



## Comparative Evaluation of Numerical Models and Machine Learning Algorithms for Groundwater Level Prediction: A Case Study of Birjand Plain

Mehran Malekinia<sup>1</sup>, Amir Hossein Javid<sup>1</sup>, Seyedeh Hoda Rahmati<sup>2</sup>✉, Alireza Jahangir<sup>3</sup>

1. Ph.D. Student, Department of Environmental Engineering, Science and Research Branch, Islamic Azad University, Tehran, Iran.
2. Professor, Department of Natural Resources and Environment, SR.C., Islamic Azad University, Tehran, Iran.
3. Assistant Professor, Department of Natural Resources and Environment, SR.C., Islamic Azad University, Tehran, Iran.
4. Assistant Professor, Department of Water Science and Engineering, Payame Noor University, Tehran, Iran.

✉Corresponding Author: [Rahmati@iau.ac.ir](mailto:Rahmati@iau.ac.ir)

Submit Date  
24 November 2025

Revise Date  
03 February 2025

Accept Date  
03 February 2025

**Keywords:**  
*Machine Learning,  
Ensemble Hybrid Model,  
LSTM Neural Network,  
Time Series,  
Birjand Plain.*

### Extended abstract

#### Introduction

Accurate estimation of groundwater storage variations is essential for sustainable water resources management in arid and semi-arid regions. The Birjand Plain has experienced persistent groundwater decline due to intensive agricultural abstraction and climatic variability, highlighting the need for reliable predictive tools. Traditional physically based numerical models such as MODFLOW provide a comprehensive representation of aquifer dynamics but often struggle to reproduce local-scale heterogeneity. Conversely, data-driven machine learning (ML) approaches offer high predictive accuracy at observation points but lack physical interpretability.

The primary objective of this study is to comparatively evaluate numerical and machine learning models for groundwater level (GWL) prediction and to develop an innovative hybrid stacking framework that integrates physically informed simulations with data-driven algorithms to enhance prediction accuracy and robustness.

**Cite this article:** Malekinia, M., Javid, A.H., Rahmati, S.H., & Jahangir, A. (2025). Comparative evaluation of numerical models and machine learning algorithms for groundwater level prediction: A case study of Birjand plain. *Journal of Aquifer and Qanat*, 6 (2), 257-281. DOI: 10.22077/jaaq.2026.10568.1133.



Copyright: © 2025 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/>).

## Materials and Methods

Groundwater level data from 17 observation wells in the Birjand Plain were collected semiannually (spring and autumn) from October 2008 to September 2023 (544 observations) from the South Khorasan Regional Water Authority. Associated hydrological variables included precipitation, air temperature, groundwater abstraction rates, and aquifer boundary inflow–outflow data.

Data preprocessing involved normalization, outlier detection using the Z-score method ( $|Z| > 3$ ), and application of the Savitzky–Golay smoothing filter to reduce noise. A conceptual hydrogeological model of the aquifer was developed and implemented in MODFLOW with a spatial resolution of  $250 \times 250$  m for calibration and validation.

For data-driven modeling, Long Short-Term Memory (LSTM) neural networks and Random Forest (RF) algorithms were applied, followed by the development of a stacking ensemble hybrid model that incorporated calibrated MODFLOW outputs as physically informed features. Sensitivity analysis was conducted using SHAP (Shapley Additive Explanations) within MATLAB to quantify the contribution of each predictor variable.

Uncertainty analysis was performed using Monte Carlo simulations by introducing Gaussian noise ( $\mu = 0$ ,  $\sigma = 5\%$  of original values) to generate 95% confidence intervals (5th–95th percentiles). Model performance was evaluated using  $R^2$ , RMSE, Nash–Sutcliffe Efficiency (NSE), Kling–Gupta Efficiency (KGE), and Mean Absolute Percentage Error (MAPE), with  $R^2$ , RMSE, and KGE serving as primary selection criteria.

## Results and Discussion

The stacking ensemble hybrid model achieved the highest predictive accuracy on the test dataset ( $R^2 = 0.978$ , RMSE = 0.032 m, NSE = 0.978, KGE = 0.946). In comparison, the standalone MODFLOW model during validation yielded  $R^2 = 0.968$ , RMSE = 3.17 m, NSE = 0.993, and KGE = 0.975.

The observed differences arise mainly from modeling scale contrasts. MODFLOW simulates groundwater flow at the aquifer scale using a structured grid, thereby smoothing local heterogeneity. In contrast, ML models directly fit time-series data at key wells, capturing localized fluctuations more precisely.

Taylor diagram analysis confirmed the superiority of the hybrid approach (correlation coefficient = 0.978; normalized standard deviation = 0.812), outperforming LSTM (correlation = 0.977; std = 0.845) and RF (correlation = 0.940; std = 1.344).

SHAP-based sensitivity analysis identified Well 1 as the most influential observation point (relative importance = 0.28). Monte Carlo uncertainty assessment demonstrated that the hybrid model exhibited the narrowest confidence intervals, indicating superior robustness under input variability.

Given the annual groundwater decline of approximately 40–60 cm and hydraulic conductivity variability (5–45 m/day), the hybrid framework effectively captured both large-scale aquifer dynamics and local-scale variability. Results suggest sustainable abstraction potential in eastern zones, while western areas require immediate management interventions, including 30–40% reduction in agricultural pumping, smart irrigation systems, and artificial recharge implementation.

## Conclusion

This study presents a robust hybrid modeling framework integrating machine learning algorithms with calibrated MODFLOW simulations for groundwater level prediction in the Birjand Plain. The hybrid approach significantly enhances predictive accuracy while preserving physical interpretability, overcoming limitations associated with standalone numerical or purely data-driven models.

Advanced preprocessing techniques—including Savitzky–Golay filtering and controlled Gaussian noise injection—improved generalization in a heterogeneous aquifer experiencing persistent decline. Sensitivity and uncertainty analyses provided practical insights for monitoring network optimization and adaptive groundwater management.

The proposed framework aligns with Integrated Water Resources Management (IWRM) principles and offers a transferable decision-support tool for sustainable groundwater management in semi-arid regions. Future research should incorporate high-resolution satellite observations and real-time monitoring systems to further enhance predictive capability and develop AI-based early warning systems for groundwater depletion.



## مقایسه مدل‌های عددی و الگوریتم‌های یادگیری ماشین با رویکرد پیش‌بینی تراز آب زیرزمینی: مطالعه‌ای موردی در دشت بیرجند

مهران ملکی نیا<sup>۱</sup> | امیرحسین جاوید<sup>۲</sup> | سیده هدی رحمتی<sup>۳</sup> | علیرضا جهانگیر<sup>۴</sup>

۱. دانشجوی دکتری، گروه مهندسی محیط زیست، دانشگاه آزاد اسلامی واحد علوم و تحقیقات، تهران، ایران.
۲. استاد، گروه مهندسی محیط زیست، دانشگاه آزاد اسلامی واحد علوم و تحقیقات، تهران، ایران.
۳. استادیار، گروه مهندسی محیط زیست، دانشگاه آزاد اسلامی واحد علوم و تحقیقات، تهران، ایران.
۴. استادیار گروه علوم و مهندسی آب، دانشگاه پیام نور، تهران، ایران.

✉ نویسنده مسئول: [Rahmati@srbiau.ac.ir](mailto:Rahmati@srbiau.ac.ir)

### چکیده

با توجه به اهمیت تعیین حجم ذخایر منابع آب زیرزمینی برای مدیریت پایدار آب در مناطق خشک و نیمه‌خشک، این مطالعه به بررسی رویکردهای پیش‌بینی و شبیه‌سازی تراز آب زیرزمینی با استفاده از مدل‌های یادگیری ماشین (LSTM, RF) و مدل ترکیبی Stacking Ensemble در مقایسه با مدل عددی MODFLOW در ۱۷ چاه مشاهده‌ای دشت بیرجند پرداخت. داده‌های سری زمانی سطح آب زیرزمینی از مهر ۱۳۸۷ تا شهریور ۱۴۰۲ در مجموع ۵۴۴ نمونه با فرکانس ۲ بار در سال (بهار و پاییز) همراه با داده‌های هیدرولوژیکی مربوطه، با معیارهای آماری تحلیل شدند. تحلیل حساسیت با روش‌های SHAP و مونت‌کارلو چاه‌های کلیدی را شناسایی کرد. مدل ترکیبی Stacking Ensemble، با تلفیق خروجی‌های کالیبره‌شده مدل عددی MODFLOW و پیش‌بینی‌های مدل‌های یادگیری ماشین، بالاترین دقت را در مجموعه آزمایشی با  $R^2 = 0.978$ ،  $RMSE = 0.032$  متر،  $NSE = 0.978$ ،  $KGE = 0.946$  نشان داد، درحالی‌که مدل عددی در دوره صحت‌سنجی به  $R^2 = 0.968$ ،  $RMSE = 3.17$  متر،  $NSE = 0.993$  و  $KGE = 0.975$  دست یافت. این تفاوت ناشی از مقیاس مدل عددی (کل آبخوان با رزولوشن  $250 \times 250$  متر) در مقابل مدل‌های یادگیری ماشین (تمرکز بر سری‌های زمانی چاه‌های کلیدی) است. مدل عددی ناهمگنی محلی را کمتر می‌گیرد، درحالی‌که مدل‌های ML مستقیماً روی داده‌های مشاهده‌ای فیت می‌شوند. نمودارهای تیلور برتری مدل ترکیبی را با ضریب همبستگی  $0.978$  و انحراف معیار نرمال شده  $0.812$  نسبت به LSTM (ضریب همبستگی  $0.977$ ، انحراف معیار  $0.845$ ) و RF (ضریب همبستگی  $0.940$ ، انحراف معیار  $1.344$ ) تأیید کردند. این رویکرد ترکیبی نیاز به داده‌های ورودی گسترده را کاهش می‌دهد، دقت پیش‌بینی را در آبخوان‌های ناهمگن با افت سالانه  $60 - 40$  سانتی‌متر بهبود می‌بخشد و ابزاری کارآمد برای مدیریت پایدار منابع آب در دشت بیرجند و مناطق مشابه با شرایط هیدرولوژیکی پیچیده فراهم می‌کند.

تاریخ دریافت: ۱۴۰۴/۰۹/۰۳  
تاریخ بازنگری: ۱۴۰۴/۱۱/۱۵  
تاریخ پذیرش: ۱۴۰۴/۱۱/۱۵

### کلیدواژه‌ها:

یادگیری ماشین،  
مدل ترکیبی Ensemble  
شبکه عصبی LSTM،  
سری زمانی،  
دشت بیرجند.

## مقدمه

نشان داد که تنها ۱۵ درصد از تحقیقات توضیحاتی برای نتایج مدل‌ها ارائه می‌دهند، این در حالی است که پذیرش یادگیری ماشینی در کاربردهای زیست‌محیطی رو به افزایش است (Torres-Martínez et al., 2024). این کمبود تفسیرپذیری به‌ویژه نگران‌کننده است، زیرا مدل‌های مبتنی بر فیزیک در شرایط نیاز به برون‌یابی عملکرد بهتری دارند، در حالی که مدل‌های یادگیری ماشینی تنها در صورت وجود داده‌های آموزشی کافی عملکرد مطلوبی از خود نشان می‌دهند (Dunnington et al., 2021). با این حال، راه‌حل‌های نوظهور نویدبخش رفع محدودیت‌های تفسیرپذیری هستند (Janssen et al., 2025). نشان دادند که مدل‌های یادگیری ماشینی با محدودیت‌های فیزیکی می‌توانند از طریق مقایسه با ادبیات هیدرولوژی آب زیرزمینی تفسیر شوند، در حالی که (Rad et al., 2024) با موفقیت از روش‌های SHAP<sup>5</sup> برای شناسایی اهمیت متغیرها در مدل‌سازی آلودگی آب زیرزمینی استفاده کردند. این رویکردها نشان می‌دهند که ترکیب محدودیت‌های دانش حوزه با تکنیک‌های هوش مصنوعی قابل تفسیر می‌تواند شکاف تفسیرپذیری در کاربردهای مبتنی بر یادگیری ماشینی برای آب زیرزمینی را پر کند. مطالعات اخیر در سال پیشرفت‌های خوبی در مدل‌های ترکیبی داده‌محور و فیزیکی برای پیش‌بینی سطح و کیفیت آب زیرزمینی نشان داده‌اند؛ برای مثال، (Cui et al., 2024) یک مدل یادگیری عمیق تجمیعی مبتنی بر تجزیه مدال ثانویه با ترکیب LSTM و CNN<sup>6</sup> با ماژول‌های وزن تطبیقی توسعه دادند که عملکرد بهتری نسبت به مدل‌های منفرد در سه حوضه آبخیز چین به دست آورد؛ (Adombi et al., 2024) با اعمال محدودیت‌های علت و معلولی در مدل‌های یادگیری عمیق آگاه از فیزیک (مانند

آب زیرزمینی منبعی حیاتی برای تأمین نیازهای کشاورزی، صنعتی و خانگی، به‌ویژه در مناطق خشک و نیمه‌خشک محسوب می‌شود که وابستگی شدیدی به این منبع برای پایداری اکوسیستم‌های کشاورزی خود دارد (Alao et al., 2024). پیش‌بینی دقیق تراز سطح آب زیرزمینی برای مدیریت پایدار این منبع ارزشمند ضروری است، اما پیچیدگی سیستم‌های هیدرولوژیکی، تنوع عوامل محیطی و محدودیت‌های محاسباتی، این وظیفه را به چالشی بزرگ تبدیل کرده است (Chang et al., 2025). مدل‌های فیزیکی مانند مدل‌های عددی، با ارائه بینش‌های ارزشمند درباره دینامیک آب زیرزمینی، به دلیل هزینه‌های محاسباتی بالا و حساسیت به عدم قطعیت پارامترها، اغلب در دستیابی به دقت بالا با دشواری مواجه هستند (Navarro-Farfán et al., 2024). در مقابل، مدل‌های داده‌محور، از جمله شبکه‌های عصبی مصنوعی<sup>1</sup> (ANN)، ماشین‌های بردار پشتیبان<sup>2</sup> (SVM) و به‌ویژه شبکه‌های حافظه کوتاه‌مدت بلند<sup>3</sup> (LSTM)، به دلیل توانایی در مدل‌سازی روابط غیرخطی و پویایی‌های پیچیده سری‌های زمانی هیدرولوژیکی، عملکرد امیدوارکننده‌ای از خود نشان داده‌اند (Igwebuike et al., 2025). با این حال، این مدل‌ها با محدودیت‌هایی نظیر بیش‌برازش روی داده‌های محدود، کمبود تفسیرپذیری فیزیکی و وابستگی به کیفیت داده‌های ورودی مواجه هستند (Dunnington et al., 2021; Torres-Martínez et al., 2024).

کاربردهای یادگیری ماشینی<sup>4</sup> در مدل‌سازی آب زیرزمینی با چالش‌های قابل توجهی در زمینه تفسیرپذیری مواجه هستند و مطالعات اخیر این شکاف حیاتی را برجسته کرده‌اند. بررسی جامع ۲۳۰ مطالعه کیفیت آب زیرزمینی

<sup>4</sup> Machine Learning

<sup>5</sup> SHapley Additive exPlanations

<sup>6</sup> Convolutional Neural Network

<sup>1</sup>Artificial Neural Network

<sup>2</sup> Support Vector Machine

<sup>3</sup> Long Short-Term Memory

یادگیری ماشین ترکیب می‌کنند، نویدبخش هستند، اما به‌ویژه در شرایط رویدادهای حدی و جریان‌های بالا با محدودیت‌هایی مواجه می‌شوند (Zarei et al., 2024) بود داده‌های هیدرومتری قابل اعتماد، ضرورت مقایسه بین مدل‌های هیدرولوژیکی متعدد با استفاده از منابع داده‌ای جهانی، منطقه‌ای، و محلی را ایجاب می‌کند (Ghebrehiwot & Kozlov., 2019). تحقیقات اخیر بر نیاز به یکپارچه‌سازی بهبودیافته منابع داده‌ای متنوع، بازنمایی بهتر فرآیندهای پیچیده هیدرولوژیکی، و توسعه روش‌های قوی برای کمی‌سازی عدم قطعیت، به‌ویژه متناسب با شرایط داده‌کم، تأکید دارند (Borzi., 2025).

رویکردهای داده‌محور در ایران به سمت استفاده از الگوریتم‌های پیشرفته یادگیری ماشین و مدل‌های ترکیبی سوق یافته‌اند تا دقت پیش‌بینی سطح آب‌های زیرزمینی بهبود یابد (Emamgholizadeh et al., 2025). این مطالعه با ارزیابی الگوریتم‌های تقویتی در دشت‌های شاهرود و بسطام نشان داد که مدل‌های مبتنی بر Boosting مانند XGBoost و CatBoost توانایی بالاتری در مدل‌سازی رفتار غیرخطی تراز آب زیرزمینی نسبت به روش‌های کلاسیک دارند. در ادامه این رویکرد، (Nohani et al., 2025) با تلفیق شبکه‌های عصبی مصنوعی و تبدیل موجک در دشت دلفان، بهبود معناداری در دقت پیش‌بینی گزارش کردند که بر نقش مؤثر مدل‌های ترکیبی در استخراج الگوهای زمانی-مکانی تأکید دارد. در مقیاس محلی تر، (Dastourani et al., 2025) با استفاده از مدل تلفیقی موجک-عصبی فازی در دشت بیرجند، کارایی بالای روش‌های هوشمند ترکیبی را در شبیه‌سازی نوسانات سطح آب زیرزمینی نشان دادند. با وجود موفقیت این مطالعات، استفاده ساخت‌یافته از خروجی مدل‌های عددی به‌عنوان منبع دانش فیزیکی و همچنین

(H-Lin و H-HBV)، تفسیرپذیری را بهبود بخشیدند و افزایش سطح آب زیرزمینی بین ۵/۶ تا ۷۰ سانتی‌متر تا سال را در منطقه کبک کانادا پیش‌بینی کردند؛ (Pourmorad et al., 2024) در یک بررسی جامع تأکید کردند که ترکیب روش‌های هوش مصنوعی مانند ANN، SVM، ANFIS و یادگیری عمیق با تکنیک‌های بهینه‌سازی، دقت پیش‌بینی را در شرایط هیدرولوژیکی خاص به طور چشمگیری افزایش می‌دهد؛ همچنین (Rammohan et al., 2024) با استفاده از الگوریتم XGBoost<sup>۱</sup> و فناوری‌های ژئوپاتیل، به‌دقت ۹۴/۶ درصدی در پیش‌بینی کیفیت آب زیرزمینی برای ۱۲۹ چاه در منطقه تامپل نادو هند دست یافتند؛ این پیشرفت‌ها نشان می‌دهند که تلفیق دانش فیزیکی با مدل‌های داده‌محور، دقت و قابلیت تعمیم‌پذیری را در آبخوان‌های پیچیده بهبود می‌بخشد و نوآوری پژوهش حاضر (ترکیب خروجی‌های مدل عددی با رویکرد استکینگ یادگیری ماشین) را در ادامه همین روند قرار می‌دهد.

مدل‌سازی آب زیرزمینی در مناطق با کمبود داده، به‌ویژه مناطق نیمه‌خشک، با چالش‌های قابل‌توجهی مواجه است که دقت و قابلیت اطمینان مدل‌ها را محدود می‌کند. فقدان داده‌های هیدرولوژیکی جامع و با کیفیت بالا مانعی اساسی برای مدیریت مؤثر منابع آب زیرزمینی ایجاد می‌کند (Borzi, 2025; Ghebrehiwot & Kozlov., 2019). کاربردهای یادگیری ماشین در تحقیقات آب زیرزمینی شکاف‌های روش‌شناختی قابل‌توجهی را نشان می‌دهند، به‌طوری که ۸۳ درصد از مطالعات مراحل پیش‌پردازش حیاتی را نادیده می‌گیرند و تنها ۱۵ درصد توضیحاتی برای تفسیرپذیری مدل‌ها ارائه می‌دهند (Torres- Martínez et al., 2024) رویکردهای مدل‌سازی ترکیبی که مدل‌های هیدرولوژیکی سنتی را با تکنیک‌های

<sup>۱</sup> eXtreme Gradient Boosting

است. افزودن این نویز با هدف جلوگیری از بیش برآزش مدل‌ها، بهبود تعمیم‌پذیری آن‌ها و شبیه‌سازی عدم قطعیت‌های اندازه‌گیری واقعی انجام شد که در نهایت منجر به کاهش - درصدی خطاهای پیش‌بینی در مجموعه آزمون گردید. مدل ترکیبی با استفاده از یک متا-یادگیر جنگل تصادفی<sup>۵</sup>، برای اولین بار پیش‌بینی‌های مدل‌های داده‌محور و فیزیکی را به صورت یکپارچه ترکیب کرده و راهکاری کارآمد برای پیش‌بینی سطح آب زیرزمینی ارائه می‌دهد. عملکرد مدل‌ها با معیارهای جامعی مانند ضریب تعیین ( $R^2$ )، خطای میانگین مربعات ریشه (RMSE)، خطای مطلق میانگین (MAE)، درصد خطای مطلق میانگین (MAPE)، کارایی نش-ساتکلیف (NSE)، و کارایی کلینگ-گوپتا (KGE) ارزیابی شده است. اهداف مطالعه شامل ارزیابی دقت پیش‌بینی مدل‌های منفرد و ترکیبی با استفاده از داده‌های مدل عددی، تحلیل نقاط قوت و محدودیت‌های تلفیق مدل‌های فیزیکی و یادگیری ماشین در چارچوب استتکینگ، و ارائه راهکارهای عملی برای مدیریت پایدار منابع آب زیرزمینی در مناطق نیمه‌خشک با وابستگی بالای کشاورزی است. نتایج این پژوهش نه تنها راهکارهایی برای مدیریت منابع آب در دشت بیرجند فراهم می‌کند، بلکه چارچوبی قابل تعمیم برای مناطق با چالش‌های هیدرولوژیکی مشابه ارائه می‌دهد. ساختار مقاله به این صورت است: بخش دوم داده‌ها و روش‌ها را توضیح می‌دهد، بخش سوم نتایج ارزیابی مدل‌ها را ارائه می‌کند، بخش چهارم یافته‌ها را تحلیل و بحث می‌کند، و بخش پنجم با نتیجه‌گیری و پیشنهادات برای تحقیقات آینده به پایان می‌رسد.

به کارگیری چارچوب‌های ترکیبی در این پژوهش‌ها مورد توجه قرار نگرفته است.

مدل‌سازی سطح آب زیرزمینی در مناطق نیمه‌خشک مانند دشت بیرجند با چالش‌هایی مانند تفسیرپذیری پایین مدل‌های داده‌محور (مانند LSTM) به دلیل نبود توضیحات مبتنی بر اصول هیدرولوژیکی (Torres-Martínez et al., 2024)، استفاده محدود از خروجی‌های مدل‌های عددی مانند به‌عنوان ورودی مدل‌های یادگیری ماشین به دلیل پیچیدگی محاسباتی و کمبود داده‌های دقیق، و نبود مقایسه جامع بین مدل‌های داده‌محور، فیزیکی، و ترکیبی با معیارهای دقیق مانند  $R^2$  و RMSE در چنین مناطقی موجه است (Dunnington et al., 2020). این مطالعه با هدف رفع این چالش‌ها، چارچوبی نوآورانه برای پیش‌بینی سطح آب زیرزمینی در دشت بیرجند ارائه می‌دهد که با تلفیق داده‌های مدل‌های عددی و مدل‌های یادگیری ماشین (LSTM, RF) و یک مدل ترکیبی<sup>۱</sup> مبتنی بر تجمیع مدل‌های یادگیری ماشین مقایسه می‌کند.

نوآوری کلیدی این پژوهش، استفاده از خروجی‌های کالیبره‌شده مدل عددی به‌عنوان ویژگی‌های ورودی اضافی برای مدل‌های یادگیری ماشین شامل شبکه حافظه کوتاه‌مدت بلند<sup>۲</sup>، جنگل تصادفی<sup>۳</sup> و ترکیبی متا-یادگیر است که دانش فیزیکی و هیدرولوژیکی را با تولنایی‌های مدل‌سازی غیرخطی ترکیب می‌کند. این چارچوب با بهره‌گیری از پیش‌پردازش پیشرفته داده‌ها، از جمله فیلتر<sup>۴</sup> صاف‌کننده چندجمله‌ای برای صاف‌سازی سری‌های زمانی و تکنیک افزایش داده از طریق افزودن نویز گاوسی کنترل‌شده (با میانگین صفر و انحراف معیار ۰/۰۵ متر)، دقت پیش‌بینی را در شرایط داده‌ای محدود افزایش داده

<sup>4</sup> Savitzky-Golay

<sup>5</sup> Stacking model using a Random Forest meta-learner

<sup>1</sup> Stacking Ensemble

<sup>2</sup> LSTM

<sup>3</sup> Random Forest

## مواد و روش‌ها

### روش پژوهش

برای انجام این پژوهش، داده‌های سطح آب زیرزمینی در ۱۷ چاه مشاهده‌ای دشت بیرجند با فرکانس دو بار در سال (بهار و پاییز) از سال ۱۳۸۷ تا پایان ۱۴۰۲ (معادل حدود ۱۶ سال کامل، یعنی ۳۲ مشاهده زمانی) از سازمان آب منطقه‌ای خراسان جنوبی اخذ شد که در مجموع ۵۴۴ نمونه را شامل می‌شود. این داده‌ها به همراه اطلاعات هیدرولوژیکی شامل دما، میزان بارندگی سالانه، دبی برداشت‌شده از چاه‌های بهره‌برداری و جبهه‌های ورودی و خروجی آبخوان، طی فرآیند داده‌کاوی و حذف داده‌های پرت با روش  $Z\text{-Score} > 3$  پردازش شدند. با استفاده از این داده‌ها، مدل مفهومی آبخوان دشت بیرجند ساخته شد. پس از مراحل واسنجی<sup>۱</sup> و صحت‌سنجی<sup>۲</sup>، سطح آب زیرزمینی با استفاده از مدل عددی مبتنی بر فیزیک در یک (شبکه  $250 \times 250$  متر) شبیه‌سازی شد. مراحل انجام پژوهش به صورت شماتیک در شکل ۱ ارائه شده است.

برای ارزیابی اهمیت متغیرهای ورودی در مدل‌های یادگیری ماشین توسعه‌یافته، از روش SHAP استفاده شد. این روش با بهره‌گیری از کتابخانه shapley در محیط MATLAB، تأثیر دقیق، منصفانه و مبتنی بر نظریه بازی‌ها را برای هر متغیر ورودی شامل بارندگی، دما، برداشت آب و به‌ویژه خروجی‌های کالیبره‌شده مدل عددی MODFLOW به‌عنوان ویژگی‌های فیزیکی-محور بر پیش‌بینی تراز آب زیرزمینی محاسبه کرد. نتایج SHAP نه تنها چاه‌های کلیدی و متغیرهای مؤثر را به‌خوبی شناسایی کرد، بلکه تفسیرپذیری مدل‌های یادگیری ماشین (به‌ویژه RF و مدل ترکیبی) را نیز به طور قابل توجهی افزایش داد.

برای ارزیابی عدم قطعیت پیش‌بینی‌ها، از روش مونت‌کارلو<sup>۳</sup> با انجام ۱۰۰۰ شبیه‌سازی تصادفی با توزیع نرمال (انحراف معیار ۵٪ مقدار داده اصلی) تغییر یافتند. سپس بازه اطمینان ۹۵٪ (محاسبه‌شده به‌عنوان فاصله بین ۹۵ تا ۵ درصد) برای پیش‌بینی‌های سطح آب زیرزمینی در هر چاه به‌دست آمد. این روش امکان بررسی پایداری مدل‌ها در برابر تغییرات احتمالی ورودی‌ها را فراهم کرد و نشان داد که مدل ترکیبی کمترین عرض بازه اطمینان را نسبت به سایر مدل‌ها دارد.

عملکرد مدل‌های یادگیری ماشین و مدل عددی مبتنی بر فیزیک با استفاده از معیارهای ارزیابی شامل ضریب تعیین ( $R^2$ )، ریشه میانگین مربعات خطا (RMSE)، شاخص کلینگ-گوپتا (KGE)، میانگین درصد خطای مطلق (MAPE) و کارایی نش-ساتکلیف (NSE) مقایسه خواهد شد.

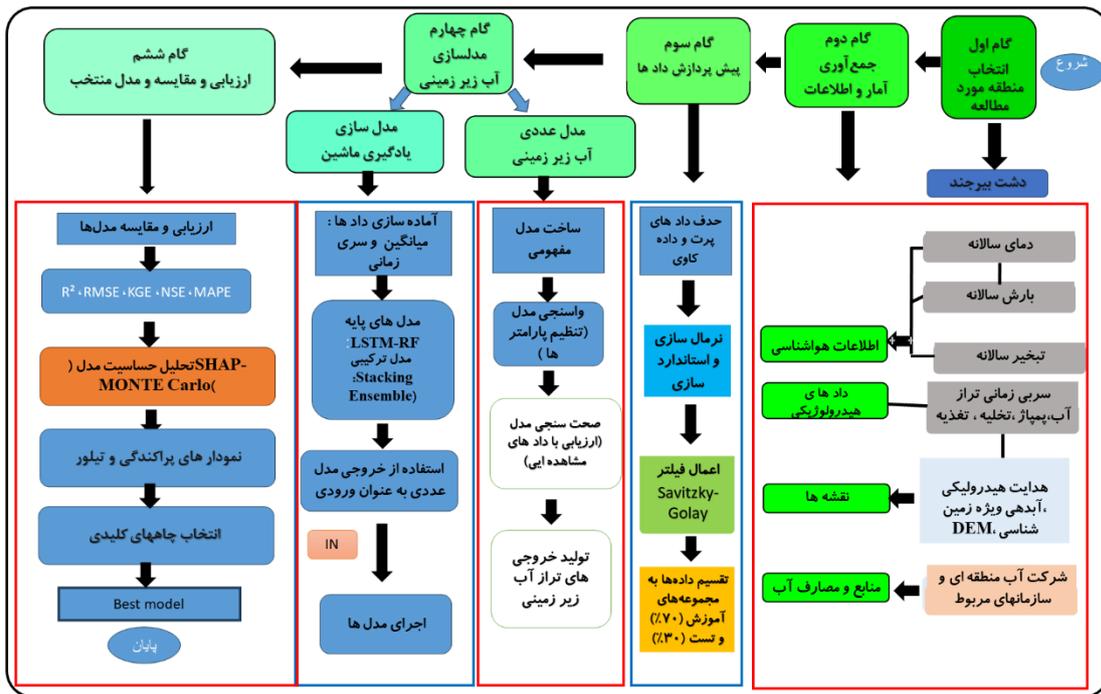
### منطقه مورد مطالعه: دشت بیرجند

دشت بیرجند در استان خراسان جنوبی در محدوده طول جغرافیایی  $41^{\circ}58'$  تا  $44^{\circ}59'$  شرقی و عرض جغرافیایی  $32^{\circ}34'$  تا  $33^{\circ}08'$  شمالی قرار دارد و مساحتی حدود  $3425$  کیلومتر مربع را دربرمی‌گیرد (شکل ۲). آبخوان بیرجند با مساحت حدود  $269.71$  کیلومتر مربع عمدتاً از رسوبات آبرفتی تشکیل شده است که از نهشته‌های دانه‌درشت مخروط‌افکنه‌ها در بخش‌های شرقی تا رسوبات ریزدانه رسی و سیلتی در بخش‌های غربی گسترش می‌یابد. به‌طور کلی، به سمت غرب منطقه اندازه ذرات کاهش یافته و در نتیجه ضریب هدایت هیدرولیکی نیز کاهش می‌یابد. متوسط افت سطح آب در این منطقه طی سال‌های ۱۹۹۶ تا ۲۰۲۰، حدود ۴۰ تا ۶۰ سانتی‌متر در سال گزارش شده است (شکل ۲).

<sup>3</sup> Monte Carlo: Monte Carlo Simulation

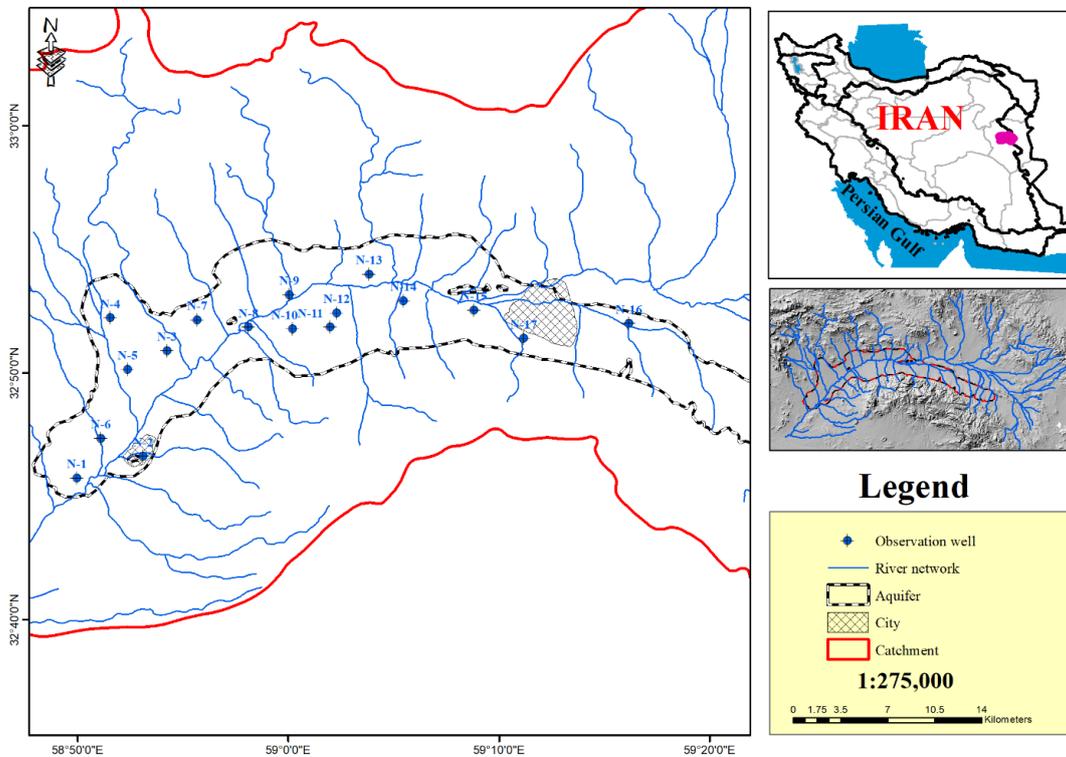
<sup>1</sup> Calibration

<sup>2</sup> Validation



شکل ۱. فلوجارت فرایند مدل سازی آب زیرزمینی دشت بیرجند.

Fig 1. Flowchart of the groundwater modeling process of Birjand Plain.



شکل ۲. منطقه مورد مطالعه حوزه دشت بیرجند.

Fig 2. Study area of the Birjand Plain watershed.

نواحی انتهایی به سمت جنوب غربی متمایل است. در این راستا، مدل مفهومی بر اساس هندسه آبخوان، منابع تغذیه (شامل رودخانه‌ها، نرخ تغذیه طبیعی و مصنوعی) و برداشت از چاه‌ها طراحی شد.

برای توسعه مدل مفهومی، داده‌های زمین‌شناسی شامل نقشه‌های مقیاس ۱:۲۵۰۰۰۰، مقاطع عرضی، لوگ‌های سنگ‌شناسی چاه‌های اکتشافی و پیزومتری، عکس‌های ماهواره‌ای، داده‌های توپوگرافی و اطلاعات هیدرودینامیکی (هدایت هیدرولیکی ۵ تا ۴۵ متر بر روز و آب‌دهی ویژه ۰/۰۳ تا ۰/۱۳۵) جمع‌آوری شد. اطلاعات رقومی توپوگرافی سطح زمین از داده‌های DEM با استفاده از نرم‌افزار سیستم اطلاعات جغرافیایی استخراج و به مدل معرفی شد. منحنی‌های رقوم سنگ کف آبخوان، به دلیل نبود داده‌های مستقیم، از طریق درونیابی دستی با استفاده از داده‌های چاه‌هایی که به سنگ بستر رسیده بودند، ایجاد شد. تراز اولیه آب زیرزمینی بر اساس داده‌های مهرماه ۱۳۸۷ به مدل معرفی شد و شرایط مرزی با تخصیص بار هیدرولیکی یا میزان جریان ترکیبی به سلول‌های مرزی تعیین گردید. بر این اساس، مدل مفهومی با استفاده از هفت پوشش اطلاعاتی (مرز سفره، پیزومترها، چاه‌های بهره‌برداری، تغذیه سطحی، زهکش، هدایت هیدرولیکی و ضریب ذخیره) و ابزارهای GIS طراحی شد و شبکه مدل‌سازی با سلول‌های ۲۵۰×۲۵۰ متر شامل ۱۹۰ چاه (۱۷ چاه مشاهده‌ای، ۱۳۹ چاه کشاورزی و ۳۴ چاه کاربری‌های دیگر مانند شرب یا صنعتی یا غیرفعال) ایجاد شد. تغذیه آبخوان عمدتاً از برگشت ۸۰٪ آب شرب (۱۱ میلیون مترمکعب در سال) و ۲۰٪ آب کشاورزی تأمین می‌شود، در حالی که تغذیه بارشی به دلیل ناچیز بودن در نظر گرفته نشد.

مدل‌سازی غیرماندگار بر اساس داده‌های یک دوره ۱۰ ساله (سال آبی ۱۳۸۷ تا ۱۳۹۸) با گام‌های زمانی ماهانه (مانند ۲۹ روز برای اسفند و ۳۱ روز برای فروردین) انجام شد.

## شبیه‌سازی آب زیرزمینی

برای شبیه‌سازی عددی جریان آب زیرزمینی، از روش تفاضل محدود استفاده شد که معادلات حاکم را در دو حالت ملندگار و غیرملندگار حل می‌کند. در حالت ملندگار، معادله بر اساس اصل جریان پیوسته (رابطه ۱) و در حالت غیرماندگار با در نظر گرفتن تغییرات زمانی تراز آب زیرزمینی فرموله شده است. این مدل با حل عددی معادلات، تراز آب زیرزمینی دشت بیرجند را شبیه‌سازی کرد و ابزاری مؤثر برای تحلیل پایداری منابع آب و مدیریت بهینه آن در منطقه فراهم آورد (Aghlmand & Abbasi, 2019). برای پیش‌بینی دقیق‌تر تراز آب زیرزمینی در دشت بیرجند، مدل‌های یادگیری ماشینی شامل شبکه LSTM، جنگل تصادفی (RF) و مدل ترکیبی Stacking Ensemble به کار گرفته شدند (Zarei et al., 2024).

$$\frac{\partial}{\partial x} \left( k_x \frac{\partial h^2}{\partial x} \right) + \frac{\partial}{\partial y} \left( k_y \frac{\partial h^2}{\partial y} \right) + \frac{\partial}{\partial z} \left( k_z \frac{\partial h^2}{\partial z} \right) = 2S_y \frac{\partial h}{\partial t} - 2R \quad (1)$$

که در آن  $k_x, k_y, k_z$  مولفه‌های تانسور هدایت هیدرولیکی،  $S_y$  آب‌دهی ویژه و  $h$  تراز آب زیر زمینی  $t$ ، زمان (روز) با پتانسل و  $R$  تغذیه و تخلیه است.

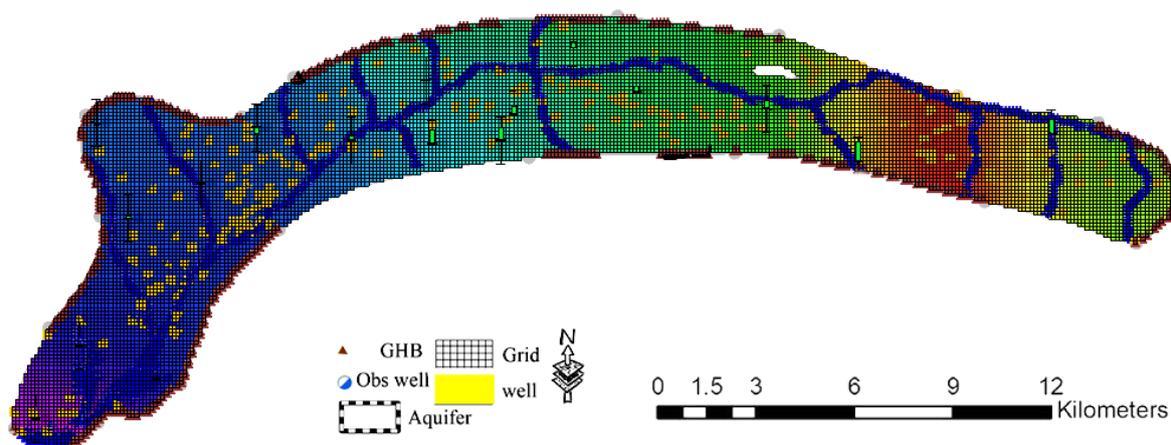
## شبیه‌سازی عددی تراز آب زیرزمینی دشت بیرجند

### با استفاده از مدل عددی MODFLOW

مدل‌سازی جریان آب زیرزمینی ابزاری کلیدی برای مدیریت پایدار منابع آب، به‌ویژه در مناطق خشک با داده‌های محدود مانند دشت بیرجند، محسوب می‌شود. برای این منظور، این فرایند با شناسایی دقیق محدوده آبخوان و تعیین شرایط مرزی مناسب آغاز شد. آبخوان دشت بیرجند، یک آبخوان آزاد با ضخامت متغیر ۵ تا ۲۲۵ متر (عمیق‌ترین بخش در شرق با ۱۵۰ تا ۲۲۵ متر)، از جبهه‌های ورودی شرقی به‌عنوان منبع اصلی تغذیه بهره می‌برد و جریان آب زیرزمینی آن از شرق به غرب و در

مستعد خطا بود؛ بنابراین، از نرم‌افزار Excel برای تسریع و دقت در ورود این اطلاعات استفاده شد. این مدل امکان تحلیل اثرات برداشت و تغذیه بر سطح آب زیرزمینی را فراهم کرد و ابزاری مؤثر برای مدیریت پایدار منابع آب در مناطق خشک مانند دشت بیرجند ارائه داد.

تمامی اطلاعات وابسته به زمان، از جمله تراز آب زیرزمینی در چاه‌های مشاهده‌ای، تخلیه توسط چاه‌های بهره‌برداری و تغذیه ناشی از جریان برگشتی آب کشاورزی، با فرمت مناسب در گام‌های زمانی به مدل وارد شد. به دلیل تعداد زیاد چاه‌ها (۱۹۰ حلقه چاه) و دوره‌های تنش، ورود داده‌های موقعیت و دبی چاه‌ها به صورت دستی وقت‌گیر و



شکل ۳. مدل مفهومی آبخوان دشت بیرجند. مدل مفهومی طراحی شده بر اساس شرایط مرزی، هندسه آبخوان، و داده‌های برداشت و تغذیه سطحی.

**Fig 3. Conceptual model of the Birjand Plain aquifer. Conceptual model designed based on boundary conditions, aquifer geometry, and surface abstraction and recharge data.**

الگوهای فصلی بودند. پیش‌پردازش داده‌ها با فیلتر Savitzky-Golay انجام گرفت تا نویزهای ناشی از خطاهای اندازه‌گیری یا نوسانات پمپاژ حذف شود. این فیلتر با فیت کردن چندجمله‌ای محلی در یک پنجره متحرک عمل می‌کند و مقدار مرکزی هر پنجره را با مقدار پیش‌بینی شده چندجمله‌ای جایگزین می‌نماید. برای انتخاب پارامترهای بهینه، آزمون حساسیت بر روی اندازه پنجره (۵، ۷، ۹، ۱۱، ۱۳ نقطه) و درجه چندجمله‌ای (۲ و ۳) انجام شد. بهترین عملکرد بر اساس کمترین RMSE و حفظ حداکثری دینامیک‌های فصلی و روند بلند مدت، با پنجره ۱۱ نقطه و درجه چندجمله‌ای ۳ حاصل گردید که تعادل مناسبی بین کاهش نویز و حفظ روندهای بلندمدت و تغییرات فصلی برقرار کرد. این فیلتر نسبت به روش‌های

### مدل‌های یادگیری ماشین

برای پیش‌بینی سطح آب زیرزمینی در دشت بیرجند، از مدل‌های LSTM، جنگل تصادفی (RF) و مدل ترکیبی استکینگ استفاده شد. مدل‌های یادگیری ماشین (Stacking Ensemble و Random Forest، LSTM) با استفاده از نرم‌افزار MATLAB و Deep Learning Toolbox (برای LSTM) و توابع داخلی یادگیری ماشین (برای RF و استکینگ) پیاده‌سازی شدند. این مدل‌ها به دلیل توانایی بالای خود در مدل‌سازی روابط غیرخطی و تحلیل سری‌های زمانی هیدرولوژیکی انتخاب شدند. داده‌های ورودی شامل خروجی‌های کالیبره‌شده مدل عددی، سری‌های زمانی سطح آب مشاهده‌ای و ویژگی‌های مهندسی شده مانند میانگین متحرک، مشتقات زمانی و

لایه‌های LSTM بر اساس روابط استاندارد گیت‌های فراموشی، ورودی، کاندیدای سلول، حالت سلول، گیت خروجی و حالت مخفی (معادلات ۲ تا ۷) است؛ پیش‌بینی‌های مدل با داده‌های آزمایشی مقایسه و ارزیابی شدند و جزئیات تنظیمات هاپرپارامترها و معماری مدل (تعداد نورون‌ها، نرخ Dropout، تابع بهینه‌ساز و غیره) در جدول ۳ ارائه شده است.

$$\text{Forget Gate: } f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \quad (2)$$

$$\text{Input Gate: } i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (3)$$

$$\text{Cell Candidate: } \tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C) \quad (4)$$

$$\text{Cell State: } C_t = f_t \cdot C_{t-1} + i_t \cdot \tilde{C}_t: \quad (5)$$

$$\text{Output Gate: } o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \quad (6)$$

$$\text{Hidden State: } h_t = o_t \cdot \tanh(C_t) \quad (7)$$

### جنگل تصادفی (RF)

مدل جنگل تصادفی با ترکیب چندین درخت تصمیم، پیش‌بینی دقیقی ارائه داد و در برابر نویز مقاوم بود (Gupta et al., 2024). داده‌های ورودی شامل سری‌های زمانی پیش‌پردازش‌شده با فیلتر Savitzky-Golay، میانگین‌ها، تغییرات زمانی و الگوهای فصلی بود که پس از انتخاب ویژگی وارد مدل شدند. ساختار مدل شامل سه بخش است: در بخش ورودی، ویژگی‌های آماده وارد می‌شوند؛ در بخش پردازش، ۲۰۰ درخت (عمق ۱۵، معیار جینی (Gini)، شاخصی برای اندازه‌گیری ناهمگنی در گره‌های درخت تصمیم است. مقدار Gini کمتر نشان‌دهنده خلوص بیشتر گره می‌باشد و در الگوریتم‌های جنگل تصادفی برای انتخاب بهترین تقسیم ویژگی استفاده می‌شود)) به صورت موازی عمل کرده و نتایج آن‌ها از طریق

ساده‌تر مانند میانگین متحرک، دینامیک‌های زمانی را بهتر حفظ می‌کند و در شرایط ناهمگنی آبخوان دشت بیرجند (هدایت هیدرولیکی ۵ تا ۴۵ متر بر روز) و محدودیت تعداد داده‌ها بسیار مؤثر است.

پیش‌پردازش با این پارامترها بر روی داده‌های تمام چاه‌های مشاهده‌ای (به‌ویژه چاه‌های ۱، ۵، ۱۰ و ۱۵ به‌عنوان نمایندگان) و خروجی‌های مدل عددی اعمال شد. کاربرد فیلتر Savitzky-Golay کیفیت داده‌های ورودی را به‌طور قابل‌توجهی بهبود بخشید و در نهایت دقت پیش‌بینی مدل‌ها را افزایش داد.

عملکرد مدل‌ها با معیارهای  $R^2$ ، RMSE، MAE، MAPE، NSE و KGE ارزیابی شد. برای مقایسه بصری، از نمودار تیلور که ضریب همبستگی، انحراف معیار نرمال شده و RMSE را نشان می‌دهد استفاده گردید. جزئیات معماری مدل‌ها در جدول ۳ ارائه شده است.

### شبکه‌های عصبی حافظه‌دار طولانی-کوتاهمدت (LSTM)

شبکه LSTM به دلیل توانایی بالای خود در حفظ و مدل‌سازی الگوهای بلندمدت در سری‌های زمانی هیدرولوژیکی انتخاب شد (Ehteram and Ghanbari, 2022; Adivi, 2023; Sheikh Khozani et al., 2022). داده‌های ورودی شامل سری‌های زمانی پیش‌پردازش‌شده با فیلتر Savitzky-Golay، ویژگی‌های مهندسی‌شده مانند میانگین‌های متحرک، تغییرات زمانی (مشتقات) و الگوهای فصلی بودند که ابتدا نرمال‌سازی شده و سپس به فرمت متوالی (Sequences) برای ورودی به شبکه تبدیل گردیدند؛ ساختار مدل LSTM شامل سه بخش اصلی است: بخش ورودی برای دریافت داده‌های متوالی سری زمانی، بخش پردازش با سه لایه LSTM پیاپی، دو لایه Dropout (برای جلوگیری از بیش‌برازش)، یک لایه کاملاً متصل (Dense) و لایه نهایی رگرسیون، و بخش خروجی برای تولید پیش‌بینی سطح آب زیرزمینی؛ عملکرد

الگوهای پیچیده موجود در سری‌های زمانی را به خوبی شناسایی کند.

داده‌های ورودی به مدل متا شامل پیش‌بینی‌های مدل‌های پایه و ویژگی‌های پیش‌پردازش شده مانند سری‌های زمانی فیلترشده با روش Savitzky-Golay، میانگین‌ها، تغییرات زمانی و الگوهای فصلی بودند. پیش‌بینی‌های مدل‌های پایه بر اساس مقدار  $R^2$  وزن‌دهی شده و سپس به مدل متا منتقل شدند. بر اساس رابطه (۹)، پیش‌بینی نهایی سطح آب زیرزمینی محاسبه می‌شود:

$$\widehat{Y}_{\text{final}} = w_1 \widehat{Y}_1 + w_2 \widehat{Y}_2 + \dots + w_K \widehat{Y}_K + b \quad (9)$$

که در آن  $\widehat{Y}_1, \widehat{Y}_2, \dots, \widehat{Y}_K$  پیش‌بینی‌های مدل‌های پایه،  $w_1, w_2, \dots, w_K$  وزن‌های مرتبط بر اساس  $R^2$  و بایاس است. در بخش خروجی، پیش‌بینی نهایی سطح آب زیرزمینی تولید و با داده‌های آزمایشی ارزیابی می‌شود. جزئیات تنظیمات مدل در جدول ۳ آمده است.

میانگین‌گیری ترکیب می‌شود؛ و در بخش خروجی، پیش‌بینی نهایی سطح آب تولید می‌گردد که با رابطه ۸ عمل می‌کند:

$$\widehat{y} = \frac{1}{T} \sum_{t=1}^T h_t(x) \quad (8)$$

در بخش خروجی، پیش‌بینی سطح آب زیرزمینی تولید و با داده‌های آزمایشی ارزیابی می‌شود. جزئیات تنظیمات مدل در جدول ۳ ارائه شده است

مدل ترکیبی استکینگ برای پیش‌بینی سطح آب

### زیرزمینی

مدل ترکیبی استکینگ با ترکیب پیش‌بینی‌های دو مدل پایه LSTM و جنگل تصادفی و استفاده از یک مدل متا (رگرسیون خطی)، دقت پیش‌بینی سطح آب زیرزمینی در دشت بیرجند را به طور چشمگیری افزایش داد (Victor & Ali, 2024; Sheikh Khozani et al., 2022). این مدل با بهره‌گیری از نقاط قوت هر یک از مدل‌های پایه، توانست

جدول ۳. مشخصات و کاربرد مدل‌های یادگیری ماشین در پیش‌بینی تراز آب زیرزمینی.

**Table 3. Specifications and Application of Machine Learning Models in Groundwater Level Prediction.**

توضیحات Description	ویژگی‌های مدل Model Features	مدل Model	شماره Number
برای مدل‌سازی الگوهای بلندمدت در سری‌های زمانی هیدرولوژیکی، با ارزیابی عملکرد بر اساس داده‌های آزمون و اعتبارسنجی متقاطع For modeling long-term patterns in hydrological time series, with performance evaluation based on test data and cross-validation	شبکه عصبی بازگشتی با سه لایه LSTM، دو لایه Dropout، بهینه‌ساز Adam، تابع فعال‌سازی ReLU و لایه Dense نهایی Recurrent Neural Network with three LSTM layers, two Dropout layers, Adam optimizer, ReLU activation function, and a final Dense layer	LSTM	1
برای پیش‌بینی دقیق با ترکیب نتایج چندین درخت و کاهش واریانس، با بهینه‌سازی هیبرید پارامترها و ارزیابی بر اساس داده‌های آزمون For accurate prediction by combining the results of multiple trees and reducing variance, with hyperparameter optimization and evaluation based on test data	مجموعه‌ای از درخت تصمیم، حداکثر عمق نامحدود، معیار تقسیم Gini، بوت‌استرپ فعال A set of Decision Trees, unlimited maximum depth, Gini splitting criterion, bootstrap enabled	Random Forest	2
برای بهبود دقت کلی پیش‌بینی با تلفیق نقاط قوت مدل‌های پایه، با ارزیابی بر اساس داده‌های آزمون و اعتبارسنجی متقاطع To improve overall prediction accuracy by combining the strengths of base models, with evaluation based on test data and cross-validation	مدل پایه: LSTM و Random Forest؛ متا-یادگیرنده: Random Forest با درخت Base models: LSTM and Random Forest; Meta-learner: Random Forest with trees	Staking Ensemble	3

## شاخص‌های ارزیابی آماری

برای ارزیابی دقت و صحت پیش‌بینی‌های مدل‌های یادگیری ماشین (LSTM، RF ترکیبی) در پیش‌بینی شاخص کیفیت آب در دشت بیرجند، از معیارهای آماری  $R^2$ ، RMSE، MAPE، NSE و KGE استفاده شده است که عمدتاً برای ارزیابی و مقایسه عملکرد مدل‌ها از معیارهای  $R^2$ ، RMSE و KGE استفاده شد. فرمول‌های مربوط به این معیارها در روابط ۱۰ تا ۱۵ نشان داده شده است.

$$R^2 = \frac{[\sum_{i=1}^n (D_i - \bar{D})(O_i - \bar{O})]^2}{\sum_{i=1}^n (D_i - \bar{D})^2 \sum_{i=1}^n (O_i - \bar{O})^2} \quad (10)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (D_i - O_i)^2} \quad (11)$$

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{O_i - D_i}{O_i} \right| \quad (12)$$

$$NSE = 1 - \frac{\sum_{i=1}^n (O_i - D_i)^2}{\sum_{i=1}^n (O_i - \bar{O})^2} \quad (13)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |O_i - P_i| \quad (14)$$

در این روابط  $O_i$  و  $D_i$  به ترتیب آمین داده شبیه‌سازی شده توسط مدل و مقدار واقعی (مشاهده شده)،  $\bar{O}$  و  $\bar{D}$  به

ترتیب میانگین کل داده‌های  $D_i$  و  $O_i$  در جامعه آماری،  $n$  تعداد نمونه‌ها است.

$$KGE = \frac{1 - \sqrt{\left(\frac{\sigma_{pred}}{\sigma_{true}} - 1\right)^2 + \left(\frac{\mu_{pred}}{\mu_{true}} - 1\right)^2 + (r - 1)^2}}{1} \quad (15)$$

$\sigma_{pred}$  انحراف استاندارد پیش‌بینی‌ها،  $\sigma_{true}$  انحراف استاندارد مقادیر واقعی،  $\mu_{pred}$  میانگین پیش‌بینی‌ها،  $\mu_{true}$  میانگین مقادیر واقعی و  $r$  ضریب همبستگی بین پیش‌بینی‌ها و مقادیر واقعی است.

## نتایج و بحث

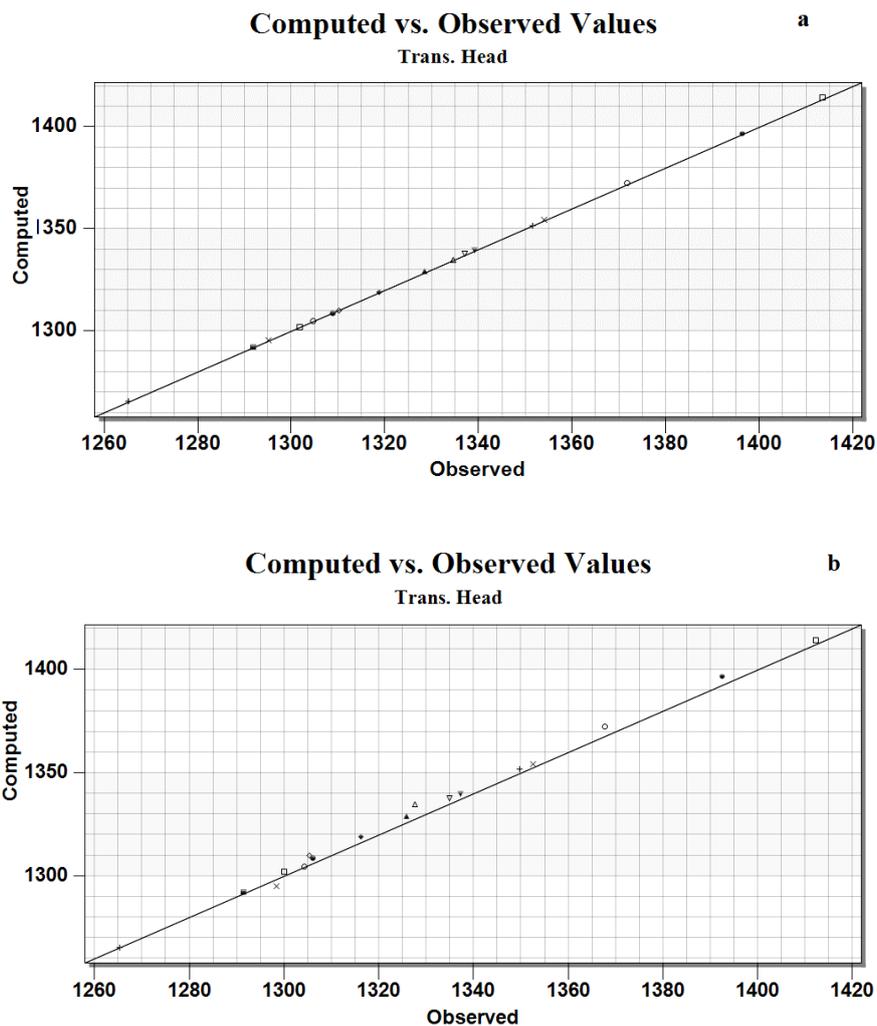
## نتایج مدل‌سازی مدل عددی

شبیه‌سازی تراز آب زیرزمینی دشت بیرجند با استفاده از مدل عددی در دو رژیم ماندگار (شهریور تا مهر ۱۳۸۷) و غیرماندگار (مهر ۱۳۸۷ تا مهر ۱۳۹۸) انجام گرفت که در رژیم غیرماندگار پس از واسنجی، مدل دقت بسیار بالایی از خود نشان داد و ضریب تعیین ۰.۹۸۹ به دست آمد. عملکرد مدل در این رژیم با مقایسه خروجی‌های شبیه‌سازی شده با مشاهدات ۱۷ چاه مشاهده‌ای ارزیابی شد و نتایج معیارهای ارزیابی شامل  $R^2$ ، RMSE، KGE، NSE و MAPE برای دوره‌های واسنجی و صحت‌سنجی به طور کامل در جدول ۴ و شکل ۴ ارائه شده است

جدول ۴. معیارهای عملکرد مدل MODFLOW در شبیه‌سازی تراز آب زیرزمینی دشت بیرجند (حالت غیرماندگار)

Table 4. Performance Metrics of the MODFLOW Model in Simulating Groundwater Levels of the Birjand Plain (Transient State)

معیار Criterion	واسنجی Calibration	صحت‌سنجی Validation
$R^2$	0.989	0.968
RMSE	0.87	3.17
KGE	0.995	0.975
NSE	0.999	0.993
MAPE	0.02%	0.20%



شکل ۴. نمودارهای پراکندگی و اسنجی (الف) و صحت سنجی مدل (ب) در حالت غیر ماندگار.

Fig 4. Scatter plots of calibration (a) and model validation (b) in the non-steady state.

استاندارد طولایی شناخته می‌شود (Gupta et al., 2009; Kling et al., 2012; Knoben et al., 2019). نتایج ارزیابی دوره آزمون (جدول ۵) نشان می‌دهد که مدل در چاه ۱ عملکرد استثنایی دارد  $R^2=0.902$ ،  $KGE=0.948$  و  $RMSE=0.093$  m، درحالی‌که در چاه‌های ۳، ۱۳ و ۱۴ به دلیل ناهمگنی شدید محلی، برداشت‌های بی‌رویه کشاورزی و صنعتی، و جریان‌های خروجی مرزی غربی، دقت کمتری مشاهده می‌شود. با این حال، عملکرد کلی مدل در مقیاس دشت (جدول ۴) بسیار مطلوب است که تأییدی بر موفقیت فرایند کالیبراسیون در سطح آبخوان است.

انتخاب چاه‌های مشاهده‌ای بر اساس معیار ارزیابی

عملکرد مدل عددی

انتخاب چاه‌های مشاهده‌ای کلیدی بر اساس ارزیابی جامع عملکرد مدل عددی MODFLOW و با اولویت‌بندی شاخص کلینگ-گوپتا (KGE) به‌عنوان معیار اصلی انجام شد. این شاخص به دلیل توانایی منحصر به فرد خود در تجزیه هم‌زمان همبستگی خطی، بایاس نسبی و نسبت واریانس بین سری‌های شبیه‌سازی شده و مشاهده شده، ارزیابی دقیق‌تری نسبت به معیارهای سنتی مانند NSE ارائه می‌دهد و در مطالعات هیدرولوژیکی مدرن به‌عنوان

مقادیر منفی NSE در نقاط محلی مذکور، پدیده‌های شناخته‌شده در مدل‌سازی آبخوان‌های ناهمگن پیچیده است و صرفاً بیانگر محدودیت مدل فیزیکی در بازتولید نوسانات شدید نقطه‌ای با رزولوشن شبکه  $250 \times 250$  متر می‌باشد، نه نقص کلی مدل. در همین شرایط، حفظ مقادیر مثبت KGE ( $0/10$ ) تا  $0/31$  در این چاه‌ها برتری آشکار این شاخص را نسبت به NSE اثبات می‌کند، زیرا KGE با تمرکز بر اجزای مستقل عملکرد، ارزیابی مقاوم‌تری ارائه می‌دهد. در دوره صحت‌سنجی، چاه ۱ با معیارهای برتر  $R^2=0.902$ ،  $RMSE=0.093m$  و  $NSE=0.897$ ،  $MAPE=0.005\%$  و

مقادیر منفی NSE در نقاط محلی مذکور، پدیده‌های شناخته‌شده در مدل‌سازی آبخوان‌های ناهمگن پیچیده است و صرفاً بیانگر محدودیت مدل فیزیکی در بازتولید نوسانات شدید نقطه‌ای با رزولوشن شبکه  $250 \times 250$  متر می‌باشد، نه نقص کلی مدل.

در همین شرایط، حفظ مقادیر مثبت KGE ( $0/10$ ) تا  $0/31$  در این چاه‌ها برتری آشکار این شاخص را نسبت به NSE اثبات می‌کند، زیرا KGE با تمرکز بر اجزای مستقل عملکرد، ارزیابی مقاوم‌تری ارائه می‌دهد.

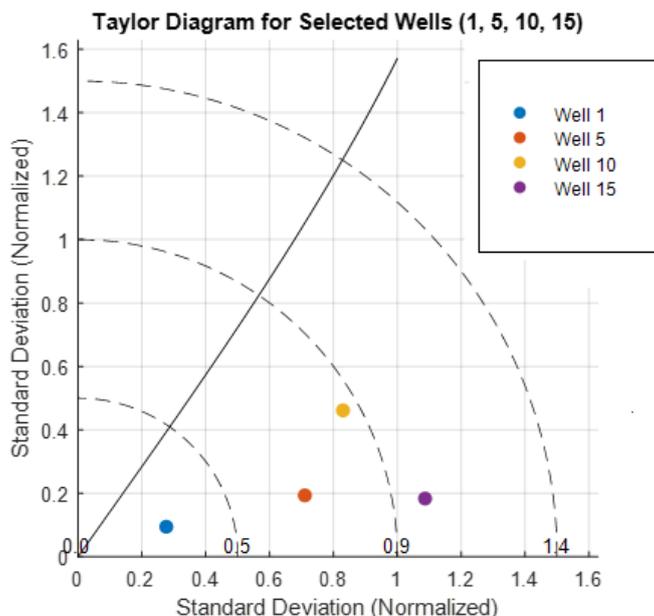
در دوره صحت‌سنجی، چاه ۱ با معیارهای برتر  $R^2=0.902$ ،  $RMSE=0.093m$  و  $NSE=0.897$ ،  $MAPE=0.005\%$  و

به‌عنوان چاه‌های کلیدی مکمل تعیین گردیدند تا نمایندگی مناسبی از دینامیک‌های فضایی دشت بیرجند ارائه دهند. این انتخاب استراتژیک نه تنها دقت ورودی‌های مدل‌های یادگیری ماشین را تضمین می‌کند، بلکه امکان تعمیم‌پذیری بهتر نتایج پیش‌بینی در کل آبخوان را فراهم می‌سازد.

جدول ۵. خلاصه معیارهای عملکرد برای تمام ۱۷ چاه (مرتب‌شده بر اساس KGE نزولی).

Table 5. Summary of Performance Metrics for All 17 Wells (Sorted in Descending Order of KGE).

Well	R <sup>2</sup>	RMSE	KGE	NSE	MAPE	MAE
well1	0.9021	0.0929	0.9479	0.8974	0.005	0.0639
well2	0.8927	0.4578	0.5633	0.3981	0.0295	0.3823
well3	0.6693	1.2857	0.3146	-5.0095	0.0898	1.1712
well4	0.9805	0.4028	0.9085	0.8917	0.0268	0.3509
well5	0.9477	0.4488	0.6747	0.8282	0.0261	0.3403
well6	0.5523	0.3795	0.7067	0.4784	0.0242	0.3121
well7	0.9879	0.6383	0.7393	0.8377	0.0432	0.565
well8	0.8274	1.1407	0.77	0.3041	0.0752	0.9923
well9	0.7832	1.5214	0.7001	-0.0565	0.1011	1.3453
well10	0.8313	0.5606	0.7406	0.7998	0.0315	0.4191
well11	0.833	0.822	0.8516	0.3831	0.0578	0.7728
well12	0.6961	0.7525	0.579	-0.2117	0.0544	0.7283
well13	0.5629	1.0578	0.0981	-2.7464	0.0697	0.9409
well14	0.6791	1.1173	0.2389	-2.4991	0.0688	0.9314
well15	0.9721	0.6973	0.7951	0.7177	0.045	0.616
well16	0.9033	0.626	0.7896	0.8783	0.035	0.4882
well17	0.3707	1.1588	0.4568	0.3683	0.0655	0.9251



شکل ۵. نمودار تیلور عملکرد مدل MODFLOW برای چاه‌های هدف (۱، ۵، ۱۰، ۱۵) در دوره صحت‌سنجی.  
 Fig 5. Taylor diagram of MODFLOW model performance for target wells (1, 5, 10, 15) during the validation period.

#### نتایج مدل‌سازی یادگیری ماشین

برای پیش‌بینی سری‌های زمانی تراز آب زیرزمینی<sup>۱</sup> (GWL) در دشت بیرجند، از مدل‌های پیشرفته یادگیری ماشین شامل شبکه عصبی بازگشتی حافظه‌دار بلند-کوتاه‌مدت (LSTM)، جنگل تصادفی و مدل ترکیبی مبتنی بر استکینگ بهره گرفته شد. این مدل‌ها به دلیل توانایی برتر در شناسایی روابط غیرخطی پیچیده، مدیریت سری‌های زمانی و عملکرد Robust در شرایط محدودیت داده (که در دشت‌های خشک مانند بیرجند رایج است)، به‌عنوان رویکرد مکمل مدل عددی MODFLOW انتخاب گردیدند؛ مدلی که علی‌رغم دقت فیزیکی بالا، به داده‌های ورودی گسترده و پارامترهای هیدروژئولوژیکی دقیق وابسته است. این استراتژی هیبریدی (ترکیب فیزیکی و داده‌محور) با روندهای نوین در تحقیقات هیدرولوژیکی هم‌خوانی دارد و امکان پیش‌بینی دقیق‌تر و کارآمدتر در سناریوهای کمبود داده را فراهم می‌سازد (Khan et al., 2023; Boo et al., 2024).

#### تحلیل حساسیت ورودی‌های مدل‌های یادگیری

ماشین با روش‌های SHAP و Monte Carlo تحلیل حساسیت با بهره‌گیری از روش‌های SHAP و مونت‌کارلو انجام شد تا نقش متغیرهای ورودی در دقت پیش‌بینی‌ها ارزیابی شود. روش SHAP با تخمین اهمیت نسبی متغیرها، چاه‌های شماره ۱، ۵، ۱۰ و ۱۵ را با مقادیر اهمیت تقریبی بین ۰/۱۰ تا ۰/۲۸ شناسایی کرد. شبیه‌سازی‌های مونت‌کارلو با ۱۰۰۰ تکرار تصادفی، پایداری همبستگی (بین ۰/۶۶۷۹ تا ۰/۹۴۹۸) و خطای مطلق میانگین (MAE) را تأیید کردند. نتایج این تحلیل‌ها در جدول ۶ و شکل‌های ۶ و ۷ ارائه شده‌اند. برای بهبود دقت مدل‌های یادگیری ماشین، از تکنیک‌های بهینه‌سازی پارامترها و اعتبارسنجی متقابل استفاده شد. این نتایج نشان‌دهنده نقش کلیدی چاه‌های منتخب در بهبود پیش‌بینی‌های سری زمانی تراز آب زیرزمینی و کارایی روش‌های یادگیری ماشین در تحلیل هیدرولوژیکی دشت بیرجند است.

<sup>1</sup> Ground water level

جدول ۶. نتایج خلاصه‌شده تحلیل حساسیت یادگیری ماشین برای چاه‌های منتخب در دشت بیرجند.

**Table 6. Summarized Results of Machine Learning Sensitivity Analysis for Selected Wells in the Birjand Plain.**

چاه	همبستگی (C با O1)	SHAP (تقریبی)	میانگین همبستگی MC	MAE MC (متر)	میانگین MAE MC	MAE (C-O)
Well	Correlation (C with O1)	SHAP (approximate)	Mean Correlation MC	Mean MAE MC (m)		
1	0.9498	0.28	$14.2 \times 10^{-3}$	20.1		0.0639
5	0.7863	0.15	$8.5 \times 10^{-3}$	45.3		0.3403
10	0.7175	0.12	$7.1 \times 10^{-3}$	60.2		0.4191
15	0.6679	0.1	$6.8 \times 10^{-3}$	80.4		0.616

همبستگی و MAE در مقیاس بدون بعد و متریک هستند، و مقادیر SHAP نرمال‌شده بین ۰ تا ۱ هستند. برای جزئیات بصری، به نمودارهای ۶ و ۷ رجوع شود.

### ارزیابی عملکرد مدل‌های یادگیری ماشین برای

#### پیش‌بینی تراز آب زیرزمینی

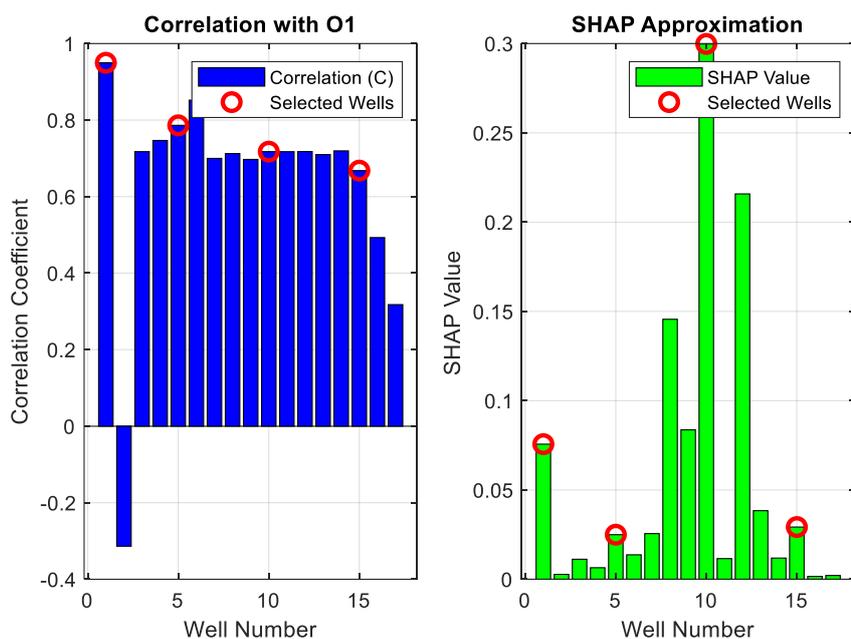
برای پیش‌بینی سری زمانی تراز آب زیرزمینی در چاه شماره ۱ دشت بیرجند (و سایر چاه‌های مشاهده‌ای)، سه مدل یادگیری ماشین شامل شبکه عصبی بازگشتی LSTM، جنگل تصادفی (Random Forest) و مدل ترکیبی استکینگ ارزیابی شدند. عملکرد این مدل‌ها با استفاده از معیارهای ضریب تعیین ( $R^2$ )، ریشه میانگین مربعات خطا (RMSE)، کارایی کلینگ-گوپتا (KGE)، کارایی نش-ساتکلیف (NSE) و درصد خطای مطلق میانگین (MAPE) بر روی مجموعه‌های (داده‌های آموزشی، اعتبارسنجی و آزمایشی) سنجیده تقسیم‌بندی داده‌ها به صورت زمانی<sup>۱</sup> و با حفظ ترتیب زمانی<sup>۲</sup> انجام گرفت تا پیش‌بینی واقعی شرایط آینده با استفاده از داده‌های گذشته شبیه‌سازی شود. نتایج این ارزیابی در جدول ۷ ارائه شده است.

جدول ۶ خلاصه‌ای از نتایج تحلیل حساسیت را برای چاه‌های منتخب نشان می‌دهد. چاه شماره ۱ با ضریب همبستگی ۰/۹۴۹۸، اهمیت SHAP=0.28 و MAE=0.0639 متر بالاترین تأثیر را بر دقت مدل داشت. چاه‌های ۵، ۱۰ و ۱۵ به ترتیب با همبستگی ۰/۷۸۶۳، ۰/۷۱۷۵ و ۰/۶۶۷۹ و مقادیر MAE برابر با ۰/۳۴۰۳، ۰/۴۱۹۱ و ۰/۶۱۶ متر عملکرد قابل قبولی ارائه کردند. شکل ۶ ضریب همبستگی (بین مقادیر محاسباتی و مشاهداتی) و اهمیت SHAP را برای ۱۷ چاه نشان می‌دهد، که در آن چاه‌های منتخب (۱، ۵، ۱۰ و ۱۵) با دایره‌های قرمز مشخص شده‌اند. شکل ۷ نیز میانگین همبستگی و MAE حاصل از شبیه‌سازی‌های مونت‌کارلو را نمایش می‌دهد، که تأثیر برجسته این چاه‌ها را تأیید می‌کند.

مقادیر همبستگی (Correlation) بین مقادیر محاسباتی (C) و مرجع (O1)، اهمیت تقریبی SHAP، میانگین همبستگی و خطای مطلق میانگین (MAE) از شبیه‌سازی‌های مونت‌کارلو (MC) با ۱۰۰۰ تکرار و MAE واقعی برای چاه‌های منتخب (۱، ۵، ۱۰، ۱۵). مقادیر

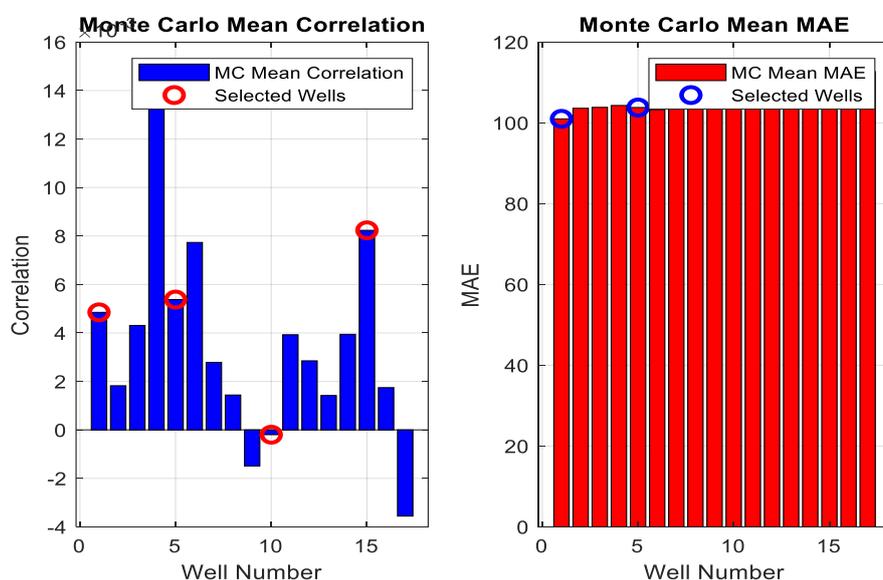
<sup>2</sup> Chronological Order

<sup>1</sup> Temporal Split



شکل ۶. تحلیل همبستگی و SHAP برای چاه‌های دشت بیرجند نمایش ضریب همبستگی مقادیر محاسباتی (C) با O1 (آبی) و اهمیت تقریبی SHAP (سبز) برای ۱۷ چاه. چاه‌های منتخب (۱، ۵، ۱۰، ۱۵) با دایره قرمز مشخص شده‌اند. مقادیر همبستگی بین -۰/۴ تا ۱ و SHAP بین ۰ تا ۰/۳ نرمال شده هستند.

Fig 6. Correlation and SHAP analysis for wells in the Birjand plain showing the correlation coefficient of calculated values (C) with O1 (blue) and approximate significance of SHAP (green) for 17 wells. Selected wells (1, 5, 10, 15) are marked with red circles. Correlation values range from -0.4 to 1 and SHAP is normalized from 0 to 0.3.



شکل ۷. تحلیل مونت کارلو و MAE برای چاه‌های دشت بیرجند نمایش میانگین همبستگی (آبی، مقیاس  $\times 10^{-3}$ ) و میانگین خطای مطلق (MAE) مونت کارلو (قرمز) برای ۱۷ چاه با ۱۰۰۰ شبیه‌سازی. چاه‌های منتخب (۱، ۵، ۱۰، ۱۵) با دایره قرمز (همبستگی) و آبی (MAE) مشخص شده‌اند. مقادیر MAE از ۲۰ تا ۱۲۰ متغیر است.

Fig 7. Monte Carlo and MAE analysis for Birjand Plain wells. Showing the mean correlation (blue, scale  $\times 10^{-3}$ ) and mean absolute error (MAE) of Monte Carlo (red) for 17 wells with 1000 simulations. Selected wells (1, 5, 10, 15) are marked with red (correlation) and blue (MAE) circles. MAE values range from 20 to 120.

جدول ۷. نتایج مدل‌های یادگیری ماشین در حالت آموزش و آزمایش.

Table 7. Results of Machine Learning Models in Training and Testing Phases.

مدل Model	مجموعه داده Dataset	R <sup>2</sup>	RMSE (m)	KGE	NSE	MAPE (%)
LSTM	آموزشی Train	0.997	0.015	0.985	0.997	0.001
	آزمایشی Test	0.977	0.034	0.952	0.977	0.002
Random Forest	آموزشی Train	0.978	0.041	0.961	0.978	0.003
	آزمایشی Test	0.94	0.053	0.946	0.94	0.003
Stacking Ensemble	آموزشی Train	0.997	0.015	0.986	0.997	0.001
	آزمایشی Test	0.978	0.032	0.946	0.978	0.002

برای افزایش اعتبار نتایج و بررسی پایداری مدل‌های یادگیری ماشین در برابر تغییرات داده‌های آموزشی، از روش اعتبارسنجی متقابل<sup>۱</sup> استفاده شد. در این روش، مجموعه داده‌های آموزشی به پنج زیرمجموعه مساوی تقسیم گردید و در هر تکرار، مدل بر روی چهار زیرمجموعه آموزش دیده و بر روی زیرمجموعه باقی‌مانده ارزیابی شد؛ این فرایند پنج بار تکرار گردید تا هر زیرمجموعه دقیقاً یک بار نقش داده آزمون را ایفا کند.

این رویکرد، به‌ویژه در شرایط محدودیت داده‌های مشاهده‌ای هیدرولوژیکی مانند دشت بیرجند، امکان ارزیابی دقیق قابلیت تعمیم‌پذیری<sup>۲</sup> و استحکام<sup>۳</sup> مدل‌ها را فراهم می‌سازد و خطر بیش‌براز را به طور مؤثر کاهش می‌دهد.

نتایج میانگین اعتبارسنجی متقابل (جدول ۸) تأیید کرد که مدل ترکیبی مبتنی بر استکینگ (Stacking Ensemble) همچنان برترین عملکرد و بالاترین پایداری را به خود اختصاص داده است (R<sup>2</sup> میانگین ≈ ۰/۹۷۲، RMSE میانگین ≈ ۰/۰۳۵ متر، NSE میانگین ≈ ۰/۹۷۲).

جدول ۷ نشان می‌دهد که مدل ترکیبی استکینگ در داده‌های آزمایشی با R<sup>2</sup>=0.978، RMSE=0.032 متر، NSE=0.978 و MAPE=0.002% بهترین عملکرد را ارائه کرد و به‌عنوان مدل منتخب برگزیده شد. مدل LSTM با R<sup>2</sup>=0.977 و RMSE=0.034 متر عملکردی نزدیک به استکینگ داشت، در حالی که جنگل تصادفی با R<sup>2</sup>=0.940 و RMSE=0.053 متر دقت کمتری نشان داد. مقدار KGE مدل استکینگ (۰/۹۴۶) کمی پایین‌تر از LSTM (0.952) بود، که می‌تواند به چالش‌های هم‌خطی در متا-مدل مربوط باشد. نمودارهای پراکندگی مدل استکینگ (شکل‌های ۸ و ۹) در داده‌های آموزشی و آزمایشی، تطابق نزدیک پیش‌بینی‌ها با مقادیر مشاهده‌شده را نشان می‌دهند. در این نمودارها، توزیع نقاط نزدیک به خط  $y=x$  با ضریب همبستگی تقریبی ۰/۹۹، دقت بالای مدل منتخب را تأیید می‌کند. این نتایج با مطالعات اخیر در زمینه کاربرد یادگیری ماشین برای پیش‌بینی تراز آب زیرزمینی هم‌خوانی دارد.

<sup>3</sup> robustness<sup>1</sup> 5-fold cross-validation)<sup>2</sup> generalizability

و KGE میانگین  $\approx 0.941$ ). نزدیکی چشمگیر این مقادیر به نتایج تقسیم آموزشی-آزمایشی اصلی (جدول ۷) نشان‌دهنده پایداری فوق‌العاده بالای مدل ترکیبی در برابر تغییرات ترکیب داده‌های آموزشی است.

جدول ۸. نتایج میانگین اعتبارسنجی متقابل برای مدل‌های یادگیری ماشین.

**Table 8. Mean Cross-Validation Results for Machine Learning Models.**

مدل	میانگین $R^2$	میانگین RMSE (m)	میانگین NSE	میانگین KGE
Model	Mean $R^2$	Mean RMSE (m)	Mean NSE	Mean KGE
Stacking Ensemble	0.972	0.035	0.972	0.941
LSTM	0.968	0.038	0.968	0.948
Random Forest	0.932	0.056	0.932	0.895

مدل LSTM نیز عملکردی نزدیک و پایدار از خود نشان داد ( $R^2$  میانگین  $\approx 0.968$ ، RMSE میانگین  $\approx 0.038$  متر) در حالی که مدل جنگل تصادفی (RF) حساسیت بیشتری به تغییرات داده‌های آموزشی داشت ( $R^2$  میانگین  $\approx 0.932$ ، RMSE میانگین  $\approx 0.056$  متر). این تفاوت، برتری مدل‌های مبتنی بر شبکه‌های عصبی بازگشتی و به‌ویژه رویکرد استکینگ را در مدیریت سری‌های زمانی پیچیده و ناهمگن هیدرولوژیکی برجسته می‌سازد.

انجام این تحلیل اعتبارسنجی متقابل اضافی، قابلیت اطمینان، تعمیم‌پذیری و استحکام مدل‌های پیشنهادی را در شرایط واقعی و پویای آبخوان دشت بیرجند به طور قاطع تقویت کرده و اعتماد به نتایج پیش‌بینی بلندمدت را به طور قابل توجهی افزایش می‌دهد.

تحلیل نمودار شکل ۱۰ نشان داد که مدل ترکیبی با ضریب همبستگی  $0.978$ ، انحراف معیار نرمال شده  $0.812$  و  $RMSE=0.032$  متر بهترین عملکرد را در میان سه مدل دارد. مدل LSTM با ضریب همبستگی  $0.978$ ، انحراف معیار نرمال شده  $0.845$  و  $RMSE=0.034$  متر عملکردی نزدیک به مدل ترکیبی ارائه داد، در حالی که مدل RF با ضریب همبستگی  $0.940$ ، انحراف معیار نرمال شده  $1/344$  و  $RMSE=0.053$  متر ضعیف‌ترین نتیجه را به دست آورد.

نتایج این مطالعه نشان داد که مدل ترکیبی در پیش‌بینی تراز آب زیرزمینی چاه مرجع (چاه ۱) دقت بالاتری نسبت به سایر مدل‌ها از خود نشان داد. این مدل با معیارهای  $R^2$  و  $RMSE=0.032$  متر را به دست آورد.

این چارچوب، تولنایی LSTM در شناسایی روندهای بلندمدت و الگوهای فصلی را با مقاومت RF در برابر نویز و بیش‌برازش ترکیب کرده و دلنش فیزیکی نهفته در خروجی‌های کالیبره‌شده MODFLOW را به‌عنوان ویژگی‌های ورودی ادغام می‌نماید (Chen et al, 2023).

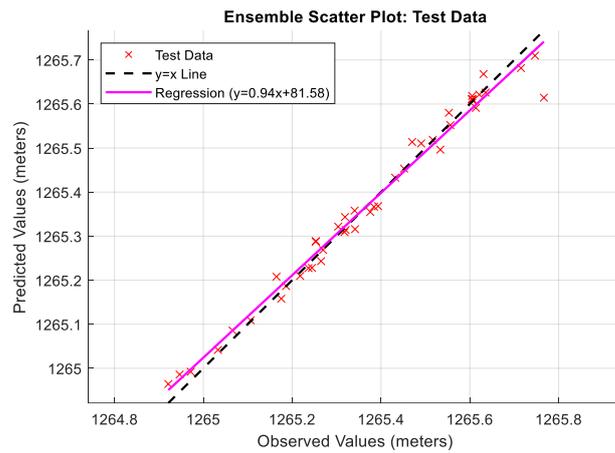
نمودارهای پراکندگی مدل ترکیبی (شکل‌های ۸ و ۹) تطابق بسیار نزدیک پیش‌بینی‌ها با داده‌های مشاهده‌شده را در هر دو مجموعه آموزشی و آزمایشی نشان می‌دهند، با ضریب همبستگی تقریبی  $0.99$ . همچنین، نمودار تیلور (شکل ۱۰) برتری این مدل را تأیید کرد؛ به طوری که ضریب همبستگی  $0.978$ ، انحراف معیار نرمال شده  $0.812$  و  $RMSE=0.032$  متر را به دست آورد.

مدل LSTM نیز عملکردی نزدیک و پایدار از خود نشان داد ( $R^2$  میانگین  $\approx 0.968$ ، RMSE میانگین  $\approx 0.038$  متر) در حالی که مدل جنگل تصادفی (RF) حساسیت بیشتری به تغییرات داده‌های آموزشی داشت ( $R^2$  میانگین  $\approx 0.932$ ، RMSE میانگین  $\approx 0.056$  متر). این تفاوت، برتری مدل‌های مبتنی بر شبکه‌های عصبی بازگشتی و به‌ویژه رویکرد استکینگ را در مدیریت سری‌های زمانی پیچیده و ناهمگن هیدرولوژیکی برجسته می‌سازد.

انجام این تحلیل اعتبارسنجی متقابل اضافی، قابلیت اطمینان، تعمیم‌پذیری و استحکام مدل‌های پیشنهادی را در شرایط واقعی و پویای آبخوان دشت بیرجند به طور قاطع تقویت کرده و اعتماد به نتایج پیش‌بینی بلندمدت را به طور قابل توجهی افزایش می‌دهد.

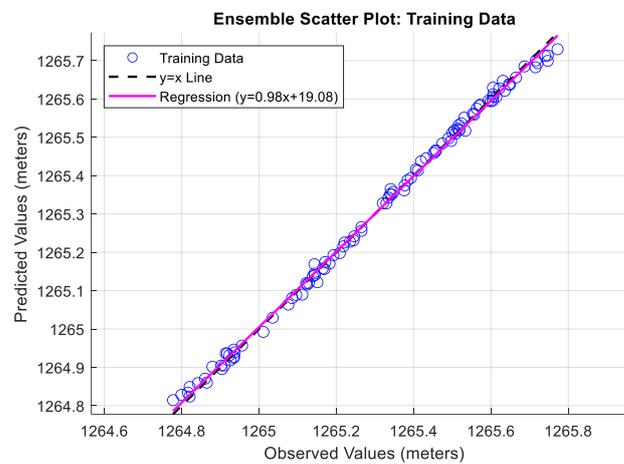
تحلیل نمودار شکل ۱۰ نشان داد که مدل ترکیبی با ضریب همبستگی  $0.978$ ، انحراف معیار نرمال شده  $0.812$  و  $RMSE=0.032$  متر بهترین عملکرد را در میان سه مدل دارد. مدل LSTM با ضریب همبستگی  $0.978$ ، انحراف معیار نرمال شده  $0.845$  و  $RMSE=0.034$  متر عملکردی نزدیک به مدل ترکیبی ارائه داد، در حالی که مدل RF با ضریب همبستگی  $0.940$ ، انحراف معیار نرمال شده  $1/344$  و  $RMSE=0.053$  متر ضعیف‌ترین نتیجه را به دست آورد.

نتایج این مطالعه نشان داد که مدل ترکیبی در پیش‌بینی تراز آب زیرزمینی چاه مرجع (چاه ۱) دقت بالاتری نسبت به سایر مدل‌ها از خود نشان داد. این مدل با معیارهای  $R^2$



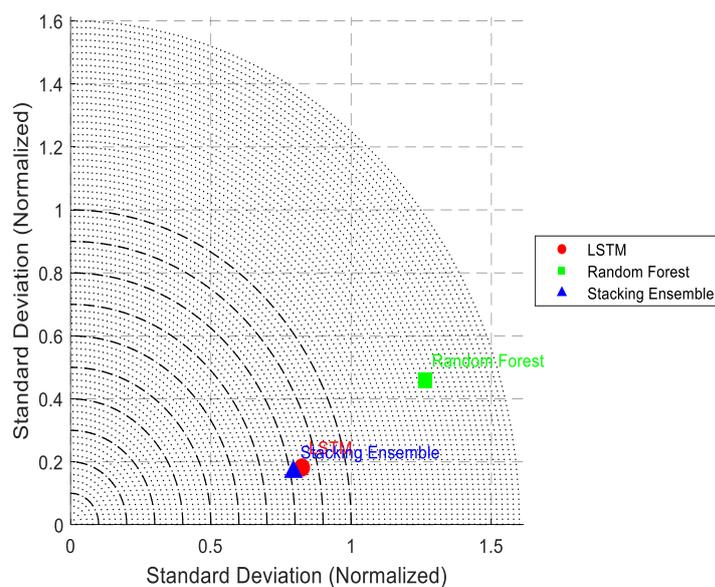
شکل ۸. نمودار پراکندگی نتایج مدل ترکیبی در حالت آموزش.

Fig 8. of the scatter plot of the results of the Stacking Ensemble model in training mode.



شکل ۹. نمودار پراکندگی نتایج مدل ترکیبی در حالت آزمایش.

Fig 9. of the scatter plot of the results of the Stacking Ensemble model in testing mode.



شکل ۱۰. نمودار نمودار تیلور برای انتخاب بهترین مدل یادگیری ماشین در مدل‌سازی مرحله آزمون.

Fig 10. Taylor diagram for selecting the best machine learning model in testing phase modeling.

مشاهداتی، چالش برانگیزتر بودند. تغییر  $\pm 10\%$  در هدایت هیدرولیکی، خطای مدل عددی را تا  $25\%$  افزایش داد، در حالی که مدل ترکیبی تنها حدود  $5\%$  تغییر نشان داد که بیانگر توانایی برتر روش‌های یادگیری ماشین در مدیریت عدم قطعیت پارامتری است (Di Salvo, 2022). همچنین، حذف چاه‌های با NSE منفی (چاه‌های ۳، ۱۳ و ۱۴) از فرایند آموزش و تحلیل حساسیت، دقت کلی مدل‌های یادگیری ماشین را بهبود بخشید و تمرکز بر مناطق نماینده‌تر آبخوان را ممکن ساخت.

مدل ترکیبی با وزن‌دهی مبتنی بر دقت، علی‌رغم سادگی نسبی، تفسیرپذیری مناسبی ارائه داد که از طریق تحلیل SHAP به خوبی آشکار شد. این تحلیل تأکید کرد که ویژگی‌های استخراج‌شده از خروجی‌های مدل عددی (به‌ویژه چاه ۱) نقش کلیدی داشتند و میانگین اهمیت SHAP آن‌ها برابر  $0.28$  بود. مقایسه این ویژگی‌ها با معادلات حاکم بر جریان در مدل MODFLOW نیز تأیید کرد که ادغام دلنش فیزیکی، الگوهای پیش‌بینی را به‌صورت معنادار تقویت کرده و به عملکرد واقع‌گرایانه در مجموعه آزمایشی منجر شده است.

در مجموعه آزمایشی، مقادیر  $NSE = 0.978$ ،  $KGE = 0.946$  و  $R^2 = 0.978$  نشان‌دهنده دقت بالا، خطای کم و تعمیم‌پذیری مناسب مدل ترکیبی هستند (Janssen et al., 2025; Torres-Martínez et al., 2024). این عملکرد با معیارهای استاندارد هیدرولوژیکی همخوانی دارد؛ به‌طوری که  $KGE \geq 0.9$  به‌عنوان «بسیار خوب» و  $NSE \geq 0.95$  به‌عنوان «عالی» طبقه‌بندی می‌شود. بهبود کلی در تعادل معیارها (به‌ویژه RMSE پایین‌تر و  $R^2$  بالاتر نسبت به LSTM) برتری رویکرد ترکیبی را در شرایط عملی تأیید با این حال، وابستگی مدل به کیفیت خروجی‌های MODFLOW و پیچیدگی محاسباتی بالاتر نسبت به

در مقایسه، مدل LSTM با ضریب همبستگی  $0.977$ ، انحراف معیار نرمال شده  $0.845$  و  $RMSE = 0.034$  متر، و مدل RF با ضریب همبستگی  $0.940$ ، انحراف معیار نرمال شده  $1.344$  و  $RMSE = 0.053$  متر عملکرد ضعیف‌تری داشتند.

هرچند مقدار KGE مدل ترکیبی ( $0.946$ ) اندکی کمتر از LSTM ( $0.952$ ) بود، اما RMSE پایین‌تر و  $R^2$  بالاتر آن، تعادل بهتری بین اجزای همبستگی، بلیاس و تطابق واریانس برقرار کرد. این تفاوت جزئی احتمالاً به پدیده هم‌خطی<sup>۱</sup> در متا-مدل جنگل تصادفی مربوط می‌شود که می‌تواند در مجموعه‌های داده محدود تأثیرگذار باشد (Chen et al., 2023).

استفاده از خروجی‌های کالیبره‌شده مدل عددی به‌عنوان ویژگی‌های ورودی، دقت پیش‌بینی را در شرایط کمبود داده‌های مشاهده‌ای طولانی‌مدت دشت بیرجند به طور قابل توجهی ارتقا داد (Chen et al., 2023). پیش‌پردازش داده‌ها با اعمال فیلترهای صاف‌سازی، نوسانات ناشی از خطاهای اندازه‌گیری یا تغییرات ناگهانی پمپاژ را کاهش داد و ویژگی‌های زمانی مانند میانگین متحرک و مشتقات سری زمانی را برای مدل LSTM تقویت کرد. مدل RF نیز با مکانیسم میانگین‌گیری پیش‌بینی‌های درخت‌های تصمیم، حساسیت کمتری به تغییرات محلی هدایت هیدرولیکی نشان داد. هرچند در شناسایی و بازتولید الگوهای فصلی پیچیده، نسبت به مدل‌های LSTM و ترکیبی عملکرد ضعیف‌تری داشت.

تحلیل حساسیت با استفاده از روش‌های SHAP و شبیه‌سازی مونت‌کارلو (۱۰۰۰ تکرار) نشان داد که چاه ۱ با ضریب همبستگی  $0.949$  و اهمیت SHAP برابر  $0.28$ ، مهم‌ترین چاه مشاهده‌ای است. چاه‌های ۵، ۱۰ و ۱۵ به دلیل ناهمگنی شدید محلی و محدودیت داده‌های

<sup>1</sup> Collinearity

مهم در مدیریت منابع آب مناطق نیمه‌خشک به شمار می‌رود.

تحلیل حساسیت مبتنی بر روش‌های SHAP و مونت‌کارلو، چاه شماره ۱ را به عنوان نقطه کلیدی شناسایی کرد که با اهمیت نسبی ۰/۲۸، مبنایی برای طراحی شبکه‌های پایش هدمند شد و نشان داد چگونه این رویکرد می‌تواند عدم قطعیت‌های محلی را در نواحی با هدایت هیدرولیکی متغیر (۵ تا ۴۵ متر بر روز) مدیریت کند. نتایج همگرایی مدل ترکیبی و عددی، ظرفیت برداشت پایدار در بخش‌های شرقی آبخوان (افت ۴۰ تا ۶۰ سانتی متر سالانه) را برجسته کرد، در حالی که بخش‌های غربی نیازمند راه‌حل‌های مدیریتی فوری مانند کاهش ۳۰ تا ۴۰ درصدی برداشت کشاورزی، پیاده‌سازی سیستم‌های آبیاری هوشمند (قطره‌ای و زیرسطحی) و ادغام تغذیه مصنوعی از طریق تزریق آب برگشتی است تا پایداری اکوسیستم و تاب‌آوری در برابر تغییرات اقلیمی تضمین شود.

این چارچوب، هم‌راستا با اصول مدیریت یکپارچه منابع آب (IWRM) و استانداردهای وزارت نیرو، ابزار عملی برای پایش مداوم و بهینه‌سازی الگوی مصرف فراهم می‌کند و می‌تواند در دشت‌های مشابه با چالش‌های هیدرولوژیکی پیچیده اعمال شود. برای پیشرفت‌های علمی آینده، پیشنهاد می‌شود از داده‌های ماهواره‌ای پیشرفته (مانند GRACE-FO و Sentinel) برای بهبود مدل‌سازی فضایی، شبکه‌بندی دقیق‌تر آبخوان با رزولوشن زیر ۱۰۰ متر، و سیستم‌های تله‌متری بلادرنگ بهره برد تا دقت پیش‌بینی‌ها فراتر رود و راه‌حل‌های مدیریتی مبتنی بر هوش مصنوعی، مانند سامانه‌های هشدار زودهنگام افت تراز، توسعه یابد و در نهایت به پایداری بلندمدت منابع آب در شرایط کمبود داده کمک کند.

مدل‌های پایه، از محدودیت‌های اصلی این روش به شمار می‌روند.

در نهایت، چارچوب ترکیبی پیشنهادی با بهبود قابل توجه دقت پیش‌بینی در مناطق خشک و نیمه‌خشک مانند دشت بیرجند، ابزاری مؤثر برای مدیریت پایدار منابع آب زیرزمینی ارائه می‌دهد. این بهبود می‌تواند به برنامه‌ریزی بهینه برداشت، کاهش نرخ افت سطح آب و حملیت از بخش کشاورزی کمک شایانی نماید (Aghlmand & Abbas, 2019; Porhemmat et al., 2019). ترکیب دانش فیزیکی با روش‌های یادگیری ماشین پیشرفته، راهکاری امیدوارکننده برای مواجهه با چالش‌های تغییر اقلیم و کمبود داده‌های مشاهداتی است و پیشنهاد می‌شود در سایر دشت‌های ممنوعه کشور نیز مورد آزمون قرار گیرد.

### نتیجه‌گیری

این پژوهش با تلفیق نوآورانه مدل‌های یادگیری ماشین و خروجی‌های کالیبره‌شده مدل عددی MODFLOW، چارچوبی ترکیبی برای پیش‌بینی تراز آب زیرزمینی در دشت بیرجند ارائه داد. این رویکرد نه تنها دقت پیش‌بینی را به سطح بالایی ارتقا داد ( $R^2=0.978$ ,  $RMSE=0.032$  متر،  $NSE=0.978$  و  $KEG=0.946$  در مجموعه آزمایشی)، بلکه محدودیت‌های مدل‌های فیزیکی سنتی (مانند چالش در بازتولید ناهمگنی محلی) و مدل‌های صرفاً داده‌محور (مانند حساسیت به نویز) را با ادغام دانش هیدروژئولوژیکی و الگوهای غیرخطی داده‌ها برطرف کرد. نوآوری کلیدی در پیش‌پردازش پیشرفته داده‌ها با فیلتر Savitzky-Golay برای کاهش نویز و افزودن نویز گاوسی کنترل‌شده برای جلوگیری از بیش‌برازش نهفته بود که تعمیم‌پذیری مدل را در آبخوان‌های ناهمگن با نرخ افت سالانه ۴۰ تا ۶۰ سانتی‌متر به طور قابل توجهی افزایش داد و دستاوردی

## منابع

- Dunnington, D. W., Trueman, B. F., Raseman, W. J., Anderson, L. E., & Gagnon, G. A. (2021). Comparing the predictive performance, interpretability, and accessibility of machine learning and physically based models for water treatment. *ACS ES&T Engineering*, 1(3), 348–356. <https://doi.org/10.1021/acsestengg.0c00053>
- Ehteram, M., & Ghanbari-Adivi, E. (2023). Self-attention (SA) temporal convolutional network (SATCN)-long short-term memory neural network (SATCN-LSTM): An advanced Python code for predicting groundwater level. *environmental science and pollution research*, 30(40), 92903-92921. <https://doi.org/10.1007/s11356-023-28771-8>
- Emamgholizadeh, S., Khoshnevisan, S., Asli Charandabi, M. R., & Khanzad, S. F. (2025). Comparing the performance of machine learning algorithms in modeling and predicting groundwater levels (Case study: Shahrood Plain). *Journal of Aquifer and Qanat*, 6(2), 77-96. <https://doi.org/10.22077/jaaq.2025.10441.1132>
- Ghebrehiwot, A. A., & Kozlov, D. V. (2019). Hydrological modelling for ungauged basins of arid and semi-arid regions: Review. *Vestnik MGSU*, 8, 1023–1036. <https://doi.org/10.22227/1997-0935.2019.8.1023-1036>
- Gupta, H. V., Kling, H., Yilmaz, K. K., & Martinez, G. F. (2009). Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. *Journal of Hydrology*, 377(1-2), 80–91. <https://doi.org/10.1016/j.jhydrol.2009.08.003>
- Janssen, J., Tootchi, A., & Ameli, A. A. (2025). Tackling water table depth modeling via machine learning: From proxy observations to verifiability. *Advances in Water Resources*, 201, 104955. <https://doi.org/10.1016/j.advwatres.2025.104955>
- Kishor, K., Aggarwal, A., Srivastava, P. K., Sharma, Y. K., Lee, J., & Ghobadi, F. (2025). A systematic literature review of MODFLOW combined with artificial neural networks (ANNs) for groundwater flow modelling. *Water*, 17(16), 23-75. <https://doi.org/10.3390/w17162375>
- Gupta, S. K., Sahoo, S., Sahoo, B. B., Srivastava, P. K., Pateriya, B., & Santosh, D. T. (2024). Prediction of groundwater level changes based on machine learning technique in highly groundwater irrigated alluvial aquifers of south-central Punjab, India. *Physics and Chemistry of the Earth*, 135, 103603. <https://doi.org/10.1016/j.pce.2024.103603>
- Igwebuike, N., Ajayi, M., Okolie, C., Kanyerere, T., & Halihan, T. (2025). Application of machine learning techniques in groundwater level prediction: A review. *Journal of Hydrology*, 637, 131370. <https://doi.org/10.1016/j.jhydrol.2024.131370>
- Adombi, A. V. D. P., Chesnaux, R., Boucher, M.-A., Braun, M., & Lavoie, J. (2024). A causal physics-informed deep learning formulation for groundwater flow modeling and climate change effect analysis. *Journal of Hydrology*, 637, 131370. <https://doi.org/10.1016/j.jhydrol.2024.131370>
- Aghlmand, R., & Abbasi, A. (2019). Application of MODFLOW with boundary conditions analyses based on limited available observations: A Case study of Birjand plain in east Iran. *Water*, 11(9), 1904. <https://doi.org/10.3390/w11091904>
- Alao, J. O., Bello, A., Lawal, H., & Abdullahi, D. (2024). Assessment of groundwater challenge and the sustainable management strategies. *Results in Earth Sciences*, 2, 100049. <https://doi.org/10.1016/j.rines.2024.100049>
- Borzì, I. (2025). Modeling groundwater resources in data-scarce regions for sustainable management: methodologies and limits. *Hydrology*, 12(1), 11. <https://doi.org/10.3390/hydrology12010011>
- Chang, Y.-W., Sun, W., Kow, P.-Y., Lee, M.-H., Chang, L.-C., & Chang, F.-J. (2025). Advanced groundwater level forecasting with a hybrid deep learning model: Tackling water challenges in Taiwan's largest alluvial fan. *Journal of Hydrology*, 655, 132887. <https://doi.org/10.1016/j.jhydrol.2025.132887>
- Chen, C., Zhang, H., Shi, W., Zhang, W., & Xue, Y. (2023). A novel paradigm for integrating physics-based numerical and machine learning models: A case study of an eco-hydrological model. *Environmental Modelling & Software*, 163, 105669. <https://doi.org/10.1016/j.envsoft.2023.105669>
- Cui, X., Wang, Z., Xu, N., Wu, J., & Yao, Z. (2024). A secondary modal decomposition ensemble deep learning model for groundwater level prediction using multi-data. *Environmental Modelling & Software*, 175, 105969. <https://doi.org/10.1016/j.envsoft.2024.105969>
- Dastourani, M., Khayat, A., & Akhondi, Z. (2024). Accuracy evaluation of hybrid wavelet-neuro-fuzzy and wavelet models in predicting groundwater level (Case study: Birjand Plain). *Journal of Aquifer and Qanat*, 5(2), 139-154. <https://doi.org/10.22077/jaaq.2025.9046.1105>
- Di Salvo, C. (2022). Improving results of existing groundwater numerical models using machine learning techniques: A review. *Water*, 14(15), 2307. <https://doi.org/10.3390/w14152307>

*Environment*, 949, 174973.

<https://doi.org/10.1016/j.scitotenv.2024.174973>

Zarei, E., Saleh, F. N., & Dalir, A. N. (2024). Comparing the hybrid-lumped-LSTM model with a semi-distributed model for improved hydrological modeling. *Journal of Water and Climate Change*, 15(8), 4099–4113.

<https://doi.org/10.2166/wcc.2024.269>

learning and deep learning for predicting groundwater levels in the West Coast Aquifer System, South Africa. *Earth Science Informatics*, 18(1), 6. <https://doi.org/10.1007/s12145-024-01623-w>

Navarro-Farfán, M. D. M., García-Romero, L., Martínez-Cinco, M. A., Hernández-Hernández, M. A., & Sánchez-Quispe, S. T. (2024). Comparison between MODFLOW groundwater modeling with traditional and distributed recharge. *Hydrology*, 11(1), 9.

<https://doi.org/10.3390/hydrology11010009>

Nohani, E., Babaali, H. R., & Dehghani, R. (2024). Groundwater level modeling of Delfan Plain in Lorestan Province using meta-heuristic hybrid models. *Journal of Aquifer and Qanat*, 5(2), 79-98. <https://doi.org/10.22077/jaaq.2025.8834.1096>

Fotovat, M., & BABAZADEH, H. (2019). Evaluation of WEAP-MODFLOW model as an integrated water resources management model for sustainable development (A case study: Gharesoo at Doab-Merek, Kermanshah, Iran). <https://doi.org/10.22059/cej.2019.260084.1495>

Pourmorad, S., Kabolizade, M., & Dimuccio, L. A. (2024). Artificial intelligence advancements for accurate groundwater level modelling: An updated synthesis and review. *Applied Sciences*, 14(16), 7358. <https://doi.org/10.3390/app14167358>

Rad, M., Abtahi, A., Berndtsson, R., McKnight, U. S., & Aminifar, A. (2024). Interpretable machine learning for predicting the fate and transport of pentachlorophenol in groundwater. *Environmental Pollution*, 345, 123449. <https://doi.org/10.1016/j.envpol.2024.123449>

Rammohan, B., Partheeban, P., Ranganathan, R., & Balaraman, S. (2024). Groundwater quality prediction and analysis using machine learning models and geospatial technology. *Sustainability*, 16(22), 9848. <https://doi.org/10.3390/su16229848>

Sheikh Khozani, Z., Barzegari Banadkooki, F., Ehteram, M., Najah Ahmed, A., & El-Shafie, A. (2022). Combining autoregressive integrated moving average with Long Short-Term Memory neural network and optimisation algorithms for predicting groundwater level. *Journal of Cleaner Production*, 348, 131224. <https://doi.org/10.1016/j.jclepro.2022.131224>

Torres-Martínez, J. A., Mahlkecht, J., Kumar, M., Loge, F. J., & Kaown, D. (2024). Advancing groundwater quality predictions: Machine learning challenges and solutions. *Science of The Total*