیل بالای مدل‌های ترکیبی در ارتقاء پیش‌بینی رواناب روزانه می‌باشد.

کلیدواژه‌ها

IHACRES، الگوریتم ژنتیک، XGBoost، شبیه‌سازی رواناب، میانگین متحرک، یادگیری ماشین، شاخص‌نش ساتکلیف، حوضه قره‌سو.

تاریخ دریافت: ۱۴۰۴/۰۴/۲۴
تاریخ پذیرش: ۱۴۰۴/۰۶/۲۹
تاریخ انتشار: ۱۴۰۴/۱۰/۲۲

ارجاع به این مقاله: کاکاپور، وحید؛ بیاتی خطیبی، مریم؛ صمدیان فرد، سعید؛ (۱۴۰۴). توسعه مدل ترکیبی IHACRES-XGBoost برای شبیه‌سازی رواناب روزانه در حوضه قره‌سو. هیدروژنومورفولوژی، ۱۲(۴۵): ۱۰۰-۱۲۲.

* نویسنده مسئول: وحید کاکاپور

رایانامه: v.kakpor@gmail.com

شناسه دیجیتال مقاله: 10.22034/hyd.2025.68164.1800



مقدمه

تخمین دقیق فرآیند بارش-رواناب^۱ یکی از مسائل اساسی و چالش برانگیز در علوم آب است که نقش کلیدی در مدیریت منابع آب، طراحی سازه‌های هیدرولیکی و پیش‌بینی سیلاب ایفا می‌کند. رواناب تحت تأثیر متقابل متغیرهای متعددی از جمله بارش، دما، تبخیر، رطوبت خاک، پوشش برف و ویژگی‌های فیزیکی حوضه مانند توپوگرافی و کاربری اراضی قرار دارد. رفتار پیچیده و غیرخطی این سیستم‌ها باعث شده است که مدل‌های مختلفی برای شبیه‌سازی و پیش‌بینی رواناب توسعه یابند. به‌طور کلی، مدل‌های مورد استفاده در این زمینه به دو دسته اصلی تقسیم می‌شوند: مدل‌های مفهومی^۲ و مدل‌های داده‌محور یا مبتنی بر یادگیری ماشین مدل‌های مفهومی^۳ نظیر IHACRES، SWAT^۴ بر مبنای اصول فیزیکی و مفاهیم هیدرولوژیکی توسعه یافته‌اند و اغلب نیازمند داده‌هایی چون فیزیوگرافی حوضه، نوع خاک و پارامترهای تبخیر و تعرق هستند (فتحیان و همکاران^۵، ۲۰۱۹: ۱۲۰۰). با این حال، هریک از این رویکردها دارای محدودیت‌هایی هستند. مدل‌های مفهومی علی‌رغم تفسیرپذیری بالا، معمولاً به داده‌های ورودی گسترده‌ای نیاز دارند و ممکن است در پیش‌بینی رواناب دقت مطلوبی نداشته باشند. از سوی دیگر، مدل‌های داده‌محور اگرچه قادر به شناسایی روابط پیچیده و غیرخطی میان متغیرها هستند و عملکرد پیش‌بینی بالاتری دارند، اما فاقد تفسیرپذیری فیزیکی می‌باشند. در پاسخ به این محدودیت‌ها، پژوهشگران به سمت توسعه مدل‌های ترکیبی^۶ گرایش یافته‌اند که در آن، مزایای دو رویکرد فوق با یکدیگر تلفیق می‌گردد (نیرومندفرد^۷، ۲۰۱۸: ۱۰۳؛ فتاحی^۸ و همکاران، ۲۰۲۲: ۳۶۰). در این میان، مدل IHACRES به دلیل ساختار ساده، نیاز کمتر به داده‌های ورودی، و قابلیت اجرا در حوضه‌های فاقد آمار بلندمدت، توجه بسیاری از محققان را به خود جلب کرده است. این مدل نخستین بار توسط جکمن و همکاران (۱۹۹۰) معرفی گردید و شامل دو مؤلفه اصلی است: یک ماژول غیرخطی برای تعیین بارش مؤثر و یک ماژول خطی برای شبیه‌سازی پاسخ جریان. نسخه به‌روزرشده این مدل نیز توسط کروک و جکمن (۲۰۰۴) با هدف کاهش هم‌خطی میان پارامترها و ارتقاء کارایی در شرایط محدودیت داده توسعه یافت (کروک و جکمن^۹، ۲۰۰۸: ۴۱). آقابیگی^{۱۰} و همکاران (۲۰۱۹: ۱۷۸) با استفاده از مدل LARS-WG و داده‌های HadCM3 تغییرات اقلیمی ۲۰۱۱ تا ۲۰۳۰ در استان اردبیل را شبیه‌سازی کردند و کاهش ۳/۶۸٪ در بارندگی، افزایش دمای حداقل ۱۶/۴۸٪ و دمای حداکثر ۵/۳۹٪ و کاهش دبی متوسط رواناب حدود ۱۶٪ را نشان دادند. در پژوهش حاضر، مدل IHACRES برای ۸ ایستگاه هیدرومتری جنوب و جنوب‌غرب اردبیل به کار گرفته شد. نتایج مشابه کاهش بارندگی و دبی متوسط و افزایش تعداد وقایع دبی اوج را نشان داد که بیشترین تغییر مربوط به ایستگاه یامچی با دبی متوسط ۲/۰۹ مترمکعب بر ثانیه و ۱۶ روز دبی اوج بالای ۶ مترمکعب بر ثانیه است. با وجود این مزایا، مطالعاتی چون ناظری تاهرودی^{۱۱} (۲۰۲۳: ۱۳۴) و ابوشاندی و مرکل^{۱۲} (۲۰۲۳: ۲۳۹۱) نشان داده‌اند که عملکرد IHACRES در اقلیم‌های خشک، به‌ویژه در نبود برخی داده‌های ورودی، نسبت به مدل‌های جایگزین ضعیف‌تر بوده است. به‌عنوان نمونه اسماعیلی و همکاران (۲۰۲۱: ۷۹)، گزارش کردند که در اقلیم خشک، دقت IHACRES نسبت به SWAT کمتر است، در حالی که در اقلیم نیمه‌مرطوب عملکرد قابل‌قبولی از خود نشان می‌دهد. این نتایج، بیانگر آن است که تکیه صرف بر مدل‌های مفهومی در مواجهه با اقلیم‌های خاص می‌تواند منجر به افت دقت در شبیه‌سازی شود. در مطالعه‌ای دیگر، نیرومند فرد و همکاران^۲ (۲۰۱۸: ۱۰۳) با

1-Rainfall-Runoff

2- white-box

3- black-box

4- Soil & Water Assessment Tool

5- Fathian et al.,

6- Hybrid Models

7- Niromandfard et al .,

8- Fattahi et al.,

9- Croke & Jakeman

10-Aghabeigi et al .,

11- Nazeri Tahroudi et al .,

12- Abushandi et al .,

به کارگیری مدل IHACRES برای بررسی اثر تغییر اقلیم در شمال شرق ایران، ضریب نش-ساتکلیف (NS) را در مراحل کالیبراسیون و اعتبارسنجی به ترتیب ۰/۵ و ۰/۴۹ گزارش کردند. در مقابل، طی سال‌های اخیر، مدل‌های داده‌محور نظیر شبکه‌های عصبی مصنوعی (ANN)، روش گروهی پردازش داده‌ها (GMDH)، جنگل تصادفی (RF) و XGBoost به دلیل توانایی بالا در استخراج روابط غیرخطی و پیچیده میان متغیرهای اقلیمی و رواناب، مورد توجه گسترده قرار گرفته‌اند (صالحپور لاقانی و همکاران ۲۰۲۰: ۱۲۶۵؛ احمدی و همکاران ۲۰۱۹: ۶۵). در این راستا، الگوریتم XGBoost به عنوان نسخه‌ای بهینه‌شده از درخت‌های تصمیم‌گیری، به دلیل سرعت بالای اجرا، انعطاف‌پذیری و توانایی مقابله با بیش‌برازش، به طور گسترده مورد استفاده قرار گرفته است (چن و گاسترین ۲۰۱۶: ۱۷۸). در پاسخ به چالش‌های مطرح‌شده، مدل‌های ترکیبی به عنوان رویکردی مؤثر برای بهبود دقت و کارایی شبیه‌سازی رواناب معرفی شده‌اند. در یک مطالعه، مدل نیمه‌مفهومی IHACRES با استفاده از محیط R و بسته Hydromad برای شبیه‌سازی جریان رودخانه شور قائن به کار گرفته شد. عملکرد مدل در بخش خطی با دو روش آماری ARMAX و EXPUH طی دوره ۱۳۷۷ تا ۱۳۹۸ مورد ارزیابی قرار گرفت و شاخص‌های عملکرد NS و KGE برای سنجش دقت مدل به کار گرفته شدند. نتایج نشان داد که روش EXPUH در مرحله واسنجی دقت بالاتری ارائه می‌دهد، در حالی که در مرحله صحت‌سنجی، روش ARMAX توانایی بهتری در برآورد دبی اوج و زمان وقوع آن دارد. (محمد فولادی نصرآبادی و همکاران ۲۰۲۳: ۱۷). در پژوهش‌های انجام‌شده بر حوضه آبریز رودخانه گاماسیاب، ترکیب مدل‌های هوش مصنوعی شامل ANN (MLP)، RBF و LSTM با مدل نیمه‌مفهومی IHACRES به کار گرفته شد. داده‌های مورد استفاده شامل دبی جریان، بارش و دمای روزانه طی ۳۱ سال بوده و عملکرد مدل‌ها با شاخص‌های NSE، RMSE و ضریب همبستگی (R) ارزیابی شد. نتایج نشان داد مدل LSTM عملکرد برتری نسبت به سایر مدل‌ها دارد و مدل‌های هوش مصنوعی به ویژه در پیش‌بینی دبی اوج، دقت بالاتری نسبت به IHACRES ارائه می‌کنند. (صادق مومنه ۲۰۲۲: ۱۳۴). مطالعات مقایسه‌ای نشان داده‌اند که مدل‌های هوش مصنوعی و ترکیبی در شبیه‌سازی رواناب عملکرد بهتری نسبت به مدل‌های سنتی دارند. برای نمونه، مقایسه سه مدل SWAT، IHACRES و ANN در غرب تهران نشان داد که مدل ANN در تمامی مقیاس‌های زمانی دقت بالاتری نسبت به دو مدل دیگر ارائه می‌دهد (احمدی و همکاران ۲۰۱۹: ۶۵). به طور مشابه، پژوهش‌های انجام‌شده در اندونزی و سوئیس نشان دادند که مدل‌های ترکیبی مبتنی بر هوش مصنوعی در اقلیم‌های متنوع عملکرد بهتری نسبت به مدل‌های کلاسیک HBV و NRCA دارند (محمدی و همکاران ۲۰۲۲: ۱۲۰). همچنین، در مطالعه‌ای بر روی ۱۹ زیرحوضه در شمال ایران، ترکیب مدل IHACRES با مدل‌های داده‌محور موجب بهبود معنادار کارایی مدل‌ها بر اساس شاخص NSE-ساتکلیف شد؛ در شش زیرحوضه، مدل ترکیبی عملکرد مناسبی ارائه داد، در حالی که IHACRES به تنهایی نتایج ضعیف‌تری داشت. (فتاحی^{۱۱} و همکاران، ۲۰۲۲: ۳۶۰). یک رویکرد چندمرحله‌ای ترکیبی مبتنی بر یادگیری عمیق (DL) با استفاده از انتخاب ویژگی عمیق (DFS) برای بهبود پیش‌بینی بلندمدت رواناب فصلی ارائه شده است (آسف اقبال و همکاران ۲۰۲۵: ۴۰۹). در این مدل، ترکیب BiLSTM-BiGRU همراه با کوپولاهای دومتغیره مبتنی بر زنجیره مارکوف (MCMC-BC) به کار گرفته شد. متغیرهای پیش‌بین بهینه از سه دسته شامل متغیرهای هواشناسی محلی، شاخص‌های جوی مقیاس بزرگ و شاخص‌های فعالیت خورشیدی با استفاده از الگوریتم ACO و روش‌های انتخاب ویژگی تعیین شدند. نتایج نشان داد که این مدل ترکیبی نسبت به سایر روش‌های یادگیری ماشین مانند MLP، XGBoost، SVM و RF عملکرد بهتری دارد و قابلیت بالایی در پیش‌بینی فصلی رواناب و پشتیبانی از تصمیم‌گیری‌های زیست‌محیطی ارائه می‌کند.

1- Fathian et al.,

2- Iqbal et al.,

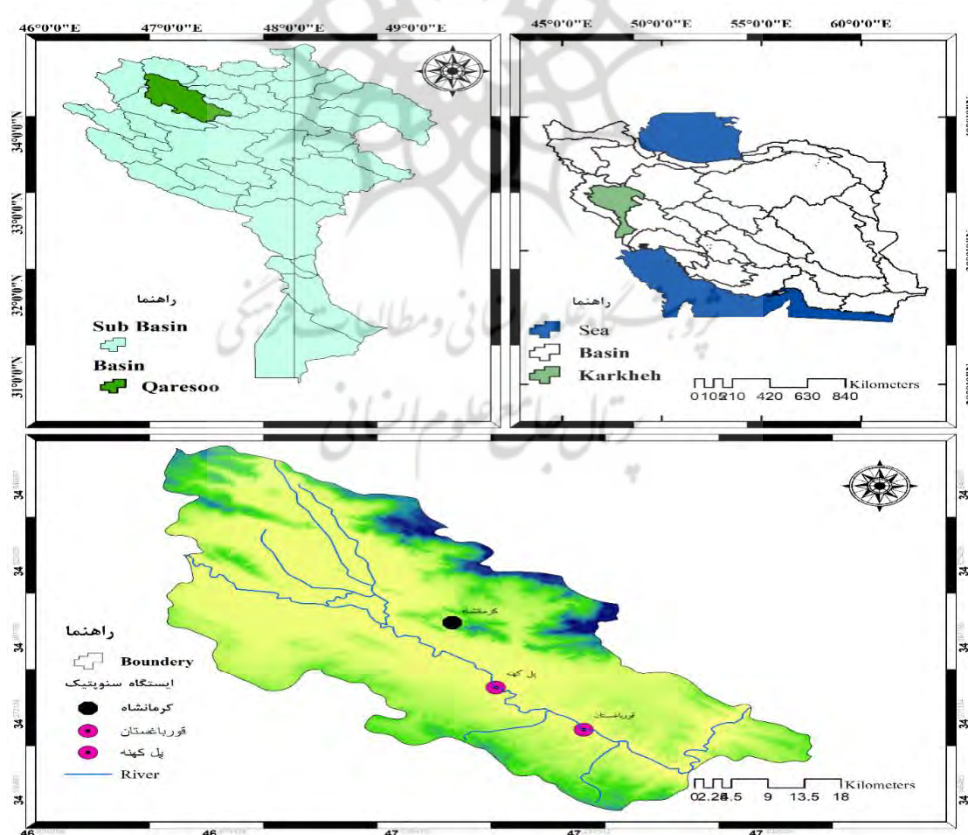
3- Ismail et al.,

مطالعات اخیر نشان می‌دهند که مدل‌های یادگیری ماشین (ML) با بهره‌گیری از ورودی‌هایی مانند بارش و دما می‌توانند دقت پیش‌بینی دبی رواناب را به‌طور قابل‌توجهی افزایش دهند (اسماعیل و امینو^۳، ۲۰۲۵: ۱۰). در این مطالعات، ایستایی داده‌ها با آزمون Augmented Dickey-Fuller و نرمال‌سازی تضمین شد و تحلیل همبستگی ترکیب بهینه ورودی‌ها را مشخص کرد. نتایج نشان داد مدل Gaussian Process Regression (GPR-M3) بهترین عملکرد را ارائه می‌دهد و استفاده از الگوریتم‌های ترکیبی (ensemble) موجب تقویت عملکرد کلی مدل‌ها می‌شود. این یافته‌ها اهمیت پیش‌پردازش داده‌ها، انتخاب مدل مناسب و بهره‌گیری از روش‌های هوش مصنوعی را در بهبود پیش‌بینی رواناب و تحلیل تعاملات غیرخطی برجسته می‌کنند. بر همین اساس، در این پژوهش، به منظور شبیه‌سازی جریان روزانه رودخانه‌های پل‌کهنه و قورباغستان در حوضه قره‌سوه کرمانشاه، از مدل نیمه‌توزیعی IHACRES استفاده شد. پس از اجرای مدل پایه، پارامترها با کمک الگوریتم ژنتیک (GA) بهینه‌سازی شدند که منجر به بهبود شاخص‌های عملکرد RMSE و NSE گردید. سپس یک مدل داده‌محور مبتنی بر XGBoost با ورودی‌هایی شامل بارش تجمعی، دمای میانگین متحرک، شاخص SPI و سایر متغیرها توسعه یافت. در نهایت، خروجی مدل IHACRES نیز به‌عنوان یکی از ورودی‌های XGBoost لحاظ شد تا یک مدل ترکیبی (Hybrid) شکل گیرد و دقت شبیه‌سازی رواناب افزایش یابد.

مواد و روش

منطقه مورد مطالعه

زیرحوضه قره‌سوه در بخش شمال‌غربی حوضه آبریز کرخه و در ناحیه غربی ایران قرار دارد. این منطقه حدود ۵۳۵۴ کیلومتر مربع وسعت دارد و یکی از زیرحوضه‌های آن کرمانشاه می‌باشد که مساحت آن ۱۹۲۰ کیلومتر مربع می‌باشد و ارتفاعات آن از ۱۲۳۹ متر در پایین‌ترین نقطه تا ۳۳۷۸ متر در بلندترین قسمت متغیر است. (شکل ۱).



شکل (۱): موقعیت جغرافیایی حوضه کرخه و زیر حوضه قره سو

Figure (1): Geographical Location of the Karkheh Basin and the Qarasu Sub-Basin

بارش سالانه در این حوضه در بازه‌ای بین ۳۰۰ تا ۸۰۰ میلی‌متر قرار دارد، اما طی دهه‌های گذشته با کاهش محسوسی روبه‌رو بوده است. قره‌سو از اتصال چند شاخه رودخانه‌ای کوچک شکل می‌گیرد، با این حال سرچشمه‌ی اصلی آن را می‌توان چشمه‌ی روانسر دانست. جریان خروجی از این سراب پس از پیمودن حدود ۲۰ کیلومتر، در حوالی روستای شاه‌گدار با چند جریان دیگر یکی شده و رودخانه‌ی قره‌سو را تشکیل می‌دهد. این رودخانه از سرچشمه‌ی روانسر که در فاصله‌ی تقریبی ۵۰ کیلومتری شمال‌غرب کرمانشاه قرار دارد، آغاز می‌شود و مسیر آن به طور کلی از شمال‌غرب به جنوب‌شرق است (بیاتی خطیبی^۱ و همکاران، ۱۰۴:۲۰۲۵). در نزدیکی ۱۵ کیلومتری شهر کرمانشاه، جریان‌های رود راز‌آور و انشعابات آن وارد قره‌سو می‌شوند. با ورود به ناحیه‌ی دشتی، رود مرگ نیز در محدوده‌ی روستای قزانچی به این رودخانه می‌پیوندد. قره‌سو با شیبی ملایم از درون شهر کرمانشاه عبور می‌کند (نقشه ۱) و در این بخش نیز دو رود دیگر با نام‌های چمبشیر و آبشوران به آن ملحق می‌شوند. در ادامه‌ی مسیر و پس از گذر از کرمانشاه، قره‌سو در نزدیکی منطقه‌ی فرمان به رودخانه‌ی گاماسیاب می‌ریزد و در نتیجه، رود سیمره شکل می‌گیرد. این رودخانه در مسیر خود به دریاچه‌ی سد کرخه وارد شده و از آن‌جا به دو شاخه تقسیم می‌شود: یکی به سمت تالاب هورالعظیم در مرز ایران و عراق روان می‌شود و شاخه‌ی دیگر از طریق رودخانه‌ی کارون به آب‌های خلیج فارس می‌رسد (بیاتی خطیبی^۱ و همکاران، ۲۰۲۵).

پژوهش حاضر در دو مرحله‌ی اصلی طراحی و اجرا شده است. در گام نخست، از مدل نیمه‌توزیعی IHACRES برای شبیه‌سازی فرآیند بارش-رواناب استفاده شده و با به‌کارگیری الگوریتم ژنتیک، پارامترهای مدل بهینه‌سازی شده‌اند. در ادامه، با بهره‌گیری از خروجی‌های مدل و متغیرهای مشتق‌شده از آن، یک مدل یادگیری ماشین از نوع XGBoost توسعه یافته است تا دقت پیش‌بینی دبی روزانه افزایش یابد. شکل (۲) مراحل انجام پژوهش را نشان می‌دهد. داده‌های مورد استفاده شامل مقادیر روزانه بارش، دمای حداقل، دمای حداکثر، دمای میانگین و جریان دبی رودخانه طی بازه زمانی ۱۹۹۵ تا ۲۰۲۳ بوده‌اند. این داده‌ها از ایستگاه‌های هیدرومتری پل کهنه و قورباغستان و همچنین ایستگاه هواشناسی سینوپتیک کرمانشاه واقع در زیرحوضه قره‌سو جمع‌آوری شدند. پس از بررسی و پاک‌سازی، مجموعه داده‌ها به ۷۰٪ برای آموزش مدل و ۳۰٪ برای آزمون و ارزیابی عملکرد تقسیم شد. این داده‌های پالایش‌شده پایه و اساس تمامی مراحل مدل‌سازی در این پژوهش را تشکیل داده‌اند.

۱. مدل‌سازی بارش-رواناب با IHACRES

ر گام نخست، مدل IHACRES، که یک مدل مفهومی و ساده‌سازی‌شده برای شبیه‌سازی پاسخ هیدرولوژیکی حوضه به بارش است، اجرا شد. این مدل شامل دو بخش اصلی می‌باشد: (چن و گاسترین^۲، ۲۰۱۶).

-برآورد ذخیره آب: در این بخش، ذخیره آب حوضه با استفاده از بارش مؤثر و تبخیر تقریبی برآورد می‌شود. بارش مؤثر از ترکیب بارش روزانه و دمای میانگین (حاصل از میانگین دمای حداقل و حداکثر) محاسبه می‌شود.

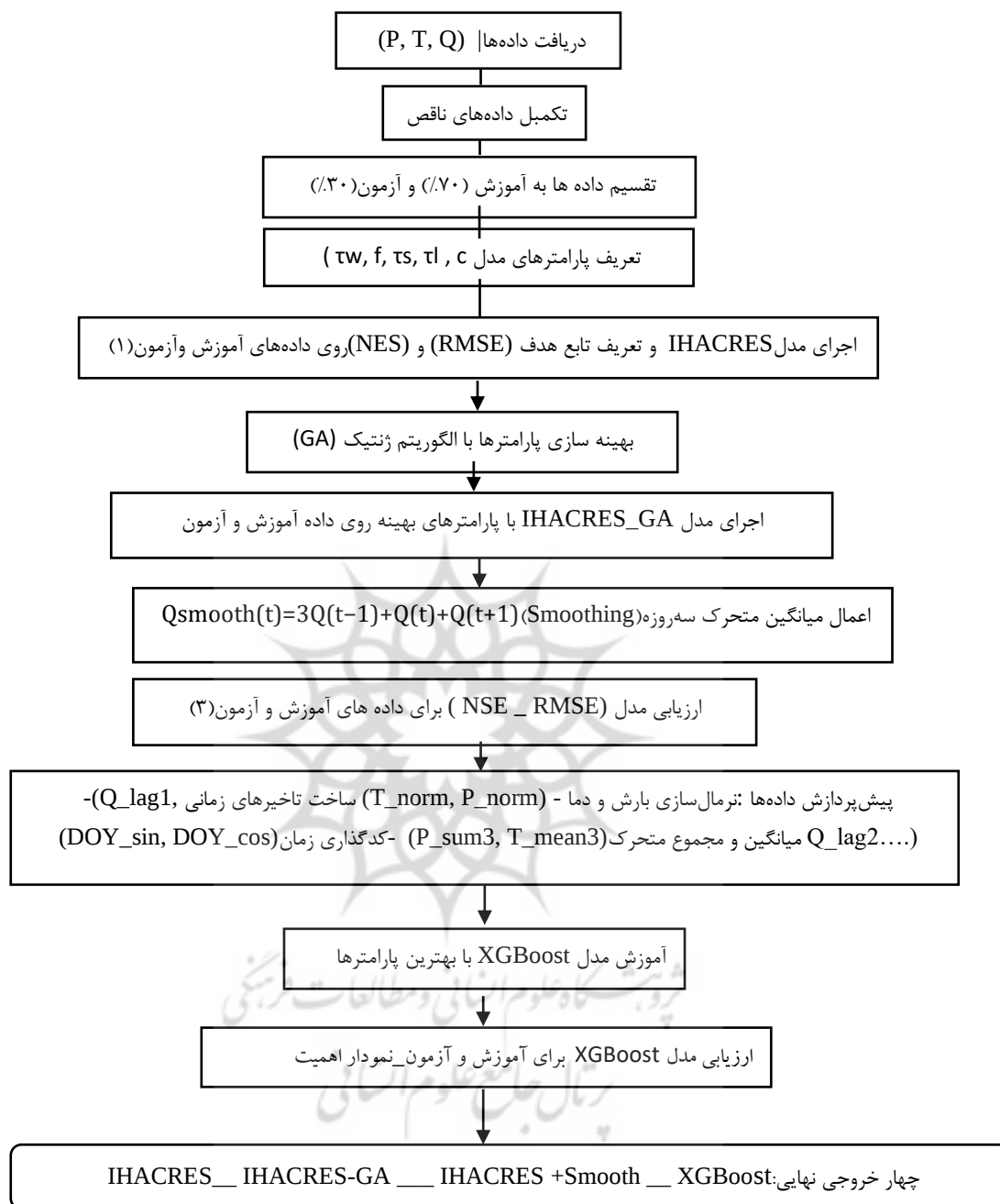
-مدلسازی رواناب: این بخش شامل مدلسازی تأخیر رواناب سریع و کند و تبدیل آن به جریان نهایی رودخانه است.

مدل IHACRES از پنج پارامتر اصلی استفاده می‌کند: زمان پاسخ اولیه رطوبت خاک (τ_w)، ضریب تبخیر (f)، ضرایب تأخیر رواناب سریع و کند τ_s و τ_l و ضریب مقیاس‌دهی جریان خروجی (c). محدوده مقادیر هر پارامتر در جدول ۱ ارائه شده است.

1- Bayati khatibi M et al

1- Chen, T.Q. and Guestrin

پیاپیاده‌سازی کلیه معادلات و فرآیندهای مدل‌سازی در محیط Python و با زبان برنامه‌نویسی پایتون انجام شده است.



شکل (۲): فلوچارت مراحل تحقیق

Figure (2): Research steps flowchart

۱-۱ برآورد ذخیره آب:

رطوبت مؤثر: $S(t)$ به صورت تابعی از بارش، دمای متوسط و پارامترهای تبخیر و تخلیه خاک تعریف می‌شود که در آن w^T : ثابت زمانی ذخیره‌ی رطوبت خاک، f : ضریب تبخیر وابسته به دما و $P(t)$: بارش روزانه

$$T_{mean}(t) = (T_{min}(t) + T_{max}(t)) / 2 \quad (1)$$

$$evap(t) = f \times T_{mean}(t) \quad (2)$$

$$S(t) = \max \left[S(t-1) + P(t) - evap(t) - \frac{S(t-1)}{w^T}, 0 \right] \quad (3)$$

۱-۲ مدل‌سازی رواناب:

پاسخ حوضه به رطوبت مؤثر: که در آن τ_s : ثابت زمانی جریان مولفه سریع، τ_l : ثابت زمانی برای جریان کند، c : ضریب تبدیل رواناب مؤثر به دبی Q_{sim} دبی شبیه‌سازی شده است.

$$Q_q(t) = Q_q(t-1) + (S(t) - Q_q(t-1)) / \tau_s \quad (4)$$

$$Q_s(t) = Q_s(t-1) + (S(t) - Q_s(t-1)) / \tau_l \quad (5)$$

$$Q_{sim}(t) = c \cdot (0.6 \cdot Q_q(t) + 0.4 \cdot Q_s(t)) \quad (6)$$

۲. بهینه‌سازی پارامترهای مدل IHACRES با الگوریتم ژنتیک (GA)

برای افزایش دقت شبیه‌سازی مدل IHACRES، پارامترهای آن با استفاده از الگوریتم ژنتیک بهینه‌سازی شدند (فرشاد احمدی^۱ و همکاران، ۲۰۱۶). در این روش، ابتدا بازه مناسب برای هر پارامتر بر اساس تجربیات پیشین و آزمون و خطا تعیین گردید. سپس با تعریف تابع هدف مبتنی بر کمینه‌سازی خطای RMSE، جمعیتی اولیه از پاسخ‌ها به صورت تصادفی ایجاد شد. روند تکاملی الگوریتم با اعمال عملگرهای انتخاب (روش تورنمنت)، تقاطع (Blend) و جهش (جهش چندجمله‌ای با حدود) در ۱۰۰ نسل اجرا شد تا بهترین ترکیب پارامترها تعیین شود (جدول ۲). خروجی این مرحله مجموعه‌ای از پارامترهای بهینه بود که منجر به تولید جریان شبیه‌سازی شده با دقت بالاتر گردید.

$$RMSE = \sqrt{(1/n) \cdot \sum (Q_{obs}(i) - Q_{sim}(i))^2} \quad (7)$$

در این رابطه، Q_{obs} و Q_{sim} به ترتیب دبی مشاهده شده و شبیه‌سازی شده در گام زمانی i هستند و n تعداد کل داده‌هاست. محدوده‌ی جست‌وجو برای پارامترهای مدل به شرح زیر است:

جدول (۱): محدوده‌ی جست‌وجو برای پارامترهای مدل IHACRES

Table (1): Search range for IHACRES model parameters

پارامتر	بازه	توضیحات
w^T	۱ تا ۱۰۰	زمان پاسخ ذخیره‌ی رطوبتی خاک
f	۰ تا ۱	ضریب تبخیر وابسته به دما
τ_s	۱ تا ۵۰	زمان پاسخ جریان سریع
τ_l	۱ تا ۱۰۰	زمان پاسخ جریان کند
c	۰/۰۱ تا ۵	ضریب تبدیل رواناب مؤثر به دبی

۴. ارزیابی عملکرد مدل

مدل در دو مرحله آموزش (با استفاده از ۷۰ درصد داده‌ها) و آزمون (با استفاده از ۳۰ درصد باقی‌مانده) مورد ارزیابی قرار گرفت. برای سنجش عملکرد مدل، از شاخص‌های آماری شامل ریشه میانگین مربعات خطا (RMSE) و ضریب کارایی نش-ساتکلیف (NSE) استفاده شد. (خرسندی کوهستانی^۱ و همکاران، ۲۰۲۵: ۶۰).

$$NSE = 1 - (\sum (Q_{obs}(i) - Q_{sim}(i))^2 / \sum (Q_{obs}(i) - Q_{obs})^2) \quad (۸)$$

$$RMSE = \sqrt{(1/n) \cdot \sum (Q_{obs}(i) - Q_{sim}(i))^2} \quad (۹)$$

۵. فیلتر میانگین متحرک سه روزه بر روی مدل IHACRES پایه

به منظور حذف نوسانات کوتاه‌مدت از سری زمانی دبی، از فیلتر میانگین متحرک سه‌روزه استفاده شد. پس از پیش‌پردازش داده‌ها، مدل در دو مرحله آموزش (با ۷۰ درصد داده‌ها) و آزمون (با ۳۰ درصد داده‌ها) ارزیابی گردید (پراکش و همکاران^۲، ۲۰۲۵: ۴۲۹).

$$Q_{smooth}(t) = (Q(t-1) + Q(t) + Q(t+1)) / 3 \quad (۱۰)$$

۶. مدل‌سازی ترکیبی با XGBoost

در مرحله نهایی، مدل XGBoost به منظور پیش‌بینی دبی روزانه (متغیر وابسته) توسعه داده شد (یانگیو و همکاران^۱، ۲۰۲۵:۱۰۲)، به منظور بهره‌برداری کامل از اطلاعات موجود، مجموعه‌ای از ویژگی‌های مشتق شده به عنوان ورودی مدل یادگیری ماشین تولید گردید. این ویژگی‌ها شامل متغیرهای تأخیری دبی‌های یک تا چهار روزه، مجموع و میانگین متحرک بارش و دما در پنجره‌های زمانی سه‌روزه و هفت‌روزه، شاخص‌های زمانی نظیر مؤلفه‌های سینوسی و کسینوسی روز سال DOY_sin و DOY_cos، فصل، بارش فصلی و همچنین

شاخص خشکسالی استاندارد (SPI) بودند همچنین، به منظور ایجاد ساختاری ترکیبی (Hybrid)، خروجی مدل IHACRES نیز به عنوان یکی از ورودی‌های مدل XGBoost در نظر گرفته شد (جدول ۲).

کلیه این ویژگی‌ها نرمال‌سازی شده و جهت آموزش به مدل XGBoost مورد استفاده قرار گرفتند. الگوریتم XGBoost (Gradient Boosting) یکی از الگوریتم‌های قدرتمند حوزه یادگیری ماشین است که بر پایه‌ی مجموعه‌ای از درخت‌های تصمیم‌گیری تقویتی و ترتیبی عمل کرده و با کاهش خطای باقیمانده از مدل‌های پیشین، عملکرد کلی را بهبود می‌بخشد. این الگوریتم با برخورداری از قابلیت‌هایی نظیر منظم‌سازی داخلی، هرس درخت‌ها، پردازش موازی و مدیریت داده‌های گمشده، دقت بالایی در مسائل رگرسیون و طبقه‌بندی داشته و در برابر بیش‌برازش مقاوم است (پراکش و همکاران^۱، ۲۰۲۵:۴۲۹). داده‌های آموزش با استفاده از اعتبارسنجی متقابل ۱۰-لایه‌ای ارزیابی شده و عملکرد نهایی مدل با شاخص‌هایی همچون ریشه میانگین مربعات خطا (RMSE) و ضریب کارایی نش-ساتکلیف (NSE) مورد سنجش قرار گرفت. تنظیمات نهایی مدل XGBoost در جدول (۳) ارائه شده است.

جدول (۲): ویژگی‌های ورودی (متغیرهای مستقل) مدل ترکیبی IHACRES-XGBoost

Table (2): Input characteristics (independent variables) of the mixed model IHACRES-XGBoost

مؤلفه	ویژگی‌ها	توضیح	مشتق شده از
رواناب تأخیری	Q_lag4, Q_lag3, Q_lag2, Q_lag1	رواناب نرمال شده در روزهای قبل (۱ تا ۴ روز قبل)	رواناب اولیه
رواناب مدل	Q_sim_IHACRES (m ³ /s)	دبی شبیه‌سازی شده توسط مدل IHACRES	خروجی مدل IHACRES
بارش	P_sum3_norm, P_norm P_sum7_norm, seasonal_precip	بارش نرمال شده روزانه و تجمعی (۳ روز، ۷ روز، و بارش فصلی)	بارش اولیه / فصل
شاخص خشکسالی	SPI	شاخص ساده شده خشکسالی بر اساس میانگین متحرک بارش ۳۰ روزه	بارش نرمال شده
دما	T_mean5, T_mean3_norm T_norm(c),	دمای نرمال شده روز جاری و میانگین دمای ۳ و ۵ روز گذشته	دمای اولیه
زمان / فصل	season, DOY_cos, DOY_sin	مؤلفه‌های سینوسی، کسینوسی روز سال و عدد فصل (۰ تا ۳) برای مدل‌سازی زمانی-فصلی	روز ژولی (DOY)

جدول (۳): تنظیمات پارامترهای مدل XGBoost به منظور شبیه سازی دبی

Table (3): Model settings XGBoost

پارامتر	مقدار
تعداد درختها (n_estimators)	۲۰۰
نرخ یادگیری (learning_rate)	۰/۰۵
منظم سازی (L1 (reg_alpha)	۱
منظم سازی (L2 (reg_lambda)	۱
روش ارزیابی	اعتبارسنجی متقابل ۱۰-لایه ای (10-fold CV)
هدف مدل	پیش بینی دبی نرمال شده بر پایه ویژگی های نرمال شده

نتایج و بحث

در این بخش، عملکرد چهار مدل و تکنیک شامل IHACRES پایه، IHACRES پایه با اعمال میانگین متحرک ۳ روزه (smooth _ihacres) و IHACRES بهینه شده با الگوریتم ژنتیک (IHACRES-GA) و مدل ترکیبی IHACRES-XGBoost مورد مقایسه و ارزیابی قرار گرفت. شاخص های آماری مورد استفاده برای سنجش عملکرد مدل ها شامل ضریب نش-ساتکلیف (NSE) و ریشه میانگین مربعات خطا (RMSE) در دوره های کالیبراسیون و صحت سنجی بودند عملاً از ضریب NSE برای ارزیابی عملکرد مدل های بارش-رواناب استفاده می شود. اگر مقدار این ضریب کمتر از ۰/۳۰ باشد، عملکرد مدل "غیرقابل قبول" محسوب می شود. مقادیر بین ۰/۳۰ تا ۰/۵۰ "قابل قبول"، بین ۰/۵۰ تا ۰/۷۵ "خوب"، و بیش از ۰/۷۵ "بسیار خوب" در نظر گرفته می شوند. در گام نخست، مدل IHACRES با استفاده از مقادیر اولیه ی پارامترها اجرا شد. مقادیر اولیه این پارامترها بر اساس مطالعات پیشین و آزمون و خطا تعیین گردید (جدول ۴).

جدول (۴): مقادیر پارامترهای اولیه مدل پایه IHACRES برای اجرا

Table (4): Initial parameter values of the IHACRES base model for execution

پارامتر	بازه	توضیحات
w^r	۵۰	زمان پاسخ دخیره رطوبتی خاک
f	۰/۲	ضریب تبخیر وابسته به دما
τ_s	۱۰	زمان پاسخ جریان سریع
τ_l	۷۰	زمان پاسخ جریان کند
c	۰/۶	ضریب تبدیل رواناب مؤثر به دبی

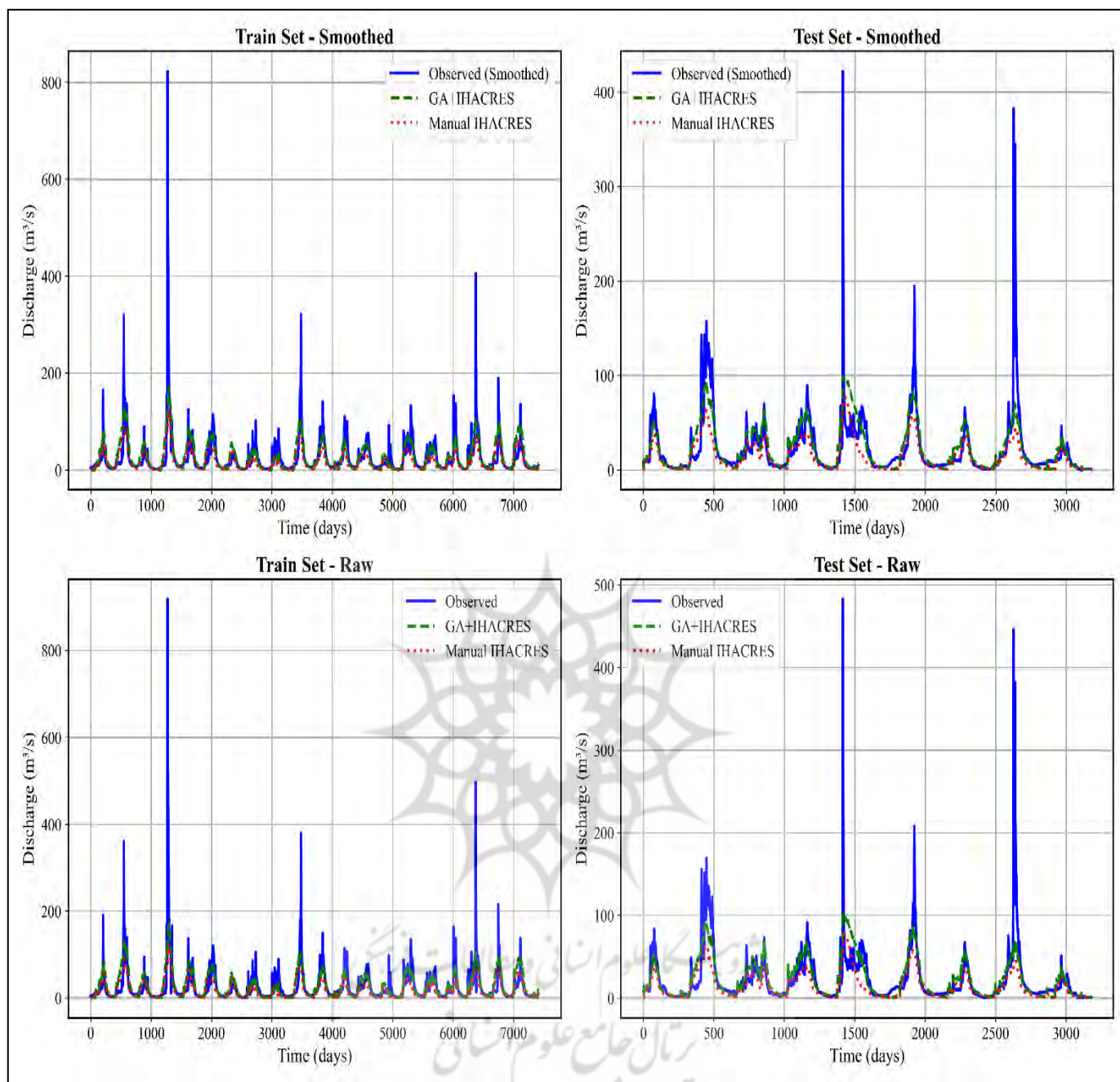
بر اساس مقادیر ارائه شده در جدول (۴)، مدل IHACRES با داده‌های ورودی روزانه اجرا شد و نتایج اولیه‌ی شبیه‌سازی رواناب در ایستگاه‌های هیدرومتری مورد مطالعه به دست آمد. این نتایج به‌عنوان مبنایی برای مقایسه با رویکردهای تکمیلی شامل IHACRES-Smooth، IHACRES-GA و مدل ترکیبی IHACRES-XGBoost در مراحل بعدی پژوهش قرار گرفت. مدل اولیه با استفاده از آزمون و خطا اجرا شد (شکل‌های ۳ و ۴). مقدار NSE در دوره آموزش برای ایستگاه‌های هیدرومتری پل کهنه و قورباغستان به ترتیب ۰/۳۹ و ۰/۴۰ و در دوره آزمون ۰/۴۴ و ۰/۴۷ به دست آمد. همچنین مقادیر RMSE در دوره آموزش به ترتیب ۲۹/۵ و ۲۷/۱ و در دوره آزمون ۲۳/۱۱ و ۲۲/۴ گزارش شد.

این نتایج بیانگر آن است عملکرد مدل IHACRES در این مطالعه، اگرچه قابل قبول بود، اما به‌طور کلی ضعیف‌تر از حد انتظار ارزیابی شد. کاهش دقت مدل احتمالاً ناشی از کیفیت پایین داده‌های ورودی (بارش و دما)، دقت محدود داده‌های جریان ثبت شده و نوسانات کوتاه‌مدت و پیک‌های ناگهانی در مقیاس روزانه بوده است. این یافته‌ها با نتایج خطیبی و همکاران (۲۰۲۵) و خرسندی و همکاران (۲۰۲۵) همخوانی دارد. با توجه به هدف اصلی پژوهش که ارتقاء دقت شبیه‌سازی رواناب روزانه بود، در مراحل بعدی از چندین رویکرد و مدل مکمل برای بهبود عملکرد مدل استفاده شد.

برای کاهش اثر نوسانات کوتاه‌مدت ناشی از عوامل غیرسیستماتیک یا خطاهای اندازه‌گیری، از فیلتر میانگین متحرک سه‌روزه به‌عنوان یک تکنیک پیش‌پردازش بر روی خروجی مدل پایه IHACRES استفاده شد. این روش موجب حذف نویزهای فرکانس بالا و هموارسازی سری زمانی دبی گردید که در نهایت با کاهش RMSE و بهبود قابل توجه NSE، عملکرد مدل را به‌طور میانگین بیش از ۳۰ درصد نسبت به حالت اولیه برای هردو ایستگاه ارتقاء داد. نتایج نشان می‌دهد که استفاده از چنین رویکرد ساده‌ای می‌تواند دقت مدل‌های مفهومی را در مقیاس روزانه به‌طور مؤثر افزایش دهد.

یکی دیگر از رویکردهای مؤثر که در این مطالعه به‌منظور ارتقاء دقت شبیه‌سازی رواناب توسط مدل IHACRES مورد استفاده قرار گرفت، بهره‌گیری از الگوریتم ژنتیک (Genetic Algorithm) برای بهینه‌سازی پارامترهای این مدل بود. در این روش، فرآیند جست‌وجوی پارامترهای بهینه با هدف یافتن ترکیبی از مقادیر که منجر به بیشترین تطابق بین داده‌های شبیه‌سازی شده و مشاهداتی شود، بر اساس سازوکارهای تکاملی الگوریتم ژنتیک مانند انتخاب، تقاطع و جهش انجام شد.

در پیاده‌سازی این الگوریتم، تنظیمات مؤثر شامل اندازه جمعیت، تعداد نسل‌ها، نرخ تقاطع و نرخ جهش با دقت انتخاب گردید (جدول ۵) تا تعادل مناسبی میان همگرایی و تنوع جمعیت برقرار گردد. سپس، مقدار تابع هدف که ترکیبی از شاخص‌های آماری NSE (ضریب کارایی نش-ساتکلیف) و RMSE (ریشه میانگین مربعات خطا) بود، در هر نسل مورد ارزیابی قرار گرفت.



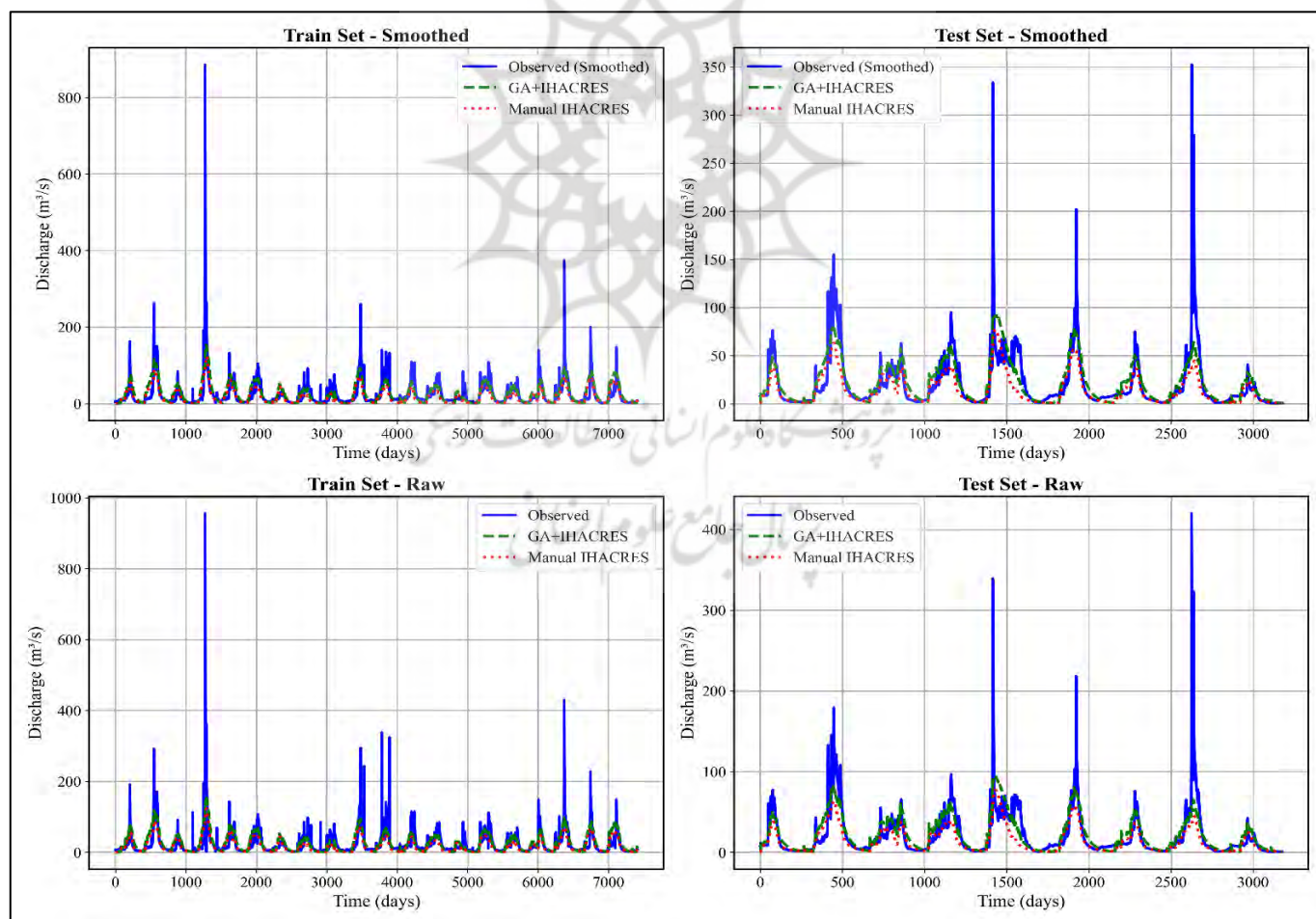
شکل (۳) رواناب شبیه سازی شده برای دوره آموزش و آزمون برای ایستگاه پل کهنه

Figure (3): Simulated runoff for training and testing period for Pol Kohneh station

جدول (۵): تنظیم متغیرهای الگوریتم ژنتیک برای بهینه سازی پارامترهای مدل IHACRES

Table (5): Adjusting genetic algorithm variables to optimize fashion parameters

پارامتر	بازه	توضیحات
اندازه جمعیت	۲۰۰	تعداد نمونه‌ها در هر نسل
تعداد نسل‌ها	۱۰۰	تعداد تکرارهای الگوریتم
احتمال تقاطع	۰/۵	نسبت افراد جدید حاصل از تقاطع
احتمال جهش	۰/۲	احتمال تغییر تصادفی در ژن‌ها
روش انتخاب	تورنمنت با ۳ عضو	انتخاب والدین بر پایه رقابت ۳ عضو
نوع تقاطع	Blend با ضریب $\alpha = 0/5$	ترکیب خطی ژن‌های والدین
نوع جهش	چندجمله‌ای محدودشده با احتمال ۰/۱	غیر مقدار ژن با در نظر گرفتن محدوده‌ها



شکل (۴): رواناب شبیه سازی شده برای دوره آموزش و آزمون برای ایستگاه هیدرومتری قورباغستان

Figure (4): Simulated runoff for the training and testing period for the Qorbaghistan hydrometric station

نتایج حاصل از اجرای مدل GA-IHACRES در جدول (۶) نشان داد که مقدار NSE در هر دو دوره آموزش و آزمون به صورت معنی داری افزایش یافته و همزمان مقدار RMSE کاهش یافته است که نشان دهنده بهبود کارایی مدل نسبت به حالت پایه (بدون بهینه سازی) می باشد.

جدول (۶): ارزیابی عملکرد model XGBoos __ IHACRES Moving average __ IHACRES-GA __ IHACRES پایه

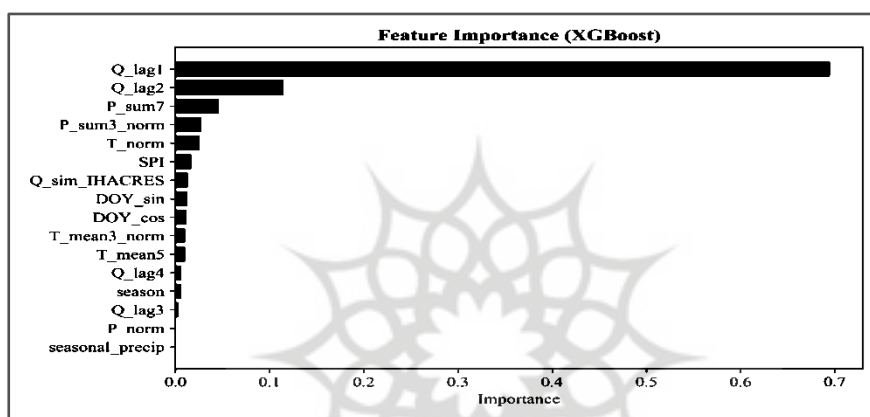
Table (6): IHACRES Performance Evaluation __ IHACRES-GA __ IHACRES Moving average __ model XGBoos base

مدل	NES	RMSE	ایستگاه	مجموعه
IHACRES پایه	۰/۳۹	۲۹/۵	پل کهنه	tarin
	۰/۴	۲۷/۱	قورباغستان	
IHACRES پایه	۰/۴۴	۲۳/۱۱	پل کهنه	test
	۰/۴۷	۲۲/۴	قورباغستان	
GA-IHACRES	۰/۴۹	۲۷/۹۹	پل کهنه	tarin
	۰/۵۳	۲۵/۱	قورباغستان	
GA-IHACRES	۰/۵۹	۲۳/۴۲	پل کهنه	test
	۰/۶۳	۲۱/۶	قورباغستان	
Moving average_Smoothed	۰/۵۴	۲۷/۰۲	پل کهنه	tarin
	۰/۵۷	۲۵/۷	قورباغستان	
Moving average_Smoothed	۰/۵۶	۲۲/۲۳	پل کهنه	test
	۰/۵۹	۲۰/۸۷	قورباغستان	
IHACRES-XGBoost	۰/۹۱	۱۰/۵۴	پل کهنه	tarin
	۰/۹۷	۷/۱۶	قورباغستان	
IHACRES-XGBoost	۰/۹	۱۱/۶۵	پل کهنه	test
	۰/۹۱	۱۲/۰۱	قورباغستان	

این بهبود در عملکرد، بیانگر توانمندی الگوریتم ژنتیک در تنظیم مؤثر پارامترهای مدل و کاهش عدم قطعیت ناشی از تنظیمات اولیه دستی است. به طور کلی، این روش نه تنها موجب افزایش دقت مدل در بازسازی رواناب شده بلکه پایداری نتایج مدل را نیز بهبود بخشیده است. همچنین، به کارگیری الگوریتم ژنتیک برای بهینه سازی پارامترهای IHACRES منجر به بهبود ۲۵ تا ۳۴ درصدی عملکرد مدل در هر دو ایستگاه و در هر دو مرحله آموزش و آزمون شده است. در مرحله نهایی این پژوهش، به منظور بهبود عملکرد مدل پایه IHACRES و رفع محدودیت های آن در شبیه سازی دقیق رواناب روزانه، از یک چارچوب مدل سازی ترکیبی مبتنی بر الگوریتم

XGBoost استفاده شد. در این رویکرد، خروجی مدل IHACRES به همراه مجموعه‌ای از متغیرهای مؤثر شامل داده‌های بارش، دما، شاخص خشکسالی، مؤلفه‌های زمانی و به‌ویژه رواناب‌های تأخیری (Lagged Flow) به‌عنوان ورودی مدل XGBoost تعریف گردید تا با بهره‌گیری از قدرت این الگوریتم در کشف روابط غیرخطی و پیچیده، دقت شبیه‌سازی رواناب افزایش یابد.

در طراحی این مدل ترکیبی، مجموعه‌ای هدفمند از ویژگی‌ها بر پایه اصول هیدرولوژیکی و تحلیل‌های آماری انتخاب شد. از میان این ویژگی‌ها، متغیرهای Q_lag1 تا Q_lag4 که نمایانگر رواناب نرمال‌شده چهار روز گذشته هستند، نقش حافظه هیدرولوژیکی حوزه آبخیز را در تولید رواناب بازتاب می‌دهند. همچنین، داده‌هایی مانند دبی شبیه‌سازی‌شده توسط IHACRES ($Q_sim_IHACRES$)، شاخص SPI، دماهای میانگین چندروزه و مؤلفه‌های زمانی DOY_sin ، DOY_cos ، $season$ نیز در مدل گنجانده شدند تا جنبه‌های اقلیمی و فصلی در رفتار جریان لحاظ شوند.

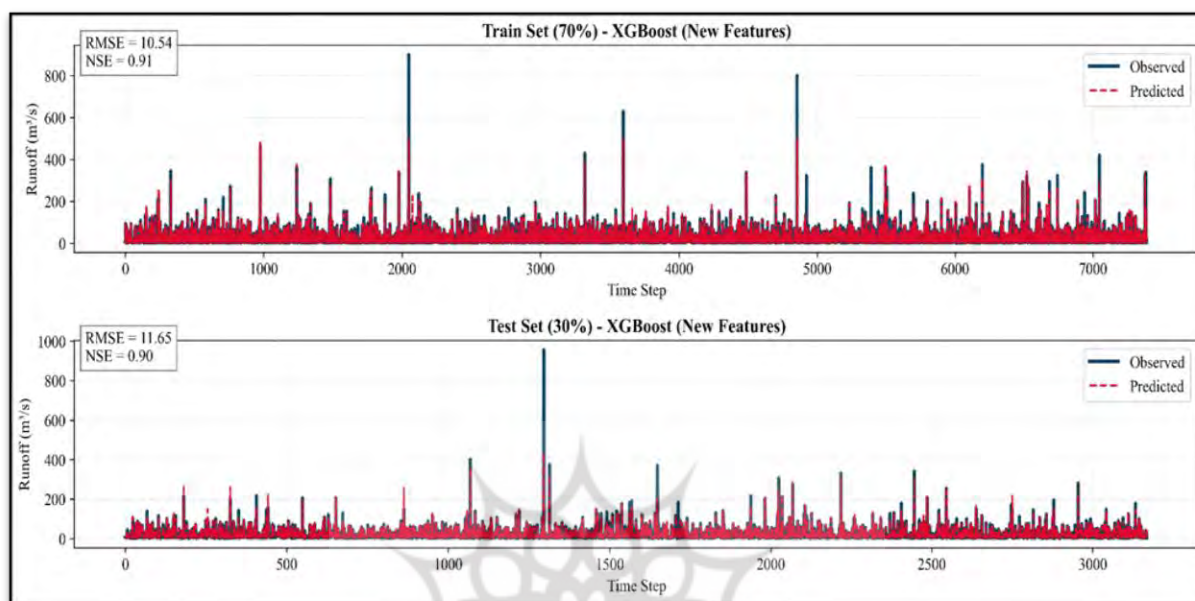


شکل (۵) اهمیت متغیرهای ورودی به مدل XGBoost

Figure (5): Importance of input variables to the XGBoost model

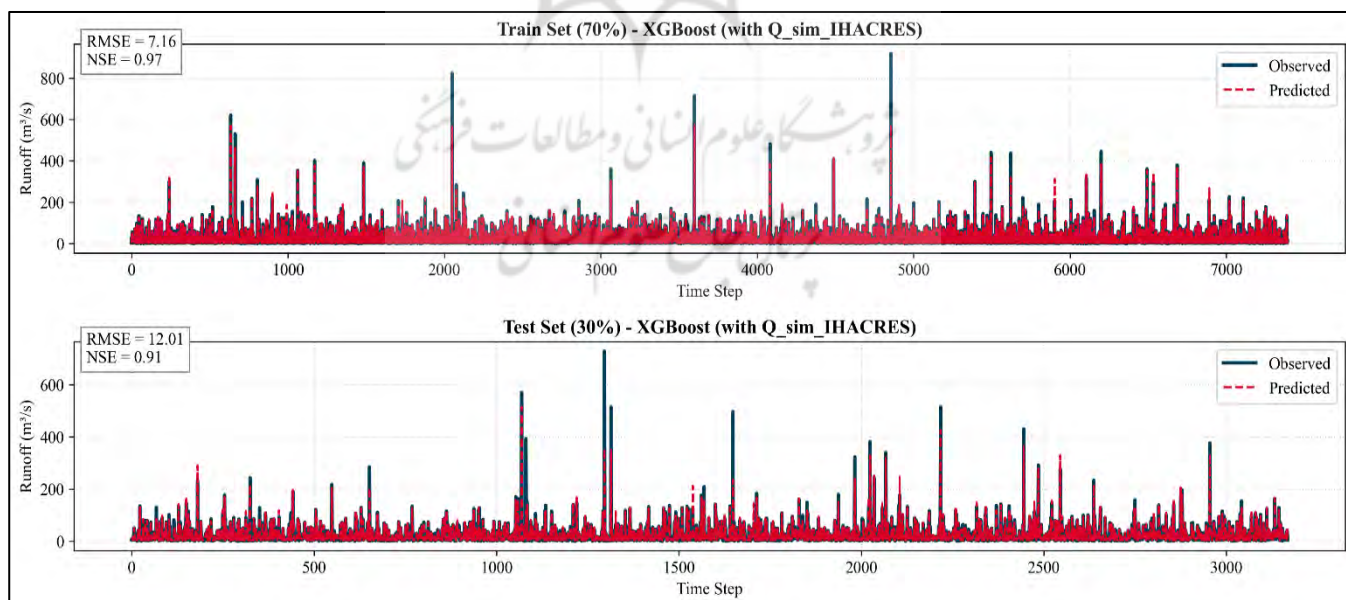
تحلیل اهمیت ویژگی‌ها شکل (۵) باروش گرادیان بوستینگ (Gradient Boosting Feature Importance) در مدل XGBoost به‌روشنی نشان داد که Q_lag1 با اختلاف زیادی نسبت به سایر متغیرها، بالاترین میزان تأثیر را در بهبود شبیه‌سازی رواناب دارد. این یافته با منطق فیزیکی سیستم هیدرولوژیکی کاملاً هم‌خوانی دارد، چراکه شرایط آبی روز پیشین معمولاً تعیین‌کننده روند جریان در روز بعد است. پس از آن، Q_lag2 و بارش تجمعی ۷ روزه (P_sum7) در رتبه‌های بعدی قرار گرفتند و اهمیت بارش‌های انباشته میان‌مدت در فرآیند تولید رواناب را برجسته کردند. $Q_sim_IHACRES$ نیز حاکی از آن است که مدل XGBoost قادر بوده خروجی مدل مفهومی IHACRES را به‌عنوان یکی از اجزای مؤثر در بازسازی نهایی جریان تحلیل و اصلاح نماید. همچنین، هرچند برخی ویژگی‌ها مانند SPI، دمای میانگین ۳ و ۵ روزه و مؤلفه‌های زمانی اهمیت کمتری داشتند، اما نقش آن‌ها در بهبود دقت مدل انکارناپذیر است و به شناخت بهتر الگوهای فصلی و اقلیمی کمک کرده‌اند. جایگاه میانی متغیر $Q_sim_IHACRES$ نیز حاکی از آن است که مدل XGBoost قادر بوده خروجی مدل مفهومی IHACRES را به‌عنوان یکی از اجزای مؤثر در بازسازی نهایی جریان تحلیل و اصلاح نماید. همچنین، هرچند برخی ویژگی‌ها مانند SPI، دمای میانگین ۳ و ۵ روزه و مؤلفه‌های زمانی اهمیت کمتری داشتند، اما نقش آن‌ها در بهبود دقت مدل

انکارناپذیر است و به شناخت بهتر الگوهای فصلی و اقلیمی کمک کرده‌اند. به‌طور کلی، مدل ترکیبی IHACRES-XGBoost با توجه شکل (۷و۶) کاهش چشم‌گیر RMSE و افزایش محسوس ضریب کارایی NSE در هر دو دوره آموزش و آزمون، دقت و پایداری مدل شبیه‌سازی رواناب را به‌طور معناداری ارتقا دهد به‌طور خاص، در ایستگاه قورباغستان، مقدار RMSE در داده‌های آموزشی از ۲۷/۱ به ۷/۲۳ کاهش یافت که نشان‌دهنده بیش از ۷۳ درصد بهبود است.



شکل (۶): عملکرد مدل IHACRES-XGBoost در شبیه‌سازی رواناب برای ایستگاه پل کهنه

Figure (6): Performance of the IHACRES-XGBoost model in runoff simulation for the polkohnah station



شکل (۷): عملکرد مدل IHACRES-XGBoost در شبیه‌سازی رواناب برای ایستگاه قورباغستان

Figure (7): Performance of the IHACRES-XGBoost model in runoff simulation for Qorbaestan station

مقدار NSE در همین ایستگاه از ۰/۴۰ به ۰/۹۷ افزایش یافته که دقت مدل را به شکل قابل توجهی افزایش داده است. به طور مشابه، در ایستگاه پل کهنه نیز بهبود قابل توجهی مشاهده شد. مطابق یافته‌های فتاحی و همکاران (۲۰۲۲) و محمدی و همکاران (۲۰۲۲) و نتایج این پژوهش نیز مؤید آن است که مدل‌های ترکیبی می‌توانند به‌ویژه در شرایط داده‌های ناقص یا اقلیم‌های متغیر مانند اقلیم نیمه‌خشک کرمانشاه، عملکرد بهتری نسبت به مدل‌های صرفاً مفهومی یا صرفاً داده‌محور داشته باشند. با توجه به نتایج به‌دست‌آمده، استفاده از ساختارهای ترکیبی، راهکاری مؤثر برای ارتقاء دقت مدل‌سازی رواناب در کاربردهای مدیریت منابع آب و پیش‌بینی سیلاب به شمار می‌رود. شکل (۷و۶)

نتیجه‌گیری

شبیه‌سازی فرآیند بارش-رواناب، به‌ویژه در مناطق نیمه‌خشک با ویژگی‌هایی چون بارندگی نامنظم، تبخیر بالا و دسترسی محدود به داده‌های دقیق، از اهمیت بسزایی در مدیریت منابع آب برخوردار است. انتخاب مدل مناسب برای یک حوضه آبریز، می‌تواند نقش تعیین‌کننده‌ای در ارتقای دقت پیش‌بینی‌ها و بهینه‌سازی برنامه‌ریزی‌ها داشته باشد. از این‌رو، ارزیابی و مقایسه مدل‌ها و تکنیک‌های مختلف در شرایط بومی حوضه‌های مطالعاتی ضروری است.

در این پژوهش، چهار رویکرد مختلف شامل: مدل پایه IHACRES، IHACRES با میانگین متحرک سه‌روزه، IHACRES بهینه‌شده با الگوریتم ژنتیک، و مدل ترکیبی IHACRES-XGBoost برای شبیه‌سازی رواناب روزانه در حوضه آبریز قره‌سو مورد بررسی و مقایسه قرار گرفت. مدل مفهومی IHACRES علی‌رغم ساختار ساده و قابل فهم، نتوانست دبی‌های اوج را با دقت کافی بازسازی کند. یکی از دلایل این امر، نادیده گرفتن فرآیندهایی چون رواناب ناشی از ذوب برف بود که در مطالعات پیشین خطیبی و همکاران و (صابر و همکاران، ۱۳۹۵: ۴۲) نیز مطرح شده است. در ادامه، با هدف کاهش نوسانات کوتاه‌مدت و افزایش پایداری خروجی‌ها، از فیلتر میانگین متحرک سه‌روزه استفاده شد که منجر به بهبود نسبی نتایج گردید. به‌دنبال آن، بهینه‌سازی پارامترهای مدل با استفاده از الگوریتم ژنتیک صورت گرفت که موجب بهبود ۲۵ تا ۳۴ درصدی عملکرد مدل در شاخص‌هایی مانند NSE و RMSE شد. این بهینه‌سازی، نقش مؤثری در کاهش عدم قطعیت ناشی از تنظیمات دستی ایفا کرد. در نهایت، با تلفیق مدل مفهومی IHACRES و الگوریتم قدرتمند XGBoost، یک چارچوب مدل‌سازی ترکیبی ارائه شد. در این ساختار، خروجی مدل مفهومی همراه با مجموعه‌ای از متغیرهای هیدرولوژیکی و اقلیمی مانند رواناب‌های تأخیری Q_{lag1} تا Q_{lag4} ، بارش تجمعی، شاخص خشکسالی SPI و مؤلفه‌های زمانی، به عنوان ورودی مدل XGBoost تعریف گردید. این مدل ترکیبی توانست با بهره‌گیری از ظرفیت الگوریتم‌های یادگیری ماشین در تحلیل روابط پیچیده و غیرخطی، دقت مدل‌سازی رواناب را به‌طور معناداری افزایش دهد. تحلیل اهمیت ویژگی‌ها نشان داد که متغیرهایی نظیر Q_{lag1} و Q_{lag2} نقشی کلیدی در بازسازی رواناب ایفا می‌کنند، که بیانگر وجود حافظه هیدرولوژیکی در رفتار جریان است. نتایج حاصل از مدل ترکیبی در ایستگاه‌های قورباغه‌ستان و پل کهنه نشان داد که RMSE بیش از ۷۰٪ کاهش یافته و مقدار NSE از ۰/۴۰ به ۰/۹۷ افزایش یافته است؛ که تأییدکننده دقت، پایداری و کارایی بالای چارچوب پیشنهادی است. در مجموع، این مطالعه نشان داد که ترکیب رویکردهای مفهومی و داده‌محور می‌تواند محدودیت‌های هر یک را جبران کرده و در شرایط اقلیمی متغیر و داده‌ناقص، ابزاری قدرتمند برای شبیه‌سازی رواناب فراهم آورد. بنابراین، چارچوب مدل‌سازی IHACRES-XGBoost را می‌توان به‌عنوان راهکاری مؤثر، دقیق و قابل اعتماد برای پیش‌بینی رواناب، مدیریت منابع آب و هشدار سیلاب در حوضه‌های آبریز نیمه‌خشک پیشنهاد کرد. چارچوب ترکیب IHACRES-XGBoost توانست دقت و

پایداری شبیه‌سازی رواناب در حوضه‌های نیمه‌خشک را بهبود دهد، اما محدودیت‌هایی دارد که باید در مطالعات آینده مورد توجه قرار گیرد. افزودن متغیرهای اقلیمی و هیدرولوژیکی بیشتر مانند ذوب برف، رطوبت خاک و شاخص‌های کاربری اراضی می‌تواند عملکرد مدل را افزایش دهد. به‌کارگیری الگوریتم‌های بهینه‌سازی پیشرفته (PSO)، (Bayesian Optimization) همراه با تحلیل عدم قطعیت مبتنی بر آنسامبل، پایداری پیش‌بینی‌ها را تقویت می‌کند. بررسی قابلیت انتقال‌پذیری مدل به دیگر حوضه‌های نیمه‌خشک و اعتبارسنجی آن با داده‌های مستقل، ضروری است. همچنین، توسعه سامانه‌های عملیاتی هشدار سیلاب مبتنی بر این چارچوب می‌تواند ابزار کارآمدی برای مدیریت منابع آب و تصمیم‌گیری زیست‌محیطی فراهم کند. این پیشنهادها مسیرهای کلیدی برای ارتقای دقت و کاربرد مدل در شرایط واقعی را فراهم می‌سازد.

References

- Abushandi, E., & Merkel, B. (2013). Modelling rainfall–runoff relations using HEC-HMS and IHACRES for a single rain event in an arid region of Jordan. *Water Resources Management*, 27(7), 2391–2409. <https://doi.org/10.1007/s11269-013-0292-3>.
- Aghabeigi, N., Esmaili Ouri, A., Abazar, N., Mostafazadeh, R., & Golshan, M. (2019). Effects of climate change on runoff using the IHACRES hydrological model in some watersheds of Ardabil Province. *Iranian Journal of Irrigation and Water Engineering*, 10(2), 178–189. <https://doi.org/10.22125/iwe.2019.100750>.
- Ahmadi, M., Moeini, A., Ahmadi, H., Motamedvaziri, B., & Zehtabiyani, G. R. (2019). Comparison of the performance of SWAT, IHACRES and artificial neural networks models in rainfall–runoff simulation: A case study of the Kan watershed, Iran. *Physics and Chemistry of the Earth*, 111, 65–77. <https://doi.org/10.1016/j.pce.2019.05.002>.
- Bayati Khatibi, M., Kakapour, V., & Sadeghi, M. (2025). Discharge prediction using IHACRES semi-conceptual model: A case study of Gharasoo Basin, Kermanshah province. *Quantitative Geomorphological Researches*, 13(3), 104–119. <https://doi.org/10.22034/gmpj.2024.384992.1417>.
- Chen, T. Q., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 785–794). <https://doi.org/10.1145/2939672.2939785>.
- Croke, B. F. W., & Jakeman, A. J. (2008). Use of the IHACRES rainfall–runoff model in arid and semi-arid regions. In H. S. Wheater, S. Sorooshian, & K. D. Sharma (Eds.), *Hydrological modeling in arid and semi-arid areas* (pp. 41–48). Cambridge University Press.
- Czado, C. (2019). Analyzing dependent data with vine copulas. *Lecture Notes in Statistics*. Springer.
- Esmali, A., Golshan, M., & Kaviani, A. (2021). Investigating the performance of SWAT and IHACRES in streamflow simulation under different climatic regions in Iran. *Atmósfera*, 34(1), 79–96. <https://doi.org/10.20937/ATM.2021.34.01.07>.
- monthly river flow prediction: Integrating artificial intelligence and non-linear time series models. *Journal of Hydrology*, 575, 1200–1213. <https://doi.org/10.1016/j.jhydrol.2019.06.025>.
- Fattahi, P., Ashrafzadeh, A., Pirmoradian, N., & Vazifiedoust, M. (2022). Integrating IHACRES with a data-driven model to investigate the possibility of improving monthly flow estimates. *Water Supply*, 22(1), 360–371. <https://doi.org/10.2166/ws.2021.342>.

- Fouladi Nasrabadi, M., Amirabadizadeh, M., Pourreza Bielandi, M., & Yaghoubzadeh, M. (2022). Evaluation of IHACRES model performance with ARMAX and EXPUH linear methods (Case study: Shoor river basin in Qaen). *Journal of Water and Soil*, 36(1), 17–30. <https://doi.org/10.22067/jsw.2022.74115.112>.
- data-driven deep learning methodology with deep feature selection. *Environmental and Ecological Statistics*, 32, 409–459. <https://doi.org/10.1007/s10651-024-00642-6>.
- Ismail, I. I., Jibril, M. M., Muhammad, U. J., Mahmoud, I. A., Aliyu, U. U., Abdullahi, A., & Malami, S. I. (2025). Ensemble machine learning technique based on Gaussian algorithm for stream flow modelling. *Techno-Computing Journal*, 1(2), 1–17. <https://doi.org/10.71170/tecoj.2025.1.2.pp1-17>.
- James K., W. (2024). Characterizing nonlinear, nonstationary, and heterogeneous hydrologic behavior using ensemble rainfall–runoff analysis (ERRA). *Hydrology and Earth System Sciences*, 28, 4427–4454. <https://doi.org/10.5194/hess-28-4427-2024>.
- Khorsandi Kohanestani, Z., & Ta'atpour, F. (2025). Comparison of the efficiency of integrated models and machine learning methods in discharge simulation. *Hydrogeomorphology*, 12(42), 60–80. <https://doi.org/10.22034/hyd.2024.63280.1754>.
- Mohammadi, B., Moazenzadeh, R., Christian, K., & Duan, Z. (2021). Improving streamflow simulation by combining hydrological process-driven and artificial intelligence-based models. *Environmental Science and Pollution Research*, 28, 65752–65768. <https://doi.org/10.1007/s11356-021-15131-6>.
- Mohammadi, B., Safari, M. J. S., & Vazifehkhah, S. (2022). IHACRES, GR4J and MISD-based multi conceptual-machine learning approach for rainfall–runoff modeling. *Scientific Reports*, 12, 12096. <https://doi.org/10.1038/s41598-022-16217-1>.
- Momeneh, S. (2022). Comparison of artificial intelligence models with IHACRES model in flow modeling of Gamasiab river basin. *Modeling and Management of Water and Soil*, 2(3), 1–16. <https://doi.org/10.22098/mmws.2022.9972.1076>.
- Nazeri Tahroudi, M., Ahmadi, F., & Mirabbas, R. (2023). Performance comparison of IHACRES, random forest and copula-based models in rainfall–runoff simulation. *Applied Water Science*, 13, 134. <https://doi.org/10.1007/s13201-023-01929-y>.
- Niromandfard, F., Zakerinia, M., & Yazerloo, B. (2018). Investigating the effect of climate change on river flow using IHACRES rainfall–runoff model. *Irrigation Sciences and Engineering*, 41(3), 103–117. <https://doi.org/10.22055/jise.2018.13750>.
- Noori, N., & Kalin, L. (2016). Coupling SWAT and ANN models for enhanced daily streamflow prediction. *Journal of Hydrology*, 533, 141–151. <https://doi.org/10.1016/j.jhydrol.2015.11.050>.
- Prakash, S., Sandilya, S. S., & Das, B. S. (2025). Discharge prediction in meandering compound channel using XGBoost and CATBoost. In M. Pandey, N. V. Umamahesh, Z. Ahmad, & G. Oliveto (Eds.), *Hydraulics and fluid mechanics, volume 2. HYDRO 2023. Lecture Notes in Civil Engineering*, 2 (pp. 429–437). Springer. https://doi.org/10.1007/978-981-97-8895-8_30.

- Saber, M., Salahi, B., & Maleki Meresht, R. (2025). Simulating the water balance of Aras basin based on the CNRM-CM6 climate model. *Hydrogeomorphology*, 12(42): 99 – 117. DOI: [10.22034/hyd.2024.63758.1759](https://doi.org/10.22034/hyd.2024.63758.1759).
- Salehpoor Laghani, J., Ashrafzadeh, A., & Moussavi, S. A. (2020). Investigating the uncertainty of data-based models in forecasting monthly flow of the Hablehroud River. *Iranian Journal of Soil and Water Research*, 51(5), 1265–1280. <https://doi.org/10.22059/ijswr.2020.286920.668288>.
- Yangyu, D., Zhang, D., Zhang, D., Wu, J., & Liu, Y. (2023). A hybrid ensemble machine learning model for discharge coefficient prediction of side orifices with different shapes. *Flow Measurement and Instrumentation*, 91, 102372. <https://doi.org/10.1016/j.flowmeainst.2023.102372>.

