

## Water quality Index (WQI) modeling using Artificial Intelligence Techniques (AIT) (Case study: Kashkan watershed, Lorestan province)

Alireza Sepahvand\* | Nasrin Beiranvand | Negar Arjmand

Department of Range and Watershed Management, Faculty of Natural Resources, Lorestan University, Khorramabad, Iran.

E-mail: [sepahvand.a@lu.ac.ir](mailto:sepahvand.a@lu.ac.ir)

### Article Info

**Article type:**  
Research Article

**Article history:**  
Received: 15 Nov. 2024  
Revised: 11 Feb. 2025  
Accepted: 12 Feb. 2025  
Published online: 01 Jun. 2025

**Keywords:**  
*Artificial Intelligence,  
Kharkheh watershed,  
Lorestan province,  
Modelling,  
Water Quality Index (WQI).*

### Abstract

Water quality (WQ) is influenced by various variables, including natural ones like rainfall and erosion and human ones like urban, agricultural, and industry operations, that plays a very important role in assessment and determining factors such as environmental conditions, public health, economic and social progress and development. Therefore, temporal and spatial trending of water quality is necessary for planning water resource management. In this research, the performance of the six soft computing techniques, including, Random Forest, Reduced Error Pruning Tree (REPt), M5P model, bagging RF, bagging REPt and bagging M5P were compared to estimate the water quality index (WQI) in Khorramabad, Biranshahr and Alashtar sub-watersheds, Lorestan province, Iran. At first, based on water quality data, water quality index (WQI) was calculated and ten distinct water quality parameters (2014 to 2023) were used as input variables and WQI as output. Total data set consists of water quality parameters of three sub-watersheds out of which 70% data used to training and 30% data were used to testing phase. Finally, the models were compared with Correlation Coefficient (C.C.), Root Mean Square Error (RMSE), Maximum Absolute Error (MAE), Taylor diagram and Violin plot box. The obtained results suggest that the BM5P is more accurate to estimate the water quality index (WQI) compared to the M5P, ReepTree and Random Forest (RF) models for the given study area. According to the results of the test part of the BM5P model, it has given us the best result, which are the correlation coefficient, the Root Mean Square Error and the Mean Absolute Error 0.99, 0.2, and 0.15, respectively. Also, the Taylor diagram and violin box plot were concluded that BM5P was the most reliable soft computing technique for the prediction of WQI. Finally, the structure of Artificial Intelligence Techniques (AIT) for modeling is very simple and very less time consumable. Thus, the BM5P model can be useful in the water quality index (WQI) modeling not only for accuracy but also for its time-saving and simple structure compared with other models.

**Cite this article:** Sepahvand, A., Beiranvand, N., Arjmand, N. (2025). Water quality Index (WQI) modeling using Artificial Intelligence Techniques (AIT) (Case study: Kashkan watershed, Lorestan province). *Journal of Range & Watershed Management*, 78 (2), 207-223. DOI: <http://doi.org/10.22059/jrwm.2025.385428.1789>



© The Author(s).

Publisher: University of Tehran Press

## پیش‌بینی شاخص کیفیت آب با استفاده از تکنیک‌های یادگیری ماشین (منطقه مورد مطالعه: بخشی از حوزه آبخیز کشکان)

علیرضا سپهوند\*<sup>ID</sup> | نسرين بیرانوند | نگار ارجمند

گروه مهندسی مرتع و آبخیزداری، دانشکده منابع طبیعی، دانشگاه لرستان، خرم آباد، ایران  
رایانامه: [sepahvand.a@lu.ac.ir](mailto:sepahvand.a@lu.ac.ir)

### چکیده

### اطلاعات مقاله

#### نوع مقاله:

مقاله پژوهشی

تاریخ دریافت: ۱۴۰۳/۰۸/۲۵

تاریخ بازنگری: ۱۴۰۳/۱۱/۲۳

تاریخ پذیرش: ۱۴۰۳/۱۱/۲۴

تاریخ انتشار: ۱۴۰۴/۰۳/۱۱

#### کلیدواژه‌ها:

استان لرستان،

حوزه آبخیز کرخه،

شاخص کیفیت آب،

مدل‌سازی،

هوش مصنوعی.

کیفیت آب تحت تأثیر متغیرهای مختلفی از جمله متغیرهای طبیعی مانند بارندگی و فرسایش و متغیرهای انسانی مانند عملیات شهری، کشاورزی و صنعتی است که در ارزیابی و تعیین عواملی مانند شرایط زیست محیطی، سلامت عمومی و پیشرفت و توسعه اقتصادی و اجتماعی نقش بسیار مهمی دارد. لذا روندیابی زمانی و مکانی کیفیت آب جهت برنامه‌ریزی مدیریت منابع آب ضروری است. این تحقیق با هدف مدل‌سازی شاخص کیفیت آب (WQI) به کمک مدل‌های هوش مصنوعی با تکیه بر مدل‌های یادگیری ماشین BM5P، M5P، RF، RepTree (ترکیب M5P و Bagging)، BRF (ترکیب RF و Bagging) و BRepTree (ترکیب RepTree و Bagging) در حوضه‌های خرم‌آباد، الشتر و بیرانشهر، استان لرستان انجام گرفته است. در این تحقیق ابتدا بر اساس داده‌های کیفیت آب، شاخص کیفیت آب (WQI) محاسبه شد و در ادامه برای مدل‌سازی، داده‌های ورودی شامل پارامترهای کیفی آب یک دوره ۱۰ ساله (۱۳۹۳-۱۴۰۳) و همچنین داده خروجی شاخص کیفیت آب رودخانه‌ها بود. در این تحقیق برای مدل‌سازی در مرحله آموزش ۷۰ درصد داده‌ها و در مرحله صحت سنجی ۳۰ درصد باقی‌مانده مورد استفاده قرار گرفتند و در نهایت بر اساس نتایج معیارهای ارزیابی خطای ضریب همبستگی (CC)، ریشه میانگین مربعات خطا (RMSE) و میانگین خطای مطلق (MAE) همچنین نمودارهای تی‌لور و ویولن باکس مدل بهینه انتخاب شد. نتایج تحقیق نشان داد که از بین مدل‌های استفاده شده، مدل BM5P با توجه به معیارهای ارزیابی مدل CC، MAE و RMSE در بخش صحت سنجی به ترتیب برابر ۰/۹۹، ۰/۱۵ و ۰/۲۰ از عملکرد بهتری نسبت به سایر مدل‌ها برخوردار بوده است. در ادامه با توجه به نتایج حاصل از نمودارهای تی‌لور و ویولن باکس به منظور صحت سنجی کارایی مدل‌ها، برتری این مدل نسبت به سایر مدل‌ها تایید شد. در نهایت نتایج نشان داد که برای صرفه‌جویی در زمان و هزینه و همچنین مدیریت بهینه پارامترهای کیفیت آب، یکی از راهکارهای مناسب، استفاده از تکنیک‌های هوش مصنوعی به منظور مدل‌سازی و ارزیابی شاخص کیفیت آب (WQI) می‌باشد.

**استناد:** سپهوند؛ علیرضا، بیرانوند؛ نسرين، ارجمند؛ نگار (۱۴۰۴). پیش‌بینی شاخص کیفیت آب با استفاده از تکنیک‌های یادگیری ماشین (منطقه مورد مطالعه: بخشی از حوزه آبخیز کشکان).

نشریه مرتع و آبخیزداری، ۷۸ (۲)، ۲۰۷-۲۲۳.

DOI: <http://doi.org/10.22059/jrwm.2025.385428.1789>



© نویسندگان.

ناشر: انتشارات دانشگاه تهران.

<sup>1</sup> Water Quality Index

## ۱. مقدمه

استفاده‌ی بهینه از منابع آب و حفظ آن از اصول توسعه پایدار در هر کشور به حساب می‌آید. فرآیندهای طبیعی و فعالیت‌های انسانی بر کیفیت آب تاثیرگذار می‌باشند (Muangthong, 2015). کیفیت آب در سلامت عمومی و محیط اکولوژیکی نقش مهمی دارد (Lu et al, 2015؛ Shi et al 2017). یکی از معضلات زیست محیطی جهان، تخریب کیفیت آب بر اثر منابع آلودگی نقطه‌ای و غیرنقطه‌ای می‌باشد (Shoemaker et al, 2017؛ Zeilinski et al, 2016). تخریب کیفیت آب‌های زیرزمینی باعث افزایش شوری آب‌های زیرزمینی می‌شود که عمدتاً در اثر فعالیت‌های انسانی و طبیعی می‌باشد (Mohammed et al, 2022). ورود مواد مغذی و آلاینده‌های تجزیه‌پذیر بیولوژیکی به رودخانه‌ها سبب آلودگی و به هم زدن توازن اکسیژن محلول می‌شود. کاربرد بسیار زیاد و بی‌رویه‌ی مواد شیمیایی در کشاورزی، اکوسیستم آب و خاک، رودخانه‌ها و منابع آب زیرزمینی را تحت تاثیر قرار می‌دهد (Baldwin et al, 2011؛ Kostka, 2016). از آنجایی که عوامل انسانی و آلاینده‌های صنعتی باعث افزایش غلظت آلاینده‌ها در آب رودخانه می‌شود همچنین باتوجه به اینکه مکانیزم‌های طبیعی مثل خصوصیات شیمیایی و فیزیکی، سهم زیادی در کم و زیاد شدن شدت این غلظت‌ها دارند، در نتیجه اولین قدم در تعیین کیفیت آب رودخانه‌ها، آگاه بودن از تغییرات کیفی آب رودخانه‌ها در ابعاد زمان و مکان و همچنین مشخص کردن منابع اصلی و انواع آلوده‌کننده‌های آب می‌باشد (Lao, 2019).

تغییر در کیفیت آب و بیابان آب از اثرات مهم تغییر کاربری اراضی در هیدرولوژی حوزه‌های آبخیز است (Zare et al, 2012). یکی از نکات مهم برای کنترل آلودگی، شناسایی عوامل و منابع آلودگی در مناطق بحرانی و آلوده و همچنین مسیر حرکت آلودگی می‌باشد که آگاهی از این اطلاعات باعث حفظ و بالا بردن کیفیت آب می‌شود (Abdelgafez et al, 2021). پارامترهای کیفیت آب زیرزمینی در استفاده‌ی موارد خاص با سطوح مجاز به شناسایی منابع آلودگی کمک می‌کنند (El-Rawy et al, 2019). امروزه از روش‌های مختلفی برای ارزیابی شاخص کیفیت آب استفاده می‌گردد که می‌توان به روش‌هایی مانند روش فازی، درجه عضویت، تحلیل عملی و مدل‌سازی خاکستری اشاره کرد. این روش‌ها به‌طور دقیق مواد آلوده‌کننده آب را توصیف نمی‌کند و نمی‌توان توضیح داد که به‌صورت دقیق پارامترهای انتخابی برای ارزیابی کیفیت آب مناسب می‌باشند یا خیر (Jian-Hua et al, 2021).

سیستم اطلاعات جغرافیایی (GIS) ابزار تکمیل‌کننده‌ی بسیار خوبی برای ایجاد و توسعه‌ی ارزیابی کیفیت آب می‌باشد (El-Rawy et al, 2021). علاوه بر روش‌های سنتی محدودی وسیعی از پارامترهای کیفی وجود دارد که اندازه‌گیری آن با توجه به محدودیت‌های زمانی و اقتصادی به راحتی امکان‌پذیر نیست. به منظور حل این مشکل می‌توان از شاخص کیفی آب استفاده کرد (Noori et al, 2019). شاخص کیفیت آب مکانیسمی است که از بیان عددی برای مشخص کردن کیفیت آب سطحی و زیرزمینی استفاده می‌کند. این روش قابلیت زیادی در بیان کامل اطلاعات کیفی آب و به‌کارگیری پارامترهای مهم و تاثیرگذار در ارزیابی و مدیریت کیفیت آب در نقاط مختلف دارد (Simoes et al, 2008). قطعی نبودن ضرایب مورد استفاده در مدل‌سازی کیفیت آب، اطمینان به نتایج مدل‌سازی را تحت تاثیر قرار می‌دهد. در بسیاری از موارد ممکن است میزان این ضرایب با مقدار واقعی خود در محیط متفاوت باشد (Saleemi and Lohani, 1982). مکانیسمی که برای سطح مشخص کیفیت آب زیرزمینی از یک عدد بیانی استفاده می‌کند را شاخص کیفی می‌گویند که روشی با قابلیت بالا در توصیف اطلاعات کیفی آب و به‌کارگیری پارامترهای مهم و تاثیرگذار در مدیریت کیفیت آب زیرزمینی در نقاط بسیاری از جهان می‌باشد (Simoes et al, 2008). شاخص کیفیت آب (WQI) روشی ممتاز به منظور درک و خلاصه کردن تعداد زیادی از داده‌های کیفیت آب برای نمایش تاثیر ترکیب متغیرهای موثر بر کیفیت آب می‌باشد (Fathi et al, 2018).

امروزه برای تخمین WQI استفاده از تکنیک‌های هوش مصنوعی افزایش یافته است (Ubah et al, 2021). ساختار غیرخطی تکنیک‌های هوش محاسباتی و تولنایی در پیش‌بینی رخدادهای پیچیده و مدیریت داده‌ها با اندازه‌های مختلف همگی از مزایای این تکنیک‌ها در پیش‌بینی میزان آب می‌باشد (Kouadri et al, 2022). در سال‌های اخیر برای ارزیابی کیفیت آب کیفیت آب‌های سطحی و زیرزمینی شاخص‌های مختلفی مانند WQI، NSFQI، GWQI و ... در زمینه بررسی کیفیت آب توسعه یافته است که مطالعات زیادی

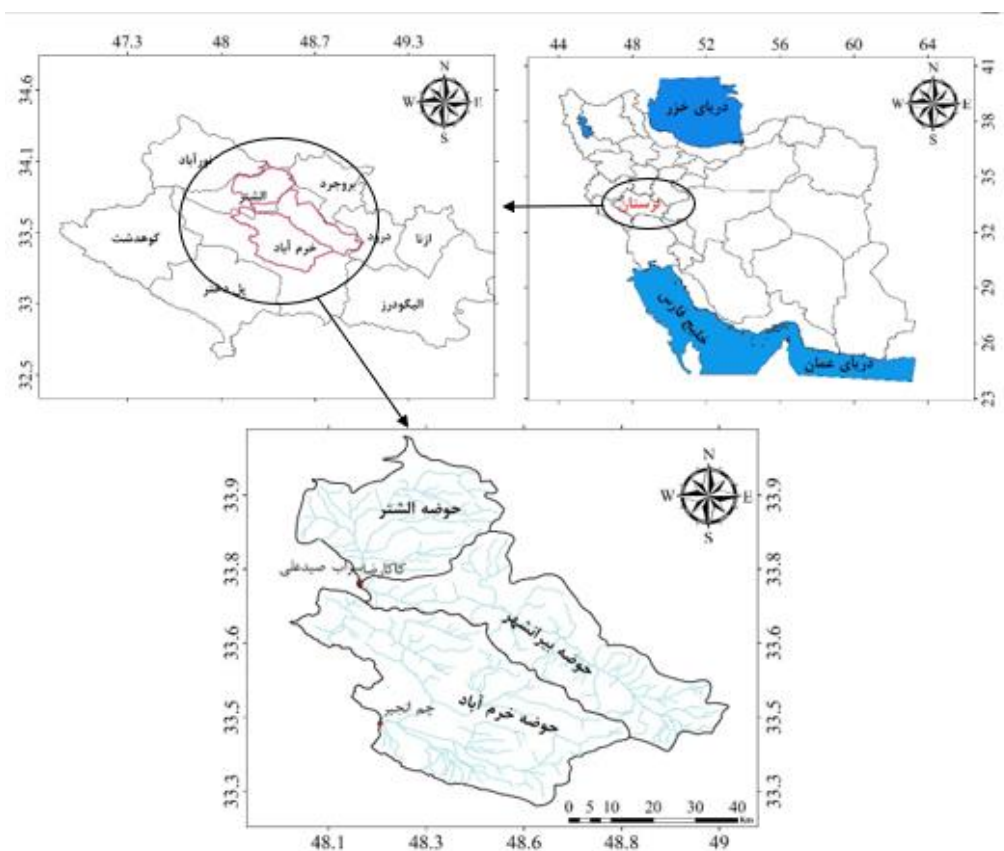
در داخل و خارج از کشور انجام شده است.

Aryen-Najad et al (2020) در تحقیق خود به مدل سازی کیفیت آب رودخانه با استفاده از مدل Qual2Kw در رودخانه شاهرود پرداختند و نتایج نشان داد که مدل Qual2Kw در شبیه سازی پارامترهای کیفیت آب رودخانه شاهرود دقتی مناسب دارد. از طرفی، دقت شبیه سازی هر پارامتر با توجه به مقدار تغییرات آن در طول رودخانه متغیر است؛ به گونه ای که هرچه تغییرات یک پارامتر در طول رودخانه و در بازه های مختلف کمتر باشد، دقت مدل در شبیه سازی این پارامتر نیز بیشتر خواهد بود. (Babakhani et al (2020 به تعیین ظرفیت خودپالایی رودخانه دیواندره با استفاده از مدل Qual2Kw پرداختند نتایج نشان داد به دلیل افزایش سکونت گاه های انسانی، تأثیر فاضلاب ها، رواناب های کشاورزی و همچنین ورود رودخانه های فرعی چم زرد و پول گشتی، وضعیت زیست محیطی رودخانه دیواندره در بازه های انتهایی آن نگران کننده است. (Eidi et al (2022 به ارزیابی کیفیت آب منطقه دهرود و تنگ ارم پرداختند. آن ها در تحقیق خود کیفیت آب را در فصل های خشک و تر اندازه گیری کردند. نتایج نشان داد با کاهش سطح آب زیرزمینی از کیفیت آب کاسته می شود. (Karimi et al (2023 به اثر سرشاخه سیول و قذح در کاهش کیفیت آب پرداختند. نتایج نشان داد رودخانه پس از جدا شدن از سازندهای کارستی دچار کاهش تدریجی کیفیت می شود. (Verma et al (2020 با استفاده از شاخص کیفیت آب (WQI) و سیستم اطلاعات جغرافیایی (GIS) کیفیت آب منطقه بوکارو در هند را بررسی کردند. نتایج اسیدی تا کمی قلیایی بودن آب زیرزمینی را در منطقه نشان داده است. (Khalili et al (2020 کیفیت آب رودخانه گرمارود را با استفاده از شاخص های مختلف بررسی کردند که نتایج نشان داد کیفیت آب در ایستگاه یک نسبت به دو ایستگاه دوم و سوم بهتر بوده است. (Mussab et al (2023 و (Mohammed et al (2023 به بررسی کیفیت آب زیرزمینی با استفاده از الگوریتم های یادگیری پرداختند. آن ها در تحقیقات خود پارامترهای کیفی شاخص کیفیت آب GWQI برآورد کردند، نتایج مدل سازی هر دو تحقیق نشان داد مدل ها در مدل سازی عملکرد مثبتی داشته اند. (El-Rawy et al (2023 به بررسی کیفیت آب زیرزمینی را با استفاده از مدل های هوش مصنوعی SVM، GP و ANN پرداختند. آن ها در این تحقیق از سیستم اطلاعات جغرافیایی و الگوریتم های یادگیری استفاده کردند. نتایج نشان دهنده عملکرد ضعیف تر مدل SVM نسبت به دو مدل دیگر در مدل سازی کیفیت آب بوده است. (Sihag et al (2024 به توسعه مدل های مبتنی بر الگوریتم های یادگیری برای پیش بینی شاخص کیفیت آب در استان لرستان پرداختند. آن ها در این تحقیق از مدل های ANN-FFA، BREPt، REPt، GP و GEP برای پیش بینی WQI استفاده کردند. نتایج تحقیق آن ها نشان داد که مدل ANN-FFA دارای بیشترین دقت برای پیش بینی شاخص کیفیت آب می باشد. لذا باتوجه به اهمیت بسیار زیاد تأثیر عوامل مختلف بر کیفیت آب و همچنین تأثیر کیفیت آب هر منطقه بر مناطق دیگر و باتوجه به مطالعات انجام شده، تعیین مدلی مناسب با دقت بالا برای تعیین شاخص کیفیت آب برای صرفه جویی در زمان و هزینه ضروری می باشد. لذا این پژوهش با هدف بررسی عملکرد مدل های مختلف بر تعیین شاخص کیفیت آب رودخانه های چمانجیر، بهرام جو، سراب صیدعلی و کاکارضا در حوزه های خرم آباد، الشتر و بیرانشهر انجام گرفته است.

## ۲. مواد و روش ها

### ۲-۱. منطقه مورد مطالعه

حوزه های آبخیز خرم آباد، الشتر و بیرانشهر از نظر ژئومورفولوژیکی در بخش مرکزی سلسله جبال زاگرس در استان لرستان قرار دارند که خود بخشی از حوزه آبخیز کرخه هستند. حوزه های آبخیز خرم آباد، الشتر و بیرانشهر در مجموع با مساحت ۳۵۶۲/۱۲ کیلومتر مربع مطابق شکل ۱ بین طول های جغرافیایی  $38^{\circ}04'02''$  تا  $48^{\circ}59'07''$  شرقی و عرض های  $33^{\circ}16'05''$  تا  $34^{\circ}01'07''$  شمالی قرار دارند. بلندترین نقطه در بخش شمالی با ارتفاع ۳۵۷۸ متر و پست ترین نقطه در بخش غربی، یعنی خروجی حوزه آبخیز با ارتفاع ۱۱۵۸ متر از سطح دریا قرار دارد. میانگین بارندگی سالانه منطقه بین ۴۰۰ تا ۵۰۰ میلی متر بوده و دارای اقلیم نیمه خشک و سرد است (Sepahvand et al, 2019). در جدول ۱ مشخصات جغرافیایی ایستگاه های مختلف حوزه آبخیز کشکان نشان داده شده است.



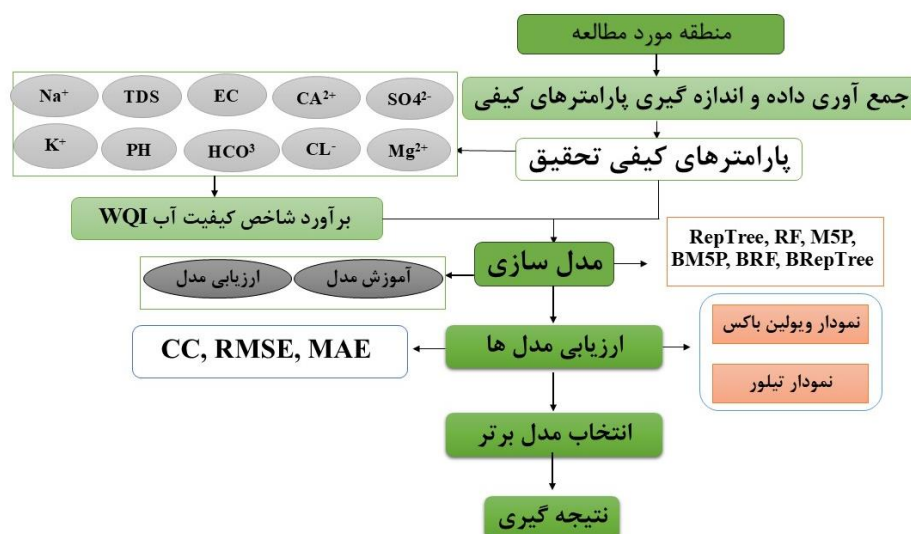
شکل ۱. نقشه موقعیت جغرافیایی منطقه مورد مطالعه در نقشه استان لرستان و ایران

جدول ۱. مشخصات جغرافیایی ایستگاه‌های مورد مطالعه

| ردیف | ایستگاه     | طول جغرافیایی    | عرض جغرافیایی    | حوزه آبخیز | منطقه |
|------|-------------|------------------|------------------|------------|-------|
| ۱    | چم انجیر    | $48^{\circ} 12'$ | $33^{\circ} 52'$ | خرم آباد   | خروجی |
| ۲    | کاکارضا     | $48^{\circ} 15'$ | $33^{\circ} 42'$ | کاکارضا    | خروجی |
| ۳    | سراب صیدعلی | $48^{\circ} 15'$ | $33^{\circ} 30'$ | الشتر      | خروجی |

## ۲-۲. جمع آوری و اندازه گیری پارامترهای کیفی

در این تحقیق در محل ایستگاه‌های هیدرومتری (معمولاً خروجی حوضه‌ها)، نمونه‌برداری آب در یک دوره ۱۰ ساله (۱۳۹۳-۱۴۰۳) در مقیاس ماهانه انجام گرفت و بعد از انتقال نمونه‌ها به آزمایشگاه مرکزی دانشگاه لرستان، خصوصیات کیفی آب تعیین شد. برای این منظور، ظروف نمونه‌برداری از جنس پلاستیک به حجم یک لیتر تهیه شد که قبل از نمونه‌برداری بایستی با یک پاک کننده به‌خوبی شسته شوند. نمونه‌برداری از جریان اصلی و در عمق حدود پنج سانتی‌متری از سطح به کف انجام گرفت. لازم به ذکر است که بخشی از داده‌های کیفیت آب ایستگاه‌های هیدرومتری مورد تحقیق از شرکت آب منطقه‌ای استان لرستان تهیه شدند. شکل ۲ نمودار جریانی تحقیق را نشان می‌دهد.



شکل ۲. نمودار جریان مراحل تحقیق

## ۲-۳. مجموعه داده‌های مدل‌سازی

داده‌های استفاده شده در این تحقیق برای مدل‌سازی کیفیت آب در مجموع شامل ۱۲۴ داده برای هر پارامتر می‌باشد که پارامترها در دو گروه داده‌های ورودی و خروجی قرار گرفته شد. داده‌های ورودی شامل  $\text{Na}^+$ ,  $\text{TDS}$ ,  $\text{EC}$ ,  $\text{PH}$ ,  $\text{HCO}_3^-$ ,  $\text{CL}^-$ ,  $\text{SO}_4^{2-}$ ,  $\text{CA}^{2+}$ ,  $\text{Mg}^{2+}$  و  $\text{K}^+$  می‌باشند همچنین داده‌های خروجی شاخص کیفیت آب (WQI) می‌باشند. در این تحقیق ۷۰ درصد کل داده‌ها مربوط به بخش آموزش (تعداد ۸۶ داده برای هر پارامتر) و ۳۰ درصد آن‌ها در مرحله صحت سنجی مدل (تعداد ۳۸ داده برای هر پارامتر) استفاده شده است. در این تحقیق از مدل‌های RepTree, RF, M5P, BM5P, BRF, BRepTree و برای مدل‌سازی و تخمین شاخص کیفیت آب (WQI) در محیط برنامه نویسی پایتون استفاده شدند. مدل‌های مبتنی بر یادگیری به لحاظ به کارگیری الگوریتم‌های تخمین و توزیع، طبیعت داده محور و تکرار بالای فرآیند مدل‌سازی، توانایی بالایی در شناسایی رفتار وقوع پدیده مورد نظر داشته و در پژوهش‌های علمی زیادی برتری خود را نسبت به دیگر مدل‌های آماری دو متغیره و چند متغیره اثبات نموده‌اند (Singh et al, 2021) که در ادامه توضیح مختصری در مورد این مدل‌ها آورده شده است.

## ۲-۴. شاخص کیفیت آب (WQI)

سه نوع پارامتر بیولوژیکی، فیزیکی و شیمیایی برای تجزیه و تحلیل کیفیت آب مورد استفاده قرار می‌گیرد که با استفاده از آن‌ها WQI محاسبه می‌شود (Dogan et al, 2009). این پارامترها شامل  $\text{Na}^+$ ,  $\text{TDS}$ ,  $\text{EC}$ ,  $\text{PH}$ ,  $\text{HCO}_3^-$ ,  $\text{CL}^-$ ,  $\text{SO}_4^{2-}$ ,  $\text{CA}^{2+}$ ,  $\text{Mg}^{2+}$  و  $\text{K}^+$  می‌باشند که با استفاده از این پارامترها WQI به صورت زیر محاسبه می‌شود:

۱- ابتدا وزن ( $Z_i$ ) هر پارامتر مشخص می‌شود.

۲- وزن نسبی پارامترها با استفاده از رابطه ۱ مشخص می‌شود:

$$Z_i = \frac{Z_i}{\sum_{i=1}^n Z_i} \quad (\text{رابطه ۱})$$

که در آن  $Z_i$  وزن هر پارامتر،  $Z_i$  وزن نسبی و  $n$  تعداد پارامترها است (Singh et al, 2021; Wagh et al, 2017; Vasant et al, 2016).

۳- مقیاس نرخ کیفیت آب برای هر پارامتر از طریق رابطه ۲ محاسبه می‌شود:

$$s_i = \left( \frac{con_i}{std_i} \right) * 100 \quad \text{رابطه ۲}$$

که در این رابطه  $s_i$  مقیاس نرخ کیفیت آب،  $con_i$  غلظت هر پارامتر در نمونه‌های آب برحسب میلی‌گرم در لیتر و  $std_i$  استاندارد هر پارامتر طبق WHO است (Singh et al, 2021).

۴- محاسبه‌ی زیر شاخص کیفیت آب (SIL)<sup>۱</sup> برای هر پارامتر با استفاده از رابطه ۳:

$$SIL_i = Z_i \times s_i \quad \text{رابطه ۳}$$

۵- در نهایت محاسبه‌ی شاخص کیفیت آب (WQI) با استفاده از رابطه ۴:

$$WQI = \sum SIL_i \quad \text{رابطه ۴}$$

## ۲-۵. مدل REP Tree<sup>۲</sup>

مدل REPTree، یک مدل از مجموعه درخت تصمیم‌گیری (DT) است که برای رفع مشکلات طبقه‌بندی و رگرسیون کارآمد است (Quinlan, 1992). الگوریتم REPTree یک درخت رگرسیون تصمیم‌گیری است که به‌وسیله تقسیم و هرس کردن درخت رگرسیون، تصمیم‌گیری را براساس بالاترین اطلاعات مقدار سود (IGR)، تشکیل می‌دهد، که مقدار IGR بر اساس تابع آنتروپی (E) و مطابق رابطه (۵) محاسبه شده است.

$$IGR(X, S) = \frac{E(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} E(S_i)}{- \sum_{i=1}^n \frac{|S_i|}{|S|} \log_2 \frac{|S_i|}{|S|}} \quad \text{رابطه ۵}$$

که در آن: T ضریب تصادفی،  $H(T|a)$  آنتروپی T برای محاسبه خصوصیات a،  $t_i$  زمان رویداد، N(t) تعداد کل رخدادها و IGR اطلاعات مقدار سود می‌باشد (Quinlan, 1992).

## ۲-۶. مدل درختی M5P

یکی از رایج‌ترین الگوریتم‌های مدل درختی، الگوریتم M5 است. این الگوریتم اولین بار توسط کوئینلان در سال ۱۹۹۲ پیشنهاد شد (Quinlan, 1992). سپس الگوریتم M5P که درواقع بازسازی منطقی و توسعه‌یافته از M5 است، توسط ونگ و ویتن در سال ۱۹۹۷ معرفی شد (Wang and Witten, 1997). این مدل‌ها یکی از روش‌های نسبتاً جدید و قدرتمند هوش محاسباتی است که به علت قابلیت مناسب آن، توسط محققان مختلفی مورد استفاده قرار گرفته است. سادگی محاسبه‌ها و دقت نتایج از مهم‌ترین ویژگی‌های این مدل است (Bonakdar et al, 2021). کاهش انحراف معیار مطابق رابطه (۶) محاسبه شده است (Wang and Witten, 1997).

$$SDR = sd(T) - \sum \frac{|T_i|}{|T|} sd(T_i) \quad \text{رابطه ۶}$$

که در آن: T بیانگر یکسری نمونه‌هایی است که به گره می‌رسند،  $T_i$  بیانگر یکسری نمونه‌هایی است که i امین برآمد تست پتانسیلی را دارند، sd بیانگر انحراف معیار و SDR کاهش انحراف معیار می‌باشد.

<sup>1</sup> sub-index level number

<sup>2</sup> Reduced Error Pruning Tree

## ۲-۷. مدل جنگل تصادفی<sup>۱</sup>

الگوریتم RF نخستین بار توسط Breiman ارائه شد و به عنوان روشی دقیق در تهیه نقشه کاربری اراضی و طبقه بندی تصاویر ماهواره ای استفاده شده است (Pal, 2005 و Prasad et al, 2006). رگرسیون RF شامل ترکیبی از متغیرها در هر گره برای رشد یک درخت یا استفاده از متغیر ورودی تصادفی انتخاب شده است. روشی است که داده ها را به طور تکراری برای به دست آوردن ارتباط بین متغیر پاسخ و متغیرهای مستقل و انجام تخمین جداسازی می کند. در روش RF برخلاف سایر روش های درختی که تعداد محدودی درخت ترسیم می کنند، صدها یا هزاران درخت طبقه بندی تولید می شود. این روش یک روش یادگیری گروهی است و برای طبقه بندی با ساختن تعداد درختان زیاد عمل می نماید. روش RF با استفاده از آنالیز حساسیت، اهمیت متغیرها در مدل سازی را نیز تعیین می کند. RF به روش میانگین کاهش حداقل (MDA)، قادر به ارائه اهمیت متغیرهای مورد استفاده در فرآیند مدل سازی می باشد. در روش MDA، مقادیر درست متغیرها با مقداری که به طور تصادفی برای هر درخت تولید شده است جایگزین می شود و اگر این جایگزینی اثری روی خطای اندازه گیری نداشته باشد اهمیت آن کم است و اگر مقدار خطای اندازه گیری افزایش یابد، آن متغیر مهم است (Breiman, 1996). برای رگرسیون RF دو پارامتر لازم توسط کاربر تعریف شود: تعداد متغیرهای ورودی (متر) استفاده شده در هر گره برای تولید یک درخت و تعداد درختان قابل رشد در هر گره، فقط متغیرهای انتخاب شده برای بهترین تقسیم جستجو می شوند (Sepahvand et al, 2020).

## ۲-۸. الگوریتم Bagging

یک تکنیک یادگیری یکپارچه است که در بسیاری از مدل ها به منظور تولید یک مدل قوی و دقیق استفاده می شود. روشی است که در مجموعه ای از داده ها به منظور کاهش اختلافات و ناسازگاری ها و افزایش توانایی تعمیم دهی الگوریتم کاربرد دارد (Deng et al, 2022). این روش در یادگیری ماشین برای ایجاد مدل های پیش بینی پایدارتر و همچنین به حداقل رساندن تغییرات استفاده می شود. ماهیت این الگوریتم نمونه برداری تصادفی از مجموعه داده های اولیه برای ایجاد زیرمجموعه ها استفاده می شوند. سپس هر مدل پایه با توجه به زیر مجموعه های نمونه ها آموزش داده می شود و در نهایت مقدار میانگین نتایج هر مدل به عنوان خروجی در نظر گرفته می شود، که بر این اساس انواع موقعیت های تصادفی را شبیه سازی می کند، و در نهایت باعث کاهش خطای مدل سازی می شود (Deng et al, 2022؛ Das et al, 2022؛ Liang et al, 2023). همچنین این روش ممکن است به همراه بسیاری از مدل های پایه مانند شبکه عصبی، جنگل تصادفی و غیره استفاده شود. در صورت اختلاف بالا در مدل های پایه و همچنین زمانی که تعداد بالایی از داده های آموزش در دسترس باشد، این روش عملکرد بسیار خوبی دارد. لذا در این تحقیق الگوریتم Bagging با مدل های Rf، RepTree و M5P ترکیب شد و برای مدل سازی شاخص کیفیت آب مورد استفاده قرار گرفتند.

## ۲-۹. معیارهای سنجش خطای مدل سازی

برای بررسی دقت نتایج حاصل از مدل سازی پارامترها از معیارهای متفاوتی می توان استفاده کرد. در این تحقیق برای تعیین مدل بهینه از معیارهای سنجش خطای ریشه میانگین مربعات خطا<sup>۲</sup> (RMSE)، میانگین خطای مطلق<sup>۳</sup> (MAE) و ضریب همبستگی<sup>۴</sup> (CC) استفاده شد. براساس معیارهای استفاده شده مدلی که دارای ضریب همبستگی بالاتر و ریشه میانگین مربعات خطا و میانگین خطای مطلق کمتری باشد از دقت بالاتری برخوردار است (رابطه های ۷ تا ۱۰).

<sup>1</sup> Random Forest

<sup>2</sup> Root Mean Squared Error

<sup>3</sup> Mean Bias Error

<sup>4</sup> Correlation Coefficient



$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (H - F)^2} \quad \text{(رابطه ۷)}$$

$$MAE = \frac{\sum_{i=1}^n |H_k - F_k|}{K} \quad \text{(رابطه ۸)}$$

$$CC = \frac{n \sum HF - (\sum H)(\sum F)}{\sqrt{n(\sum H^2) - (\sum H)^2} \sqrt{n(\sum F^2) - (\sum F)^2}} \quad \text{(رابطه ۹)}$$

در روابط فوق،  $F$  مقدار برآورد شده،  $H$  مقدار مشاهده شده،  $\bar{H}$  میانگین مقادیر مشاهده شده و  $n$  تعداد داده‌ها می‌باشد. علاوه بر معیارهای سنجش خطا، از نمودارهای تیلور و ویولین باکس برای مقایسه نتایج مدل‌ها استفاده شد که این نمودارها در نرم‌افزار Origin 2024 SR1 استفاده شد.

## ۲-۱۰. نمودار تیلور<sup>۱</sup>

نمودارهای تیلور، نمودارهای ریاضی هستند و برای این منظور طراحی شدند تا به صورت گرافیکی نشان دهند که کدام یک از چندین مدل تقریبی یک سیستم، فرآیند یا پدیده واقعی‌تر است. نمودار تیلور در سال ۲۰۰۱ توسط Taylor توسعه یافت، یکی از پرکاربردترین و مدرن‌ترین روش‌هایی که نشان می‌دهد یک الگو یا مجموعه‌ای از الگوها را از نظر همبستگی، تغییرات استاندارد و ریشه میانگین مربعات خطا بررسی می‌کند که برای حل مسئله‌های مربوط به مدل‌های پیچیده مفید است (Taylor, 2001). نمودار تیلور، مقایسه‌ای مدل‌های مختلف را بر اساس ضریب همبستگی پیرسون، خطای ریشه میانگین مربعات (RMSE) و انحراف معیار استاندارد تسهیل می‌کند.

## ۲-۱۱. نمودار ویولین باکس<sup>۲</sup>

نمودار ویولین باکس یک نمایش گرافیکی منحصر به فرد است که به طور یکپارچه طرح ویولن و طرح جعبه را با هم ادغام می‌کند. در نتیجه نمودار ویولن باکس ترکیبی از نمودار ویولن و نمودار جعبه است که مزایای هر دو طرح را دارد. نمودار جعبه مانند ویولن دارای چند لایه است که میانگین داده‌ها را به صورت عمودی روی نمودار شبیه نمودارهای ویولن و جعبه نشان می‌دهد (Singh et al, 2021; Hameed et al, 2024).

## ۳. نتایج

جدول ۲ خصوصیات آماری داده‌های ورودی و خروجی در بخش آموزش و صحت‌سنجی فرآیند مدل‌سازی را نشان می‌دهد. مطابق نتایج این جدول مشخص شد که بیشترین میزان شاخص کیفیت آب (WQI) در بخش آموزش و صحت‌سنجی به ترتیب برابر ۵۶/۶۳ و ۴۹/۵۷ برآورد شده و کمترین مقدار شاخص کیفیت آب به ترتیب برابر ۲۴/۸۱ و ۲۳/۶۵ می‌باشد. جدول ۳ میزان وزن و وزن نسبی پارامترهای ورودی و خروجی مدل و جدول ۴ بیان‌کننده اطلاعات کیفی شاخص کیفیت آب (WQI) می‌دهد.

جدول ۵ نتایج حاصل از مدل‌سازی داده‌ها با استفاده از مدل‌های یادگیری ماشین BRepTree، RepTree، RF، M5P، BM5P و BRF استفاده شده در تحقیق را نشان می‌دهد و نتایج به دست آمده نشان می‌دهد کدام یک از مدل‌های به کار گرفته شده در مقایسه با سایر مدل‌ها از عملکرد بهتری برخوردار بوده است. مطابق نتایج جدول ۵ مشخص شد که مدل BM5P بر اساس میزان معیارهای ارزیابی خطای MAE، CC و RMSE در مرحله آموزش به ترتیب برابر ۰/۹۹، ۰/۱۲ و ۰/۱۷ و در مرحله صحت‌سنجی به ترتیب برابر ۰/۹۹، ۰/۱۵ و ۰/۲۰، نسبت به نتایج مدل‌های RepTree، RF، M5P، BRF و BRepTree عملکرد بهتری داشته است و به عنوان مدل برتر برای تخمین میزان

<sup>۱</sup> Taylor diagram

<sup>۲</sup> Violin box diagram

شاخص کیفیت آب (WQI) انتخاب شد. همچنین مدل RepTree نسبت به سایر مدل‌ها عملکرد ضعیف‌تری داشته که نتایج این مدل در بخش آموزش باتوجه به معیارهای ارزیابی کارایی به‌ترتیب برابر ۰/۹۸، ۰/۹۰ و ۱/۴۶ بوده و در مرحله صحت‌سنجی به‌همین ترتیب برابر ۰/۹۶، ۱/۴۵ و ۱/۷۲ برآورد شده است.

جدول ۲. خصوصیات آماری داده‌های ورودی و خروجی آموزش و صحت سنجی

| آموزش                         |         |        |         | صحت سنجی     |         |        |         |
|-------------------------------|---------|--------|---------|--------------|---------|--------|---------|
| پارامتر                       | بیشترین | کمترین | میانگین | انحراف معیار | بیشترین | کمترین | میانگین |
| TDS                           | ۵۷۳     | ۱۶۰    | ۳۱۷/۶۲  | ۹۳/۵۱        | ۴۹۸     | ۱۵۰    | ۳۱۲/۸۶  |
| EC                            | ۸۸۲     | ۲۴۱    | ۴۹۲/۴۸  | ۱۴۴/۵۹       | ۷۶۷     | ۲۴۹    | ۴۹۰/۸۹  |
| PH                            | ۸/۳۹    | ۶/۴۷   | ۷/۷۰    | ۰/۵۱۳        | ۸/۳۶    | ۶/۸    | ۷/۶۸    |
| HCO <sub>3</sub> <sup>-</sup> | ۱۸۰     | ۵۶/۴۴  | ۱۱۴/۲۶  | ۲۶/۲۲        | ۱۸۰     | ۶۴/۰۷  | ۱۱۸/۸۸  |
| CL <sup>-</sup>               | ۹۲/۱۷   | ۳/۵۴   | ۲۷/۲۳   | ۲۰/۳۴        | ۷۰/۹    | ۳/۵۴   | ۲۵/۷۸   |
| SO <sub>4</sub> <sup>2-</sup> | ۷۴/۴۴   | ۰      | ۱۷/۴۵   | ۱۵/۰۶        | ۵۵/۷۱   | ۰/۴۸   | ۱۴/۵۸   |
| Ca <sup>2+</sup>              | ۱۱۰/۲۲  | ۲۴/۰۴  | ۵۹      | ۱۶/۲۵        | ۸۲/۱۶   | ۲۶/۰۵  | ۶۱/۷۲   |
| Mg <sup>2+</sup>              | ۳۸/۸۹   | ۳/۶۴   | ۱۸/۴۱   | ۸/۷۴         | ۴۸/۶۲   | ۲/۴۳   | ۹۶/۱۸   |
| Na <sup>+</sup>               | ۴۵/۹۸   | ۶۸۰    | ۹/۳۹    | ۹/۹۹         | ۳۱/۹۵   | ۰/۶۸   | ۷/۲۲    |
| K <sup>+</sup>                | ۸/۶۰    | ۰      | ۲/۵۶    | ۲/۲۶         | ۱۱/۷۳   | ۰      | ۲/۶۵    |
| WQI                           | ۵۶/۶۳   | ۲۴/۸۱  | ۳۵/۶۰   | ۷/۴۸         | ۴۹/۵۷   | ۲۳/۶۵  | ۳۵/۶۲   |

جدول ۳. جزئیات وزن و وزن نسبی پارامترها

| پارامتر                       | z <sub>i</sub> | Z <sub>i</sub> | WHO     |
|-------------------------------|----------------|----------------|---------|
| TDS                           | ۵              | ۰/۱۵۶۲۵        | ۵۰۰     |
| EC                            | ۲              | ۰/۰۶۲۵         | ۵۰۰     |
| PH                            | ۴              | ۰/۱۲۵          | ۶/۵-۸/۵ |
| HCO <sub>3</sub> <sup>-</sup> | ۳              | ۰/۰۹۳۷۵        | ۵۰۰     |
| CL <sup>-</sup>               | ۴              | ۰/۱۲۵          | ۲۵۰     |
| SO <sub>4</sub> <sup>2-</sup> | ۴              | ۰/۱۲۵          | ۲۵۰     |
| Ca <sup>2+</sup>              | ۳              | ۰/۰۹۳۷۵        | ۷۵      |
| Mg <sup>2+</sup>              | ۳              | ۰/۰۹۳۷۵        | ۵۰      |
| Na <sup>+</sup>               | ۲              | ۰/۰۶۲۵         | ۲۰۰     |
| K <sup>+</sup>                | ۲              | ۰/۰۶۲۵         | ۱۰      |
| TDS                           | ۳۲             | ۱              | -       |
| جمع                           | ۵              | ۰/۱۵۶۲۵        | ۵۰۰     |

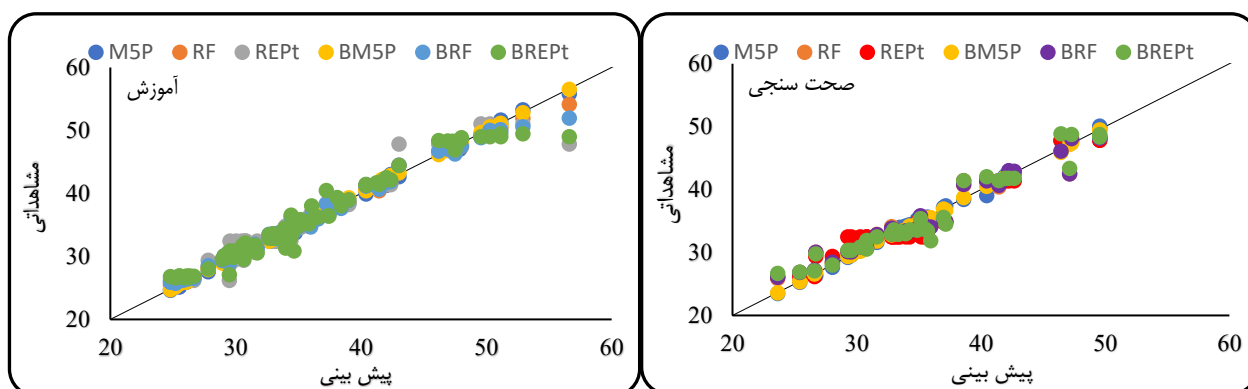
جدول ۴- توصیف کیفی شاخص WQI (Eidi et al, 2022)

| کیفیت آب | شاخص WQI      |
|----------|---------------|
| خیلی خوب | $50 >$        |
| خوب      | $50 - 100 >$  |
| بد       | $100 - 200 >$ |
| خیلی بد  | $200 - 300 >$ |
| نامناسب  | $300 <$       |

شکل ۳ نمودار پراکنش داده‌های مشاهداتی و پیش‌بینی مدل‌های RepTree، RF، M5P، BRF، BM5P و BRepTree در مرحله آموزش و صحت‌سنجی مدل‌سازی را نشان می‌دهد.

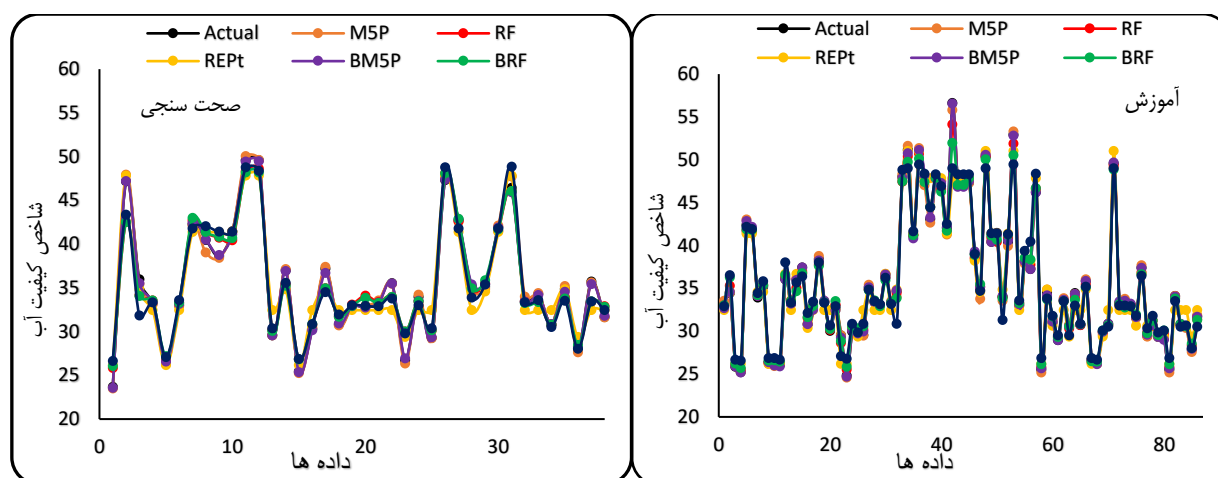
جدول ۵. عملکرد مدل‌های مورد استفاده

| مدل‌ها   | BRepTree | BRf  | BM5p | RepTree | Rf   | M5p  |
|----------|----------|------|------|---------|------|------|
| آموزش    |          |      |      |         |      |      |
| CC       | ۰/۹۸     | ۰/۹۹ | ۰/۹۹ | ۰/۹۸    | ۰/۹۹ | ۰/۹۹ |
| MAE      | ۰/۹۴     | ۰/۵۸ | ۰/۱۲ | ۰/۹۰    | ۰/۳۹ | ۰/۲۴ |
| RMSE     | ۱/۴۲     | ۰/۸۴ | ۰/۱۷ | ۱/۴۶    | ۰/۵۳ | ۰/۳۲ |
| صحت‌سنجی |          |      |      |         |      |      |
| CC       | ۰/۹۶     | ۰/۹۸ | ۰/۹۹ | ۰/۹۶    | ۰/۹۷ | ۰/۹۹ |
| MAE      | ۱/۲۲     | ۱/۰۷ | ۰/۱۵ | ۱/۴۵    | ۱/۰۲ | ۰/۲۷ |
| RMSE     | ۱/۶۲     | ۱/۴۴ | ۰/۲۰ | ۱/۷۲    | ۱/۴۰ | ۰/۴۱ |

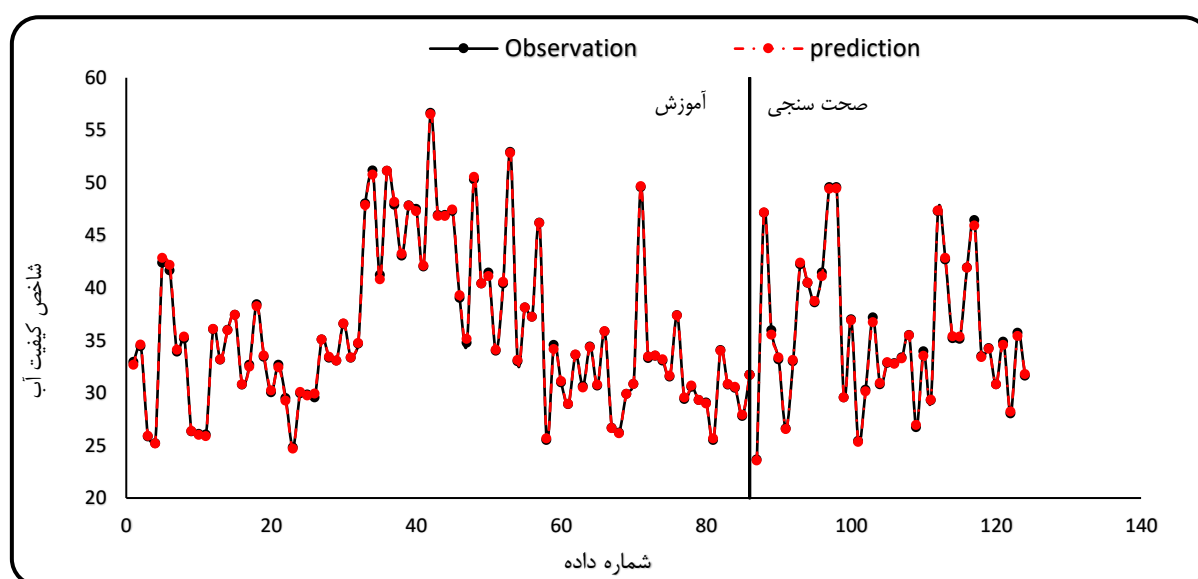


شکل ۳. نمودار پراکنش داده‌های مشاهداتی و پیش‌بینی مدل‌ها در مرحله آموزش و صحت‌سنجی

شکل ۴ نمودار داده‌های مشاهداتی و داده‌های پیش‌بینی شده مدل‌های RepTree، RF، M5P، BRF، BM5P و BRepTree در مرحله آموزش و صحت‌سنجی نشان می‌دهد که مطابق نتایج این نمودار تقریباً تطابق خوبی بین داده‌های مشاهداتی و داده‌های پیش‌بینی وجود دارد. همچنین شکل ۵ نمودار داده‌های مشاهداتی و تخمینی مدل BM5P به‌عنوان مدل برتر را در مرحله آموزش و صحت‌سنجی نشان می‌دهد.



شکل ۴. نمودار داده‌های مشاهداتی و پیش‌بینی مدل‌های استفاده شده در مرحله‌ی آموزش و صحت‌سنجی

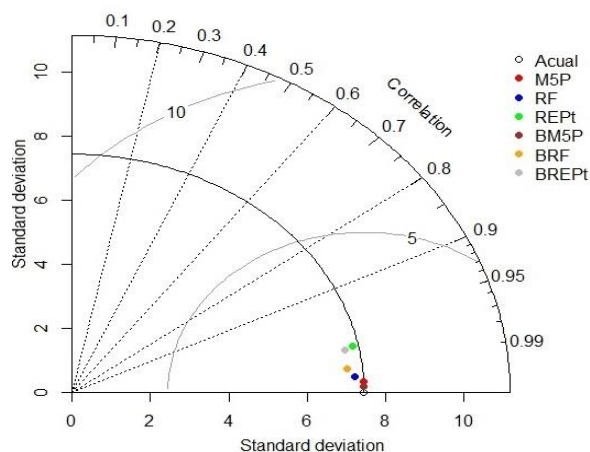


شکل ۵. نمودار داده‌های مشاهداتی و پیش‌بینی مدل BM5P

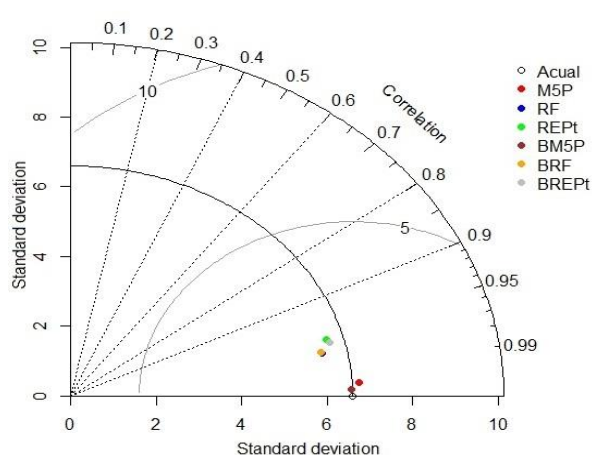
شکل‌های ۶ و ۷ نمودار تیلور را بر اساس نتایج داده‌های تخمینی مدل‌ها برای شاخص کیفیت آب (WQI) در مرحله آموزش (شکل ۶) و مرحله صحت‌سنجی (شکل ۷) نشان می‌دهد. مطابق نتایج شکل‌های ۶ و ۷ در بخش آموزش همه‌ی مدل‌ها در تخمین میزان WQI به خوبی عمل کرده‌اند، اما در بخش صحت‌سنجی مدل BM5P در تخمین میزان WQI بر اساس داده‌های مشاهداتی دقیق‌تر از سایر مدل‌های به کار برده شده عمل کرده است. بنابراین طبق نتایج نمودار تیلور نیز می‌توان به این نتیجه رسید که BM5P مدلی با دقت بیشتر برای تعیین میزان شاخص کیفیت آب (WQI) می‌باشد.

شکل ۸ نمودار ویولن را برای مقایسه نتایج مدل‌های استفاده شده در بررسی و پیش‌بینی شاخص کیفیت آب WQI در بخش آموزش و صحت‌سنجی داده‌ها نشان می‌دهد. مطابق نتایج شکل‌های ۸ و ۹ در بخش آموزش این نمودار تقریباً در همه‌ی مدل‌ها مقادیر پیش‌بینی

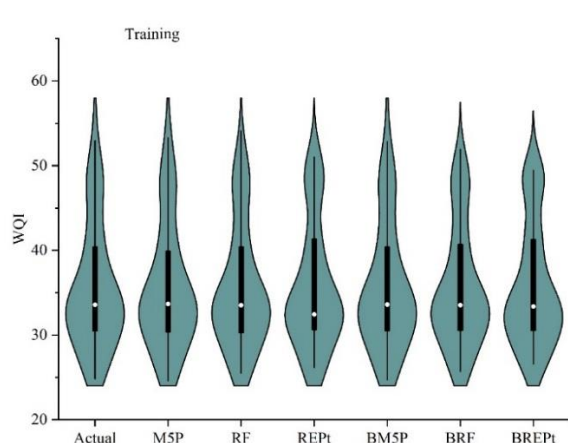
با مقادیر واقعی همخوانی دارد و در بخش صحت‌سنجی مقادیر پیش‌بینی مدل BM5P با مقادیر واقعی همخوانی بیشتری دارد. بنابراین با توجه به نتایج این نمودار عملکرد بسیار خوب مدل BM5P در پیش‌بینی WQI تایید می‌شود.



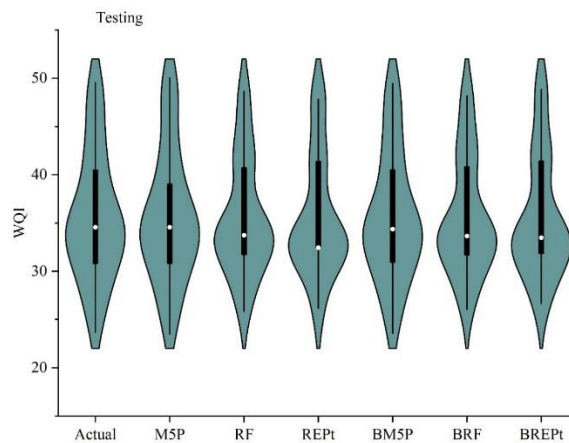
شکل ۷. نمودار تیلور در بخش صحت‌سنجی



شکل ۶. نمودار تیلور در بخش آموزش



شکل ۹. نمودار ویولین باکس نتایج مدل‌ها در بخش صحت‌سنجی



شکل ۸. نمودار ویولین باکس نتایج مدل‌ها در بخش آموزش

## ۴. بحث و نتیجه‌گیری

ارزیابی کیفیت آب رودخانه از مهم‌ترین عوامل برای بهبود و پیشرفت مدیریت منابع آبی می‌باشد. مدل‌های کیفیت آب با استفاده از شبیه‌سازی ساده و دقیق راه‌حل‌های بهبود کیفیت آب می‌توانند ابزارهای قدرتمندی در فرآیند مدیریت کیفیت آب مورد استفاده قرار گیرند. مدل‌های یادگیری در سال‌های اخیر برای مدل‌سازی داده‌های کیفیت آب استفاده‌های زیادی داشته‌اند (Zhang et al, 2012). در این پژوهش به مدل‌سازی داده‌های کیفیت آب با استفاده از مدل‌های RF، M5P، BM5P، BRF و BRepTree پرداخته شد.

معیارهای ارزیابی این تحقیق شامل MAE, CC و RMSE و همچنین نمودارهای تیلور و ویولین باکس می‌باشند. داده‌های استفاده شده در این تحقیق شامل داده‌های ورودی  $\text{Na}^+$ ,  $\text{Mg}^{2+}$ ,  $\text{Ca}^{2+}$ ,  $\text{SO}_4^{2-}$ ,  $\text{CL}^-$ ,  $\text{HCO}_3^-$ , PH, EC, TDS و  $\text{K}^+$  و همچنین داده‌های خروجی WQI می‌باشند. مدل‌سازی کیفیت آب با استفاده از مدل‌های ذکر شده در مجموع نتایج خوب و عملکرد مثبتی داشته‌اند که نشان می‌دهد این مدل‌ها برای پیش‌بینی کیفیت آب مناسب بوده‌اند. بین مدل‌های استفاده شده BM5P باتوجه به معیارهای ارزیابی در بخش صحت سنجی که به ترتیب برابر ۰/۹۹، ۰/۱۵ و ۰/۲۰ می‌باشد از عملکرد بهتری نسبت به سایر مدل‌های RF, M5P, BRF و BRepTree داشته است. همچنین با نتایج به دست آمده از نمودارهای تیلور و ویولین باکس، عملکرد مثبت مدل‌های استفاده شده در تحقیق به منظور مدل‌سازی شاخص کیفیت آب WQI و همچنین برتری مدل BM5P نسبت به سایر مدل‌ها تایید شده است. نتایج تحقیق نشان دهنده قدرت بالای الگوریتم‌های یادگیری در ارزیابی و پیش‌بینی شاخص کیفیت آب با استفاده از پارامترهای کیفی استفاده شده در این حوضه می‌باشد. یکی از محدودیت‌های این تحقیق تعداد سال‌های آماری تحقیق می‌باشد، بر این اساس می‌توان بیان کرد که اگر طول دوره آماری بیشتر از ۱۰ سال بود، توانایی این مدل‌ها در تخمین شاخص کیفیت آب هم بالاتر بود و با میزان خطای پایین‌تری شاخص کیفیت آب محاسبه می‌شد، هرچند نتایج این تحقیق با این دوره آماری هم بیان کننده دقت و توانایی بالای اینگونه مدل‌ها در تخمین شاخص کیفیت آب می‌باشد. (Mohammadinejad et al (2022 در بررسی کیفیت آب زیرزمینی با استفاده از شبکه عصبی به این نتیجه رسیدند که استفاده از مدل ابزار مناسبی برای شبیه‌سازی پارامترهای کیفی آب زیرزمینی در حجم بالا بدون نیاز به اندازه‌گیری و کارهای آزمایشگاهی می‌باشد در تحقیق انجام شده نیز نتایج نشان دهنده کارایی بالای مدل‌ها در ارزیابی کیفیت آب بوده است. (Eidi et al (2022 در بررسی شاخص کیفیت آب با استفاده از WQI به این نتیجه رسیدند که WQI با سایر روش‌های دیگر ارزیابی کیفیت آب همخوانی دارد و می‌توان از آن به عنوان یک شاخص مورد اطمینان برای بررسی تغییرات مکانی و زمانی کیفیت آب استفاده کرد که در نتایج تحقیق انجام شده نیز می‌توان به همین نتیجه دست یافت. (Verma et al (2021 در بررسی شاخص کیفیت آب به این نتیجه رسیدند که غلظت یون‌های کلسیم، منیزیم، بی کربنات و نمک‌های جامد محلول بیشتر از مقدار مجاز بوده و کیفیت آب منطقه پایین بوده است. (Gupta et al (2021 از الگوریتم‌های یادگیری ماشین برای ارزیابی WQI در هند استفاده کردند. نتایج نشان دهنده قدرت بالای این الگوریتم‌ها در ارزیابی کیفیت آب زیرزمینی بوده است. (Sihag et al (2021 به بررسی کیفیت آب با استفاده از تکنیک‌های هوش مصنوعی پرداختند. کارایی مدل‌های استفاده شده با استفاده از معیارهای ارزیابی NSE, MAE, RMSE, MAPE و CC بررسی شد که نتایج نشان دهنده عملکرد خوب این الگوریتم‌ها در ارزیابی کیفیت آب بوده است. (Mokhtar et al (2022 از ۳ مدل هوش مصنوعی و ۴ مدل رگرسیون چندگانه برای پیش‌بینی ۶ معیار کیفیت آب آبیاری استفاده کردند. نتایج نشان داد این مدل‌ها توانایی تصمیم‌گیری سریع در مورد کیفیت آب آبیاری را دارا می‌باشند. همچنین (Masaab et al (2023 به مدل‌سازی شاخص کیفیت آب با استفاده از مدل‌های هوش مصنوعی پرداختند. پارامترهای کیفی این تحقیق  $\text{Na}$ ,  $\text{Mg}^{2+}$ ,  $\text{Ca}^{2+}$ ,  $\text{NO}_3$ ,  $\text{SO}_4^{2-}$ , Cl, TH, TDS, EC, PH<sup>+</sup> و  $\text{NaCO}_3$  بوده است. نتایج عملکرد بالای مدل‌های هوش مصنوعی را در مدل‌سازی GWQI نشان داد که این نتایج با نتایج تحقیق انجام شده همخوانی دارند. نتایج مشابه را می‌توان در مطالعات (Saleem et al (2016, (Karuppannan and Kawo (2018, (Latha and Rao (2019 و (Jalili et al (2019 مشاهده کرد که با نتایج این تحقیق همخوانی دارد. در مجموع باتوجه به نتایج این تحقیق و همچنین بر اساس نتایج تحقیقات پیشین، می‌توان بیان کرد که تکنیک‌های هوش مصنوعی در ارزیابی و پیش‌بینی کیفیت آب سطحی از عملکرد خوب و قدرت بالایی برخوردار هستند و با توجه به میزان هزینه و زمان کم، می‌توان این مدل‌ها را برای تخمین میزان شاخص کیفیت آب (WQI) در بازه‌های زمانی مختلف و پارامترهای کیفی متفاوت، بکار برد. همچنین می‌توان از این شاخص جهت مدیریت بهینه کیفیت آب‌های سطحی و زیرزمینی در حوزه‌های آبخیز استفاده کرد. در نهایت بر اساس نتایج این تحقیق می‌توان بیان کرد که شاخص کیفیت آب (WQI) همچون سایر شاخص‌ها در بررسی کیفیت آب دارای عملکرد خوبی است.

## References

- Abdelhefaz, A., Abbas, M. H. H., Kenawy, M. H., Ewis, A. M., & Hamed, M. H. (2021). Evaluation of underground water quality for drinking and irrigation purposes in New Valley Governorate, Egypt. *Environmental Technology and Innovation*, 22(3): 101486.
- Aryen-Najad, R., Sarai Tabrizi, M., & Babazadeh, H. (2020). Modeling Water Quality of Rivers Using QUAL2Kw Model. *Journal of Environmental Science and Technology*, (7) 21: 1-13. (In persian).
- Babakhani, Z., Sarai Tabrizi, M., & Babazadeh, H. (2020). Determination of River Self-Purification Capacity Using Qual2kw Mode Case Study: divandare River. *Journal Ecohydrology*, (3)6: 673-684. (In Persian).
- Baldwin, R., Cave, M., & Lodge, M. (2011). Understanding Regulation: Theory, Strategy and Practice. 2nd ed. Oxford: Oxford University Press.
- Bonakdar, L., & Etemad Shahidi, A. (2011). Predicting wave run-up on rubble-mound structures using M5 model tree. *Ocean Engineering*, (38), 111-118.
- Breiman, L. (1996). Bagging predictors. *Machin Learning*, 24(2), 123–140. doi:10.1007/BF00058655.
- Das, B., Rathore, P., Roy, D., Chakraborty, D., Singh Jatav, R., Deepak Sethi, D., & Praveen Kumar, P. (2022) Comparison of bagging, boosting and stacking algorithms for surface soil moisture mapping using optical-thermal-microwave remote sensing synergies. *CATENA* 217: 106485.
- Deng, X., Ye, A., Zhong, J., Xu, D., Yang, W., Song, Zh., Zhang, Z., Guo, J., Wang, T., Tian, Y., Pan, H., Zhang, Zh., Wang, H., Wu, Ch., Shao, J., & Chen, X. (2022). Bagging-XGBoost algorithm based extreme weather identification and short-term load forecasting model. *Energy Reports*, 8: 8661-8674.
- Dogan, E., Sengorur, B., & Koklu, R. (2009). Modeling biological oxygen demand of the Melen River in Turkey using an artificial neural network technique. *Journal of Environmental Management*, 90 (2), 1229–1235.
- Duc, H., Nguyen, H. Q., Quang, N. X., Duy, H. N., & Thang, L. (2019). Spatio-temporal pattern of water quality in the Saigon-Dong Nai River system due to waste water pollution sources. *International Journal of River Basin Management*, 17:1-34.
- Eidi, M., & Omiri, F. (2022). Groundwater quality assessment of Dehroud and Tang Eram regions of Dashtestan using water quality index (WQI). *Journal of Renewable Natural Resources Research*, 12: 2(36), 141-153. (in persian).
- El-Rawy, M., Abdalla, F., & Negm, A.M. (2021). Groundwater Characterisation and Quality Assessment in Nubian Sandstone Aquifer, Kharga Oasis, Egypt. In *Groundwater in Egypt's Deserts*; Springer: Cham, Switzerland; pp. 177–199.
- El-Rawy, M., Batelaan, O., Alshehri, F., Almadani, S., Ahmed, M. S., & Elbeltagi, A. (2023). An Integrated GIS and Machine-Learning Technique for Groundwater Quality Assessment and Prediction in Southern Saudi Arabia. *Water*, 15, 2448. <https://doi.org/10.3390/w15132448>.
- El-Rawy, M., Ismail, E., & Abdalla, O. (2019). Assessment of groundwater quality using GIS, hydrogeochemistry, and factor statistical analysis in Qena Governorate, Egypt. *Desalination and Water Treatment*, 162, 14–29.
- Fathi, H., & El-Rawy, M. (2018). GIS-based evaluation of water quality index for groundwater resources nearby wastewater treatment plants, Egypt. *Pollution Research*, 37, 105–116.
- Gupta, A. N., Kumar, D., & Singh, A. (2021). Evaluation of Water Quality Based on a Machine Learning Algorithm and Water Quality Index for Mid Gangetic Region (South Bihar plain), India. *Journal of the Geological Society of India*, 97, 1063–1072.
- Hameed, M. M., Masood, A., Srivastava, A., Rahman, N. A., Razali, S. F. M., & Elbeltagi, A. (2024). Investigating a hybrid extreme learning machine coupled with Dingo Optimization Algorithm for modeling liquefaction triggering in sand-silt mixtures. *Scientific Reports* 14. <https://doi.org/10.1038/s41598-024-61059-6>.
- Jalili, M., Hosseini, M. S., Ehrampoush, M. H., Sarlak, M., Abbasi, F. & Fallahzadeh, R. A. (2019) Use of water quality index and spatial analysis to assess groundwater quality for drinking purpose in Ardakan, Iran. *Journal of Environmental Health and Sustainable Development*, 4(3): 834-842.
- Jian-Hua, W., Pei-Yue, L., & Hui, Q. (2011). Groundwater Quality in Jingyuan County, a Semi Humid Area in Northwest China. *Journal of Chemistry*, 8(2):787-793.
- Karimi, H., Rostamizad, Q., Moghadsifar, S. & Vakrim, A. (2023). The contribution of the two arms of Seyol and Qadh to the reduction of water quality of the Mime River; Determining the crisis points and providing solutions. *Soil and Water Modeling and Management*, 2(3), 79-93. (in Persian).

- Kawo, N.S., & Karuppannan, S. (2018). Groundwater quality assessment using water quality index and GIS technique in Modjo River Basin, central Ethiopia. *Journal of African Earth Sciences*, 147: 300-311.
- Khalili, R., Parvinnia, M., & Zali, A. (2020). Water quality assessment of Garmarood River using the national sanitation foundation water quality index (NSFWQI), river pollution index (RPI) and weighted arithmetic water quality index (WAWQI). *Environment and Water Engineering*, 6(3), 274-284.
- Kostka, G. (2016). Command without control: the case of China's environmental target system, *Regulation and Governance*, Wiley. 10(1): 58-74.
- Kouadri, S., Pande, C.B., Panneerselvam, B., Moharir, K. N., & Elbeltagi, A. (2022). Prediction of irrigation groundwater quality parameters using ANN, LSTM, and MLR models. *Environmental Science and Pollution Research*. 29, 21067-2109.
- Liang, J., Wang, Ch., Zhang, D., Xie, Y., Zeng, Y., Li, T., Zuo, Zh., Ren, J., & Zhao, Q. (2023). VSOLassoBag: a variable-selection oriented LASSO bagging algorithm for biomarker discovery in omic-based translational research, *Journal of Genetics and Genomics*, 50(3): 151-162.
- Lohani, B. N., & Saleemi, A. R. (1982). Recent developments of stochastic programming model for water quality management, *Water Supply Management*, 6: 511-520.
- Lu, Y., Song, S., Wang, R., Liu, Z., Meng, J., Sweetman, A. J., Jenkins, A., Ferrier, R. C., Li, H., Luo, W., & Wang, T. (2015). Impacts of soil and water pollution on food safety and health risks in China. *Environment International*, 77: 5-15.
- Mohammadinejad, S. A., & Eghderanjad, A. (2020). Modeling groundwater qualitative changes using optimized artificial neural network model from a case study (Zidoon Plain). *Environmental Health Research Quarterly*, 7(4), 311-322. (in persian).
- Mohammed, A. A. M., Szabo, N. P., & Szucs, P. (2022) Multivariate statistical and hydrochemical approaches for evaluation of groundwater quality in north Bahri city-Sudan. *Heliyon* 8(11): e11308. <https://doi.org/10.1016/j.heliyon.2022.e11308>.
- Mohammed, M. A., Khleel, N. A., Szabo, N. P., & Szucs, P. (2023). Modeling of groundwater quality index by using artificial intelligence algorithms in northern Khartoum State, Sudan. *Modeling Earth Systems and Environment*, 9(2), 2501-2516.
- Mokhtar, A., Elbeltagi, A., Gyasi-Agyei, Y., Al-Ansari, N., & Abdel-Fattah, M. K. (2022). Prediction of irrigation water quality indices based on machine learning and regression models. *Applied Water Science*, 12(76):1-14.
- Muangthong, S. (2015). Assessment of surface quality using multivariate statistical techniques. *Environmental Monitoring & Assessment*, 11(1), 25-73.
- Musaab, A. A. M., Nasraldeen, A. A. K., Norbert, P. S., & Peter, S. (2023). Modeling of groundwater quality index by using artificial intelligence algorithms in northern Khartoum State, Sudan. *Modeling Earth Systems and Environment*. 9:2501-2516. <https://doi.org/10.1007/s40808-022-01638-6>.
- Noori, R., Berndtson, R., Hosseinzadeh, M., Adamowski, J. F., & Rabiee Abyaneh, M. (2019). A critical review on the application of the National Sanitation Foundation Water Quality Index. *Environmental Pollution*, 244: 575-587.
- Quinlan, J. R. (1992). Learning with continuous classes. in Proceedings of the 5th Australian joint Conference on Artificial Intelligence. Hobart: 343-348.
- Rao, K. N. & Latha, P. S. (2019). Groundwater quality assessment using water quality index with a special focus on vulnerable tribal region of Eastern Ghats hard rock terrain, Southern India. *Arabian Journal of Geosciences*, 12(8): 1-16.
- Saleem, M., Hussain, A. & Mahmood, G. (2016). Analysis of groundwater quality using water quality index: A case study of greater Noida (Region), Uttar Pradesh (UP), India. *Cogent Engineering*, 3(1): 1237927.
- Sepahvand, A., Nazari Samani, A. A., Mohammadian, H., Ahmadi, H., & Feiz Nia, S. (2020). Seasonal variation of the solute and determine the solubility of limestone formations. *Iranian Journal of Watershed Management Science and Engineering*, 14(48), 21-32.
- Shi, P., Zhang, Y., Li, Z. B., Li, P., & Xu, G. C. (2017). Influence of land use and land cover patterns on seasonal water quality at multispatial scales. *Catena*, 151, 182-190.
- Shoemaker, C. M., Ervin, G. N., & Diorio, E. W. (2017). Interplay of water quality and vegetation in restored wetland plant assemblages from an agricultural landscape. *Ecological Engineering*, 108, 255-262.
- Simoes, F. S., Moriera, A. B., Bisinoti, M. C., Gimenez, S. M. N., & Yabe, M. J. S. (2008). Water quality index as a simple indicator of aquaculture effects on aquatic bodies. *Ecological indicators*. 8: 476-484.
- Singh, B., Sepahvand, A., Sihag, P., Singh, K., Prabha, Ch., Nag, A., Hassan, M., Vimal, S., & Kang, D. (2024). Development of soft computing-based models for forecasting water quality index of Lorestan Province, Iran, *Scientific Reports*, 14, Article number: 25980. <https://doi.org/10.1038/s41598-024-76894-w>



- Singh, B., Sihag, P., Singh, V. P., Sepahvand, A. & Singh, K. (2021). Soft computing technique-based prediction of water quality index. *Water Supply*, 21(8), 4015-4029. doi: 10.2166/ws.2021.157.
- Taylor, K. E. (2001). Summarizing multiple aspects of model performance in a single diagram. *Journal of Geophysical Research: Atmospheres* 106 (D7), 7183–7192.
- Ubah, J. I., Orakwe, L. C., Ogbu, K. N., Awu, J. I., Ahaneku, I. E., & Chukwuma, E. C. (2021). Forecasting water quality parameters using artificial neural network for irrigation purposes. *Scientific Reports*, 11, 24438.
- Vasant, W., Dipak, P., Aniket, M., Ranjitsinh, P., Shrikant, M., Nitin, D., Manesh, A., & Abhay, V. (2016). GIS and statistical approach to assess the groundwater quality of Nanded Tehsil, India. In: *Proceedings of First International Conference on Information and Communication Technology for Intelligent Systems: Volume 1*, Cham. Springer, pp. 409–417.
- Verma, P., Singh, P. K., Sinha, R. R. & Tiwari, A. K. (2020). Assessment of groundwater quality status by using water quality index (WQI) and geographic information system (GIS) approaches: A case study of the Bokaro district, India. *Applied Water Science*, 10(1): 1-16.
- Wagh, V. M., Panaskar, D. B., Muley, A. A., & Mukate, S. V. (2017). Groundwater suitability evaluation by CCME WQI model for Kadava river basin, Nashik, Maharashtra, India. *Modeling Earth Systems and Environment*, 3 (2), 557–565.
- Wang, Y., & Witten, I. H. (1997). Inducing model trees for continuous classes. *Proceedings of the Ninth European Conference on Machine Learning*. Prague, Czech Republic: Springer.
- Zare, A. G., Sadoddin, A., Sheikh, V., & Mahini, A. (2012). Long-term trend analysis of water quality variables for the Chehelchay River, *Iranian Water Research Journal*, 6(1(Serial number 10)): 155-165 (In Persian).
- Zhang, R., Qian, X., Li, H., Yuan, X., & Ye, R. (2012). Selection of optimal river water quality improvement programs using QUAL2K: A case study of Taihu Lake Basin, China. *Science of the Total Environment*, 431:278-285.
- Zielinski, M., Dopieralska, J., Belka, Z., Walczak, A., Siepak, M., & Jakubowicz, M. (2016). Sr isotope tracing of multiple water sources in a complex river system, Notec River, central Poland. *Science of the Total Environment*, 548, 307–316.