



Application of Machine Learning in Sustainable Qanat Management: Qanat Discharge Estimation Using the Random Forest Algorithm and Sensitivity Analysis (SHAP and Permutation Importance) in Qaen Plain

Sepideh Zeraati Neyshabouri¹, | Mohammad Fouladi Nasrabad¹ | Mahna Javaheri¹

1. Department of Water Engineering, Faculty of Agriculture, University of Birjand, Birjand, Iran.

✉Corresponding Author: sepideh_zeraatineyshabouri@birjand.ac.ir

Received:
13 December 2025

Revised:
04 March 2026

Accepted:
05 May 2026

Keywords:

*Qanat discharge prediction;
Climatic data;
Spatiotemporal variations;
Sensitivity analysis.*

Extended abstract

Introduction

Qanats are traditional and sustainable groundwater systems that play an important role in water supply and agriculture in arid and semi-arid regions of Iran. However, drought, climate variability, and increasing pressure on groundwater resources have affected qanat discharge, making accurate discharge estimation important for sustainable water management. This study aimed to evaluate the performance of the Random Forest (RF) algorithm for estimating qanat discharge in the Qaen Plain, South Khorasan Province, and to identify the most influential variables using SHAP and Permutation Importance.

Materials and Methods

A 12-year dataset (2007–2018) consisting of monthly qanat discharge, precipitation, mean temperature, and elevation was used. After data preprocessing, including normalization, outlier removal, and imputation of missing values, 438 records were retained. The data were divided chronologically into 70% training and 30% testing datasets.

Cite this article: Zeraati Neyshabouri, S., Fouladi Nasrabad, M. & Javaheri, M. (2026). Application of machine learning in sustainable qanat management: Qanat discharge estimation using the Random Forest algorithm and sensitivity analysis (SHAP and Permutation Importance) in Qaen Plain. *Journal of Aquifer and Qanat*, 7 (1), 25-40. <https://doi.org/10.22077/jaaq.2026.10686.1135>.



Copyright: © 2026 by the authors. Licensee: Journal of Aquifer and Qanat. This article is an open-access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

The RF model was optimized using grid search and five-fold cross-validation. The optimal model included 500 trees, a maximum depth of 10, a maximum feature ratio of 0.5, and a minimum of five samples for node splitting. Model performance was evaluated using RMSE, NSE, KGE, Pearson correlation coefficient (r), and NRMSE. SHAP and Permutation Importance were applied to determine the contribution of the input variables.

Results and Discussion

The RF model showed satisfactory performance on the testing dataset, with $RMSE = 4.27$ L/s, $NSE = 0.54$, $KGE = 0.51$, $r = 0.76$, and $NRMSE = 15.19\%$. The relatively small difference between training and testing performance indicated that severe overfitting did not occur. Both SHAP and Permutation Importance identified elevation as the most important predictor of qanat discharge, followed by temperature and precipitation. The mean SHAP values for elevation, temperature, and precipitation were approximately 2.16, 0.55, and 0.44, respectively. The dominant role of elevation can be related to its influence on groundwater recharge and hydraulic conditions. The relatively weak negative contribution of precipitation may be associated with delayed groundwater responses to rainfall, which was not explicitly considered in the model.

Conclusion

The results demonstrate that the Random Forest algorithm can provide an effective approach for estimating qanat discharge in arid and semi-arid regions. The model achieved acceptable predictive accuracy and successfully identified elevation as the most influential variable. Integrating RF with explainable methods such as SHAP and Permutation Importance can provide both accurate predictions and useful information for understanding the factors controlling qanat discharge. The proposed approach can support sustainable groundwater management, drought preparedness, and improved planning of qanat water resources in Qaen Plain and similar dryland regions.



کاربرد یادگیری ماشین در مدیریت پایدار قنات: برآورد دبی با الگوریتم جنگل تصادفی و تحلیل حساسیت (SHAP و Permutation Importance) در دشت قاین

سپیده زراعتی نیشابوری^۱ | محمد فولادی نصرآباد^۱ | مهنا جواهری^۱

۱. گروه مهندسی آب، دانشکده کشاورزی، دانشگاه بیرجند، بیرجند، ایران.

✉ نویسنده مسئول: sepideh_zeraatineyshabouri@birjand.ac.ir

چکیده

مناطق خشک و نیمه خشک ایران، به ویژه دشت قاین در استان خراسان جنوبی، با محدودیت‌های جدی منابع آبی مواجه هستند. قنات به عنوان سیستم‌های سنتی تأمین آب پایدار با قدمت بیش از ۳۰۰۰ سال، نقش حیاتی در کشاورزی و زندگی محلی ایفا می‌کنند، اما برآورد دقیق دبی آن‌ها تحت تأثیر تغییرات اقلیمی و محدودیت‌های داده‌ای دشوار است. بنابراین، این پژوهش با استفاده از الگوریتم جنگل تصادفی، یک روش یادگیری ماشین قدرتمند، به برآورد دبی قنات دشت قاین بر اساس داده‌های ۱۲ ساله (۱۳۸۶-۱۳۹۷ شمسی) می‌پردازد. متغیرهای ورودی شامل داده‌های تاریخی دبی مشاهده شده، دمای متوسط، بارش ماهانه و ارتفاع مکانی هر قنات انتخاب شدند. پس از پیش‌پردازش شامل نرمال‌سازی و پر کردن داده‌های گمشده، داده‌ها به بخش‌های آموزش (۷۰٪) و آزمون (۳۰٪) تقسیم گردیدند. در ادامه مدل با بهینه‌سازی پارامترها ($\text{max_depth}=10$ ، $\text{max_features}=0.5$ ، $\text{min_samples_split}=5$ و $n_estimators=500$) آموزش داده شد. نتایج با شاخص‌های $\text{RMSE}=4.27$ L/s، $\text{NSE}=0.54$ ، $r=0.76$ و $\text{KGE}=0.51$ و $\text{NRMSE}=15.19\%$ نشان‌دهنده عملکرد شایسته مدل است. تحلیل حساسیت با SHAP و Permutation Importance نیز، ارتفاع را به عنوان متغیر کلیدی (میانگین $\text{SHAP}=2.16$) شناسایی نمود، در حالی که بارش تأثیر منفی جزئی (-0.13) نشان داد. این یافته‌ها کارایی مطلوب RF را تأیید کرده و چارچوبی عملی برای مدیریت منابع آب در مناطق خشک ارائه می‌دهد.

تاریخ دریافت: ۱۴۰۴/۰۹/۲۲
تاریخ بازنگری: ۱۴۰۴/۱۲/۱۳
تاریخ پذیرش: ۱۴۰۵/۰۲/۱۴

کلیدواژه‌ها:

پیش‌بینی دبی قنات،
داده‌های اقلیمی، تغییرات
مکانی-زمانی، ارزیابی
حساسیت.

مقدمه

فشار روزافزون بر منابع آب در اقلیم‌های خشک، تهدیدی بی‌سابقه برای پایداری محیط‌زیست و معیشت جوامع انسانی به وجود آورده‌اند. ایران، به‌عنوان کشوری واقع در فلات مرکزی آسیا با گستره آب و هوایی عمدتاً خشک، بیش از ۹۰ درصد از آب مصرفی خود را از منابع زیرزمینی تأمین می‌کند و تغییرات اقلیمی مانند کاهش بارش تا حدود ۲۰ درصد در دهه‌های اخیر و افزایش دما به میزان ۱.۵ تا ۲ درجه سانتی‌گراد، این بحران را تشدید کرده است (Maghrebi et al., 2023).

قنات، به‌عنوان یکی از قدیمی‌ترین سیستم‌های مهندسی آب با قدمت بیش از ۳۰۰۰ سال، نقش کلیدی در مدیریت پایدار آب در این مناطق برعهده دارند (UNESCO World Heritage Centre, 2016). این سیستم‌های زیرزمینی، با هدایت آب از سفره‌های آب زیرزمینی کوهستانی به دشت‌ها از طریق شیب ملایم (کمتر از ۱٪)، بدون نیاز به انرژی خارجی، آب را برای کشاورزی، شرب و اکوسیستم‌ها فراهم می‌کنند (Ramezani Feriz and Ramezani, 2025; Farzaneh and Banimostafaarab, 2025).

در دشت قاین، واقع در استان خراسان جنوبی با بارش سالانه کمتر از ۱۵۰ میلی‌متر و ارتفاع متوسط ۱۴۰۰ متر، قنات بیش از ۶۰ درصد آب کشاورزی را تأمین و پایداری جوامع محلی را تضمین می‌نمایند، اما کاهش دبی آن‌ها به دلیل پمپاژ چاه‌های مدرن و خشک‌سالی، تهدیدی جدی محسوب می‌شود. برآورد دقیق دبی قنات در این مناطق با چالش‌های متعددی روبه‌رو است؛ از جمله محدودیت داده‌های اندازه‌گیری (اغلب ماهانه و ناقص)، روابط غیرخطی بین متغیرهای اقلیمی (بارش، دما) و هیدرولوژیکی (دبی، ارتفاع) و تأثیرات اقلیمی مانند ال‌نینو که نوسانات را افزایش می‌دهد.

روش‌های سنتی، مانند مدل‌های تحلیلی بر اساس معادلات دارسی، اغلب ناکارآمد هستند زیرا پارامترهای هیدرولیکی (مانند نفوذپذیری) را نادیده می‌گیرند. در مقابل، الگوریتم‌های یادگیری ماشین، به‌ویژه جنگل تصادفی^۱ (به اختصار RF)، با توانایی مدل‌سازی روابط پیچیده و مدیریت داده‌های نویزی، جایگزینی قدرتمند ارائه می‌دهند.

RF، توسعه‌یافته توسط بریمن^۲ در سال ۲۰۱۱، با ترکیب صدها درخت تصمیم‌گیری و تکنیک بگینگ^۳ دقت بالا و مقاومت در برابر بیش‌برازش را تضمین می‌کند و در هیدرولوژی برای پیش‌بینی دبی رودخانه‌ها و سطح آب زیرزمینی موفق بوده است (Breiman, 2001; Kratzert et al., 2019; Prieto et al., 2019).

اهداف این پژوهش عبارتند از ارزیابی دقت الگوریتم RF در برآورد دبی قنات دشت قاین با در نظر گرفتن متغیرهای اقلیمی و مکانی، بهینه‌سازی پارامترهای مدل و انجام تحلیل حساسیت ویژگی‌ها با استفاده از روش‌های ارزش شاپلی^۴ (به اختصار SHAP) و اهمیت ویژگی بر پایه جابه‌جایی^۵، بررسی کاربرد عملی مدل در مدیریت منابع آب و مقایسه نتایج آن با مطالعات مشابه پیشین و در نهایت ارائه راهکارهای عملی و علمی برای بهبود مدیریت منابع آب زیرزمینی در مناطق خشک و نیمه‌خشک ایران. فرضیه اصلی این پژوهش بر این موضوع استوار است که الگوریتم RF قادر است دبی قنات را با دقت قابل قبول ($NSE > 0.5$ و $RMSE < 5 \text{ L/s}$) برآورد نماید. این امر به پیش‌بینی پایدار منابع آب در مناطق خشک کمک خواهد کرد.

¹ Random Forest

² Breiman

³ Bagging

⁴ Shapley Additive Explanations

⁵ Permutation Importance

برآورد دبی در مناطق خشک و نیمه خشک، موضوعی محوری در هیدرولوژی مدرن محسوب می شود. مطالعات کلاسیک، مانند مدل های نیمه تحلیلی برای تغییرات دبی قنات در سفره های آب غیرمحدود، بر تأثیر بارش و پمپاژ چاه ها تأکید دارند و نشان می دهند که دبی قنات با کاهش سطح آب زیرزمینی تا ۳۰ درصد افت می کند (Sedghi and Zhan, 2024).

در ایران، پژوهش با استفاده از شبکه عصبی مصنوعی^۱ (به اختصار ANN) برای پیش بینی دبی ماهانه در حوضه های نیمه خشک جنوب غربی، $R^2=0.85$ به دست آورد اما محدودیت در داده های محدود را برجسته کرد (Rezaeian-Zadeh et al., 2013). در مطالعه ای مقایسه ای از روش های مختلف یادگیری ماشین برای پیش بینی دبی قنات ها انجام دادند که نتایج آنها نشان داد مدلهایی مانند الگوریتم های درخت تصمیم و روش های تقویتی دارای عملکرد بهتری هستند و دقت پیش بینی تا حد قابل توجهی بهبود یافته است. این مطالعه اهمیت استفاده از روش های متنوع یادگیری ماشین در مدیریت بهینه منابع آب در مناطق خشک را تأیید کرد (Samani et al., 2024).

در سطح بین المللی، مدل های مفهومی مانند هایدروس^۲ برای شبیه سازی جریان در قنات دقت خوبی در پیش بینی دبی ($RMSE \approx 5 \text{ L/s}$) نشان داد، اما نیاز به پارامترهای دقیق خاک داشت (Sedghi and Zhan, 2024). کاربرد یادگیری ماشین در هیدرولوژی روبه رشد است، RF به عنوان یک روش یادگیری جمعی^۳ در پیش بینی دبی سیلاب با $NSE=0.7$ در حوضه های بزرگ کانادا، برتر از مدل های مفهومی مانند هیدرومد^۴ عمل کرد

(Kratzert et al., 2019). تانگ و همکاران (۲۰۲۲) از الگوریتم جنگل تصادفی (RF) برای برآورد جریان رودخانه ای در منطقه نیمه خشک حوضه رودخانه Qiushui در چین استفاده کردند و نشان دادند که این مدل نسبت به روش های سنتی عملکرد بهتری دارد (Tang et al., 2020).

نوآوری اصلی پژوهش حاضر در چند جنبه کلیدی قابل تبیین است: اول، این مطالعه برای اولین بار در ایران، از ترکیب دو روش پیشرفته تفسیرپذیری (Interpretability) یعنی SHAP و Permutation Importance برای باز کردن جعبه سیاه مدل جنگل تصادفی و شناسایی دقیق مکانیسم های فیزیکی حاکم بر جریان قنات استفاده می کند. این رویکرد ترکیبی، بینش عمیق تری نسبت به روابط علی و معلولی میان متغیرهای اقلیمی و مکانی با دبی قنات ارائه می دهد. دوم، تمرکز بر قنات به عنوان سیستم های سنتی و پایدار تأمین آب که علیرغم اهمیت حیاتی در مناطق خشک، کمتر مورد توجه پژوهش های یادگیری ماشین قرار گرفته اند. سوم، ارائه چارچوبی عملی برای مدیریت پایدار منابع آب زیرزمینی و سیاست گذاری مبتنی بر داده که نتایج آن مستقیماً در حفاظت از آبخوان های بالادست و مدیریت بهره برداری در شرایط خشکسالی قابل کاربرد است.

مواد و روش ها

منطقه مورد مطالعه

دشت قاین، واقع در استان خراسان جنوبی، یکی از مناطق خشک و نیمه خشک ایران است که با محدودیت های جدی در مدیریت منابع آب مواجه است. این دشت در موقعیت جغرافیایی ۳۳ درجه و ۴۳ دقیقه عرض شمالی و ۵۹ درجه و ۱۱ دقیقه طول شرقی قرار دارد (شکل ۱). با میانگین بارندگی سالانه کمتر از ۱۰۰ میلی متر و دمای

¹ Artificial Neural Networks

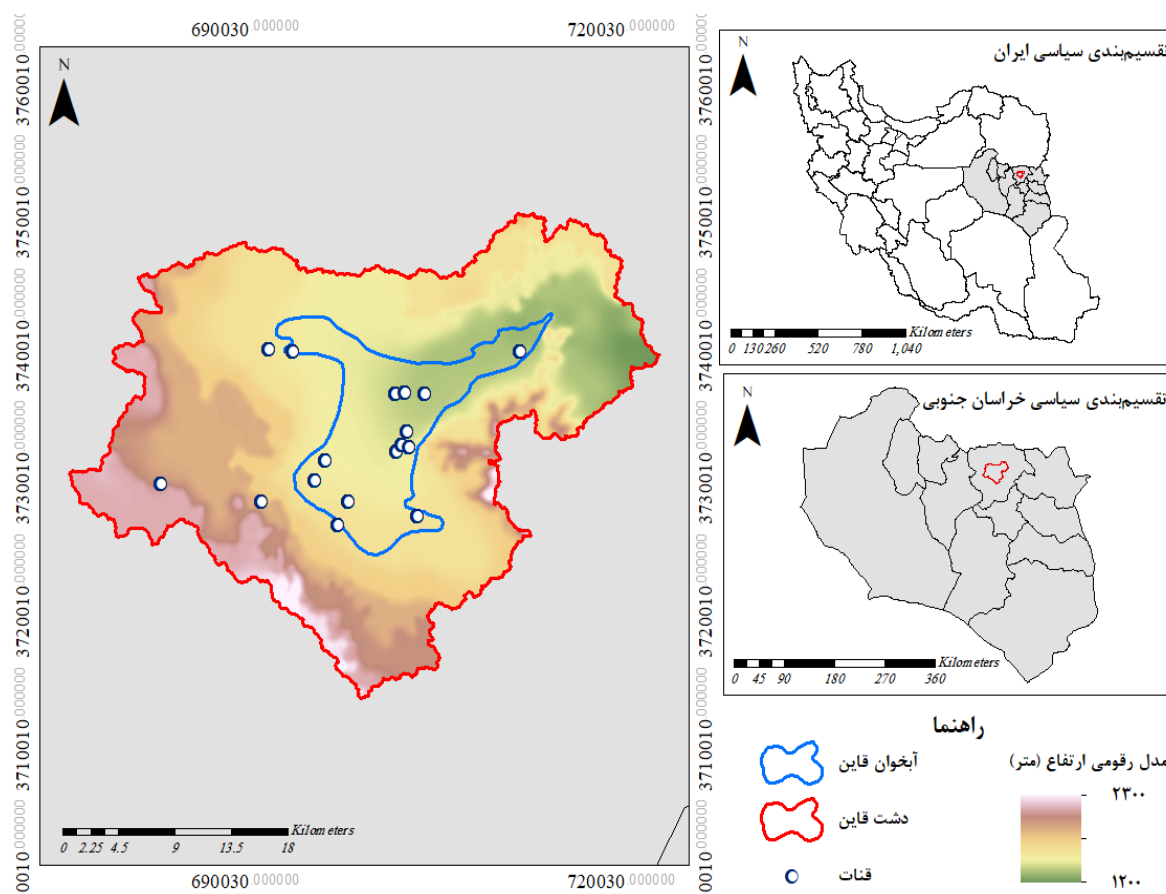
² Hydrus

³ Ensemble

⁴ Hydromad

تبخیر، این دشت را به شدت به منابع آب زیرزمینی، به ویژه قنات، وابسته کرده است (Dastourani et al., 2025; Foruzanmehr and Khozaymehnezhad, 2023).

متوسط سالانه حدود ۲۰ درجه سانتی‌گراد، این منطقه شرایط اقلیمی سختی را تجربه می‌کند. خاک‌های غالب منطقه، که عمدتاً از نوع رسی-شنی با ظرفیت پایین نگهداری آب هستند، همراه با نرخ بالای



شکل ۱. موقعیت مکانی دشت و آبخوان قاین در استان خراسان جنوبی کشور ایران
Fig 1. Location of Qaen plain and aquifer in South Khorasan Province, Iran

پاک‌سازی، کنترل کیفیت و یکپارچه‌سازی، در مجموع ۴۳۸ رکورد کامل به‌عنوان ورودی نهایی برای فرآیند مدل‌سازی مورد استفاده قرار گرفت.

پیش‌پردازش داده‌ها

مراحل پیش‌پردازش داده‌ها در این پژوهش به‌گونه‌ای طراحی شد که ضمن حفظ کیفیت اطلاعات، قابلیت اطمینان مدل‌سازی افزایش یابد. در گام نخست تمامی

داده‌های مورد استفاده

داده‌های مورد استفاده در این پژوهش شامل یک سری زمانی دوازده‌ساله طی دوره ۱۳۸۶ تا ۱۳۹۷ شمسی است. داده‌های اصلی شامل دبی ماهانه قنات بر حسب لیتر بر ثانیه، متغیرهای اقلیمی نظیر بارش ماهانه بر حسب میلی‌متر، دمای متوسط ماهانه بر حسب درجه سانتی‌گراد، داده‌های مربوط به موقعیت مکانی و ارتفاع (متر) برای هر یک از قنات بوده که از طریق سازمان آب منطقه‌ای خراسان جنوبی تأمین گردیده است. پس از انجام مراحل

گره^۸ با مقادیر ۲، ۵ و ۱۰ و همچنین تعداد درخت‌ها^۹ با مقادیر ۱۰۰، ۳۰۰ و ۵۰۰ بودند.

نتایج بهینه‌سازی نهایی نشان داد بهترین پارامترهای مدل شامل تعداد ۵۰۰ درخت، عمق حداکثر ۱۰، نسبت ویژگی‌ها برابر با ۰/۵ و حداقل ۵ نمونه برای تقسیم گره است. معیار ارزیابی به کار رفته در مدل، معیار خطای مربعات متوسط^{۱۰} بوده است. چنین تنظیماتی به مدل این امکان را می‌دهد که پیش‌بینی دقیق‌تر و تعمیم‌پذیرتری نسبت به داده‌های آموزشی و آزمون ارائه دهد.

شاخص‌های ارزیابی

برای تضمین جامعیت ارزیابی، از چند معیار تکمیلی استفاده شده است که شامل ریشه میانگین مربعات خطا^{۱۱} (به اختصار RMSE)، کارایی نش-ساتکلیف^{۱۲} (به اختصار NSE)، کارایی کلینگ-گوپتا^{۱۳} (به اختصار KGE)، همبستگی پیرسون^{۱۴} (به اختصار r) و ریشه میانگین مربعات خطای نرمال‌شده^{۱۵} (به اختصار NRMSE) می‌باشند. (روابط ۱ تا ۴):

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (1)$$

$$NSE = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (2)$$

$$KGE = 1 - \frac{1}{\sqrt{(r-1)^2 + (\alpha-1)^2 + (\beta-1)^2}} \quad (3)$$

متغیرها با استفاده از روش مقیاس‌بندی کمینه-بیشینه^۱ در بازه صفر تا یک نرمال‌سازی شدند تا امکان مقایسه مستقیم میان مقادیر فراهم گردد. در ادامه، ناهنجاری‌ها و داده‌های پرت با معیار نمره استاندارد زد^۲ بزرگ‌تر از سه شناسایی و حذف شدند. همچنین حدود ده درصد داده‌های گمشده موجود با بهره‌گیری از میانگین متحرک سه‌ماهه تکمیل گردید تا پیوستگی سری زمانی حفظ شود. در مرحله بعد داده‌ها بر اساس ترتیب زمانی به دو بخش آموزش شامل هفتاد درصد تا پایان سال ۱۳۹۳ و آزمون شامل سی درصد باقی‌مانده تقسیم شدند. نتایج تحلیل هم‌بستگی اولیه نشان داد که میزان ارتباط میان دبی قنات و ارتفاع برابر با ۰/۶۵ و میان دبی و بارش معادل ۰/۴۲ بوده که بیانگر نقش پررنگ‌تر عامل ارتفاع در تغییرپذیری جریان قنات نسبت به بارش است.

مدل جنگل تصادفی

الگوریتم جنگل تصادفی با استفاده از کتابخانه scikit-learn در زبان برنامه‌نویسی پایتون^۳ نسخه ۳.۹ پیاده‌سازی شد. به‌منظور بهبود دقت و عملکرد مدل، بهینه‌سازی پارامترها از طریق جستجوی شبکه‌ای^۴ همراه با اعتبارسنجی متقابل پنج مرحله‌ای^۵ انجام گرفت. پارامترهایی که طی فرآیند بهینه‌سازی مورد بررسی قرار گرفتند شامل حداکثر عمق درخت‌ها^۶ با مقادیر ۵، ۷ و ۱۰، نسبت ویژگی‌های انتخابی در هر گره^۷ با مقادیر ۰/۴۸، ۰/۵ و ۰/۵۲، حداقل تعداد نمونه لازم برای تقسیم

^۸ Min_Samples_Split

^۹ N_Estimators

^{۱۰} Squared_Error

^{۱۱} Root Mean Square Error

^{۱۲} Nash-Sutcliffe Efficiency

^{۱۳} Kling-Gupta Efficiency

^{۱۴} Pearson Correlation Coefficient

^{۱۵} Normalized Rmse

^۱ Min-Max Scaler

^۲ Z-score

^۳ Python

^۴ Grid Search Cv

^۵ Fold Cross-Validation-5

^۶ Max_Depth

^۷ Max_Features

تمامی این تحلیل‌ها با استفاده از نسخه ۳.۹ زبان برنامه‌نویسی پایتون و کتابخانه‌های scikit-learn نسخه ۱.۲ و SHAP نسخه ۰.۴۲ اجرا شدند که ترکیب آن‌ها امکان تحلیل جامع و تفسیرپذیری بالای مدل را فراهم می‌آورد.

نتایج و بحث

شاخص‌های عملکرد مدل

میانگین دبی مشاهده‌شده داده‌های نهایی که شامل ۴۳۸ رکورد بود برابر ۱۰ لیتر بر ثانیه محاسبه گردیده است. نتایج ارزیابی مدل جنگل تصادفی در مرحله‌های آموزش و آزمون در جدول ۱ ارائه شده است. نتایج نشان می‌دهند که مدل عملکرد قابل قبولی داشته و تفاوت معقولی بین مرحله آموزش و آزمون وجود دارد که نشان‌دهنده عدم وجود بیش‌برازش شدید است.

بهینه‌سازی پارامترها

تحلیل تأثیر پارامترهای مختلف بر عملکرد مدل نشان داد که بهترین عملکرد برای حداکثر عمق درخت در مقدار ۱۰ حاصل شد که باعث کاهش خطای RMSE از ۴/۹ به ۴/۷ گردید. افزایش تعداد درختان تا ۵۰۰ نیز بهبود جزئی در عملکرد مدل ایجاد کرد که RMSE را از ۴/۷۵ به ۴/۷۴ کاهش داد. همچنین، مقدار ۰/۵ برای پارامتر نسبت ویژگی‌های انتخابی بهترین نتایج را به همراه داشت و برای حداقل نمونه‌های لازم برای تقسیم گره، مقدار ۵ بهینه تشخیص داده شد. این تنظیمات پارامتری بهینه مدل را قادر می‌سازد تا پیش‌بینی‌های دقیق‌تر و قابل اعتمادتری ارائه دهد.

تحلیل اهمیت متغیرها

تحلیل نتایج SHAP و Permutation Importance ترتیب اهمیت متغیرها را به شکل زیر نشان داد. جدول ۲ نتایج تحلیل SHAP را ارائه می‌کند که در آن متغیر ارتفاع با

$$R = \frac{n \sum xy - (\sum x)(\sum y)}{\sqrt{[n \sum x^2 - (\sum x)^2][n \sum y^2 - (\sum y)^2]}} \quad (۴)$$

$$NRMSE = \frac{\sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}}{y_{max} - y_{min}} \quad (۵)$$

در معادلات بالا x و y ، مقادیر مجموعه داده‌ها، y_{max} بیشینه مقدار مشاهده شده، y_{min} کمینه مقدار مشاهده شده، y_i مقدار مشاهده شده (واقعی) برای نمونه i ام، $\hat{y}_i$ مقدار پیش‌بینی شده برای نمونه i ام، $\bar{y}$ میانگین مقادیر مشاهده شده، $\bar{\hat{y}}$ میانگین مقادیر پیش‌بینی شده، α ضریب همبستگی پیرسون بین مقادیر مشاهده شده و پیش‌بینی شده، α نسبت ضریب تغییرات پیش‌بینی به ضریب تغییرات مشاهده شده (معیار تغییرپذیری) و β نسبت میانگین پیش‌بینی به میانگین مشاهده شده (معیار بایاس) هستند.

تحلیل حساسیت و تفسیرپذیری

برای درک بهتر رفتار مدل و تعیین اهمیت متغیرها در پیش‌بینی، از سه روش استفاده شد. نخست، روش ارزش شاپلی با بهره‌گیری از تبیین گر کرنل^۱ برای تحلیل سهم هر متغیر در خروجی مدل به کار گرفته شد که محاسبه مقادیر شاپلی را در چارچوب نظری بازی‌ها امکان‌پذیر می‌سازد و تفسیر دقیقی از تأثیرات فردی متغیرها ارائه می‌دهد. دوم، اهمیت متغیرها با روش اهمیت ویژگی بر پایه جابه‌جایی ارزیابی شد که تأثیر هر ویژگی را بر عملکرد کلی مدل با اندازه‌گیری کاهش دقت پس از جابه‌جایی مقادیر آن مشخص می‌کند. سوم، نمودارهای وابستگی جزئی متغیر^۲ برای نمایش رابطه بین متغیرهای ورودی و دبی پیش‌بینی شده رسم گردید تا شیوه تغییرات پیش‌بینی در پاسخ به تغییرات پارامترها بهتر درک شود.

^۱ Kernel Explainer

^۲ Partial Dependence

میانگین ارزش SHAP برابر ۲/۱۶ و انحراف معیار ۰/۲۹، بالاترین اهمیت را دارد. متغیرهای دما و بارش به ترتیب با میانگین‌های ۰/۵۵ و ۰/۴۴ اهمیت قابل توجهی دارند. همچنین جدول ۳ نتایج Permutation Importance را نشان می‌دهد که در آن نیز ارتفاع با میانگین ۰/۶۷ و انحراف معیار ۰/۱۶ اهمیت برجسته‌ای دارد، در حالی که بارش و دما اهمیت کمتری داشته‌اند. این نتایج هر دو روش را تأیید می‌کند که ارتفاع مهم‌ترین متغیر در برآورد دبی قنات محسوب می‌شود.

جدول ۱. شاخص‌های ارزیابی مدل جنگل تصادفی

Table 1. Evaluation metrics of the random forest model

شاخص Metric	آموزش Training	آزمون Test
RMSE	3.46	4.27
r	0.85	0.76
NSE	0.67	0.54
KGE	0.58	0.51
NRMSE	10/48%	15/19%

جدول ۲. نتایج تحلیل SHAP

Table 2. Results of SHAP analysis

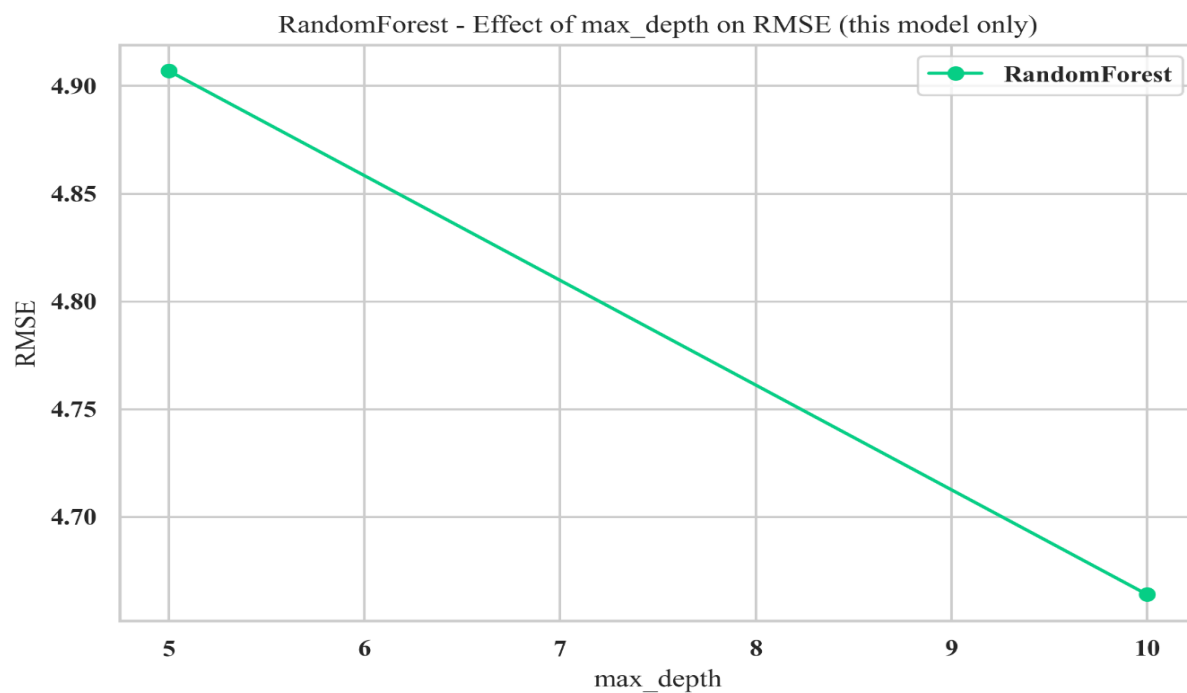
متغیر Variable	میانگین Mean	انحراف معیار Standard Deviation
elevation	2.16	0.29
Temp	0.55	0.07
rain	0.44	0.10

جدول ۳. نتایج Permutation Importance

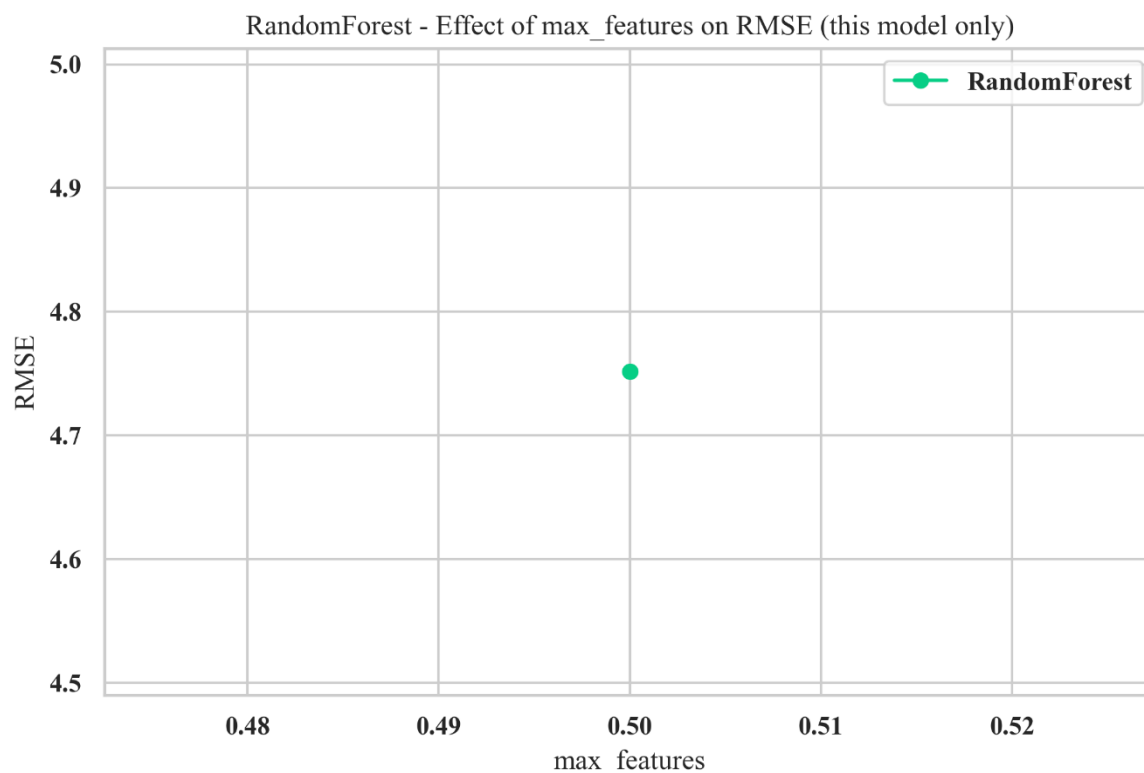
Table 3. Results of Permutation Importance

متغیر Variable	میانگین Mean	انحراف معیار Standard Deviation
elevation	0.67	0.016
rain	-0.013	0.011
Temp	0.001	0.042

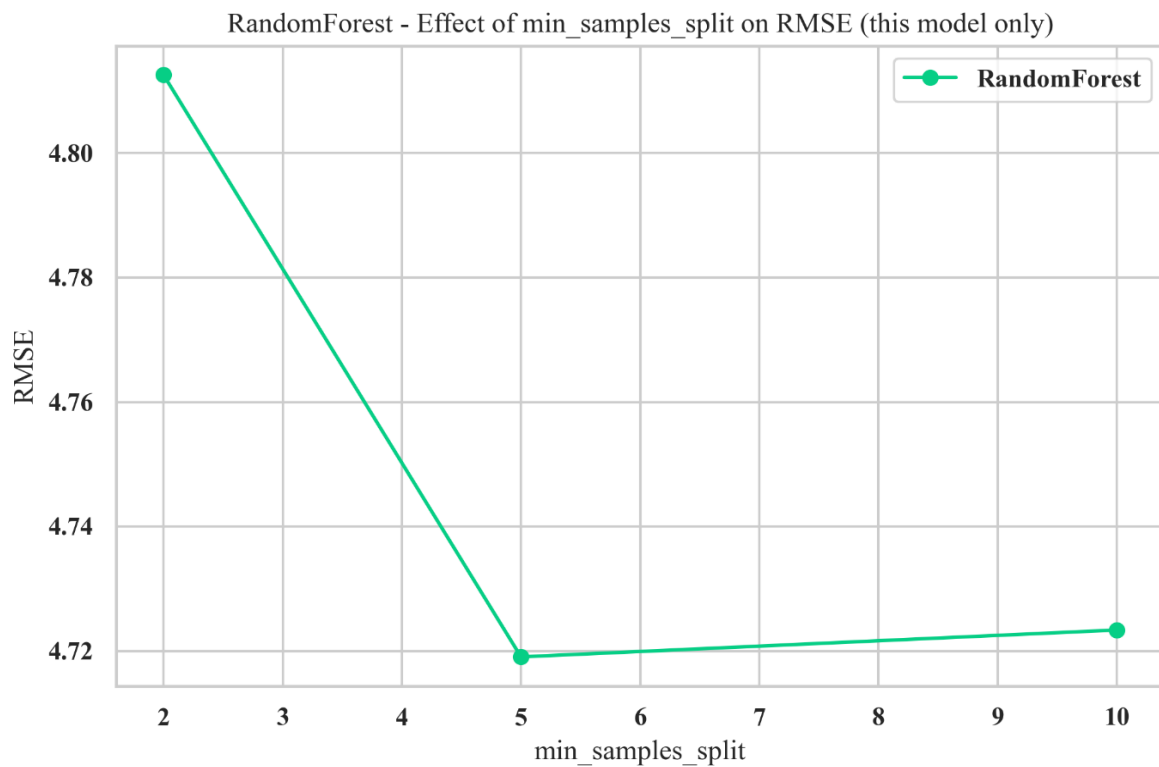
در تحلیل سری زمانی نیز مقایسه مقادیر مشاهده‌شده و پیش‌بینی شده نشان داد مدل توانسته است الگوهای فصلی و روند کلی تغییرات دبی را به خوبی دنبال کند، به‌طوری که تطابق مناسبی بین مقادیر واقعی و پیش‌بینی شده مشاهده شده است و خطا در دوره‌های خشک نسبتاً کم بوده است. شکل‌های ۲ تا ۹ نتایج یادشده را نمایان می‌کند.



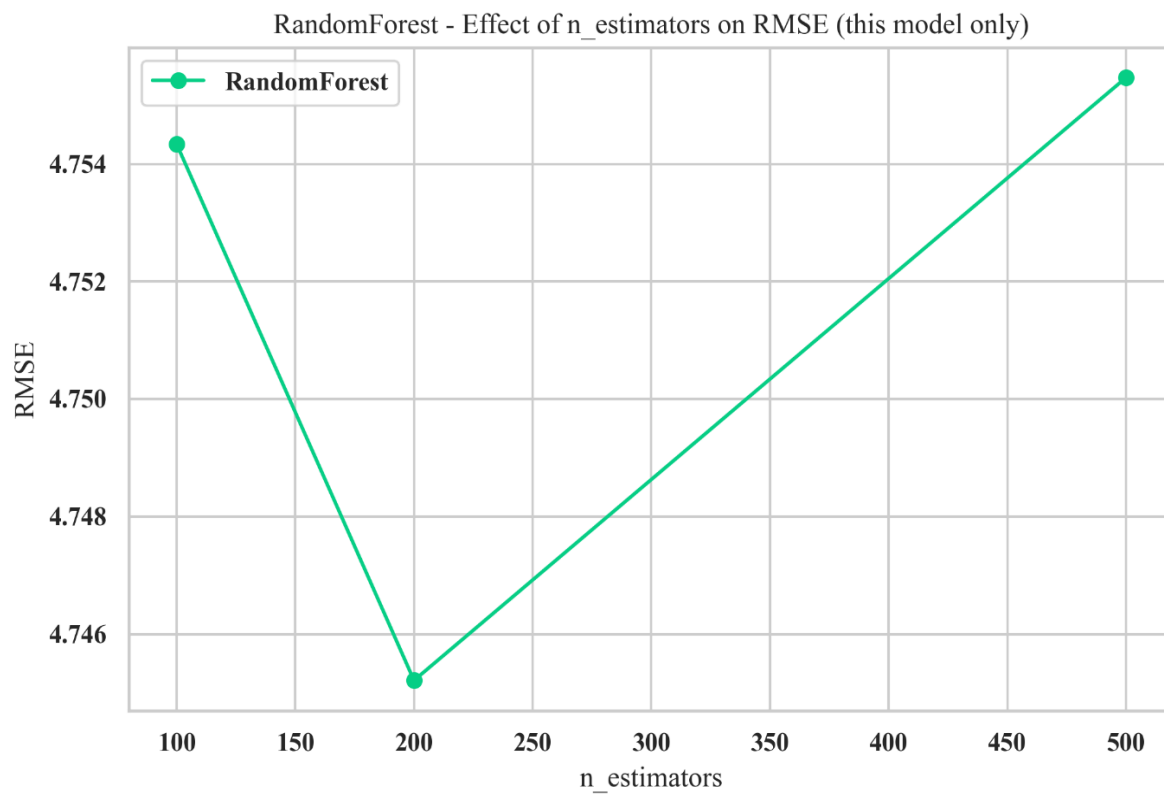
شکل ۲. اثر max_depth بر RMSE
Fig 2. Effect of max_depth on RMSE



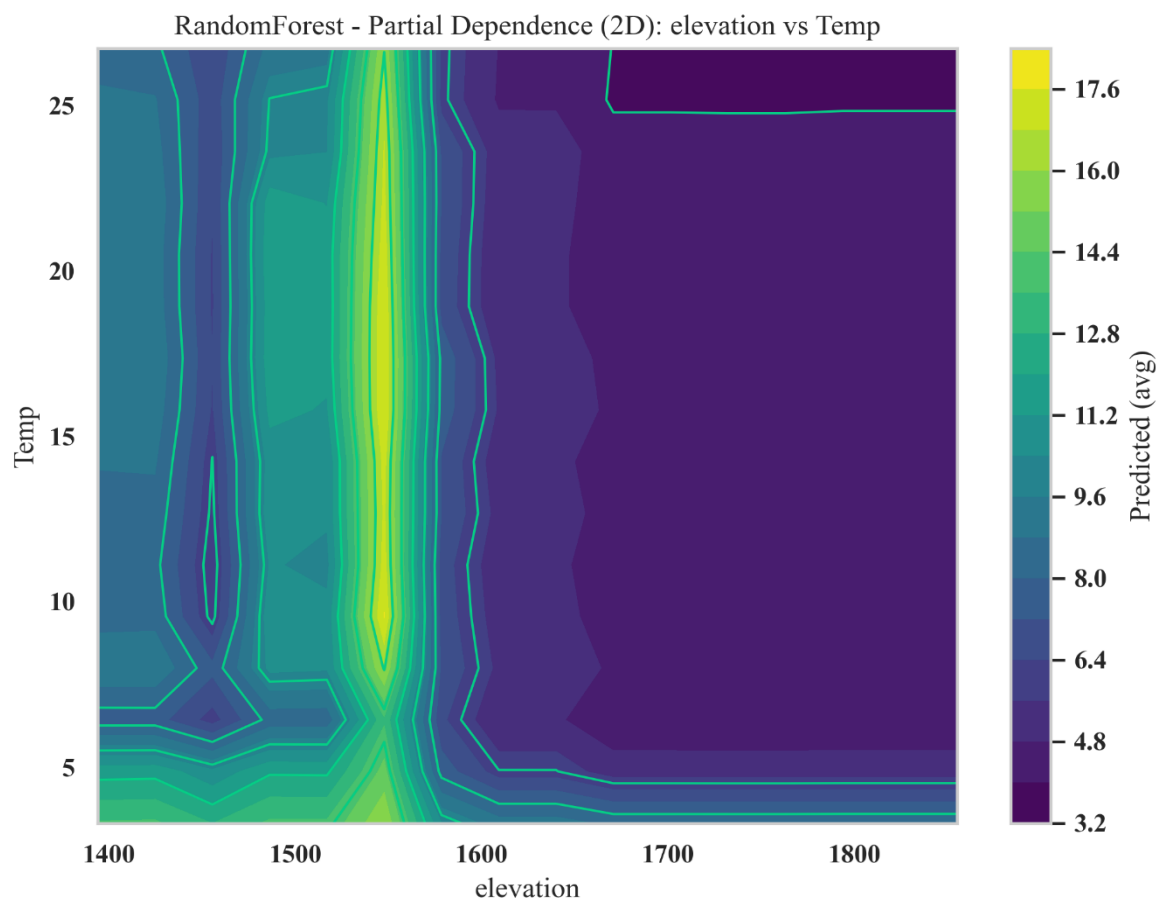
شکل ۳. اثر max_features بر RMSE
Fig 3. Effect of max_features on RMSE



شکل ۴. اثر min_samples_split بر RMSE
Fig 4. Effect of min_samples_split on RMSE

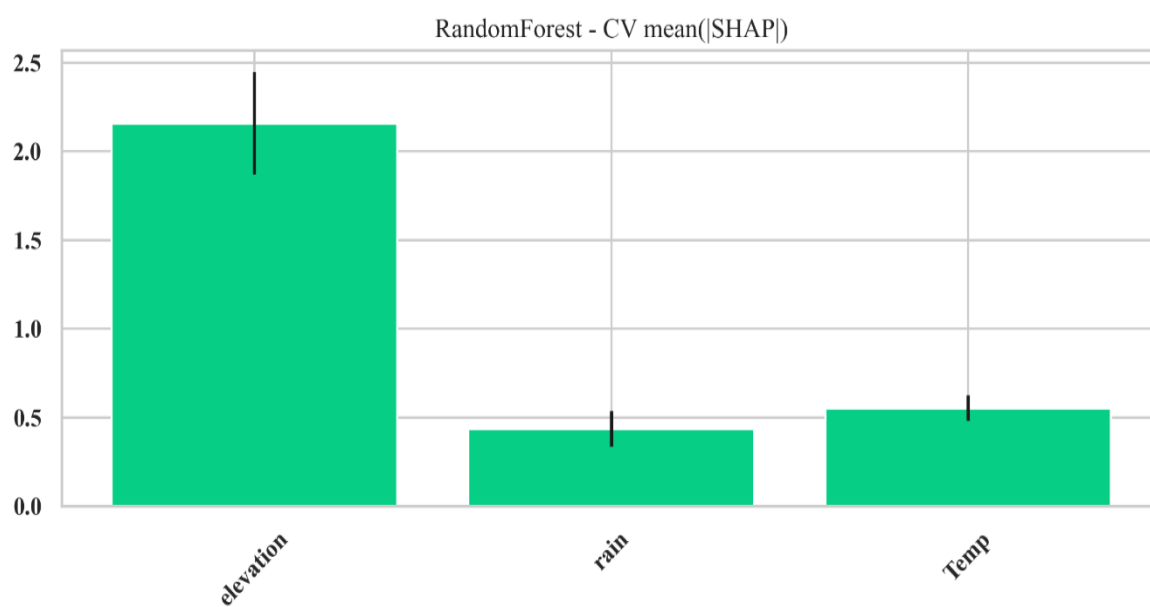


شکل ۵. اثر n_estimators بر RMSE
Fig 5. Effect of n_estimators on RMSE



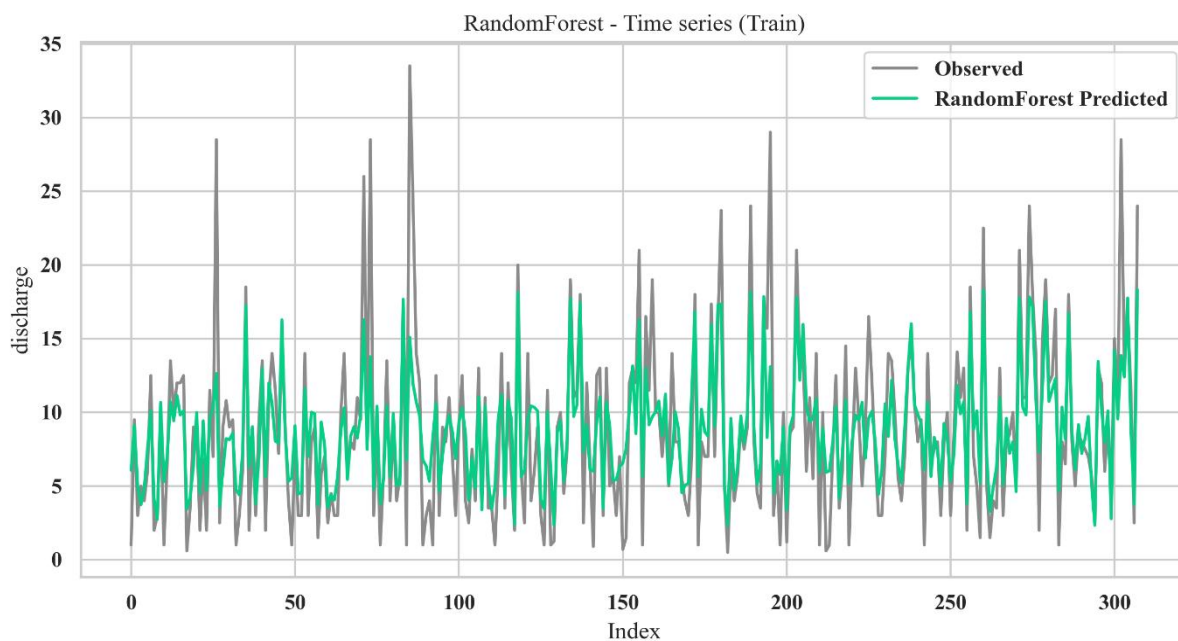
شکل ۶. Partial Dependence (elevation vs Temp)

Fig 6. Partial Dependence (elevation vs Temp)

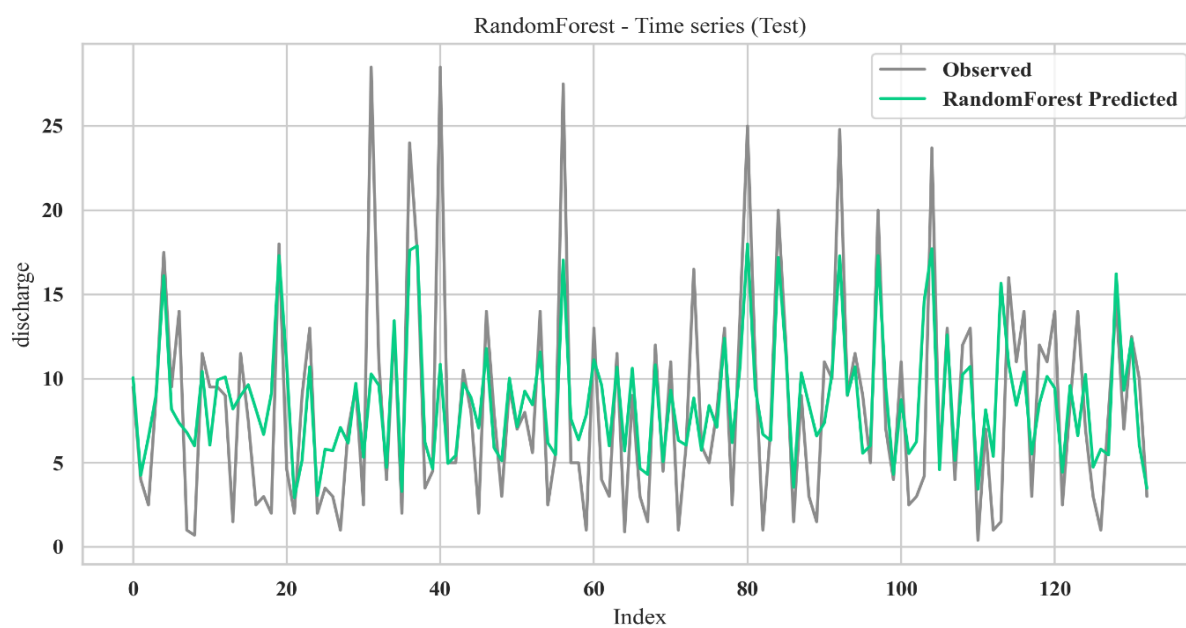


شکل ۷. CV mean (SHAP)

Fig 7. CV mean (SHAP)



شکل ۸. سری زمانی آموزش
Fig 8. Time series of the training



شکل ۹. سری زمانی آزمون
Fig 9. Time series of the test

تأثیر فوری چندانی بر سفره‌های عمیق آب زیرزمینی ندارند.

مدل RF در این تحقیق دارای نقاط قوتی از جمله مقاومت در برابر بیش‌برازش بود، به‌طوری که تفاوت معقولی بین عملکرد در فاز آموزش ($NSE=0.67$) و آزمون ($NSE=0.54$) وجود داشت. همچنین تفسیرپذیری مدل با استفاده از مقادیر SHAP بسیار تقویت شد که اطلاعات ارزشمندی درباره مکانیزم‌های فیزیکی کنترل‌کننده دبی قنات ارائه نمود. عملکرد پایدار مدل در شرایط اقلیمی مختلف، از جمله در دوره خشک‌سالی سال ۱۳۹۶ با حفظ هم‌بستگی قابل قبول ۰/۷۶، نشان‌دهنده قدرت تطبیق آن است. علی‌رغم محدودیت داده‌ای در بازه ۱۲ سال، مدل توانست عملکرد رضایت‌بخشی از خود نشان دهد.

با این حال، این مطالعه محدودیت‌هایی نیز دارد که از جمله آن‌ها می‌توان به حساسیت مدل به کیفیت و کمیت داده‌های ورودی، عدم در نظر گرفتن تأخیرهای زمانی بین بارش و تغییرات دبی، محدودیت مکانی مطالعه که تنها به دشت قاین اختصاص یافته، و حذف متغیرهای زمین‌شناسی مؤثر مانند ویژگی‌های لیتولوژیکی و ساختاری آبخوان اشاره کرد که می‌تواند در دقت پیش‌بینی تأثیرگذار باشد.

از منظر کاربردهای عملی، این پژوهش برای مدیریت منابع آب در مناطق خشک اهمیتی ویژه دارد. از جمله می‌تواند در برنامه‌ریزی منابع آب برای برآورد دقیق‌تر دبی قنات در بلندمدت و پیش‌بینی روند برداشت، که با کاهش ۲۰ تا ۳۰ درصدی دبی در دشت قاین همراه است (Maghrebi et al., 2023)، کمک کند. همچنین مدل قابلیت پیش‌بینی تأثیرات خشک‌سالی بر عملکرد قنات را دارد که به مدیران آب امکان اتخاذ تدابیر پیشگیرانه را می‌دهد. بهینه‌سازی بهره‌برداری از منابع آب، از طریق تخصیص بهتر و کاهش میزان پمپاژ چاه‌ها تا ۱۵ درصد، یکی دیگر از کاربردهای مهم این مدل است.

نتایج این مطالعه با مقادیر NSE برابر با ۰/۵۴ و RMSE معادل ۴/۲۷ بیانگر عملکرد مطلوب الگوریتم جنگل تصادفی است که در مقایسه با سایر پژوهش‌های داخلی و خارجی برجسته می‌باشد. این نتایج نسبت به مطالعه رضائیان‌زاده و همکاران (۲۰۱۳) که در جنوب غربی ایران با NSE برابر ۰/۴۵ گزارش‌شده بود، برتری دارد (Rezaeian-Zadeh et al., 2013). در مقایسه با مطالعات بین‌المللی، نتایج حاضر قابل قیاس با پژوهش کراتزرت و همکاران (۲۰۱۹) است که در حوضه‌های بزرگ کانادا با NSE برابر ۰/۷ گزارش شده، هرچند باید به تفاوت در مقیاس مطالعه و نوع منبع آبی (رودخانه در برابر قنات) توجه داشت (Kratzert et al., 2019). در مقایسه با مدل‌های مفهومی مانند هایدروس که RMSE تقریباً ۶ لیتر بر ثانیه را گزارش کرده بودند (Sedghi and Zhan, 2024)، الگوریتم جنگل تصادفی با استفاده از داده‌های محدود دوازده‌ساله عملکرد بهتری نشان داده است که احتمالاً ناشی از توانایی RF در مدیریت عدم قطعیت و روابط غیرخطی پیچیده میان ورودی‌ها است.

یکی از یافته‌های مهم این پژوهش شناسایی ارتفاع به‌عنوان مهم‌ترین متغیر مؤثر بر دبی قنات با میانگین ارزش SHAP برابر ۲/۱۶ است که با مطالعات پیشین (Smith and Samuel, 2009) همخوانی دارد. از نظر علمی این امر قابل توجه است زیرا ارتفاع بر تغذیه آبخوان‌ها تأثیرگذار است و قنات واقع در ارتفاعات بالاتر معمولاً به منابع آبخوانی با تغذیه بهتر دسترسی دارند. همچنین ارتفاع بالاتر موجب ایجاد شیب هیدرولیکی مؤثرتری برای جریان آب شده و به دلیل کاهش دما و افزایش رطوبت، تبخیر کمتر در این مناطق رخ می‌دهد. نکته غیرمنتظره تأثیر منفی جزئی بارش با مقدار ۰/۱۳- بود که احتمالاً به دلیل تأخیر زمانی یک تا سه ماهه بین بارش و تأثیر آن بر دبی قنات است که در این مطالعه مدل‌سازی نشده است. علاوه بر این، در مناطق خشک بارش‌های کوتاه‌مدت

منابع آب و بهبود برنامه‌ریزی و مدیریت بهره‌برداری از قنات کمک نمایند و به حفظ پایداری منابع آبی و تقویت امنیت غذایی در مناطق خشک بیانجامند. علاوه بر این، چارچوب ارائه شده می‌تواند توسط محققان برای مطالعات مشابه در سایر مناطق خشک مورد استفاده قرار گیرد.

پیشنهادهای پژوهش‌های آینده

بر اساس نتایج و محدودیت‌های شناسایی شده، پیشنهاد زیر برای مطالعات آتی ارائه می‌شود. در زمینه توسعه مدل‌ها، استفاده از مدل‌های هیبریدی همچون ترکیب الگوریتم جنگل تصادفی با شبکه‌های عصبی بلندمدت^۱ (به اختصار LSTM) برای بهبود پیش‌بینی‌های بلندمدت توصیه می‌شود. همچنین کاربرد شبکه‌های عصبی عمیق^۲ برای مدل‌سازی روابط پیچیده‌تر و مدل‌های فضایی-زمانی که وابستگی‌های مکانی و زمانی را در نظر می‌گیرند، از دیگر رویکردهای مهم آینده‌اند. در راستای بهبود داده‌ها، ادغام داده‌های ماهواره‌ای مانند مودیس^۳ از طریق سنجش از دور برای پوشش مکانی بهتر، گنجاندن متغیرهای اضافی شامل داده‌های زمین‌شناسی، کیفیت آب و اثرات انسانی، و جمع‌آوری سری‌های زمانی بلندمدت جهت افزایش قابلیت تعمیم داده‌ها توصیه می‌شود.

از نظر توسعه کاربردها، گسترش مطالعه به سایر دشتهای خراسان جنوبی و توسعه سیستم‌های پشتیبان تصمیم برای مدیریت قنات اهمیت دارد. همچنین بررسی تأثیرات تغییر اقلیم و سناریوهای آن بر عملکرد آینده قنات می‌تواند به مدیریت بهتر منابع آب کمک نماید. در حوزه جنبه‌های اجتماعی-اقتصادی، تحلیل اقتصادی هزینه-فایده به‌کارگیری مدل‌های پیشرفته، مطالعه پذیرش فناوری‌های نوین توسط جوامع محلی و توسعه

مقایسه با مطالعات مشابه در مناطق خشک دیگر، مانند پژوهش نوروزی قوش‌بلغ و ندری (۱۳۹۷) که با استفاده از الگوریتم جنگل تصادفی توانسته‌اند پیش‌بینی قابل قبولی از تراز آب زیرزمینی با ضریب تعیین $R^2 \approx 0.8$ ارائه دهند، نشان‌دهنده انطباق بالای این الگوریتم در شرایط اقلیمی خشک و قابلیت کارآمدی آن در زمینه پیش‌بینی پارامترهای هیدرولوژیکی است (Norouzi Qoshblagh and Naderi, 2018).

نتیجه‌گیری

این پژوهش کارایی الگوریتم جنگل تصادفی را در برآورد دبی قنات دشت قاین با دقت قابل قبول ($NSE > 0.5$)، $RMSE = 4.27 \text{ l/s}$ تأیید نمود و فرضیه اصلی مطالعه را به اثبات رساند. مهم‌ترین یافته‌ها شامل عملکرد مطلوب مدل با شاخص‌های NSE برابر 0.54 ، ضریب همبستگی 0.76 و $NRMSE$ معادل $15/19$ درصد بود. همچنین ارتفاع با مقدار میانگین SHAP برابر $2/16$ به‌عنوان مهم‌ترین متغیر شناسایی شد که پس از آن دما و بارش به‌ترتیب با مقادیر 0.55 و 0.44 قرار گرفتند.

این مطالعه مشارکت‌های علمی قابل توجهی را در حوزه هیدرولوژی مناطق خشک ارائه می‌دهد از جمله اولین کار جامع در ایران که الگوریتم RF را به‌طور یکپارچه برای برآورد دبی قنات به کار برده است. همچنین تحلیل حساسیت به‌کمک روش‌های پیشرفته SHAP و Permutation Importance در تفسیر مدل انجام شده و رویکردی چند متغیره شامل متغیرهای اقلیمی، مکانی و زمانی ادغام گردیده است. تمرکز ویژه‌ای بر داده‌های محلی دشت قاین که پیش‌تر کمتر مطالعه شده بود صورت گرفته است.

پیامدهای عملی این پژوهش برای سیاست‌گذاران، مدیران منابع آب و جوامع محلی حائز اهمیت است. از جمله، نتایج می‌توانند در اتخاذ تصمیمات آگاهانه برای تخصیص

¹ Long Short-Term Memory

² Deep Learning

³ MODIS

behaviors via machine learning applied to large-sample datasets. *Hydrology and Earth System Sciences*, 23(12), 5089-5110.

<https://doi.org/10.5194/hess-23-5089-2019>

Maghrebi, M., Noori, R., Sadegh, M., Sarvarzadeh, F., Akbarzadeh, A. E., Karandish, F.,... & Taherpour, H. (2023). Anthropogenic decline of ancient, sustainable water systems: qanats. *Groundwater*, 61(1), 139-146.

<https://doi.org/10.1111/gwat.13248>

Norouzi Qoshblagh, H., & Naderi, A. (2018). Prediction of groundwater level in the Bukan plain using fuzzy logic, random forest, and neural network models. *Journal of Range and Watershed Management*, 71(3), 829-

845. <https://doi.org/10.22059/jrwm.2018.68924>. [In Persian]

Prieto, C., Le Vine, N., Kavetski, D., García, E., & Medina, R. (2019). Flow prediction in ungauged catchments using probabilistic random forests regionalization and new statistical adequacy tests. *Water Resources Research*, 55(5), 4364-4392.

<https://doi.org/10.1029/2018WR023254>

Ramezani Feriz, A. H. and Ramezani, Y. (2025). Evaluation of the accuracy of artificial intelligence models in predicting qanat discharge in arid and semi-arid regions (Case study: Qaen Plain). *Journal of Aquifer and Qanat*, (), . <https://doi.org/10.22077/jaaq.2025.10072.1124>. [In Persian]

Rezaeian-Zadeh, M., Tabari, H., & Abghari, H. (2013). Prediction of monthly discharge volume by different artificial neural network algorithms in semi-arid regions. *Arabian Journal of Geosciences*, 6(7), 2529-2537. <https://doi.org/10.1007/s12517-011-0517-y>

Samani, S., Vadiati, M., Kisi, O., Ghasemi, L., & Farajzadeh, R. (2024). Qanat discharge prediction using a comparative analysis of machine learning methods. *Earth Science Informatics*, 17(5), 4597-4618. <https://doi.org/10.1007/s12145-024-01409-0>

Sedghi, M. M., & Zhan, H. (2024). Discharge variations of the qanat near an ephemeral stream. *Journal of Hydrology*, 637, 131367.

<https://doi.org/10.1016/j.jhydrol.2024.131367>

Tang, T., Liang, Z., Hu, Y., Li, B., & Wang, J. (2020). Research on flood forecasting based on flood hydrograph generalization and random forest in the Qiushui River basin, China. *Journal of Hydroinformatics*, 22(6), 1588-1602.

<https://doi.org/10.2166/hydro.2020.147>

UNESCO World Heritage Centre. (2016). Persian Qanat. Retrieved from

<https://whc.unesco.org/en/list/1506>

چارچوب‌های سیاست‌گذاری برای مدیریت پایدار قنات از موضوعات مورد توجه است.

در پایان، این پژوهش نشان داد که الگوریتم جنگل تصادفی ابزاری مؤثر و قابل اعتماد برای برآورد دبی قنات در مناطق خشک و نیمه‌خشک است. کاربرد این مدل در مواجهه با چالش‌های بحران آب و تغییرات اقلیمی در خراسان جنوبی می‌تواند به تقویت پایداری جوامع محلی کمک کند. نتایج به‌دست‌آمده الگویی برای سایر مناطق مشابه در جهان فراهم کرده و امید است گامی مؤثر در جهت مدیریت هوشمند منابع آب محدود در مناطق خشک برداشته شود.

منابع

Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5-32.

<https://doi.org/10.1023/A:1010933404324>

Dastourani, M., Mamirabadizadeh, M., & Mohammadie, K. (2025). The Impact Assessment of Changes in Agricultural Cropping Patterns on the Quantitative Status of Groundwater Resources in the Qaen Plain Aquifer Using the MODFLOW Model. *Iranian Journal of Irrigation & Drainage*, 19(4), 537-554. [In Persian]

Farzaneh, M. and Banimostafaarab, F. (2025). Qanat and climate resilience Investigating the climate legislations in developing countries from disaster risk management perspective. *Journal of Aquifer and Qanat*, 6(1), 67-86.

<https://doi.org/10.22077/jaaq.2025.8869.1097>. [In Persian]

Foruzanmehr, M. and Khozaymehnezhad, H. (2023). Investigation of spatial and temporal changes of qualitative parameters of Qaen plain aquifer using interpolation methods. *Journal of Aquifer and Qanat*, 4(1), 1-15. <https://doi.org/10.22077/jaaq.2023.5130.1045>. [In Persian]

Hejazi, A., & Marzbani, M. (2015). Estimation of maximum runoff height and discharge using curve number method case study: Sarab Darehshahr Watershed. *Hydro Geomorphol*, 2(5), 63-81.

Kratzert, F., Klotz, D., Shalev, G., Klambauer, G., Hochreiter, S., & Nearing, G. (2019). Towards learning universal, regional, and local hydrological