



Application of machine learning algorithms in groundwater level prediction in the Ardabil aquifer

Seyed Mahdi Hoseini¹ | Morteza Khoshsimaie Chenar²

1. Department of Irrigation and Reclamation Engineering, College of Agriculture and Natural Resources, University of Tehran, Karaj, Iran.

E-mail: Mahdi.hoseini@ut.ac.ir

2. Corresponding Author, Department of Irrigation and Reclamation Engineering, College of Agriculture and Natural Resources, University of Tehran, Karaj, Iran. E-mail: khoshsima.mortaza@ut.ac.ir

Article Info

ABSTRACT

Article type: Research Article

Article history:

Received: Dec. 30, 2024

Revised: Feb. 16, 2025

Accepted: Feb. 23, 2025

Published online: June. 2025

Keywords:

Random Forest,
Support Vector Machine,
Variance Inflation Factor,
Water Management.

Accurate prediction of groundwater levels is of great importance in water resource management, especially in arid regions. This research, with the aim of replacing traditional models with machine learning methods, has examined two algorithms: Support Vector Machine and Random Forest for predicting groundwater levels. Modeling was conducted using 20 years of data on precipitation, air temperature, evaporation, groundwater extraction, and groundwater level as input variables. After examining the normality and correlation of the data, 70% of the data were used for training and 30% for testing. The evaluation results of the R^2 , RMSE, MAE, and MSE metrics showed that the Support Vector Machine -RBF algorithm had values of 0.57, 1.05, 0.61, and 1.11 in the training phase, and values of 0.74, 0.84, 0.61, and 0.71 in the testing phase, respectively. The Random Forest algorithm, using all hyperparameters, provided values of 0.85, 0.61, 0.44, and 0.37 in the training phase and 0.71, 0.93, 0.66, and 0.86 in the testing phase, and showed better performance due to its high accuracy and resistance to multicollinearity. Additionally, using the Permutation Feature Importance method, the number of input variables for the Random Forest model was reduced from six to one, and its results, without a significant decrease in model accuracy, included values of 0.83, 0.66, 0.45, and 0.43 in the training phase and 0.71, 0.93, 0.66, and 0.86 in the testing phase. The research findings indicate that machine learning models, particularly the Random Forest algorithm, can be a suitable alternative to traditional methods for predicting groundwater levels and managing water resources sustainably.

Cite this article Hosseini, S. M., & Khoshsimaie Chenar, M. (2025). Application of machine learning algorithms in groundwater level prediction in the Ardabil aquifer, *Iranian Journal of Soil and Water Research*, 56 (4), 1041-1057. <https://doi.org/10.22059/ijswr.2025.387848.669863>

© The Author(s).

Publisher: The University of Tehran Press.

DOI: <https://doi.org/10.22059/ijswr.2025.387848.669863>



EXTENDED ABSTRACT

Introduction:

Groundwater is considered one of the most vital sources of drinking and agricultural water, especially in arid and semi-arid regions like Iran, accounting for nearly 55 percent of the country's annual water consumption. However, various factors such as population growth and hydrogeological and meteorological changes have caused an imbalance in the extraction and recharge of aquifers, leading to a decrease in groundwater levels and an increase in related crises. These crises include reduced natural recharge, resource pollution, and ecosystem instability, which threaten food security and economic development. In this context, sustainable management of groundwater resources requires tools that can accurately predict water level changes and facilitate planning and decision-making.

The use of artificial intelligence algorithms, such as Support Vector Machines (SVM) and Random Forests (RF), in predicting groundwater levels has proven highly efficient due to their ability to analyze complex patterns and nonlinear environments. These models are more flexible than traditional numerical methods and perform well even with limited data. Various studies, including the research of Yoon et al., Gupta et al., and Cordoi Milan in Iran, demonstrate the high accuracy of these algorithms in simulating groundwater level fluctuations. Specifically, the RF model has shown a strong capability in analyzing large datasets and reducing overfitting, and it has been identified as one of the most effective tools for predicting groundwater resources.

Purpose:

This study is designed to identify the most accurate regression model for predicting groundwater levels and specifically compares the performance of two advanced machine learning algorithms, Random Forest (RF) and Support Vector Machine (SVM). Using monthly historical data, this research aims to analyze the complex patterns of groundwater level fluctuations and provide a model with higher predictive accuracy. The importance of this issue is particularly pronounced in arid and semi-arid regions, where water crises and climate change exert greater pressure on groundwater resources. The results of this comparison can not only improve the accuracy of predictions but also aid in developing sustainable management strategies and informed decision-making for the protection and optimal utilization of groundwater resources.

Method:

Parts of Ardabil Province in northwest Iran—including the 4,704 square kilometer Ardabil Plain—are the subject of this study. Covering around 90% of the 969 square kilometer Ardabil Plain, the aquifer has an average depth of 53 meters. From October 1998 to September 2020, meteorological data, groundwater levels, and well information were gathered. These records comprise precipitation, evaporation, temperature, groundwater extraction, and groundwater levels over several time periods. Support Vector Machine (SVM) and Random Forest (RF) models were applied to replicate the groundwater level; the data was randomly split into training (70%) and testing (30%). Using statistical measures such as the R^2 correlation coefficient, root mean square error (RMSE), mean absolute error (MAE), and mean square error (MSE), one can assess the performance of the models.

Results and Discussion:

In this study, two machine learning algorithms, SVM and RF, were used to predict the groundwater level in the Ardabil plain. The results showed that the RF model performed better than the SVM; during the testing phase, the performance indicators of RF (R^2 , RMSE, MAE, and MSE) improved compared to SVM, providing higher prediction accuracy. The RF model outperformed the SVM model due to its resistance to multicollinearity and its ability to manage complex data, while the SVM was sensitive to multicollinearity. Additionally, RF was able to enhance prediction accuracy by reducing the number of inputs from 5 to 2 variables. These results emphasize the effectiveness of the Random Forest algorithm in predicting groundwater levels and managing groundwater resources, and they indicate that for higher accuracy, the use of spatial data and more complex models in future research will be beneficial.

Author Contributions

All authors contributed equally to the conceptualization of the article and the writing of the original and subsequent drafts.

Ethical considerations

The authors avoided data fabrication, falsification, plagiarism, and misconduct.

Conflict of interest

The author declares no conflict of interest.

کاربرد الگوریتم‌های یادگیری ماشین در پیش‌بینی تراز آب زیرزمینی در آبخوان اردبیل

سید مهدی حسینی^۱ | مرتضی خوش‌سیمای چنار^۲^۱ گروه مهندسی آبیاری و آبادانی، دانشکده‌گان کشاورزی و منابع طبیعی، دانشگاه تهران، تهران، ایران. رایانامه: Mahdi.hoseini@ut.ac.ir^۲ نویسنده مسئول، گروه مهندسی آبیاری و آبادانی، دانشکده‌گان کشاورزی و منابع طبیعی، دانشگاه تهران، تهران، ایران. رایانامه: khoshshima.mortaza@ut.ac.ir

اطلاعات مقاله	چکیده
نوع مقاله: مقاله پژوهشی	پیش‌بینی دقیق تراز آب زیرزمینی در مدیریت منابع آبی، به‌ویژه در مناطق خشک، از اهمیت بالایی برخوردار است. این پژوهش باهدف جایگزینی روش‌های یادگیری ماشین به‌جای مدل‌های سنتی، دو الگوریتم ماشین بردار پشتیبان و جنگل تصادفی را برای پیش‌بینی تراز آب زیرزمینی بررسی کرده است. مدل‌سازی با استفاده از داده‌های ۲۰ ساله بارش، دمای هوا، تبخیر، برداشت از آبخوان و تراز آب زیرزمینی به‌عنوان متغیرهای ورودی انجام شد. پس از بررسی نرمالیده و همبستگی داده‌ها، ۷۰ درصد داده‌ها به‌عنوان آموزش و ۳۰ درصد به‌عنوان تست استفاده شدند. نتایج ارزیابی معیارهای R^2 ، RMSE، MAE و MSE نشان داد که الگوریتم ماشین بردار پشتیبان در مرحله آموزش به‌ترتیب دارای مقادیر ۰/۵۷، ۱/۰۵، ۰/۶۱ و ۱/۱۱ و در مرحله آزمایش مقادیر ۰/۷۴، ۰/۸۴، ۰/۶۱ و ۰/۷۱ است. الگوریتم جنگل تصادفی با استفاده از تمامی ویژگی‌ها (ورودی‌ها) به‌ترتیب مقادیر ۰/۸۵، ۰/۶۱، ۰/۴۴ و ۰/۳۷ در مرحله آموزش و ۰/۷۱، ۰/۹۳، ۰/۶۶ و ۰/۸۶ در مرحله آزمایش را ارائه داد و به‌دلیل دقت بالا و مقاومت در برابر هم‌خطی، عملکرد بهتری نشان داد. همچنین با استفاده از روش اهمیت ویژگی جایگشت، تعداد متغیرهای ورودی الگوریتم جنگل تصادفی از شش به یک کاهش یافت و نتایج آن بدون کاهش محسوس دقت مدل شامل مقادیر ۰/۸۳، ۰/۶۶، ۰/۴۵ و ۰/۴۳ در مرحله آموزش و ۰/۷۱، ۰/۹۳، ۰/۶۶ و ۰/۸۶ در مرحله آزمایش بود. یافته‌های پژوهش نشان می‌دهند که مدل‌های یادگیری ماشین، به‌ویژه الگوریتم جنگل تصادفی، می‌توانند جایگزین مناسبی برای روش‌های سنتی در پیش‌بینی تراز آب زیرزمینی و مدیریت پایدار منابع آبی باشند.
واژه‌های کلیدی: جنگل تصادفی، شاخص تورم واریانس، ماشین بردار پشتیبان، مدیریت آب.	
استناد: حسینی، سید مهدی و خوش‌سیمای چنار، مرتضی (۱۴۰۴). کاربرد الگوریتم‌های یادگیری ماشین در پیش‌بینی تراز آب زیرزمینی در آبخوان اردبیل، مجله تحقیقات آب و خاک/ایران، ۵۶ (۴)، ۱۰۵۷-۱۰۴۱. https://doi.org/10.22059/ijswr.2025.387848.669863	
ناشر: مؤسسه انتشارات دانشگاه تهران.	© نویسندگان.
DOI: https://doi.org/10.22059/ijswr.2025.387848.669863	



مقدمه

آب زیرزمینی یکی از مهم‌ترین منابع تأمین آب شرب و کشاورزی در جهان، به‌ویژه در کشورهای خشک و نیمه‌خشک نظیر ایران است، به‌گونه‌ای که نزدیک به ۵۵ درصد از کل مصرف آب سالانه ایران را تأمین می‌کند (Karimi Alavijeh et al., 2021; Rasouli & Cheraghi, 2009). با این حال، عدم تعادل در شارژ و برداشت از آبخوان‌ها تحت عواملی نظیر رشد جمعیت، هیدروژئولوژیکی و هواشناسی سبب کاهش جدی سطح آب زیرزمینی و افزایش بحران‌های مرتبط شده است (Lancia et al., 2022). این بحران‌ها شامل کاهش تغذیه طبیعی آبخوان‌ها، افزایش آلودگی منابع زیرزمینی، و ناپایداری اکوسیستم‌ها است که در نهایت امنیت غذایی و توسعه اقتصادی را تهدید می‌کنند (Kløve et al., 2014; Kumar, 2012; Perez et al., 2024). پاسخ به بحران‌های نمایان در آبخوان‌ها بسیار پیچیده و غیر خطی بوده که درک مکانیزم پاسخ پیچیده سطح آب زیرزمینی با در نظر گرفتن ناهمگونی‌های اجتماعی-هیدروژئولوژیکی برای برنامه‌ریزی و مدیریت پایدار امری حیاتی می‌باشد (Minnig et al., 2018; Rathnayaka et al., 2016).

در شرایط کنونی، مدیریت پایدار منابع آب زیرزمینی به ابزارهایی نیاز دارد که بتوانند تغییرات تراز آب را به‌دقت پیش‌بینی کنند (Sun et al., 2022). این پیش‌بینی‌ها به تصمیم‌گیری‌های مدیریتی کمک کرده و امکان برنامه‌ریزی بلندمدت برای حفاظت از منابع را فراهم می‌کنند (Gleeson et al., 2012). مدل‌سازی جریان آب زیرزمینی یکی از ابزارهای حیاتی برای این منظور است (Ghourdoyee Milan et al., 2020). مدل‌های عددی سنتی، مانند MODFLOW، اگرچه قادر به شبیه‌سازی دقیق جریان آب در بسیاری از شرایط هستند، اما محدودیت‌هایی نظیر نیاز به داده‌های گسترده و دقیق، زمان‌بر بودن پردازش، و فرضیات ساده‌کننده دارند که کارایی آنها را در مسائل پیچیده و غیرخطی کاهش می‌دهد (Feng et al., 2011; Hu & Jiao, 2015; Rajaei et al., 2019). با توجه به این چالش‌ها، استفاده از الگوریتم‌های هوش مصنوعی که قادر به شناسایی الگوهای پیچیده در داده‌ها هستند، می‌تواند کمک شایانی به دقت پیش‌بینی‌ها کند. این الگوریتم‌ها به‌ویژه در مواجهه با داده‌های غیرخطی و محیط‌های پیچیده، توانایی بالایی در افزایش دقت پیش‌بینی‌ها دارند.

در سال‌های اخیر، استفاده از مدل‌های هوش مصنوعی در مدل‌سازی مسائل آب‌های زیرزمینی رو به گسترش است، زیرا این مدل‌ها توانایی بالایی در شناسایی الگوهای پیچیده و مسائل غیرخطی دارند که باعث افزایش دقت در حل این مسائل می‌شود (Y. Chen & Zhang, 2020; Poursaeid et al., 2020; Zhao et al., 2019). همچنین، مدل‌های هوش مصنوعی انعطاف‌پذیری زیادی در مواجهه با تغییر شرایط و داده‌های جدید دارند، درحالی‌که مدل‌های عددی نیازمند بازنگری کامل می‌باشند که این امر سبب صرفه‌جویی در زمان می‌شود (Rajaei et al., 2019). از انواع الگوریتم‌های یادگیری ماشین می‌توان به الگوریتم‌های ماشین بردار پشتیبان^۱ (SVM) و جنگل تصادفی^۲ (RF) به عنوان یکی از پرکاربردترین الگوریتم‌های یادگیری ماشین در شبیه‌سازی‌های مختلف اشاره کرد (Foody, 1995; Friedl & Brodley, 1997). SVM به دلیل کارایی بالا در طبقه‌بندی و پیش‌بینی دقیق، به‌ویژه در مواجهه با داده‌های غیرخطی و RF با قابلیت تحلیل داده‌های بزرگ و کاهش خطر بیش‌برازش، به عنوان دو گزینه برجسته در این زمینه شناخته می‌شوند (Cao et al., 2020; Catani et al., 2013). این الگوریتم‌ها به‌ویژه در پیش‌بینی تراز آب زیرزمینی کاربرد دارند، زیرا قادرند الگوهای پیچیده موجود در نوسانات سطح آب را شناسایی کنند. در ادامه به نتایج برخی از محققان داخلی و خارجی در زمینه پیش‌بینی تراز آب زیرزمینی پرداخته شده است که نتایج تحقیقات بیان‌کننده عملکرد قابل قبول SVM و RF می‌باشد.

گوپتا و همکاران (۲۰۲۴) با استفاده از الگوریتم‌های RF، درخت هرس با خطای پاک‌کننده^۳ (Bagging-REPTree) و درخت بگینگ-تصمیم‌گیری^۴ (Bagging-DSTree) اقدام به پیش‌بینی تراز آب زیرزمینی جنوب پنجاب کردند، مناطقی که در این منطقه عوامل اقلیمی-آبی، جریان آب زیرزمینی و آب سطحی در تقابل غیرمستقیم با هم بوده و آب آبیاری غالباً از منابع زیرزمینی تأمین می‌شود. نتایج نشان داد که الگوریتم RF نتایج بهتری در پیش‌بینی سطح آب زیرزمینی در چنین آبخوان‌هایی دارد (Gupta et al., 2024). قوردویی میلان و همکاران (۱۳۹۸) در آبخوان امام‌زاده جعفر استان کهگیلویه و بویراحمد، با استفاده از الگوریتم‌های رگرسیون خطی چند متغیره، ANN و حداقل مربعات بردار پشتیبان^۵ (LS-SVM)، اقدام به شبیه‌سازی تراز آبخوان امام‌زاده با استفاده از متغیرهای تراز آب زیرزمینی ماه قبل، بارش، برداشت از آب زیرزمینی، دما و تبخیر کردند. نتایج نشان داد که الگوریتم LS-SVM نسبت به سایر الگوریتم‌ها دقت

1 Support Vector Machine

2 Random Forest

3 Bagging-Reduce Error Pruning Tree

4 Bagging-Decision Stump Tree

5 Last Squares Support Vector Machine

بیشتری را در شبیه‌سازی تراز آب زیرزمینی به خود اختصاص داده است (Ghourdoyee Milan et al., 2020). نوروزی قوشبلاغ و همکاران (۱۳۹۷) به منظور پیش‌بینی سطح آب زیرزمینی دشت بوکان از الگوریتم‌های شبکه ANN، منطق فازی و RF با استفاده از متغیرهای بارندگی، دما، دبی جریان و تراز سطح ایستابی ماه قبل پرداختند. نتایج نشان داد که الگوریتم RF نسبت به سایر الگوریتم‌ها، سطح ایستابی را با دقت قابل قبولی پیش‌بینی می‌کند (norouzi & nadiri, 2018).

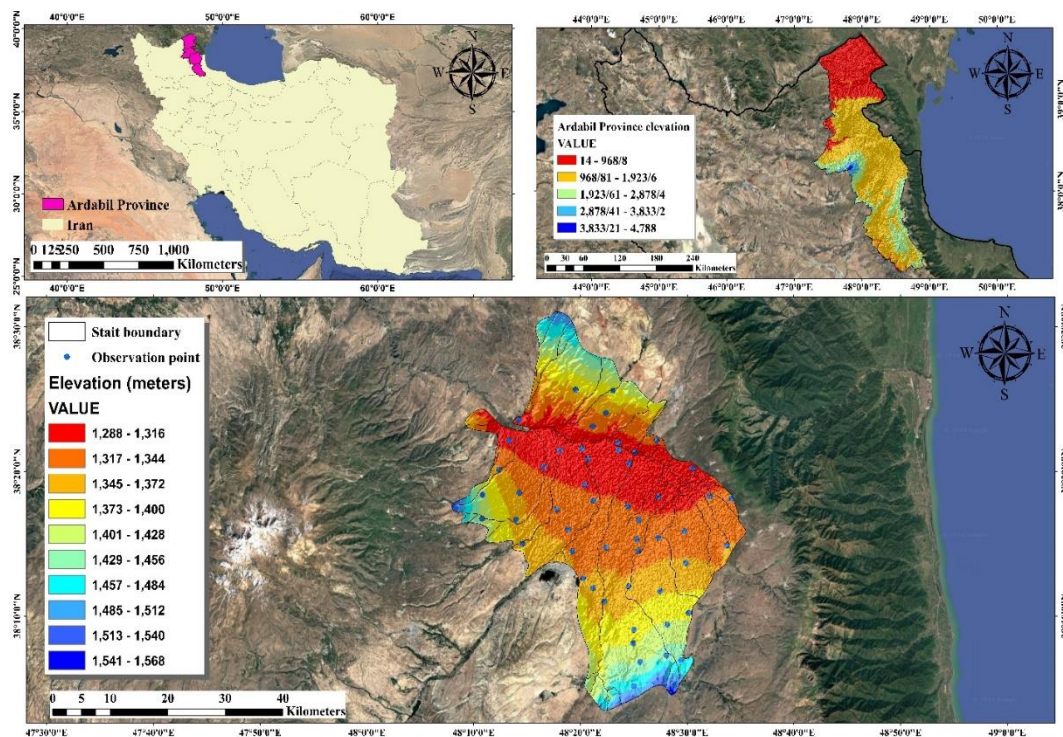
بر اساس مطالعات صورت گرفته در زمینه پیش‌بینی سطح آب زیرزمینی، الگوریتم‌های SVM و RF از پتانسیل قابل قبولی نسبت به سایر الگوریتم‌ها در زمینه پیش‌بینی تراز آب زیرزمینی برخوردارند. لذا هدف از مطالعه حاضر، یافتن دقیق‌ترین مدل رگرسیونی برای پیش‌بینی سطح آب زیرزمینی با استفاده از داده‌های تاریخی ماهانه است. این مطالعه با انجام یک تحلیل مقایسه‌ای بین SVM و RF طراحی شده است تا به شناسایی الگوهای پیچیده در نوسانات تراز آب زیرزمینی کمک کند. در شرایط تغییرات اقلیمی و بحران‌های آبی، پیش‌بینی دقیق تراز آب زیرزمینی به‌ویژه در مناطق خشک و نیمه‌خشک اهمیت ویژه‌ای می‌یابد. این پیش‌بینی‌ها می‌توانند به تصمیم‌گیری‌های استراتژیک در مدیریت منابع آب کمک کرده و از بحران‌های آبی در آینده جلوگیری کنند. لذا هدف از مطالعه حاضر، یافتن دقیق‌ترین مدل رگرسیونی برای پیش‌بینی سطح آب زیرزمینی با استفاده از داده‌های تاریخی ماهانه است. نتایج این مقایسه نه تنها به بهبود دقت پیش‌بینی‌ها کمک خواهد کرد، بلکه می‌تواند در توسعه راهکارهای مدیریتی مؤثر و پایدار برای حفاظت و بهینه‌سازی منابع آب زیرزمینی در شرایط تغییرات اقلیمی و بحران‌های آبی مؤثر باشد.

مواد و روش

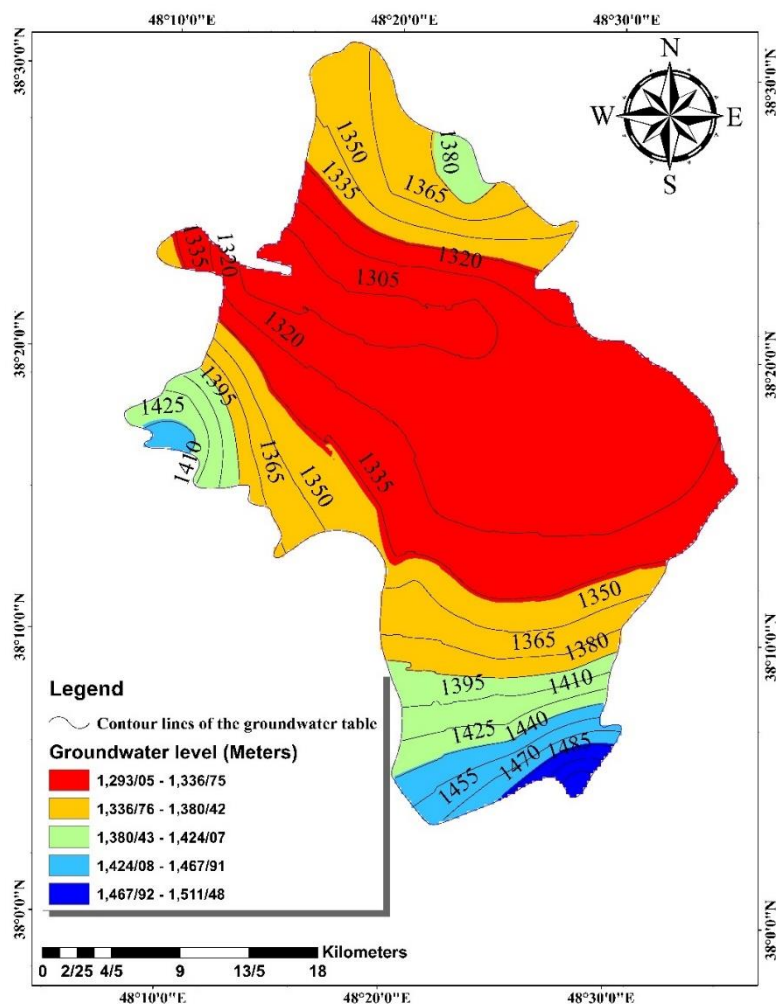
منطقه مورد مطالعه و داده‌های مورد استفاده

منطقه مورد مطالعه بخشی از استان اردبیل در شمال غربی کشور ایران می‌باشد و بین عرض جغرافیایی ۳۸ درجه و ۳ دقیقه تا ۳۸ درجه و ۳۷ دقیقه شمالی و طول جغرافیایی ۴۷ درجه و ۵۵ دقیقه تا ۴۸ درجه و ۲۰ دقیقه شرقی واقع شده است (وثوقی و همکاران، ۱۳۹۰). از نظر فیزیوگرافی، این منطقه در قسمت جنوبی حوزه رودخانه ارس واقع شده است. محدوده مطالعاتی اردبیل دارای وسعت ۴۷۰۴/۷ کیلومتر مربع بوده که مساحت ارتفاعات و دشت در این محدوده مطالعاتی به ترتیب ۳۷۳۰/۰۷ و ۱۰۷۴ کیلومتر مربع می‌باشد (جعفرزاده و همکاران، ۱۳۹۶). بر اساس داده‌های هواشناسی، متوسط بارندگی سالانه دشت اردبیل حدود ۲۷۳ میلی‌متر است. پرباران‌ترین و کم باران‌ترین ماه در این دشت به ترتیب، اردیبهشت و مرداد با متوسط بارندگی ۴۴/۵ و ۴/۵ میلی‌متر است. میانگین دمای سالانه ۹/۹ درجه سلسیوس بوده که بیشترین دمای تابستانه ۴۴ درجه و کمترین دما زمستانه ۳۳/۸- درجه سلسیوس می‌باشد. منطقه مطالعاتی دشت اردبیل در زون ساختاری البرز-آذربایجان واقع می‌باشد. این منطقه به دلیل شرایط کوهستانی از تنوع متنوع زمین‌شناسی برخوردار است که در این زمینه می‌توان در بخش‌های شمال شرقی به واحدهای کرتاسه، جنوب غرب به واحدهای پلئوسن، شمال به برش‌های آتشفشانی و غرب تا جنوب غرب تئوژن اشاره کرد (Aghazadeh et al., 2017). بر همین اساس، دشت اردبیل منطقه پست بوده که توسط ارتفاعات اطراف دشت محاصره شده و تنها خروجی آن رودخانه قره‌سو با مجموع جریان سالانه ۵۵۴ میلیون متر مکعب می‌باشد (اکبری نیاری و همکاران، ۱۴۰۳). باید توجه داشت که این دشت از آبرفت‌های رودخانه‌های جاری در خود تشکیل شده است (فضابخش و همکاران، ۱۴۰۱). آبخوان دشت اردبیل با مساحت ۹۶۹ کیلومتر حدود ۹۰ درصد از دشت اردبیل را شامل می‌شود. بر اساس مشاهدات انجام شده بر روی لگ چاه‌ها، آبخوان دشت اردبیل از نوع آبخوان آزاد بوده و عمق لایه آبرفت به طور متوسط به ۵۳ متر می‌رسد که حداکثر ضخامت آن در قسمت جنوبی با عمق ۲۲۰ متر است. شکل ۱ موقعیت جغرافیایی آبخوان دشت اردبیل و چاه‌های مشاهداتی منتخب را نشان می‌دهد.

جهت جریان آب زیرزمینی در این آبخوان از سمت جنوب، غرب تا شمال غرب می‌باشد که عمق آب در بازه ۲ تا ۷۲ متر متغیر است. آبخوان دشت اردبیل دارای ۲۴۸۳ حلقه چاه بهره‌برداری، ۱۴ دهنه چشمه و ۲۶ رشته قنات است که نیاز آبی منطقه از قبیل کشاورزی، شرب و صنعت را برآورد می‌کند (اداره آب منطقه‌ای اردبیل ۱۳۹۸). آبخوان اردبیل دارای ۵۲ چاه مشاهداتی بود که آمار ۴۸ حلقه چاه که دارای دوره آماری مشترک در مقیاس ماهانه بین سال‌های آبی ۷۸-۱۳۷۷ تا ۹۹-۱۳۹۸ بود مورد استفاده قرار گرفت. شکل ۲ تراز سطح ایستابی آبخوان دشت اردبیل را نشان می‌دهد.



شکل ۱. محدوده آبخوان دشت اردبیل و چاه‌های مشاهداتی منتخب



شکل ۲. نقشه هم‌تراز سطح ایستابی آبخوان دشت اردبیل

داده‌های مورد استفاده

به منظور شبیه‌سازی تراز آب زیرزمینی با استفاده از الگوریتم یادگیری ماشین نیاز به داده‌های مشاهداتی خواهد بود تا به صورت متغیر مستقل در تولید تراز آب زیرزمینی یعنی متغیر وابسته عمل کند. با توجه به اطلاعات دریافتی از شرکت آب منطقه‌ای استان اردبیل، داده‌های ماهانه سطح آب زیرزمینی از مهر ۱۳۷۷ تا شهریور ۱۳۹۹ (۲۲ سال) برای ۴۸ چاه مشاهداتی و آمار برداشت چاه‌های بهره‌بردار^۱ (D) در وسعت منطقه به دست آمد. همچنین داده‌های ماهانه بارندگی^۲ (P)، تبخیر^۳ (E)، دما^۴ (T) در طی دروره آماری جمع‌آوری شد. در فرآیند مدل‌سازی، ۷۰ درصد داده‌ها را به عنوان آموزش (Train) و ۳۰ درصد مابقی به عنوان آزمایش (Test) در نظر گرفته شد لذا با توجه ماهیت دو روش SVM و RF داده‌های آموزش و شبیه‌سازی به صورت تصادفی انتخاب گردید. در ادامه خصوصیات آماری داده‌های ورودی در جدول ۱ ذکر شده است. طبق جدول ۱، هر کدام از متغیرها دارای ۲۶۴ گام زمانی می‌باشد. دامنه نوسان تراز سطح ایستابی بین ۱۳۳۴/۱۹ تا ۱۳۴۶/۱۹ متر می‌باشد و همچنین بیشترین برداشت صورت گرفته از آبخوان ۱۷ میلیون متر مکعب بوده است. نواسانات بارش در آبخوان مورد نظر زیاد بوده به گونه‌ای که متوسط بارندگی ماهانه میلی‌متر در ماه و متوسط میانگین دمای ماهانه ۹/۹ میلی‌متر در ماه می‌باشد.

جدول ۱. خصوصیات آماری متغیرهای ورودی

متغیر	تعداد مشاهدات	میانگین	کمینه	بیشینه	انحراف معیار	ضریب تغییرات
تراز آب زیرزمینی (متر)	۲۶۲	۱۳۳۶/۸۹	۱۳۳۴/۱۹	۱۳۴۶/۱۹	۱/۶۳	۰/۰۰۱
برداشت از چاه بهره‌بردار (میلیون متر مکعب)	۲۶۲	۹/۱۹	۵/۵۷	۱۷/۱۹	۴/۶۱	۰/۵
بارش (میلی‌متر)	۲۶۲	۲۳/۱۵	۰	۱۰۱/۰۶	۱۹/۱۱	۰/۸۲
دما (درجه سلسیوس)	۲۶۲	۹/۸۸	-۱۱/۹۱	۲۶/۴۴	۱۹/۱	۱/۹۳
تبخیر (میلی‌متر)	۲۶۲	۸۷/۸	۰	۲۴۳/۹۶	۷۸/۹۲	۰/۸۹

قبل از اقدام به انجام فرآیند مدل‌سازی، طبق شکل ۳، ضرائب همبستگی متغیرهای مستقل با وابسته مورد بررسی قرار گرفت. مقدار ضریب همبستگی بین ۱ تا -۱ متغیر بوده و که هر چقدر ضریب به عدد یک نزدیک‌تر باشد، نشان‌دهنده رابطه مستقیم بین متغیر مستقل و وابسته می‌باشد. به طوریکه با افزایش متغیر، متغیر دیگر افزایش یافته و برعکس. به گونه‌ای که با کاهش یک متغیر، متغیر دیگر کاهش می‌یابد. همچنین همبستگی منفی نشان‌دهنده معکوس بودن رابطه بین یک متغیر و متغیر دیگر است به طریقی که افزایش یک متغیر موجب کاهش متغیر دیگر می‌شود و برعکس با کاهش اندازه یک متغیر، متغیر دیگر افزایش می‌یابد. همچنین اگر عدد همبستگی ۱ یا -۱ باشد نشان از همبستگی کاملاً مستقیم یا معکوس و اگر صفر باشد نشان از عدم وجود هرگونه رابطه همبستگی بین دو متغیر می‌باشد. طبق شکل ۲، تراز سطح آب زیرزمینی^۵ ماه قبل (GWL-1) نسبت به سایر متغیر، دارای همبستگی با تراز آب زیرزمینی ماه جاری (GWL) دارد؛ لذا انتظار می‌رود الگوهای با ترکیب متغیر فوق، نتایج قابل قبولی در مقایسه با سایر متغیر داشته‌باشد. متغیرهای دما و تبخیر با متغیر برداشت از آب زیرزمینی به ترتیب کمترین ضریب همبستگی را دارند. متغیر تراز آب زیرزمینی ماه قبل و بارش دارای همبستگی مثبت، بارش دو ماه قبل دارای همبستگی خنثی و بقیه پارامترها همبستگی منفی دارند. لذا انتظار می‌رود الگوهای متشکل از دو متغیر تراز آب زیرزمینی و بارش بهترین دقت را بین سایر الگوها داشته‌باشد.

بر اساس مطالعات انجام شده در آبخوان دشت اردبیل بیلان هیدرولوژیکی آبخوان دشت اردبیل محاسبه گردید. بر این اساس، مجموع جریان‌های ورودی و خروجی به آبخوان در سال آبی ۱۳۹۸-۹۹ به ترتیب برابر با ۱۵۳/۲۱ و ۱۶۹/۴۹ میلیون متر مکعب بوده و مقدار کسری مخزن آبخوان ۱۶/۲۸ میلیون متر مکعب می‌باشد. در جدول ۲ اجزا بیلان آبخوان دشت اردبیل ذکر شده است.

جدول ۲. بیلان آبخوان دشت اردبیل در سال آبی ۱۳۹۸-۹۹

تغییرات حجم ذخیره آبخوان (میلیون متر مکعب)	عوامل خروجی (میلیون متر مکعب)					عوامل ورودی (میلیون متر مکعب)					مساحت ناحیه آبخوان (کیلومتر مربع)
	مجموع	تبخیر از آبخوان	زهکش از آبخوان	جریان زیرزمینی خروجی	تخلیه و برداشت (چاه، چشمه و قنات)	مجموع	شرب و صنعتی	جریان‌های سطحی	نفوذ از بارندگی	نفوذ از جریان زیرزمینی	
۱۶/۲۸	۱۶۹/۴۹	۰/۸۲	۲/۸۷	۳/۵	۱۶۲/۳	۱۵۳/۲۱	۵۳/۸۷	۲۵/۳۵	۰/۰۰۹	۷۳/۹۸	۹۶۹

1 Derivation

2 Precipitation

3 Evaporation

4 Temperature

5 Ground Water Level



شکل ۳. ماتریس ضرایب همبستگی بین متغیرهای مستقل و وابسته

به منظور انتخاب بهترین الگو، الگوهای مختلف با توجه به ورودی‌ها طراحی شده و پس از ارزیابی با استفاده از روش‌های ارزیابی، الگو مد نظر انتخاب شد. لذا تعداد ۶۴ (۲^۶) الگو با در نظر گرفتن ورودی‌های دو، سه، چهار و پنج متغیره طراحی شده تا هر الگو بعد از اعمال در الگوریتم‌ها، با استفاده از معیارهای ارزیابی مورد سنجش قرار گیرد. در ادامه به توضیح الگوریتم‌های مورد استفاده در این پژوهش پرداخته خواهد شد.

الگوریتم ماشین بردار پشتیبان

الگوریتم SVM که در سال ۱۹۹۵ توسط واپینک معرفی شد، یکی از روش‌های یادگیری ماشین نظارت‌شده است که به طور گسترده در مسائل طبقه‌بندی و رگرسیون کاربرد دارد (Vapnik, 1998). هدف اصلی این الگوریتم یافتن بهترین مرز جداسازی (Hyperplane) برای تفکیک داده‌های متعلق به کلاس‌های مختلف یا پیش‌بینی داده‌ها با حداقل خطا است (Chauhan et al., 2019). در مسائل رگرسیون، SVM با تعریف یک منطقه حاشیه‌ای به نام لوله اپسیلون (Epsilon Tube)، تلاش می‌کند خطاهای پیش‌بینی خارج از این محدوده را کاهش دهد. حاشیه (Margin) بین مرز و نزدیک‌ترین نقاط هر کلاس به گونه‌ای تعیین می‌شود که بیشینه شود، و این نقاط کلیدی که به مرز نزدیک‌تر هستند، بردارهای پشتیبان نامیده می‌شوند. این طراحی باعث افزایش توانایی تعمیم‌دهی الگوریتم شده و احتمال وقوع بیش‌برازش را کاهش می‌دهد (Chauhan et al., 2019).

در بسیاری از مسائل عملی، داده‌ها اغلب به صورت خطی قابل تفکیک نیستند. برای مقابله با این محدودیت، SVM از توابع کرنل مانند کرنل خطی، چندجمله‌ای یا شعاعی (RBF) استفاده می‌کند. این توابع داده‌ها را به فضای با ابعاد بالاتر نگاشت می‌کنند، جایی که تفکیک خطی داده‌ها آسان‌تر می‌شود. این روش که به ترفند کرنل (Kernel Trick) معروف است، توانایی الگوریتم را برای حل مسائل طبقه‌بندی و رگرسیون غیرخطی به طور قابل توجهی افزایش می‌دهد (Cao et al., 2020).

الگوریتم جنگل تصادفی

الگوریتم RF که در سال ۱۹۹۶ توسط بریمن به عنوان یک روش یادگیری تجمعی نظارت‌شده معرفی شد، به طور ویژه برای حل مسائل رگرسیون و طبقه‌بندی طراحی شده است (Breiman, 2001). این روش با گسترش مفهوم درخت تصمیم، دقت پیش‌بینی را افزایش داده و احتمال بیش‌برازش (Overfitting) را کاهش می‌دهد (Tian et al., 2024; Uddin et al., 2024). RF شامل مجموعه‌ای از درختان

تصمیم هرس نشده است که هر یک با استفاده از الگوریتم تقسیم‌بندی بازگشتی ایجاد می‌شوند (Breiman, 2001; Sahour et al., 2023). در این الگوریتم، هر درخت بر اساس نمونه‌برداری تصادفی از داده‌های آموزشی و ویژگی‌ها ساخته می‌شود. در نهایت، نتایج با استفاده از رأی‌گیری (در مسائل طبقه‌بندی) یا میانگین‌گیری (در مسائل رگرسیونی) ادغام می‌شوند (X. Chen et al., 2011; Friedman & Meulman, 2003). ماهیت تصادفی این الگوریتم باعث شده است که بتواند روابط پیچیده بین متغیرهای مستقل و وابسته را به خوبی مدل‌سازی کند و در مواجهه با داده‌های نویزی عملکردی مطلوب ارائه دهد. علاوه بر این، RF با توانایی در شناسایی روابط غیرخطی میان متغیرها، کیفیت پیش‌بینی‌ها را بهبود می‌بخشد (Breiman, 2001; Tian et al., 2024).

یکی از مزیت‌های اصلی RF، قابلیت آن در مدیریت مجموعه داده‌های بزرگ، پیچیده و ناهمگن است (Catani et al., 2013). این ویژگی به پژوهشگران اجازه می‌دهد تا تحلیل‌های دقیق‌تری روی منابع آب زیرزمینی و متغیرهای مرتبط انجام دهند و اطلاعات ارزشمندی برای تصمیم‌گیری‌های مدیریتی فراهم کنند (K. Chen et al., 2024; Naghibi et al., 2017). RF به دلیل انعطاف‌پذیری بالا می‌تواند انواع مختلف داده‌ها را در تحلیل‌ها لحاظ کند، بدون اینکه کارایی مدل کاهش یابد (Pham et al., 2022). در این مطالعه، متغیرهای هیدرولوژیکی مانند بارش، تبخیر-تعرق و برداشت منابع آب به عنوان ورودی مدل استفاده شدند. با ترکیب پیش‌بینی‌های حاصل از چندین درخت تصمیم، RF توانست تأثیر نویز و تغییرات محلی داده‌ها را کاهش داده و دقت پیش‌بینی را به میزان قابل توجهی افزایش دهد.

مراحل مدل‌سازی

در پژوهش حاضر، تراز آب زیرزمینی به عنوان متغیر وابسته در نظر گرفته شده و داده‌های بارش، تبخیر، دما، برداشت از آب زیرزمینی، تراز آب زیرزمینی ماه قبل و بارش دو ماه قبل به عنوان متغیر مستقل وارد مدل‌سازی شد. به منظور تحلیل داده‌های آماری فرضیاتی مانند نرمالیت، خلا آماری و وجود داده‌های پرت مورد سنجش قرار گرفت. پس از شناسایی خلا آماری در هر سری داده‌ها، با استفاده از روش‌ها حداقل مربعات و رگرسیون خطی اقدام به تکمیل داده‌ها شد. به منظور بررسی نرمالیت داده از آزمون‌های کلموگروف-اسمیرنوف و ران تست در زبان برنامه‌نویسی python استفاده شد. سپس با توجه به نتایج حاصله جهت بررسی رفتار متغیرهای مستقل در مواجهه با متغیر وابسته، اقدام به بررسی ضرایب همبستگی پیرسون داده‌های آماری شد. با در دست داشتن نتایج همبستگی متغیرها (شکل ۲) ترکیب‌های متخف متغیرهای مستقل جهت ورد به مدل شناسایی می‌شود. سپس برای توسعه مدل با استفاده از الگوریتم‌های SVM و RF، داده‌ها در قالب اکسل گردهم آوری شده و سپس برای مدل‌سازی با زبان برنامه‌نویسی python آماده‌سازی شد. ۷۰ درصد داده‌ها برای آموزش مدل و ۳۰ درصد آن برای آزمایش انتخاب شد. به منظور تقسیم داده‌ها از تابع train_test_split موجود در کتابخانه Scikit-learn استفاده شد. در ابتدا با استفاده از کتابخانه sklearn کرنل‌های مختلف الگوریتم SVM اندازه‌گیری شد. به منظور جلوگیری از بروز هم‌خطی، شاخص تورم واریانس^۱ VIF مورد ارزیابی قرار می‌گیرد، به گونه‌ای که اگر شاخص مذکور از ۸ کمتر شد این عامل خود نمایانگر عدم هم‌خطی می‌باشد (Pawlins et al., 1998). در ادامه الگوریتم RF با در نظر گرفتن تمامی متغیرها مدل‌سازی شده و نتایج مورد ارزیابی قرار گرفت. سپس با استفاده از روش اهمیت ویژگی جایگشت، متغیر با بیشترین ویژگی شناسایی شده و مدل‌سازی دوباره انجام شد. شکل ۴ بیان‌کننده مراحل مدل‌سازی می‌باشد.

ارزیابی مدل

در این پژوهش به دلیل تحلیل عملکرد مدل‌های مختلف، از شاخص‌های آماری مختلف استفاده شده است. شاخص‌های آماری ضریب همبستگی R^2 ، نسبت انحراف استاندارد RMSE، میانگین خطای مطلق MAE و خطای میانگین مربعات MSE برای بررسی تناسب میان داده‌های مشاهداتی و پیش‌بینی شده مورد استفاده قرار گرفته شده است. در ادامه در روابط (۱)، (۲)، (۳) و (۴) به ترتیب بیانگر روابط شاخص‌های آماری می‌باشد.

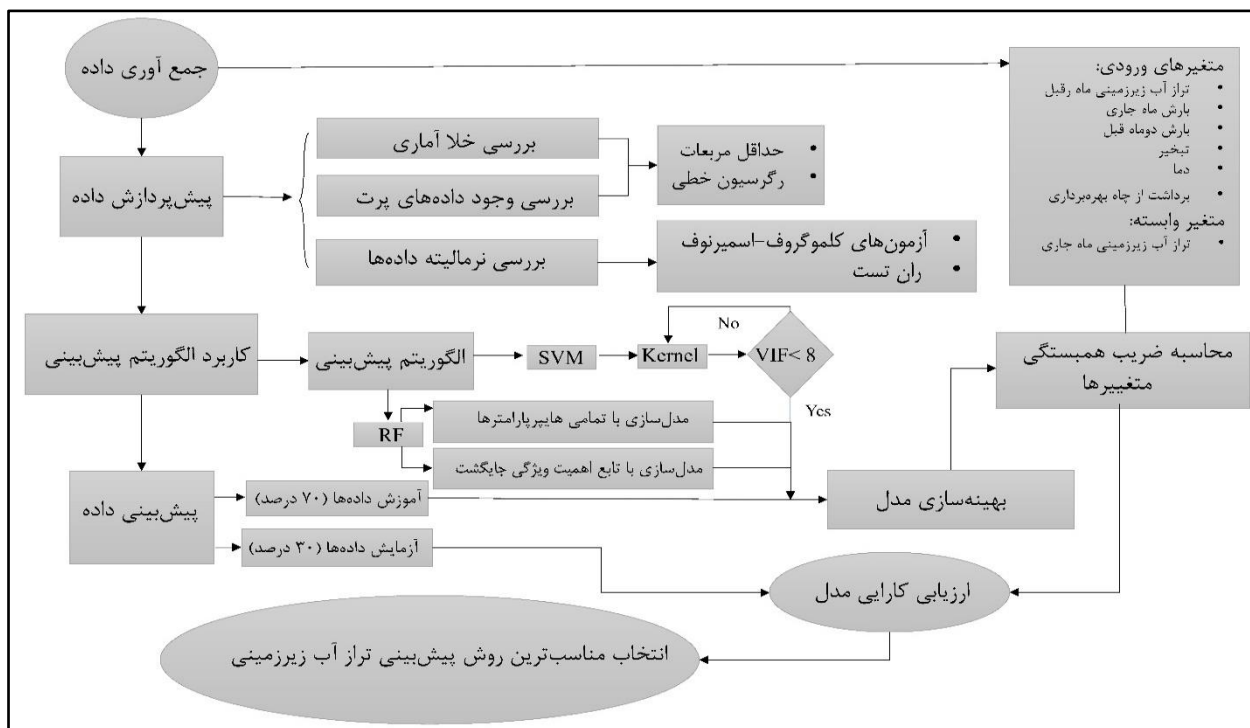
$$R^2 = 1 - \frac{\sum_{i=1}^n (S_i - P_i)^2}{\sum_{i=1}^n (S_i - \bar{P})^2} \quad \text{رابطه ۱}$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^n (S_i - P_i)^2} \quad \text{رابطه ۲}$$

$$MAE = \frac{1}{N} \sum_{i=1}^n |S_i - P_i| \quad \text{رابطه ۳}$$

$$MSE = \frac{1}{N} \sum_{i=1}^n (S_i - P_i)^2 \quad \text{رابطه ۴}$$

در روابط ۲-۴ مقادیر S_i ، P_i و $\bar{P}$ به ترتیب مقادیر مشاهداتی، پیش‌بینی شده و مقدار متوسط مقادیر پیش‌بینی شده می‌باشد.



شکل ۴. مراحل مدل‌سازی و پیش‌بینی تراز آب زیرزمینی با استفاده از دو الگوریتم یادگیری ماشین

نتایج

در این پژوهش جهت پیش‌بینی تراز آب زیرزمینی از دو الگوریتم SVM و RF با استفاده از برنامه پایتون و کتابخانه‌های sklearn، pandas و numpy استفاده گردیده‌است. به همین منظور داده‌های ۴۸ چاه مشاهداتی آبخوان دشت اردبیل در بین سال‌های ۱۳۷۷ تا ۱۳۹۹ به صورت ماهانه استخراج و مورد مدل‌سازی قرار گرفته است. در ادامه به بررسی نتایج هر یک از مدل‌های مذکور پرداخته می‌شود.

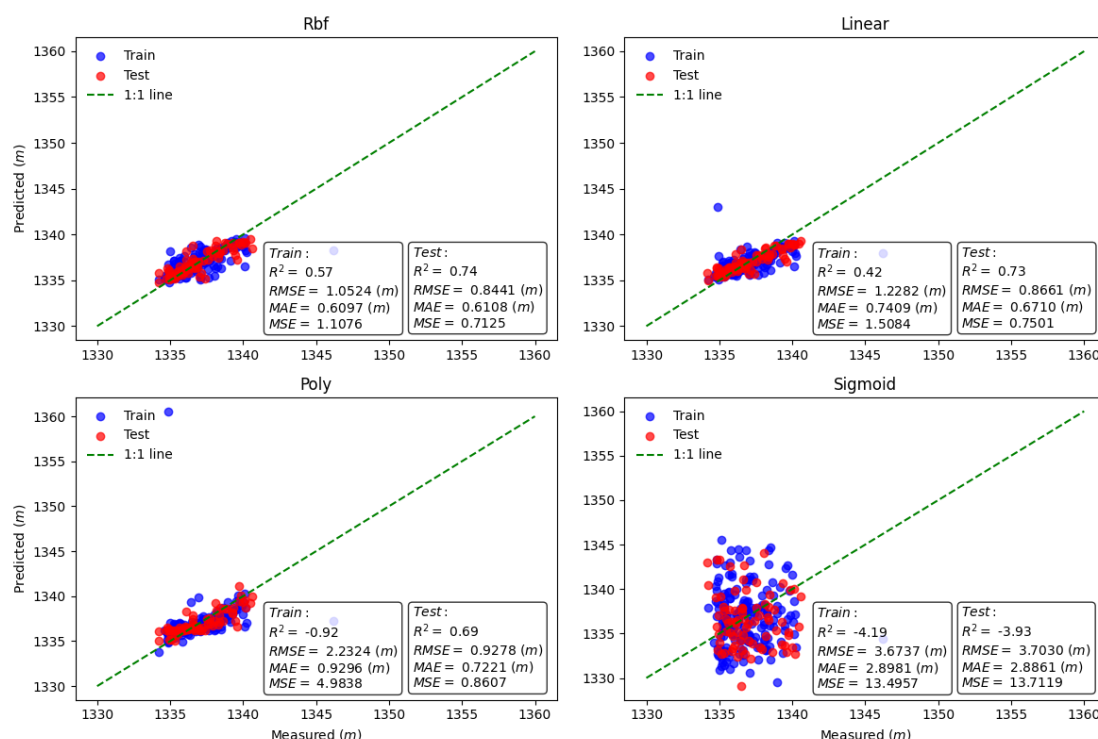
الگوریتم ماشین بردار پشتیبان

در این پژوهش، SVM با محاسبه کرنل‌های مختلف و VIF هر کدام از کرنل‌ها، کرنل RBF با VIF کمتر از ۸ و الگوی شامل متغیرهای D، P، (-2) و (-1) به عنوان کرنل با الگوی بهینه مورد انتخاب قرار داد. طبق جدول ۳، مقادیر R^2 ، RMSE، MAE و MSE الگوی منتخب در مرحله Train به ترتیب ۰/۵۷، ۱/۰۵، ۰/۶۱ و ۱/۱۱ بوده و همچنین در مرحله Test نیز به ترتیب ۰/۷۴، ۰/۸۴، ۰/۶۱ و ۰/۷۱ می‌باشد. متغیرهای مستقل ورودی در الگوی منتخب همانند حالت حقیقی آبخوان بوده زیرا که مقادیر تغذیه و تخلیه نقش به‌سزایی در تراز آب زیرزمینی داشته و در الگو مذکور مقادیر بارش دو ماه قبل و تراز آب زیرزمینی ماه قبل به عنوان عامل تغذیه‌کننده و برداشت از چاه‌های بهره‌برداری به عنوان عامل تخلیه‌کننده عمل می‌کند. عامل تبخیر و دما تاثیر مستقیم در تراز آب زیرزمینی نداشته و به صورت غیر مستقیم بر مقادیر بارش و به طور کلی تغذیه تاثیر گذاشته که این دو عامل در الگو منتخب به عنوان ورودی نتایج مطلوبی را حاصل نداشته است.

در شکل ۵ مقادیر مشاهداتی و محاسباتی هر کدام از کرنل‌های مختلف الگوریتم SVM در مرحله Train و Test ذکر شده‌است. بر اساس شکل ۵، در مرحله Train و Test مدل Sigmoid به عنوان بدترین کرنل و RBF به عنوان بهترین کرنل شناخته می‌شود زیرا که در مدل Sigmoid مقادیر محاسباتی و مشاهداتی برازش نامتناسبی بر خط ۴۵ درجه داشته ولی مدل RBF برازش مناسبی داشته که این عمل توانایی بالای الگوریتم SVM در پیش‌بینی تراز آب زیرزمینی را بیان می‌کند.

جدول ۳. نتایج ارزیابی کرنل‌های مختلف الگوریتم SVM با الگوهای منتخب

VIF	Feature	Kernel
۴/۰۳	E	Linear
۳/۴۳	D	
۱/۶۲	GWL (-1)	
۱/۹۷	P (-2)	Rbf
۱/۵۳	D	
۱/۹	GWL (-1)	
۱۵/۹۱	T	Sigmoid
۲/۴۱	P	
۲/۵۷	P (-2)	
۸/۴۳	E	Poly
۱/۶۱	E	
۱/۶۱	GWL (-1)	

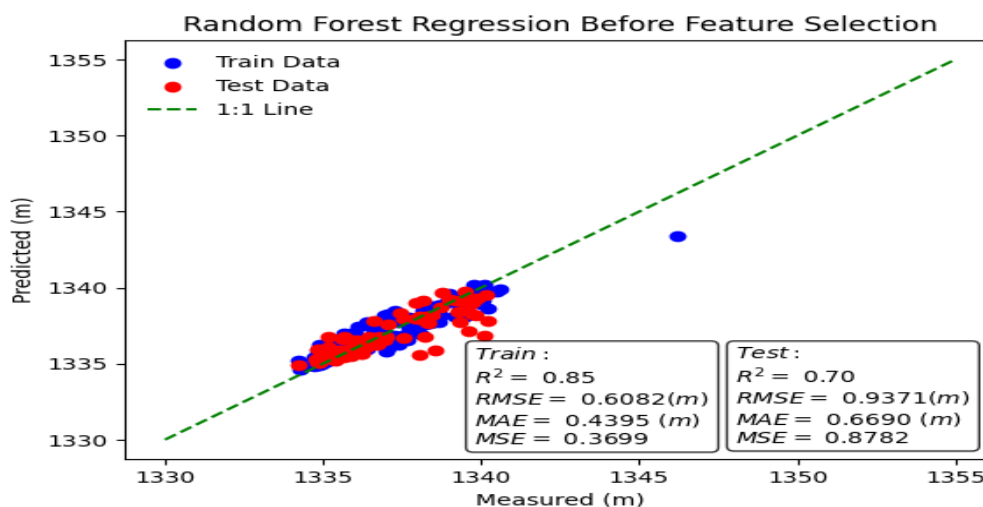


شکل ۵. مقادیر مشاهداتی و پیش‌بینی تراز آب زیرزمینی آبخوان اردبیل در کرنل مختلف الگوریتم SVM در دو مرحله آموزش و آزمایش

الگوریتم جنگل تصادفی

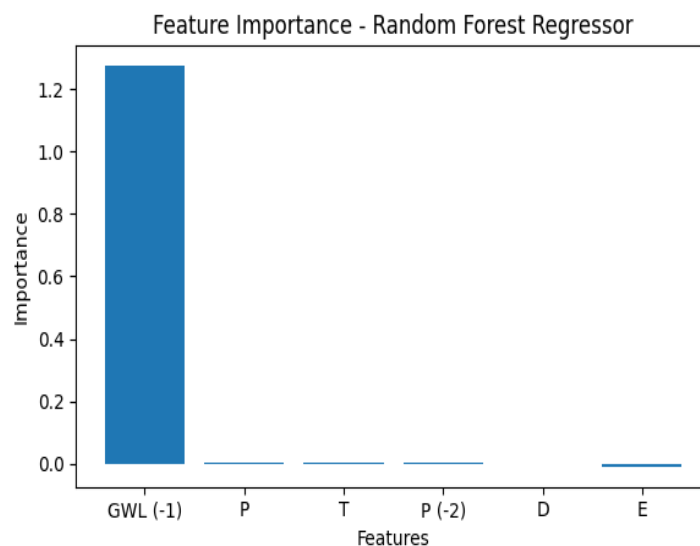
به منظور ایجاد الگوریتم به وسیله درختان تصمیم، در ابتدا به تولید متغیرهای ورودی پرداخته شد. لازم به ذکر است درخت‌های تصمیم باید به گونه‌ای تنظیم شود که با مقادیر خطا همگرا بوده تا اعتمادپذیری پیش‌بینی حاصل شود. در الگوریتم حاضر، درختان تصمیم بین ۱ تا ۲۰۰ درخته انتخاب شده و پارامترهای الگوریتم به وسیله تعداد متغیرهای بین یک تا ماکزیمم متغیر بهینه شد. همچنین به منظور کاهش ابعاد، افزایش دقت و قابلیت تفسیر الگوریتم از روش شاخص ویژگی جایگشت^۱ (PFI) استفاده شد. با توجه به این موضوع که RF بر خلاف SVM نسبت به هم‌خطی بودن مقاوم است، لذا در این الگوریتم شاخص تورم واریانس VIF مورد بررسی قرار نگرفت که این گواهی بر قدرت RF برای آموزش صحیح‌تر با داده‌های دارای نویز می‌باشد (Marcoulides & Raykov, 2019). در ادامه با در نظر گرفتن کلیه متغیرها به عنوان ورودی طبق شکل ۶، نتایج شاخص‌های عملکرد R^2 ، RMSE، MAE و MSE در بخش Train به ترتیب ۰/۸۵، ۰/۶۱،

۰/۴۴ و ۰/۳۷ شد. همچنین نتایج شاخص‌های عملکردی مدل در بخش Test به ترتیب ۰/۷، ۰/۹۴، ۰/۶۷ و ۰/۸۸ بوده که نشان از قدرت مناسب RF در بحث پیش‌بینی تراز آب زیرزمینی دارد.



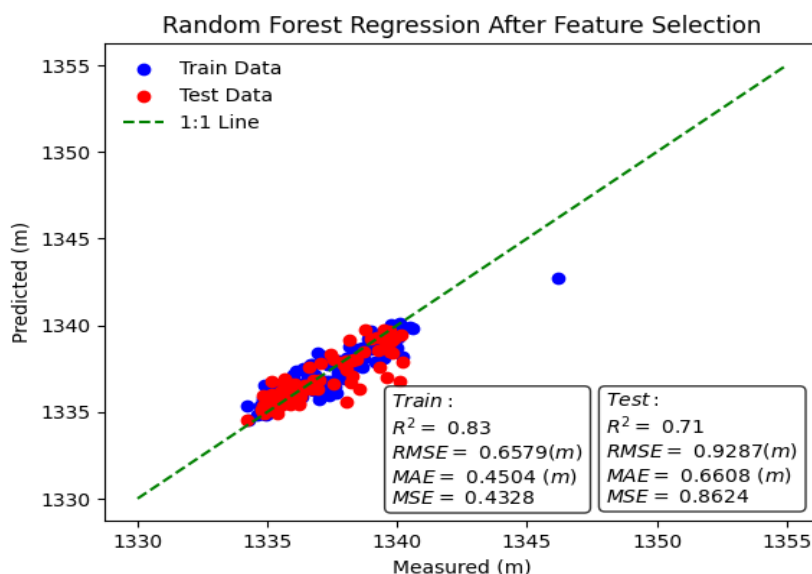
شکل ۶. مقادیر مشاهداتی و پیش‌بینی تراز آب زیرزمینی آبخوان اردبیل در الگوریتم RF قبل از انتخاب ویژگی‌ها در دو مرحله آموزش و آزمایش با در نظر گرفتن تمامی متغیرها به عنوان ورودی

بعد از استفاده از همه داده‌های به عنوان ورودی، برای بررسی رفتار و دقت الگوریتم RF در هنگام تخمین سطح ایستابی، با استفاده از PFI اقدام به انتخاب ورودی‌های مختلف شد. طبق نتایج PFI که در شکل ۷ ارائه شده است؛ متغیر (-1) GWL بیشترین تاثیر و متغیر D کمترین تاثیر را در RF را دارد.



شکل ۷. خروجی روش PFI به منظور شناسایی متغیرهای تأثیرگذار بر مدل و کاهش تعداد آن‌ها و مدل‌سازی مجدد با استفاده از الگوریتم RF

همانند روش قبل اقدام به ساخت الگوریتم RF با درخت تصمیم ۱ تا ۲۰۰ می‌شود. طبق نتایج حاصل از الگوریتم که در شکل ۸ نمایان است، شاخص‌های عملکرد R^2 ، RMSE، MAE و MSE در بخش Train به ترتیب ۰/۶۶، ۰/۸۳، ۰/۴۵ و ۰/۴۳ شد. همچنین نتایج شاخص‌های عملکردی در بخش Test به ترتیب ۰/۷۱، ۰/۹۳، ۰/۶۶ و ۰/۸۶ می‌باشد. با مقایسه معیارهای ارزیابی در شکل ۶ و ۸ که نشان‌دهنده پیش‌بینی RF با ۵ متغیر ورودی و دو متغیر ورودی می‌باشد می‌توان نتیجه گرفت تغییر تعداد متغیر ورودی، کمترین تاثیر بر روی نتایج RF را دارد.



شکل ۸. مقادیر مشاهداتی و پیش‌بینی تراز آب زیرزمینی آبخوان اردبیل در الگوریتم RF بعد از انتخاب ویژگی‌ها در دو مرحله آموزش و آزمایش بر اساس روش PFI

نتایج حاصل از دو الگوریتم SVM و RF نشان از توانایی قابل قبول هر دو الگوریتم در پیش‌بینی تراز آب زیرزمینی دارد ولی SVM حساسیت بالایی در مواجهه با هم‌خطی از خود نشان می‌دهد که این عمل حساسیت مدل‌سازی را افزایش می‌دهد. همچنین نتایج نشان داد که با در نظر که سه متغیر می‌توان الگو منتخب را انتخاب کرد در حالیکه الگوریتم RF، مقاومت بالایی را در مقابل هم‌خطی از خود نشان داده و همچنین طبق نتایج، حساسیت زیادی در مقابل افزایش ورودی‌ها نداشته و فقط با استفاده از نمودار PFI می‌توان به الگو منتخب دست پیدا کرد. با توجه به نتایج حاصله باید دقت داشت که تراز آب زیرزمینی ماه قبل، مهم‌ترین ورودی در مدل‌سازی بود که تاثیر زیادی بر خروجی ایفا می‌کند و فقدان این متغیر مهم سبب کاهش دقت نتایج مدل می‌شود. طبق نتایج، RF با توجه به شاخص‌های عملکردی R^2 ، RMSE، MAE و MSE در بخش Test به ترتیب ۰/۷، ۰/۹۴، ۰/۶۷ و ۰/۸۸ و نسبت به SVM با توجه به نتایج شاخص‌های عملکردی R^2 ، RMSE، MAE و MSE در بخش Test به ترتیب ۰/۷۴، ۰/۸۴، ۰/۶۱ و ۰/۷۱، عملکرد بهتری را به ثبت رسانده‌است. بر این اساس RF به عنوان بهترین الگوریتم جهت پیش‌بینی تراز آب زیرزمینی با استفاده از دو متغیر بارش و تراز آب زیرزمینی ماه قبل تشخیص داده‌شد.

بحث

پژوهش حاضر به بررسی و مقایسه عملکرد دو مدل یادگیری ماشین، یعنی جنگل تصادفی (RF) و ماشین بردار پشتیبان (SVM)، در پیش‌بینی تراز آب زیرزمینی در مناطق خشک و نیمه‌خشک پرداخته است. نتایج این مطالعه نشان داد که مدل RF به دلیل مقاومت بالا در برابر هم‌خطی متغیرها، انعطاف‌پذیری در کاهش تعداد ورودی‌ها و دقت قابل توجه در پیش‌بینی داده‌ها، عملکرد بهتری نسبت به مدل SVM دارد. به طور خاص، شاخص R^2 در مرحله تست مدل RF به مقدار ۰/۷۰ و RMSE به مقدار ۰/۹۴ رسید که نشان‌دهنده تطابق قابل قبول با داده‌های واقعی بود. در مقابل، مدل SVM علی‌رغم حساسیت بالاتر به هم‌خطی، با استفاده از کرنل RBF به شاخص R^2 معادل ۰/۷۴ در مرحله تست دست یافت.

این نتایج با مطالعات مشابه نیز مقایسه شد. به عنوان مثال، مطالعه Yadav و همکاران (۲۰۲۰) با استفاده از مدل‌های ترکیبی (Hybrid) و پیش‌پردازش داده، نشان داد که دقت مدل‌ها به انتخاب متغیرهای ورودی وابسته است (Yadav et al., 2020). همچنین مطالعه Liu و همکاران (۲۰۲۲) در حوضه رودخانه تاریم، عملکرد دقیق مدل RF را در پیش‌بینی‌های مکانی-زمانی به ویژه در مناطق داده‌پراکنده تایید کرد (Liu et al., 2022). علاوه بر این، مطالعه Salehi و همکاران (۲۰۲۳) که در مناطق حساس به آتش‌سوزی در اندونزی انجام شد، برتری RF و XGBoost را در پیش‌بینی تغییرات تراز آب زیرزمینی نشان داد، اما مشابه پژوهش حاضر، تأکید بر نقش پارامترهایی مانند بارش و فاصله از کانال‌های زهکشی داشت (Hikouei et al., 2023). در تحلیل متغیرهای ورودی، پژوهش حاضر نشان

داد که متغیر تراز آب زیرزمینی ماه قبل (GWL-1) و بارش دو ماه قبل (P-2) بیشترین تأثیر را در دقت پیش‌بینی دارند، در حالی که متغیرهای دما و تبخیر تأثیر غیرمستقیم و کمتری داشتند. این نتایج نشان‌دهنده اهمیت داده‌های تاریخی در بهبود عملکرد مدل‌ها است. کاربردهای مدیریتی این پژوهش شامل بهبود مدیریت پایدار منابع آب و استفاده از یافته‌ها برای تدوین استراتژی‌های عملیاتی است. همچنین پیشنهاد می‌شود در مطالعات آینده از داده‌های مکانی (GIS) و مدل‌های پیشرفته‌تر مانند LSTM استفاده شود تا دقت پیش‌بینی افزایش یابد. با این حال، پژوهش حاضر محدودیت‌هایی مانند عدم لحاظ تغییرات کاربری اراضی و ویژگی‌های توپوگرافی را داراست. ارزیابی مدل‌ها در شرایط هیدرولوژیکی متفاوت و افزودن داده‌های متنوع‌تر می‌تواند به بهبود نتایج کمک کند. به طور کلی، این مطالعه نشان داد که مدل RF به دلیل توانایی در مدیریت داده‌های متنوع و دقت بالا، ابزاری مناسب برای پیش‌بینی تراز آب زیرزمینی است و می‌تواند در مدیریت پایدار منابع آب در مناطق خشک و نیمه‌خشک به کار رود.

نتیجه‌گیری

در این پژوهش، با هدف پیش‌بینی دقیق تراز آب زیرزمینی در دشت اردبیل، از دو الگوریتم یادگیری ماشین پرکاربرد، یعنی SVM و RF استفاده شد. نتایج نشان داد که الگوریتم RF در دو مرحله Train و Test با MAE، RMSE و MSE کمتر از SVM، با دقت بهتری قادر به پیش‌بینی تغییرات تراز آب زیرزمینی نسبت به الگوریتم SVM بوده است. این امر به دلیل توانایی بالای RF در مدیریت داده‌های ناهمگن و پیچیده و همچنین مقاومت آن در برابر هم‌خطی بین متغیرها می‌باشد. همچنین الگوریتم RF با کاهش ورودی‌های مدل از ۵ متغیر به یک متغیر، باعث کاهش پیچیدگی محاسبات، افزایش دقت بهینه‌سازی مدل و پیش‌بینی شد. یکی از نوآوری‌های اصلی این پژوهش، مقایسه دقیق عملکرد دو الگوریتم SVM و RF در شرایط خاص دشت اردبیل بوده است. نتایج حاصل از این مقایسه می‌تواند به عنوان یک راهنمای عملی برای انتخاب الگوریتم مناسب در پیش‌بینی تراز آب زیرزمینی عمل کرده و جایگزین مناسبی برای مدل‌های فیزیکی و ریاضی در سایر مطالعات مشابه باشد. همچنین، شناسایی متغیرهای مؤثر بر تراز آب زیرزمینی و اهمیت نسبی هر یک از این متغیرها، می‌تواند در تدوین استراتژی‌های بهتر برای مدیریت پایدار منابع آب زیرزمینی کمک کند.

با این حال، این پژوهش دارای محدودیت‌هایی نیز می‌باشد. به عنوان مثال، تنها از داده‌های هیدرولوژیکی استفاده شده است و تأثیر عوامل دیگر مانند تغییرات کاربری اراضی، توپوگرافی و ویژگی‌های خاک بر تراز آب زیرزمینی در این مطالعه بررسی نشده است. بنابراین، پیشنهاد می‌شود در مطالعات آینده، از داده‌های مکانی و مدل‌های رقومی ارتفاع زمین استفاده شود تا بتوان به نتایج دقیق‌تری دست یافت. در نهایت، نتایج این پژوهش نشان می‌دهد که استفاده از الگوریتم‌های یادگیری ماشین، به ویژه RF می‌تواند ابزار قدرتمندی برای پیش‌بینی تراز آب زیرزمینی و مدیریت پایدار منابع آب در مناطق خشک و نیمه‌خشک باشد. با این حال، برای بهبود دقت پیش‌بینی‌ها و گسترش کاربرد این روش‌ها، نیاز به انجام مطالعات بیشتری با استفاده از داده‌های متنوع‌تر و مدل‌های پیچیده‌تر می‌باشد.

"هیچ‌گونه تعارض منافع بین نویسندگان وجود ندارد"

منابع

- اکبری نیازی، موسی؛ وردی‌نژاد، وحیدرضا؛ بهمنش، جواد و نیک پور، محمدرضا (۱۴۰۲). شبیه‌سازی کیفیت آب رودخانه قره‌سو با استفاده از مدل Kw2QUAL. *هیدروژئولوژی*، ۱۸(۱)، ۱۳۷-۱۵۵.
- جعفرزاده، جعفر؛ رستم زاده، هاشم و اسدی، اسماعیل (۱۳۹۶). مدل‌سازی زمانی تراز آب زیرزمینی با استفاده از روش‌های پایه تحلیل سری‌های زمانی (مطالعه موردی: دشت اردبیل). *دانش آب و خاک*، ۲۷(۴)، ۱۸۵-۱۹۶.
- دانشور وثوقی، فرناز؛ دین‌پژوه، یعقوب و اعلمی، محمدتقی (۱۳۹۰). تأثیر خشکسالی بر تراز آب زیر زمینی در دو دهه اخیر (مطالعه موردی: دشت اردبیل). *دانش آب و خاک*، ۲۱(۴)، ۱۶۵-۱۷۹.
- فضابخش، شاهین؛ رسول‌زاده، علی؛ رضانی‌مقدم، جواد و اسماعیلیان، مهدی (۱۴۰۱). بررسی کیفیت آب زیرزمینی با استفاده از روش‌های آماری چند متغیره در آبخوان دشت اردبیل. *نشریه علمی پژوهشی مهندسی آبیاری و آب ایران*، ۱۳(۲)، ۳۵۳-۳۳۳.
- قوردویی میلان، سامی؛ آریاآذر، ناصر؛ جواد، سامان و رازدار، بابک (۱۳۹۹). شبیه‌سازی تراز سطح آب زیرزمینی با استفاده از مدل حداقل مربعات ماشین بردار پشتیبان و مقایسه آن با شبکه عصبی مصنوعی و رگرسیون خطی چند متغیره. *هیدروژئولوژی*، ۱۸(۱)، ۱۱۸-۱۳۳.
- نوروزی قوشبلاغ، حسین و ندیری، عطاالله (۱۳۹۷). پیش‌بینی سطح آب زیرزمینی دشت بوکان با استفاده از مدل‌های منطق فازی، جنگل تصادفی و

شبکه عصبی. نشریه علمی - پژوهشی مرتع و آبخیزداری، ۷۱ (۳)، ۸۲۹-۸۴۵.

REFERENCE

- Aghazadeh, N., Chitsazan, M., & Golestan, Y. (2017). Hydrochemistry and quality assessment of groundwater in the Ardabil area, Iran. *Applied Water Science*, 7(7), 3599–3616. <https://doi.org/10.1007/s13201-016-0498-9>
- Akbari Niari, M., Rezaverdinejad, V., Behmanesh, J., & Nikpour, R. (2023). Simulating the Water Quality of Qarasu River Using the QUAL2Kw Model. *Hydrogeology*, 8(1), 137–155. <https://doi.org/10.22034/hydro.2023.57504.1299> [In Persian]
- Breiman, L. (2001). Random forests. *Machine Learning*, 45, 5–32.
- Cao, Y., Yin, K., Zhou, C., & Ahmed, B. (2020). Establishment of landslide groundwater level prediction model based on GA-SVM and influencing factor analysis. *Sensors (Switzerland)*, 20(3). <https://doi.org/10.3390/s20030845>
- Catani, F., Lagomarsino, D., Segoni, S., & Tofani, V. (2013). Landslide susceptibility estimation by random forests technique: sensitivity and scaling issues. *Natural Hazards and Earth System Sciences*, 13(11), 2815–2831.
- Chauhan, V. K., Dahiya, K., & Sharma, A. (2019). Problem formulations and solvers in linear SVM: a review. In *Artificial Intelligence Review* (Vol. 52, Issue 2, pp. 803–855). Springer Netherlands. <https://doi.org/10.1007/s10462-018-9614-6>
- Chen, K., Liu, Q., Yang, T., Ju, Q., & Zhu, M. (2024). Risk assessment of nitrate groundwater contamination using GIS-based machine learning methods: A case study in the northern Anhui plain, China. *Journal of Contaminant Hydrology*, 261, 104300. <https://doi.org/https://doi.org/10.1016/j.jconhyd.2024.104300>
- Chen, X., Wang, M., & Zhang, H. (2011). The use of classification trees for bioinformatics. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 1(1), 55–63.
- Chen, Y., & Zhang, D. (2020). Physics-constrained deep learning of geomechanical logs. *IEEE Transactions on Geoscience and Remote Sensing*, 58(8), 5932–5943.
- Daneshvar Vousoughi, F., Dinpashoh, Y., & Aalami, M. T. (2011). Effect of Drought on Groundwater Level in the Past Two Decades (Case study: Ardebil Plain). *Water and Soil Science*, 21(4), 165–179. https://water-soil.tabrizu.ac.ir/article_1195.html [In Persian]
- Fazabakhsh, S., Rasoulzadeh, Ali., Ramezani moghadam, Javad., & Esmailian, M. (2022). Investigation of Groundwater Quality Using Multivariate Statistical Methods in Ardabil Plain Aquifer. *Irrigation and Water Engineering*, 13(2), 333–353. <https://doi.org/10.22125/iwe.2022.162670> [In Persian]
- Feng, S., Huo, Z., Kang, S., Tang, Z., & Wang, F. (2011). Groundwater simulation using a numerical model under different water resources management scenarios in an arid region of China. *Environmental Earth Sciences*, 62(5), 961–971. <https://doi.org/10.1007/s12665-010-0581-8>
- Foody, G. M. (1995). Land cover classification by an artificial neural network with ancillary information. *International Journal of Geographical Information Systems*, 9(5), 527–542. <https://doi.org/10.1080/02693799508902054>
- Friedl, M. A., & Brodley, C. E. (1997). Decision tree classification of land cover from remotely sensed data. *Remote Sensing of Environment*, 61(3), 399–409. [https://doi.org/https://doi.org/10.1016/S0034-4257\(97\)00049-7](https://doi.org/https://doi.org/10.1016/S0034-4257(97)00049-7)
- Friedman, J. H., & Meulman, J. J. (2003). Multiple additive regression trees with application in epidemiology. *Statistics in Medicine*, 22(9), 1365–1381.
- Ghourdoyee Milan, S., Aryaazar, N., Javadi, S., & Razdar, B. (2020). Simulation of groundwater head using LS-SVM and comparison with ANN & MLR. *Hydrogeology*, 5(1), 118–133. <https://doi.org/10.22034/hydro.2020.10455> [In Persian]
- Gleeson, T., Alley, W. M., Allen, D. M., Sophocleous, M. A., Zhou, Y., Taniguchi, M., & Vandersteen, J. (2012). Towards sustainable groundwater use: Setting long-term goals, backcasting, and managing adaptively. *Ground Water*, 50(1), 19–26. <https://doi.org/10.1111/j.1745-6584.2011.00825.x>
- 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. <https://doi.org/10.1016/j.pce.2024.103603>
- Hikouei, I. S., Eshleman, K. N., Saharjo, B. H., Graham, L. L. B., Applegate, G., & Cochrane, M. A. (2023). Using machine learning algorithms to predict groundwater levels in Indonesian tropical peatlands. *Science of the Total Environment*, 857. <https://doi.org/10.1016/j.scitotenv.2022.159701>

- Hu, L., & Jiao, J. J. (2015). Calibration of a large-scale groundwater flow model using GRACE data: a case study in the Qaidam Basin, China. *Hydrogeology Journal*, 23(7), 1305–1317. <https://doi.org/10.1007/s10040-015-1278-6>
- Jafarzadeh, J., Rostamzadeh, H., & Asadi, E. (2017). Modeling temporal of groundwater level using basic techniques of time series analysis (Case Study: Ardabil Plain). *Water and Soil Science*, 27(4), 185–196. https://water-soil.tabrizu.ac.ir/article_6999.html [In Persian]
- Karimi Alavijeh, N., Falahi, M. A., Ahmadi Shadmehri, M. T., Salehnia, N., Larsen, M. A. D., & Drews, M. (2021). Perspectives of current and future urban water security in Iran. *Journal of Cleaner Production*, 321, 129004. <https://doi.org/10.1016/j.jclepro.2021.129004>
- Kløve, B., Ala-Aho, P., Bertrand, G., Gurdak, J. J., Kupfersberger, H., Kværner, J., Muotka, T., Mykrä, H., Preda, E., Rossi, P., Uvo, C. B., Velasco, E., & Pulido-Velazquez, M. (2014). Climate change impacts on groundwater and dependent ecosystems. *Journal of Hydrology*, 518, 250–266. <https://doi.org/10.1016/j.jhydrol.2013.06.037>
- Kumar, C. P. (2012). Climate Change and Its Impact on Groundwater Resources. In *RESEARCH INVENTY: International Journal of Engineering and Science* (Vol. 1, Issue 5). www.researchinventy.com
- Lancia, M., Yao, Y., Andrews, C. B., Wang, X., Kuang, X., Ni, J., Gorelick, S. M., Scanlon, B. R., Wang, Y., & Zheng, C. (2022). The China groundwater crisis: A mechanistic analysis with implications for global sustainability. *Sustainable Horizons*, 4, 100042. <https://doi.org/10.1016/j.horiz.2022.100042>
- Liu, Q., Gui, D., Zhang, L., Niu, J., Dai, H., Wei, G., & Hu, B. X. (2022). Simulation of regional groundwater levels in arid regions using interpretable machine learning models. *Science of the Total Environment*, 831. <https://doi.org/10.1016/j.scitotenv.2022.154902>
- Marcoulides, K. M., & Raykov, T. (2019). Evaluation of variance inflation factors in regression models using latent variable modeling methods. *Educational and Psychological Measurement*, 79(5), 874–882.
- Minnig, M., Moeck, C., Radny, D., & Schirmer, M. (2018). Impact of urbanization on groundwater recharge rates in Dübendorf, Switzerland. *Journal of Hydrology*, 563, 1135–1146.
- Naghibi, S. A., Ahmadi, K., & Daneshi, A. (2017). Application of support vector machine, random forest, and genetic algorithm optimized random forest models in groundwater potential mapping. *Water Resources Management*, 31, 2761–2775.
- Norouzi, hossein, & nadiri, ataallah. (2018). Groundwater Level Prediction of Boukan Plain using Fuzzy Logic, Random Forest and Neural Network Models. *Journal of Range and Watershed Managment*, 71(3), 829–845. <https://doi.org/10.22059/jrwm.2018.68924> [In Persian]
- Perez, N., Singh, V., Ringler, C., Xie, H., Zhu, T., Sutanudjaja, E. H., & Villholth, K. G. (2024). Ending groundwater overdraft without affecting food security. *Nature Sustainability*, 7(8), 1007–1017. <https://doi.org/10.1038/s41893-024-01376-w>
- Pham, Q. B., Tran, D. A., Ha, N. T., Islam, A. R. M. T., & Salam, R. (2022). Random forest and nature-inspired algorithms for mapping groundwater nitrate concentration in a coastal multi-layer aquifer system. *Journal of Cleaner Production*, 343, 130900.
- Poursaeid, M., Mastouri, R., Shabanlou, S., & Najarchi, M. (2020). Estimation of total dissolved solids, electrical conductivity, salinity and groundwater levels using novel learning machines. *Environmental Earth Sciences*, 79, 1–25.
- Rajaei, T., Ebrahimi, H., & Nourani, V. (2019). A review of the artificial intelligence methods in groundwater level modeling. *Journal of Hydrology*, 572, 336–351. <https://doi.org/10.1016/j.jhydrol.2018.12.037>
- Rasouli, F., & Cheraghi, A. (2009). Evaluation of the quality of groundwater used in irrigation in Fars province. 1 St Conference of Groundwater, Behbahan.
- Rathnayaka, K., Malano, H., & Arora, M. (2016). Assessment of sustainability of urban water supply and demand management options: a comprehensive approach. *Water*, 8(12), 595.
- Rawlings, J. O., Pantula, S. G., & Dickey, D. A. (Eds.). (1998). *Applied regression analysis: a research tool*. New York, NY: Springer New York.
- Sahour, S., Khanbeyki, M., Gholami, V., Sahour, H., Kahvazade, I., & Karimi, H. (2023). Evaluation of machine learning algorithms for groundwater quality modeling. *Environmental Science and Pollution Research*, 30(16), 46004–46021.
- Sun, J., Hu, L., Li, D., Sun, K., & Yang, Z. (2022). Data-driven models for accurate groundwater level prediction and their practical significance in groundwater management. *Journal of Hydrology*, 608, 127630. <https://doi.org/10.1016/j.jhydrol.2022.127630>
- Tian, Y., Liu, Q., Ji, Y., Dang, Q., Sun, Y., He, X., Liu, Y., & Su, J. (2024). Prediction of sulfate concentrations in groundwater in areas with complex hydrogeological conditions based on machine learning. *Science of*

- The Total Environment*, 923, 171312. <https://doi.org/https://doi.org/10.1016/j.scitotenv.2024.171312>
- Uddin, M. G., Nash, S., Rahman, A., Dabrowski, T., & Olbert, A. I. (2024). Data-driven modelling for assessing trophic status in marine ecosystems using machine learning approaches. *Environmental Research*, 242, 117755.
- Vapnik, V. (1998). Statistical learning theory. *John Wiley & Sons Google Schola*, 2, 831–842.
- Yadav, B., Gupta, P. K., Patidar, N., & Himanshu, S. K. (2020). Ensemble modelling framework for groundwater level prediction in urban areas of India. *Science of the Total Environment*, 712. <https://doi.org/10.1016/j.scitotenv.2019.135539>
- Zhao, W. L., Gentine, P., Reichstein, M., Zhang, Y., Zhou, S., Wen, Y., Lin, C., Li, X., & Qiu, G. Y. (2019). Physics-constrained machine learning of evapotranspiration. *Geophysical Research Letters*, 46(24), 14496–14507.