

Application of machine learning algorithms for estimating surface water quality in the Cham Anjir Watershed

Maryam Aryasadr¹, Dariush Rahimi^{2*}, Mehran Zand³ and Hadi Amiri⁴

¹ Ph.D Graduate of Climatology, University of Isfahan, Isfahan, Iran

² Professor of Department of Physical Geography, University of Isfahan, Isfahan, Iran

³ Associate Professor, Soil Conservation and Watershed Management Research Institute, Agricultural Research, Education and Extension Organization (AREEO), Tehran, Iran

⁴ Associate Professor of Department of Economics, University of Isfahan, Isfahan, Iran

Received: 11 December 2024

Accepted: 08 June 2025

Extended abstract

Introduction

The complexity of aquatic systems, their spatiotemporal variations, the high cost and time-consuming nature of traditional testing methods, and the need for continuous monitoring all contribute to the difficulty of monitoring and evaluating water quality. Therefore, artificial intelligence, machine learning, and deep learning approaches are useful for predicting water quality given the diverse parameters involved and are more cost-effective compared to traditional testing methods. There is no uniform algorithm that performs optimally for predicting water quality, and different algorithms exhibit superior performance in different contexts. The quality of rivers in the Zagros mountainous region has decreased due to the erosion of karst formations, their passage through residential areas, changes in land use, drought, and climate change. Some of these rivers, such as the Karkheh, Karun, and Dez, serve as the source of drinking water for a population of over 10 million people. Therefore, changes in water quality pose increased risks and threats to the drinking water supply of these settlements. Assessing the quality of water resources and developing a suitable model will play an effective role in managing the required water supply. The present study aims to identify the best machine learning algorithm for estimating the water quality parameters of the Cham Anjir watershed while facilitating ongoing monitoring of the resource.

Materials and methods

Data from 321 samples of discharge and water quality parameters at the Cham Anjir hydrometric station during the period 1969–2021 were used to select a suitable artificial intelligence model and to assess the quality of surface water resources in the Cham Anjir watershed. These samples included physical and chemical indicators: TDS, TH, SAR, Na, Mg, Ca, and Cl. Machine learning algorithms were employed to model the water quality of the Cham Anjir watershed. Through iterative testing of different models, the Support Vector Machine (SVM) and Classification and Regression Tree (CART) models were selected as the best performers. A correlation matrix was used to evaluate relationships among the variables, and based on these correlations, monthly discharge and monthly water quality indices—including TDS, TH, SAR, Na, Mg, Ca, Cl, EC, %Na, pH, HCO₃, and temporary hardness—were analyzed. A total of 321 monthly samples of the two key indices, TDS and TH, were studied over the statistical period of 1969–2021. To evaluate the accuracy of the machine learning algorithms in water quality modeling, the following performance indices were used: coefficient of determination (R^2), mean squared error (MSE), and mean absolute error (MAE).

Results and discussion

The p-value from trend tests confirmed the existence of an increasing trend at the 95% confidence level for the two variables TDS and TH (surface water quality parameters of the Cham Anjir watershed) since 1985. The average TDS in the first period (1969–1984) was 286.6 mg/L, and in the second period (1985–2021) it was 422.08 mg/L. The average TH in the first period (1969–1984) was 181.5 mg/L, and in the second period (1985–2021) it was 278.6 mg/L.

* Corresponding author: d.rahimi@geo.ui.ac.ir

The results of the correlation matrix indicated that TDS has a strong correlation with EC and TH. TH has a positive correlation with TDS, EC, HCO_3 , Ca, Cl, and Mg, and an inverse correlation with pH. The machine learning algorithms SVM and CART successfully captured the increasing trend for the two parameters of total hardness and total dissolved solids in the discharge of the Cham Anjir watershed. The SVM algorithm with the linear kernel exhibited the best performance. In addition, the root mean square error (RMSE) validation index also showed lower values for the SVM algorithm compared to the CART algorithm. Therefore, SVM is more accurate in predicting water quality indicators.

Conclusion

Machine learning algorithms are effective for water quality modeling. The results indicated an increasing trend in TDS and TH concentrations in the Cham Anjir watershed since 1985. SVM performed better than CART in predicting TDS and TH. The findings suggest that decreasing river flow, increasing water consumption, and karst geology influence TH and TDS concentrations in the quality of this watershed. Water quality monitoring is essential for water resource management. For future estimates of water quality in this context, the SVM algorithm is recommended.

Keywords: Decision tree, Support vector machine, Total hardness, Validation, Water quality

Cite this article: Aryasadr, M., Rahimi, D., Zand, M., Amiri, H., 2026. Application of machine learning algorithms for estimating surface water quality in the Cham Anjir Watershed. *Water. Eng. Manag.* 17(4), 553-570.

© 2026, The Author(s). Published by Soil Conservation and Watershed Management Research Institute (SCWMRI). This is an open-access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0>)



تخمین کیفیت آب سطحی و ارزیابی الگوریتم‌های یادگیری ماشین در آبخیز چم انجیر

مریم آریاصدر^۱، داریوش رحیمی^{۲*}، مهران زند^۳ و هادی امیری^۴

^۱ دانش‌آموخته دکتری آب و هواشناسی، دانشگاه اصفهان، اصفهان، ایران

^۲ استاد هیدرواقلیم، دانشکده علوم جغرافیایی و برنامه‌ریزی، دانشگاه اصفهان، اصفهان، ایران

^۳ دانشیار پژوهشکده حفاظت خاک و آبخیزداری، سازمان تحقیقات، آموزش و ترویج کشاورزی، تهران، ایران

^۴ دانشیار دانشکده اقتصاد، دانشگاه اصفهان، اصفهان، ایران

تاریخ پذیرش: ۱۴۰۴/۰۳/۱۸

تاریخ دریافت: ۱۴۰۳/۱۰/۱۱

چکیده مبسوط

مقدمه

کیفیت آب مصرفی در بخش‌های شرب و کشاورزی از شاخص‌های مهم توسعه پایدار است. فعالیت‌های کشاورزی و صنعتی انسان منجر به کاهش کیفیت آب شده است. از سوی دیگر، محدودیت منابع آب شیرین و با کیفیت مناسب از دیگر محدودیت‌های منابع آب است. پیچیدگی سامانه‌های آبی، تغییرات آن در طول زمان و مکان، افزایش هزینه و زمان بر بودن آزمایش کیفیت آب و نیاز به پایش مداوم بر دشواری پایش و ارزیابی کیفیت آب اثرگذار است. استفاده از الگوریتم‌های هوش مصنوعی، یادگیری ماشین و یادگیری عمیق، برای پیش‌بینی کیفیت آب به‌دلیل پارامترهای متنوع، مفید بوده، در مقایسه با روش‌های آزمایش سنتی، مقرون به صرفه است. هیچ الگوریتم یکنواختی برای پیش‌بینی کیفیت آب وجود ندارد. در این پژوهش تلاش می‌شود، کارایی این الگوریتم در تخمین عناصر فیزیکی و شیمیایی مورد ارزیابی قرار گیرد. در دهه‌های اخیر کیفیت رودخانه‌های رشته کوه زاگرس به‌دلیل فرسایش سازندهای کارستیک، عبور از مناطق سکونتگاهی، تغییرات کاربری اراضی، خشکسالی و تغییر اقلیم کاهش یافته است. برخی از این رودخانه‌ها مانند کرخه، کارون و دز منبع تأمین آب شرب جمعیتی بیش از ۱۰ میلیون نفر هستند. بنابراین تغییر در کیفیت آب منجر به افزایش خطر و تهدید آب شرب این سکونتگاه‌ها می‌شود. بنابراین بررسی کیفیت منابع آب و ارائه مدل مناسب آن نقش مؤثری در مدیریت تأمین آب مورد نیاز خواهد داشت. پژوهش حاضر، بر آن است، ضمن پایش کیفیت آب آبخیز چم انجیر، بهترین الگوریتم یادگیری ماشین در تخمین آن را ارزیابی کند.

مواد و روش‌ها

حوزه آبخیز چم انجیر واقع در استان لرستان و در جنوب شرق حوزه آبخیز کرخه است و بین طول‌های $48^{\circ} 40'$ تا $48^{\circ} 48'$ شرقی و عرض‌های $33^{\circ} 10'$ تا $33^{\circ} 36'$ قرار دارد. سازندهای زمین‌شناسی یکی از منابع مؤثر در غلظت متغیرهای شیمیایی منابع آب سطحی به‌ویژه در مناطق کارستیک است. این سازند در تغییر کیفیت آبخیز چم انجیر بسیار مؤثر است. سه زیرحوضه هم‌جوار الشتر، خرم‌آباد و بیرانشهر تحت تأثیر سنگ‌شناسی منطقه، میزان ترکیب شیمیایی سنگ‌های آهک و دولومیت، ۸۹-۸۷ درصد و انحلال گچ ۱۱-۱۳ درصد و میزان انحلال کربنات کارست (بین ۰/۰۱ تا ۰/۴۰ میلی‌متر) است. آهکی بودن جنس زمین و وجود سنگ‌های نفوذی در زیرحوضه‌های بالادست رودخانه دز منجر به تغییر

* مسئول مکاتبات: d.rahimi@geo.ui.ac.ir

متغیرهای کلسیم، پتاسیم، EC و TDS شده است. همچنین خشکسالی اقلیمی و هیدرولوژی کیفیت منابع آب را تحت تأثیر قرار داده، باعث افزایش درصد متغیرهای شیمیایی آب شده است. برای انتخاب مدل مناسب هوش مصنوعی و بررسی کیفیت منابع آب سطحی آبخیز چم انجیر، از ۳۲۱ نمونه از متغیر دبی چم انجیر در دوره ۲۰۲۱-۱۹۶۹ استفاده شد. این نمونه‌ها، از متغیرهای شیمیایی Na، SAR، TH، TDS، Cl، Ca و Mg ایستگاه هیدرومتریک چم انجیر انتخاب شده‌اند. برای مدل‌سازی کیفیت آب سطحی چم انجیر از الگوریتم‌های یادگیری ماشین استفاده شد. در پژوهش حاضر، متغیرهای شیمیایی با الگوریتم‌های مختلف یادگیری ماشین بررسی شده، اعتبار و دقت مقادیر برآورد شده با مقادیر مشاهداتی متغیرهای شیمیایی آب با شاخص‌های r ، MSE و MAE ارزیابی شده است. بر اساس تکرار و آزمون مدل‌های مختلف الگوریتم یادگیری ماشین، الگوریتم‌های ماشین بردار پشتیبان و رگرسیون درختی به‌عنوان بهترین مدل‌ها انتخاب شده‌اند. برای ارزیابی و ارتباط‌سنجی بین متغیرها از ماتریس همبستگی استفاده و بر اساس همبستگی، از میان دبی رودخانه و متغیرهای ماهانه کیفیت آب، HCO_3 ، pH، Na، EC، Cl، Ca، Mg، TH، TDS و سختی موقت، ۳۲۱ نمونه ماهانه از دو متغیر TDS و TH در دوره آماری ۲۰۲۱-۱۹۶۹ بررسی شد. برای بررسی دقت الگوریتم‌های یادگیری ماشین در مدل‌سازی کیفیت آب، از شاخص‌های همبستگی، میانگین مربعات خطاها و میانگین خطای مطلق استفاده شده است.

نتایج و بحث

نتایج P-test وجود روند افزایشی در سطح اطمینان ۹۵ درصد دو متغیر TDS و TH از سال ۱۹۸۵ را تأیید کرد. متوسط TDS در دوره اول (۱۹۸۴-۱۹۶۹)، $286/6$ میلی‌گرم در لیتر و در دوره دوم (۲۰۲۱-۱۹۸۵) $422/08$ میلی‌گرم در لیتر و متوسط TH در دوره اول (۱۹۸۴-۱۹۶۹) $181/5$ میلی‌گرم در لیتر و در دوره دوم (۲۰۲۱-۱۹۸۵) $278/6$ میلی‌گرم در لیتر است. نتایج ماتریس همبستگی نشان داد که TDS همبستگی قوی با EC و TH نشان می‌دهد. TH همبستگی مثبت با متغیرهای شیمیایی TDS، EC، HCO_3 ، Ca، Cl و Mg و همبستگی معکوس با pH نشان می‌دهد. الگوریتم‌های یادگیری ماشین SVM و CART روند افزایشی را برای دو متغیر سختی کل و غلظت مواد محلول آب در دبی آبخیز چم انجیر تخمین می‌زنند. الگوریتم SVM با مدل KernalLinear بهترین عملکرد را داشت. علاوه بر این، شاخص صحت‌سنجی RMSE نیز الگوریتم SVM را با مقادیر پایین‌تری نسبت به الگوریتم CART نشان داد. بنابراین، SVM در پیش‌بینی متغیرهای کیفیت آب دقیق‌تر است.

نتیجه‌گیری

الگوریتم‌های یادگیری ماشین برای مدل‌سازی کیفیت آب مؤثر هستند. نتایج نشان داد، در آبخیز چم انجیر، روند رو به رشد در غلظت‌های TDS و TH از سال ۱۹۸۵ وجود دارد. SVM در پیش‌بینی TDS و TH از CART بهتر عمل کرد. یافته‌ها نشان داد کاهش دبی رودخانه، افزایش مصرف آب و زمین‌شناسی کارست در میزان غلظت TH و TDS تأثیر دارند. پایش کیفیت آب برای مدیریت منابع آب ضروری است. برای تخمین‌های آینده کیفیت آب در این مطالعه، الگوریتم SVM توصیه می‌شود.

واژه‌های کلیدی: درخت تصمیم، سختی آب، صحت‌سنجی، کیفیت آب، ماشین بردار ویژه

مقدمه

چالش‌های مالی، با چالش‌های فنی و نهادی بزرگی در مدیریت پایدار کمیت و کیفیت منابع آب روبرو هستند (Okafor et al., 2024). خشکسالی و سیلاب و عوامل اکولوژیکی و انسانی آسیب‌پذیری منابع آب را افزایش داده است. مطالعات، ادامه روند فعلی را در آینده نشان

سالانه بیش از پنج میلیون نفر بر اثر بیماری‌های مرتبط با آب جان خود را از دست می‌دهند. مشکل کمبود آب آشامیدنی سالم به یک تنش جهانی تبدیل شده است. کشورهای در حال توسعه، همزمان با

منجر به صرفه‌جویی در هزینه و زمان شده، و به مدیریت بهتر منابع آب در دوره‌های آبی کمک می‌کند.

Al-Adhaileh et al., (2022) با کمک هوش مصنوعی و مدل‌های SES ANFIS و SES BiLSTM و Bui et al., (2020)، الگوریتم ترکیبی BA-TR و Ahmed et al., (2019)، الگوریتم تقویت گرادیان، Ditsas and Trigka, (2023)، الگوریتم Stacking را پس از SMOTE به‌عنوان بهترین الگوریتم؛ Dezfooli et al., (2018)، با استفاده از عملکرد سه مدل PNN، KNN، SVM و در نهایت با مدل PNN به‌عنوان بهترین مدل، به طبقه‌بندی و پیش‌بینی کیفیت آب با دقت بالا، در کمینه زمان و با کمینه هزینه نظارت پرداختند. یافته‌ها نشان می‌دهد که رویکرد پیشنهادی (استفاده از مدل‌های یادگیری ماشین) یک روش قابل اعتماد و قابل تفسیر برای پیش‌بینی کیفیت آب ارائه می‌کند که می‌تواند به‌طور قابل‌توجهی برای متخصصان و تصمیم‌گیرندگان آب مفید باشد. عملکرد مدل‌ها با استفاده از معیارهایی مانند MAE، RMSE، R^2 و R ارزیابی شد. نتایج دقت پیش‌بینی بالایی را با الگوریتم XGBoost² نشان داد. درحالی‌که RF و LightGBM عملکرد خوبی داشت که نشان‌دهنده کاربرد ML در این زمینه است.

پیش‌بینی‌های مدل استفاده شده در تحقیق نشان داد، COD و BOD تأثیرگذارترین عوامل در تعیین کیفیت آب بوده‌اند (Randika et al., 2024). بر اساس نتایج Nallakaruppan et al., (2024) از میان مدل‌های درخت تصمیم (DT)، ماشین بردار ویژه (SVM)، جنگل‌های تصادفی (RF)، الگوریتم RF را با دقت ۰/۹۹۹ برای توضیح مهم‌ترین پارامترهای مؤثر در قابل شرب بودن آب و تخمین ناخالصی‌ها به‌کار بردند. Kaur et al., (2024) در تحقیق خود نشان دادند که رگرسیون تقویت گرادیان (GBoost) از روش‌های دیگر بهتر عمل می‌کند و به دقت بالایی بر اساس سنج‌های MAE، MSE و RMSE به پیش‌بینی کیفیت آب دست می‌یابد. دما و رسانایی مؤثرترین عوامل در کیفیت آب بوده‌اند. نتایج پژوهش Asadollah et al., (2021)، نشان داد که الگوریتم درخت تصمیم اضافی (ETR) پیش‌بینی

می‌دهند (Barbieri et al., 2000؛ Murdoch et al., 2023). تحت تأثیر تغییر اقلیم، کیفیت منابع آب در رودخانه ارس کاهش یافته و مقدار آن در شرایط بد (مقدار شاخص WQI مساوی ۱۵/۶۹) قرار دارد (Farid Giglou et al., 2022). زمین‌شناسی از طریق هوازدگی سنگ و انحلال کربنات بر کیفیت آب تأثیر می‌گذارد. دبی رودخانه بر تغذیه و کیفیت آب‌های زیرزمینی تأثیر می‌گذارد، درحالی‌که کاهش بارندگی منجر به افزایش نفوذ آب شور می‌شود و کیفیت آب را در دره رودخانه‌های کوهستانی بدتر می‌کند (Li et al., 2023).

بنابراین برای شناخت بهتر کیفیت آب، استفاده از مدل‌هایی که بتواند خروجی‌هایی با دقت مناسب ارائه دهند، الزامی است. مدل‌های مبتنی بر هوش مصنوعی مانند شبکه عصبی، یادگیری ماشین و یادگیری عمیق ازجمله این مدل‌ها هستند. این مدل‌ها در شناسایی منابع آلودگی آب، پیش‌بینی تأثیر فعالیت‌های انسانی، نظارت و طبقه‌بندی پارامترهای کیفیت آب استفاده می‌شوند. تنوع پارامترهای مؤثر بر کیفیت آب مانند زمین‌شناسی، کاربری زمین، اقلیم، ترکیب متغیرهای شیمیایی آب و دبی گویا این است که مدل‌های هوش مصنوعی می‌تواند در ارائه مدل مناسب برای پیش‌بینی کیفیت آب مفید باشند (Fernández del Castillo et al., 2022؛ Cojbasic et al., 2023؛ Yan et al., 2024؛ Khan and see, 2016؛ Omambia et al., 2022).

علم مهندسی آب ازجمله علوم است که علم داده در آن نقش مهمی دارد، این علم در استخراج و تحلیل اطلاعات و فهم روابط بین داده‌های بزرگ مؤثر است این علم به یادگیری ماشین متکی است. با توجه به شکل‌های مختلف منابع آب که منجر به ویژگی‌های متفاوت آب شده است، بر پیچیدگی بیشتر بررسی کیفیت آب می‌افزاید؛ بنابراین، مدل‌های مختلف یادگیری ماشین به بررسی کیفیت آب سرعت می‌بخشند (Delbaz and Ebrahimian, 2024). با توجه به هزینه‌های مربوط به آزمایش کیفیت آب، وقت گیر بودن پایش و نظارت بر کیفیت منابع آب و نیز با توجه به حجم بالای اطلاعات، مدل‌سازی و پیش‌بینی کیفیت منابع آب با استفاده از الگوریتم‌های یادگیری ماشین^۱

² Gradient boosting¹ Machine learning

و طبقه‌بندی گرادیان (با دقت ۱۰۰ درصد)، El-Rawy et al., (2023) بین شش الگوریتم هوش مصنوعی، رگرسیون گاوسی GPR، مدل رگرسیون و ANN را با بالاترین همبستگی و خطای کمتر نسبت به سایر الگوریتم‌ها، Hamada et al., (2024) بین چهار الگوریتم مختلف یادگیری ماشین (GPR، RF، XGB و LightGBM)، مدل GPR بهترین عملکرد پیش‌بینی کننده را برای پارامترهای پساب TSS، COD و BOD در مقایسه با سایر مدل‌های یادگیری ماشین تشخیص دادند.

Beiranvand et al., (2023) بین پنج الگوریتم یادگیری ماشین الگوریتم‌های GP-PUK و GP-RBF تکنیک‌هایی بهینه برای مدل‌سازی انحلال رسوبات رودخانه و Kouadri et al., (2021) با الگوریتم رگرسیون درختی اضافی به پیش‌بینی متغیر کیفیت آب با دقت بالا و کمترین زمان پرداختند. Ghaemi et al., (2024) با استفاده از روش تلفیقی حاصل از GEP و ANN با درصد خطا RMSE کمتر از روش مستقل GEP و ANN، به مدل‌سازی مواد جامد محلول در رودخانه تجن پرداختند (به ترتیب خطای مدل‌ها ۴۸/۳۵ و ۱۴/۰۲ درصد کمتر است).

حدود ۲۵ درصد از حجم بارش سالانه ایران به صورت رواناب است. دو منطقه کوهستانی البرز (در شمال، غرب به شرق) و زاگرس (شمال غربی-جنوب شرقی) مهم‌ترین کانون‌های آبریز رودخانه‌های ایران هستند. کیفیت رودخانه‌های ناحیه کوهستانی زاگرس به دلیل فرسایش سازندهای کارستیک، عبور از مناطق سکونتگاهی، تغییرات کاربری اراضی، خشکسالی و تغییر اقلیم کاهش یافته است. برخی از این رودخانه‌ها مانند کرخه، کارون و دز منبع تأمین آب شرب جمعیتی بیش از ۱۰ میلیون نفر هستند. بنابراین تغییر در کیفیت آب منجر به افزایش خطر و تهدید آب شرب این سکونتگاه‌ها می‌شود. بنابراین بررسی کیفیت منابع آب و ارائه مدل مناسب آن نقش مؤثری در مدیریت تأمین آب مورد نیاز خواهد داشت.

در این پژوهش تلاش بر آن است که مدل مناسبی برای تحلیل کیفیت منابع آب سطحی در غرب و جنوب

های دقیق‌تری را برای شاخص کیفیت آب (WQI) در رودخانه هنگ‌کینگ طی هر دو مرحله آموزش و آزمایش ایجاد می‌کند. نتایج تحقیقات Sayahi et al., (2024) نشان داد، الگوریتم جنگل تصادفی (RF) در پیش‌بینی کیفیت آب رودخانه مارون بهترین عملکرد را داشته، نیرات در کاهش کیفیت آب مهم‌ترین متغیر است.

به توجه به حجم داده و خطای مدل هنگام پردازش و طبقه‌بندی متغیرهای کیفیت آب، الگوریتم‌های مختلف یادگیری ماشین عملکرد متفاوتی دارند. شبکه عصبی مصنوعی ساده و ترکیبی ابزار مناسب برای پیش‌بینی و مدل‌سازی پارامترهای کیفیت آب و تغییرات سطح آب است و بهترین عملکرد را در حفظ کیفیت آب و مدیریت منابع آب دارد (Rahimi and Oğuz, 2017; Hasheminasab, 2017; Sahoo et al., 2017; Eze et al., 2023; and Ertugrul, 2023).

در پژوهشی Khullar and Singh, (2021) مقایسه نه الگوریتم یادگیری ماشینی در مدل‌سازی کیفیت آب پرداختند. طبق نتایج، مدل‌های هوش مصنوعی و ترکیب ماشین‌های بردار پشتیبان و جنگل تصادفی روش مناسب برای بهبود مدیریت پایدار کیفیت و کمیت منابع آب با دقت بالا و هزینه کم هستند (Qie et al., 2023; Nuanmeesri et al., 2023; Zaresefat and Derakhshani, 2023).

علاوه بر آن یافته‌ها نشان داد ماشین‌های بردار پشتیبان (SVMs) مانند بردار پشتیبان رگرسیون^۱ (Granata et al., 2017; Hussein et al., 2020) و ماشین بردار ویژه^۲ (Haghiabi et al., 2018; Arabgol et al., 2016) با دقت بیشتری متغیرهای کیفیت آب را مدل‌سازی و پیش‌بینی می‌کنند. نتایج مطالعات و دقت سنجی مدل‌ها نشان داد الگوریتم یکسانی در مدل‌ها برای پیش‌بینی کیفیت آب وجود ندارد. به عنوان مثال Sahour et al., (2023) الگوریتم جنگل تصادفی، Shirazi et al., (2023) الگوریتم M5tree، Goodarzi et al., (2023) الگوریتم MARS را دقیق‌تر از سایر مدل‌ها معرفی کرده‌اند. (Murphy and Chanat, 2023) مدل رگرسیون، Abuzir and Abuzir, (2022) الگوریتم بیز ساده و پرسپترون چند لایه و Khan et al., (2021) الگوریتم رگرسیون مؤلفه‌های اصلی (با دقت ۹۵ درصد)

² Support Vector Machine (SVM)

¹ Support Vector Regression (SVR)

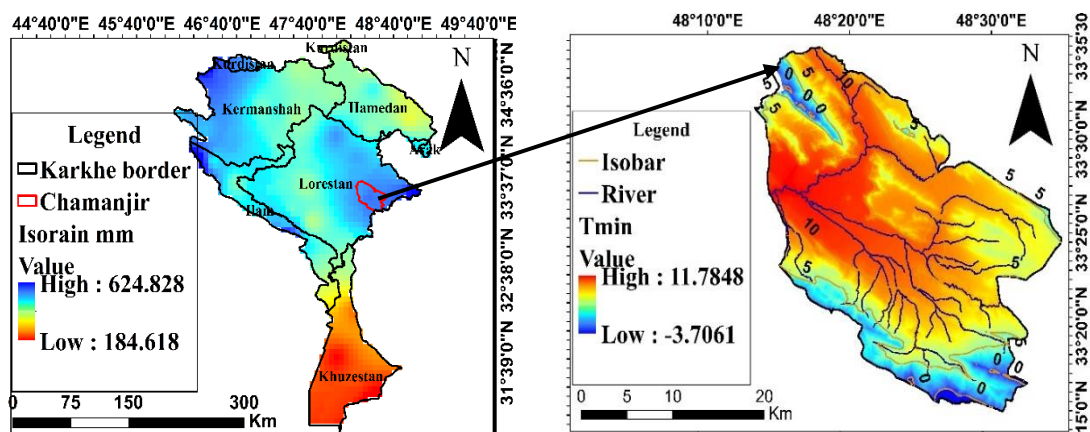
۹/۲ مترمکعب بر ثانیه است. اوج آبدهی در ماه فروردین ۲۲/۳ مترمکعب بر ثانیه و کمینه آن در ماه شهریور ۳/۶ مترمکعب بر ثانیه بوده است.

دی ماه، آغاز افزایش حجم دبی و ماه‌های بهمن، اسفند، فروردین و اردیبهشت اوج آن است (شکل ۲). در ماه‌های گرم سال علاوه بر نبود بارش، مصرف آب در تمام بخش‌ها افزایشی بوده، آبدهی ایستگاه هیدرومتری، روند کاهشی را نشان می‌دهد. به‌طوری‌که یافته‌ها نشان می‌دهند که با رخداد تغییر اقلیم در این منطقه، طبقه کیفیت آب تغییر کرده، کیفیت آب شرب و کشاورزی از خوب به قابل قبول و از طبقه C_2S_1 به طبقه C_3S_1 تغییر یافته است (Soleimani Motlagh et al., 2016).

غرب ایران ارائه شود. بدین منظور متغیرهای شیمیایی آب سطحی در یکی از زیرحوضه‌های رودخانه کرخه برای کالیبره کردن مدل مناسب در حوزه هوش مصنوعی برای تحلیل کیفیت آب انتخاب شده است.

مواد و روش‌ها

منطقه مورد مطالعه: آبخیز چم انجیر در غرب کشور و جنوب شرق حوزه آبخیز کرخه واقع شده است. این حوضه بین طول‌های $48^{\circ} 40'$ تا $48^{\circ} 10'$ شرقی و عرض‌های $33^{\circ} 10'$ تا $33^{\circ} 36'$ شمالی در بخش میانی رشته کوه زاگرس قرار دارد (شکل ۱). بارش سالانه $549/3$ میلی‌متر، دمای سالانه $17/1$ سلسیوس است. دبی سالانه چم انجیر در دوره ۱۹۹۱-۲۰۲۱ به میزان

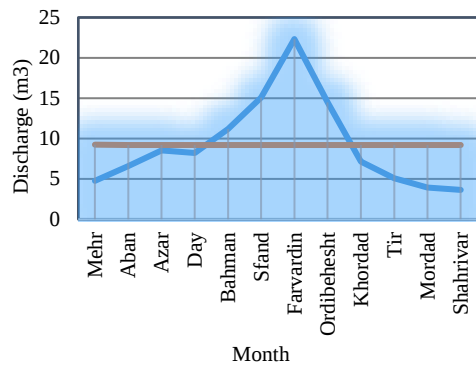


شکل ۱- موقعیت جغرافیایی آبخیز چم انجیر
Fig. 1. Geographical location of Cham Anjir Watershed

انحلال کربنات کارست (بین $0/1$ تا $0/40$ میلی‌متر) است (Sepahvand et al., 2021).

ساختار زمین شناسی، رودخانه دز را نیز تحت تأثیر قرار داده است. خشکسالی و عبور رودخانه از مسیر سنگ‌های نفوذی مانند گرانیت و گرانودیوریت منجر به افزایش کلسیم و پتاسیم و اثرات منفی بر شاخص‌های کیفی منابع آب رودخانه تیره شده است (Farhadinejad et al., 2024). با توجه به کارست بودن منطقه، افزایش دبی و بارش بر انحلال‌پذیری و افت کیفیت آب حوضه چم انجیر مؤثر است (Beiranvand et al., 2023).

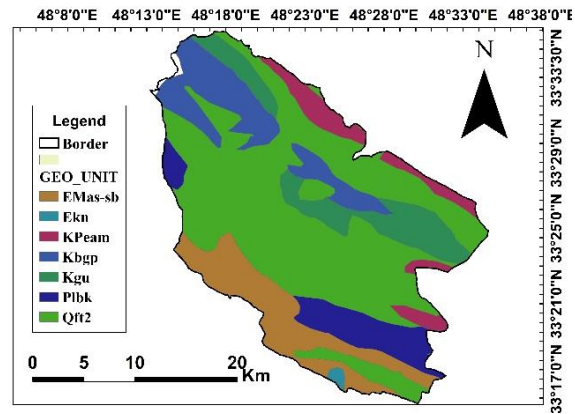
سازندهای زمین‌شناسی یکی از منابع مؤثر در غلظت متغیرهای شیمیایی منابع آب سطحی به‌ویژه در مناطق کارستیک است. حوضه چم انجیر از سازندهای ایلام- سورگاه، سروک، امیران، گورپی، کشکان، آسماری، گچساران، شهبازان، کشکان و بختیاری تشکیل شده است. سازند گچساران (گچ و نمک)، به شدت فرسایش‌پذیر بوده، منجر به تغییر سختی آب (TH) و افزایش بار محلول آب (TDS) می‌شود. این سازند در تغییر کیفیت آب حوزه آبخیز چم انجیر بسیار مؤثر است (شکل ۳). سه زیرحوضه هم‌جوار الشتر، خرم‌آباد و بیرانشهر، تحت تأثیر سنگ‌شناسی منطقه، میزان ترکیب شیمیایی سنگ‌های آهک و دولومیت، ۸۷-۸۹ درصد و انحلال گچ ۱۱-۱۳ درصد و میزان



شکل ۲- هیدروگراف میانگین بلندمدت آبخیز چم

انجیر (۱۹۹۱-۲۰۲۱)

Fig. 2. Monthly long-term hydrograph of Cham Anjir Watershed during 1991-2021



شکل ۳- زمین‌شناسی آبخیز چم انجیر

Fig. 3. Geology of Cham Anjir Watershed

برآورد شده با مقادیر مشاهداتی متغیرهای شیمیایی آب با شاخص‌های صحت‌سنجی (در ادامه تشریح می‌شوند) ارزیابی شده است. طبق صحت‌سنجی الگوریتم‌های یادگیری ماشین، الگوریتم‌های ماشین بردار پشتیبان و رگرسیون درختی به‌عنوان بهترین مدل‌ها انتخاب شده‌اند.

ماشین بردار پشتیبان (SVM): این الگوریتم، نوع پیشرفته‌ای از تحلیل رگرسیون است که در رسم خطوط مرزی بین دسته‌ها عملکرد بهتری دارد. غیرخطی بودن داده‌ها از کرنل‌های ماشین بردار پشتیبان؛ هسته لیرینر (رابطه ۲)، سیگموئید (رابطه ۳)، تابع پایه شعاعی (RBF) (رابطه ۴)، چندجمله‌ای (رابطه ۵) استفاده می‌شود (جدول ۱). به توجه به نتایج ارزیابی مدل‌ها در این پژوهش، بهترین برازش در مدل کرنل Linear حاصل شد.

داده‌ها: برای انتخاب مدل مناسب هوش مصنوعی و بررسی کیفیت منابع آب سطحی آبخیز چم انجیر، از ۳۲۱ نمونه از متغیرهای دبی چم انجیر در دوره ۲۰۲۱-۱۹۶۹ استفاده شد. این نمونه‌ها، از متغیرهای شیمیایی Na، SAR، TH، TDS، Cl، Ca و Mg ایستگاه هیدرومتریک چم انجیر انتخاب شده‌اند.

روش: روش تحقیق در این پژوهش، روش تجربی است. با آزمون‌های ناپارامتری روند تغییرات متغیرهای TH و TDS مشخص شد (آزمون پتیت) (رابطه ۱).

$$K_T \max |U_{t,T}|, \quad U_{t,T} = \sum_{i=1}^t \sum_{j=t+1}^T \text{sgn}(X_i - X_j) \quad (1)$$

$$p \approx 2 \exp\left(\frac{-6 K_T^2}{T^3 + T^2}\right)$$

برای مدل‌سازی کیفیت آب سطحی چم انجیر از الگوریتم‌های یادگیری ماشین استفاده شد. در پژوهش حاضر، متغیرهای شیمیایی با الگوریتم‌های مختلف یادگیری ماشین بررسی شده، اعتبار و دقت مقادیر

جدول ۱- روابط کرنل‌های ماشین بردار ویژه

Table 1. Kernel relationships of Support vector machine

Linear	sigmoid	RBF	Polynomial
$k = x - x' ^2$ (2)	$k = \theta h(\gamma x - x' + r)$ (3)	$e^{-(\gamma x - x')^2}$ (4)	$(\gamma x - x' + r)^d$ (5)

$\gamma = \text{gamma}$

الگوریتم رگرسیون و درخت طبقه‌بندی CART^۲: الگوریتم رگرسیون و درخت طبقه‌بندی یک ساختار درختی مانند نمودار جریان است که از روش انشعاب برای نشان دادن هر نتیجه ممکن یک تصمیم استفاده می‌کند. در این الگوریتم، جمعیت به دو یا چند مجموعه همگن تقسیم می‌شوند. تمایز بین دو مجموعه، بر اساس

گاما، ضریب کرنل برای rbf و poly و sigmoid است. هرچه مقدار گاما بیشتر باشد، الگوریتم تلاش می‌کند برازش را دقیقاً بر اساس مجموعه داده‌های تمرینی انجام دهد و این امر موجب تعمیم‌یافتن خطا و وقوع مشکل بیش برازش^۱ می‌شود.

² Classification and regression trees

¹ Over-Fitting

برای ارزیابی و ارتباط‌سنجی بین متغیرها از ماتریس همبستگی استفاده و بر اساس همبستگی، از میان دبی رودخانه و متغیرهای ماهانه کیفیت آب، pH ، HCO_3 ، Na ، EC ، Cl ، Ca ، Mg ، TH ، TDS و سختی موقت، ۳۲۱ نمونه ماهانه از دو متغیر TDS و TH در دوره آماری ۲۰۲۱-۱۹۶۹ بررسی شد. برای بررسی دقت الگوریتم‌های یادگیری ماشین در مدل‌سازی کیفیت آب، از شاخص‌های همبستگی، مربعات خطاها، میانگین خطای مطلق^۲ (رابطه‌های ۹ تا ۱۱) استفاده شده است.

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \cdot \sum_{i=1}^n (y_i - \bar{y})^2}} \quad (9)$$

$$MSE = \frac{\sum_{i=1}^n (Y_i - X_i)^2}{n} \quad (10)$$

$$MAE = \frac{\sum_{i=1}^n |Y_i - X_i|}{n} \quad (11)$$

در رابطه‌های فوق، X_i مقادیر مشاهداتی، Y_i مقادیر برآورد شده و n تعداد داده‌ها است. براین اساس، رابطه (۷) رابطه ضریب همبستگی است هرچه عدد محاسبه شده نزدیک به یک باشد، نشان‌دهنده ارتباط قوی بین عناصر مورد آزمون است (Hashminasab et al., 2022). مدل مفهومی تحقیق در ادامه ارائه شده است (شکل ۴).

مهم‌ترین متغیرهای مستقل تا حد امکان انجام می‌شود. درختان تصمیم می‌توانند داده‌های با ابعاد بالا را مدیریت کنند. بهترین تقسیم در هر گره بر اساس معیارهایی مانند ناخالصی Gini یا افزایش اطلاعات کار می‌کند (Lakshmi et al., 2013).

الگوریتم درخت طبقه‌بندی و رگرسیون به صورت رابطه (۶) است.

$$G(Q_m, t_m) = \frac{n_m^{Left}}{n_m} H(Q_m^{Left}(t_m)) + \frac{n_m^{Right}}{n_m} H(Q_m^{Right}(t_m)) \quad (6)$$

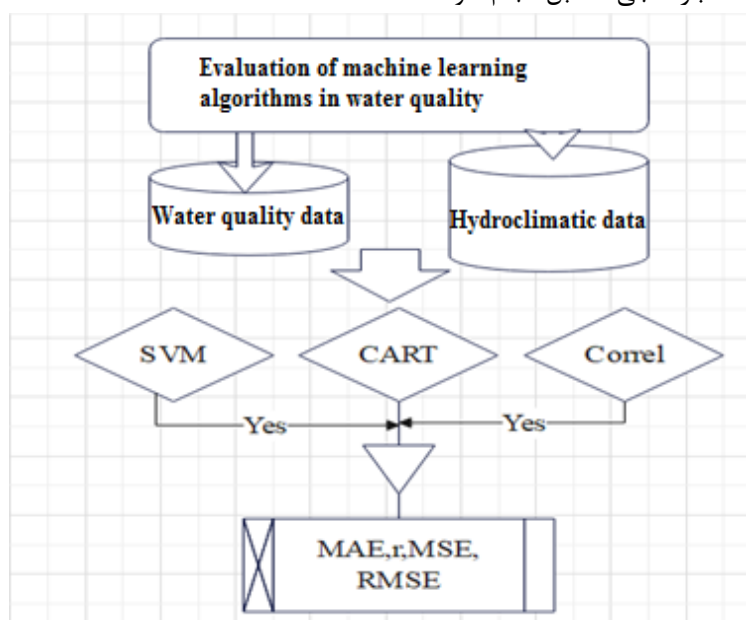
که در آن، M گره، Q_m داده‌های موجود در گره (رابطه ۸)، n_m تعداد نمونه‌های موجود در گره، H ناخالصی‌های زیرمجموعه‌های در چپ و راست گره، t_m به عنوان آستانه گره (رابطه ۷) است.

$$t_m = t_m \text{MSE}(Q_m, t_m) \quad (7)$$

$$\text{MSE } Q_m = \text{Were, } \bar{y}_m = \frac{1}{n_m} \sum_{y \in Q_m} y \quad (8)$$

$$\sum_{y \in Q_m} (y - \bar{y}_m)^2$$

ارزیابی الگوریتم: برای ارزیابی توانایی ماشین در اخذ تصمیمات، پیش‌بینی‌های آموزش داده‌شده به مدل آزمایش می‌شوند. بدین طریق میزان دقت الگوریتم و آموزش داده ارزیابی می‌شود. به منظور پیش‌گیری از بیش‌برازش حتماً باید اعتبارسنجی متقابل انجام شود.



شکل ۴- مدل مفهومی تحقیق

Fig. 4. Conceptual research model

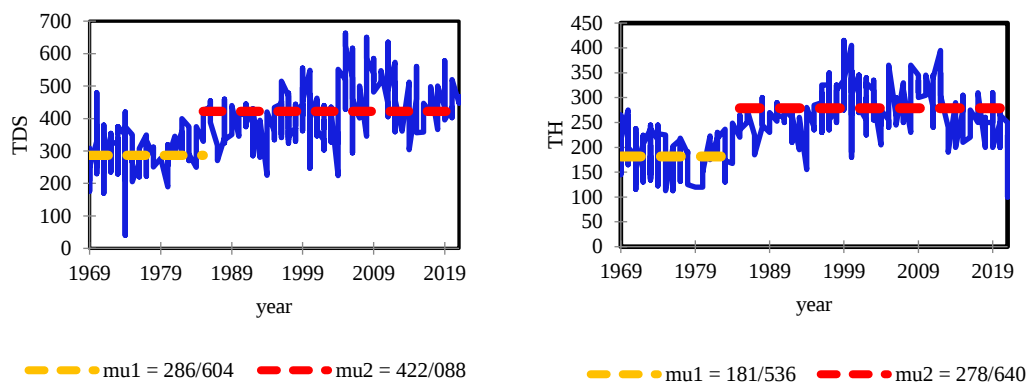
² Mean Absolute Error

¹ Mean Square Error

نتایج و بحث

تحلیل روند: در ارزیابی روند تغییر غلظت متغیرهای شیمیایی آب در آبخیز چم انجیر از دو پارامتر، بار محلول در آب و سختی آب آزمون ناپارامتری Pettitt's-Test در سطح اطمینان ۹۵ درصد استفاده شد (جدول ۲). نتایج نشان داد که دو متغیر TDS و TH دارای روند

معنی‌دار در سطح ۹۵ درصد بوده که از سال ۱۹۸۵ این تغییرات شروع شده است. میانگین TDS در دوره اول (۱۹۸۴-۱۹۶۹) ۲۸۶/۶ و دوره دوم (۲۰۲۱-۱۹۸۶) به مقدار ۴۲۲/۰۸ افزایش و میانگین TH در دوره اول ۱۸۱/۵ و در دور دوم به ۲۷۸/۶ افزایش یافته است (شکل ۵).



شکل ۵- روند پارامترهای کیفیت آب بر اساس آزمون Pettitt's-Test

Fig. 5. Trend of water quality parameters based on Pettitt's-Test

جدول ۲- سطح معنی‌داری آزمون P-test کیفیت آب آبخیز چم انجیر در دوره ۱۹۶۹-۲۰۲۱

Table 2. Significance level of P test of water quality in Cham Anjir Watershed in the period of 1969-2021

Water Quality index	TDS	TH
t	1985	1985
p-value (Two-tailed)	< 0.0001	< 0.0001

متغیرهای pH، Na و SAR رابطه معکوس بین ۰/۴۱- تا ۰/۵- دارد. یعنی با افزایش سختی آب، نسبت جذب و درصد سدیم و نیز قلیابیت آب کاهش می‌یابد. بر اساس ارتباط اسنجی، ضریب همبستگی دو متغیر غلظت بار محلول آب و سختی کل با سایر متغیرها بیشتر است. **مدل ماشین بردار ویژه:** به توجه به ماتریس همبستگی متغیرهای شیمیایی کیفیت آب در آبخیز چم انجیر در پژوهش حاضر، دو عامل TDS و TH با ۳۲۱ نمونه برای آموزش به مدل پشتیبان ماشین انتخاب شده است. در این مدل سازی ۱۶ درصد متغیرها به عنوان یادگیری و مابقی برای تست انتخاب شد. یافته‌ها نشان داد، بهترین مدل KernalLinear بوده است. این مدل با Epsilone=0.1 و Tolerance=0.001

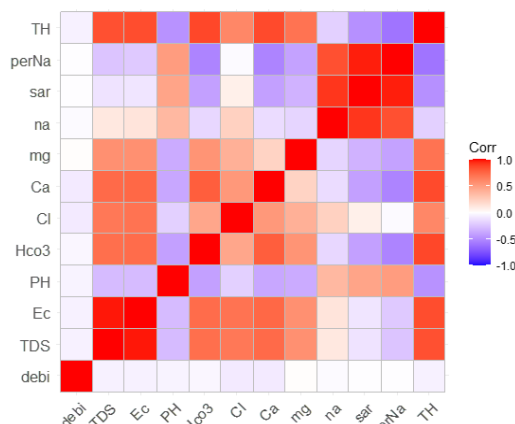
ماتریس همبستگی: برای مدل سازی داده‌های کیفیت آب در یادگیری ماشین، میزان همبستگی و روابط بین متغیرها باید مشخص شود. در شکل ۶ ماتریس همبستگی متغیرهای کیفیت آب با جریان رودخانه چم انجیر نمایش داده شده است. بررسی جدول نشان می‌دهد، در قطر ماتریس بیشترین همبستگی وجود دارد که هر متغیر با خودش همبستگی قوی و عدد یک را نشان می‌دهد. دبی رودخانه با متغیرهای کیفیت آب همبستگی معنی‌دار نشان نمی‌دهد البته در مواردی رابطه معکوس و ضعیف (۰/۸۹- تا ۰/۴۶-) وجود دارد، با افزایش دبی رودخانه در آبخیز چم انجیر، غلظت متغیرهای شیمیایی کاهش یافته است.

بین متغیرهای TDS با EC، ۰/۹۷+ و با متغیر TH رابطه مستقیم و قوی ۰/۷۳+ وجود دارد. TH علاوه بر TDS و EC با متغیرهای HCO_3 ، Ca، Cl و Mg نیز رابطه مثبت و قوی بین ۰/۵۵+ تا یک دارد. افزایش هر یک از متغیرهای مقدار TH نیز افزایش می‌یابد. TH با

خطا در جدول ۳ نمایش داده شده است. به‌طور متوسط مقدار $MSE=22.4$ ، $r=0.99$ و $MAE=2.6$ برای طبقه‌بندی مدل نشان می‌دهد که عملکرد الگوریتم SVM داده‌ها در یادگیری داده‌ها بسیار عالی است. طبق جدول ۳، نمونه‌های تصادفی شماره‌های ۹، ۷ و ۴ نشان‌دهنده دقت بسیار بالای مدل‌سازی به‌وسیله الگوریتم SVM هستند. به‌عنوان مثال، در نمونه شماره ۹، مقادیر $MAE=2.59$ ، $MSE=6.4$ و $r=0.998$ است.

است. برای اعتبارسنجی^۱ مدل، بهترین نتیجه با Number of Folds = 10 به‌دست آمد.

به‌طور تصادفی، ۱۰ نمونه برای ارزیابی عملکرد الگوریتم SVM در یادگیری ۳۱۶ مشاهدات TH انتخاب شده است. متوسط سختی کل آب در آبخیز چم انجیر طی دوره مطالعاتی ۱۹۹۶-۲۰۲۱، ۲۳۶/۵۹ میلی‌گرم در لیتر است. نتایج عملکرد نمونه‌ها و متوسط آن‌ها با ضرایب میانگین مربعات خطا، ضریب تعیین و قدر مطلق



شکل ۶- همبستگی بین متغیرهای شیمیایی کیفیت آب و دبی روانه در آبخیز چم انجیر (۱۹۶۹-۲۰۲۱)

Fig. 6. Correlation between water quality elements and discharge in Cham Anjir Watershed in 1969-2021

جدول ۳- نتایج اعتبارسنجی TH

Table 3. Cross-validation results TH

	Mean	1	2	3	4	5	6	7	8	9	10
MSE	22.4	10.08	11.05	18.005	8.4	10.8	9.2	8.3	12.8	6.4	128.8
r	0.99	0.998	0.998	0.997	0.998	0.998	0.998	0.998	0.998	0.998	0.98
MAE	2.6	2.17	2.911	2.8۶	2.12	2.79	2.44	2.209	2.437	2.059	4.14

از مقدار بالای متوسط TDS در این دوره است (۳۶۳/۴۲ میلی‌گرم در لیتر). نمونه شماره ۳ و ۵ اعتبار و دقت بالاتری در یادگیری داشته‌اند. به‌طوری‌که در نمونه شماره ۳ مقادیر ضرایب صحت‌سنجی برابر $MSE=48.2$ ، $MAE=5.59$ و $r=1$ بدین شرح است.

نتایج صحت‌سنجی متغیر TDS در الگوریتم SVM در جدول ۴ نشان می‌دهد، میانگین مربعات خطا به‌طور متوسط ۳۵۰/۵، ضریب تعیین ۰/۹۸ و قدر مطلق خطا ۸/۸۳ است. به‌جز ضریب MSE دو ضریب r و MAE دقت بالای مدل را نشان می‌دهند مقدار بالا MSE ناشی

جدول ۴- نتایج اعتبارسنجی TDS

Table 4. Cross-validation results TDS

	Mean	1	2	3	4	5	6	7	8	9	10
MSE	530.5	186.5	151.5	48.2	116.2	81.3	726.6	207.3	3499.1	230.1	58.6
r	0.98	0.98	0.98	1	0.99	0.99	0.95	0.98	0.89	0.98	0.99
MAE	8.83	8.10	8.17	5.59	5.63	6.28	12.77	7.67	18.88	9.15	6.02

¹ Accuracy

ر و MSE برای متغیرهای TH و TDS محاسبه و با یکدیگر مقایسه شده‌اند. برآورد مدل در اعتبارسنجی ضرایب MAE، r و MSE برای داده‌های آموزشی متغیر TDS به ترتیب ۰/۹۷۶، ۸/۵ و ۴۷۳/۲ و برای ارزیابی دقت داده‌های آموزشی به ترتیب ۱۰/۷۹، ۰/۹۸۸ و ۲۲۶/۷ است. مقادیر برآوردی آن در داده‌های آموزشی متغیر TH به ترتیب ۰/۹۹۹، ۲/۴ و ۸/۲ و برای ارزیابی دقت داده‌های آموزشی به ترتیب ۰/۹۹۹، ۲/۸ و ۱۲/۴ سنجیده شده است.

جدول ۵- معیارهای عملکرد (TH، TDS)

Table 5. Performance metrics (TH and TDS)

Index	Training set		Validation set	
	TH	TDS	TH	TDS
MSE	8.228	473.226	12.378	226.705
r	0.999	0.976	0.999	0.988
MAE	2.386	8.485	2.885	10.799
Bias(Estimation)	- 0.025	- 0.007		
Number of support vectors	11	53		

مشاهدات TH با شماره‌های ۲۶۶، ۱۸۸، ۹۰، ۵۷ و ۴۴ دارای مقادیر آموزشی ۳۶۵، ۳۱۰، ۱۵۷/۵، ۱۵۹ و ۱۴۷/۵ بوده که مقدار آن‌ها به وسیله الگوریتم SVM به ترتیب ۳۷۱/۱، ۳۰۵/۸، ۱۵۶، ۱۵۷/۰۶ و ۱۴/۹ میلی گرم در لیتر پیش‌بینی شده است (جدول ۶ و شکل‌های ۷-b و ۷-d). در شکل‌های ۷-a و ۷-c مقدار ۳۱۶ داده‌های مشاهداتی و پیش‌بینی شده TDS و TH در دوره ۲۰۲۱-۱۹۶۹ مدل‌سازی شده است، نمودار روند افزایشی هر دو متغیر را مطابق با مشاهدات تخمین زده است. الگوریتم ماشین بردار ویژه مقادیر آموزشی را به خوبی پوشش می‌دهد. دقت الگوریتم در مدل‌سازی متغیر TH بیشتر است.

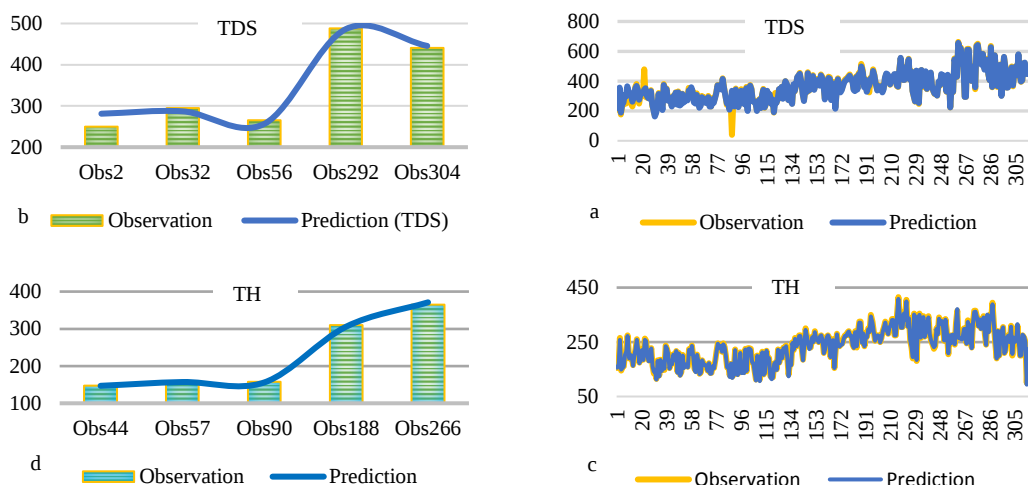
از مجموع ۳۲۱ پارامتر آموزش طبقه‌بندی TDS و TH، تعداد ۳۱۶ مشاهدات برای آموزش مورد استفاده قرار گرفت که از بین آن‌ها برای متغیر TDS، ۵۳ مورد بردار پشتیبان شناسایی شده و بایاس نزدیک به صفر (۰/۰۰۷-) است و برای متغیر TH، ۱۱ مورد بردار پشتیبان شناسایی شده و بایاس نزدیک به صفر (۰/۰۲۵-) است (جدول ۴). داده‌های آموزشی این دو متغیر، با کمینه خطا طبقه‌بندی شده‌اند. در جدول ۵، میزان دقت داده‌های آموزشی و نمونه با ضرایب MAE،

برآورد داده‌های سختی و میزان بار محلول آب با ماشین پشتیبان: مقادیر برآوردی بر اساس مشاهدات متغیرهای سختی کل و بار محلول آب در شکل ۷ و جدول ۶ نمایش داده است. از آنجا که شرح ۳۱۶ داده آموزشی مندرج در جدول در این بحث نمی‌گنجد، به شرح پنج نمونه آن پرداخته می‌شود. مشاهدات آموزشی متغیر TDS در شماره‌های ۳۰۴، ۲۹۲، ۵۶، ۳۲ و ۲ به ترتیب ۴۴۱، ۴۸۸، ۲۶۵، ۲۹۵ و ۲۵۰ است که مقادیر پیش‌بینی شده به وسیله مدل SVM به ترتیب ۴۴۵/۵، ۴۸۶/۴، ۲۵۶/۴، ۲۸۶/۹ و ۲۸۱/۲ میلی‌گرم در لیتر است.

جدول ۶- پیش‌بینی (اعتبارسنجی نمونه)

Table 6. Predictions (Validation sample)

TDS			TH		
OBS	TDS	Prediction (TDS)	OBS	TH Total	Prediction (TH)
2	250	281.2	44	147.5	146.9
32	295	286.9	57	159	157.06
56	265	256.4	90	157.5	156
292	488	486.4	188	310	305.8
304	441	445.5	266	365	371.1



شکل ۷- پیش‌بینی کیفیت آب (TDS و TH) نمونه‌های آموزشی a-c و اعتبارسنجی b-d در SVM
Fig. 7. Prediction of water quality (TDS, TH) training samples a-c and validation b-d in SVM

مقادیر متغیرهای انتخابی برآورد شده است. مجموعه قواعد برای برآورد TDS، ۱۰ قاعده اگر... آنگاه و برای برآورد TH، ۱۱۹ قاعده ارائه شده است (جدول‌های ۷ و ۸). از ۳۲۱ داده آموزشی، الگوریتم، بهترین عملکرد و ویژگی‌های مدل‌سازی را در ۱۰ نمونه و در سطح معنی‌داری ۹۵ درصد ($P\text{-value} < 0.0001$) طبقه‌بندی کرده است. جدول ۷، رابطه قوی بین TDS با EC را نشان می‌دهد.

کارایی مدل رگرسیون و طبقه‌بندی درختی: در الگوریتم طبقه‌بندی و رگرسیون درختی از روش CHAID مبتنی بر ماتریس همبستگی بین متغیرهای شیمیایی، در سطح اطمینان و آستانه ترکیب ۹۵ درصد، در دو عامل TDS و TH استفاده شد. از کل داده‌های ورودی به سامانه (۳۲۱) به‌وسیله الگوریتم C&RT، ۳۱۱ داده مشاهداتی برای آموزش و ۱۰ نمونه برای اعتبارسنجی مدل‌ها انتخاب شد.

براساس متغیرهای ذکر شده مجموعه قواعد مبتنی بر اگر.....آنگاه..... ارائه شده و براساس قواعد تهیه شده

جدول ۷- قانون حاکم بر مدل‌سازی متغیر TDS در الگوریتم رگرسیون و درخت طبقه‌بندی

Table 7. The rule governing the modeling of the TDS element in the CART algorithm

Nodes	TDS(Pred)	Objects	Rules
Node 1	361.44	311	
Node 2	221.32	25	If Ec ≤ 380 then TDS = 221.3 in 8/0% of cases
Node 3	260.16	37	If Ec (380; 425] then TDS = 260.2 in 11.9% of cases
Node 4	285.29	31	If Ec (425; 466] then TDS = 285.3 in 10.0% of cases
Node 5	314.42	31	If Ec (466; 513] then TDS = 314.4 in 10.0% of cases
Node 6	347.71	63	If Ec (513; 578] then TDS = 347.7 in 20.3% of cases
Node 7	394.06	31	If Ec (578; 632] then TDS = 394.1 in 10.0% of cases
Node 8	427.16	31	If Ec (632; 683] then TDS = 427.2 in 10.0% of cases
Node 9	459.61	31	If Ec (683; 750] then TDS = 459.6 in 10.0% of cases
Node 10	549.9	31	If Ec > 750 then TDS = 549.9 in 10.0% of cases

در شکل ۸ رابطه متغیرهای غلظت رسوبات محلول و سختی کل با ۳۲۱ داده آموزشی با الگوریتم رگرسیون و درخت طبقه‌بندی به‌خوبی مدل‌سازی شده است. به‌طوری‌که مطابق با مشاهدات، روند افزایشی و معنی‌دار متغیر TDS تخمین زده شده است. در سال ۱۹۹۶ میزان TDS طبق داده‌های مشاهداتی ۲۰۸ و بر این اساس مدل آن را ۲۲۱/۳ مدل‌سازی کرده است. در سال

با توجه به همبستگی قوی و مثبت (شکل ۶) متغیر TH با TDS و EC و HCO_3 ، Ca، Cl و Mg در جدول ۸، در الگوریتم درخت رگرسیون و طبقه‌بندی تأثیر تغییر در هر یک از متغیرهای فوق بر متغیر سختی کل آب در ۱۱۹ حالت از رابطه‌ها و تغییرات این متغیرها پیش‌بینی شده است (جدول ۷).

آن، از ۱۵۰ در سال ۱۹۹۶ به ۱۰۰ در سال ۲۰۱۹ و مجدداً در سال ۲۰۲۱ روند افزایشی (عدد ۲۰۰) دارد. که این امر می‌تواند تحت تأثیر افزایش دبی و افزایش بارندگی در این سال و کاهش مقدار غلظت مواد محلول آب باشد.

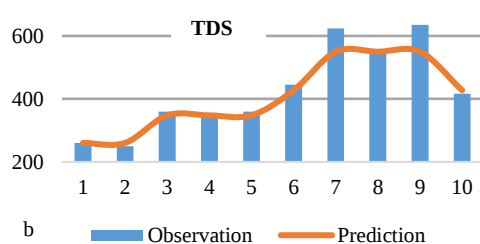
۲۰۲۱ به عدد ۴۴۸ افزایش یافته و مدل آن را ۴۵۹/۶ شبیه‌سازی کرده است.

الگوریتم درخت رگرسیون سختی کل آب را برای ۳۲۱ داده‌های مشاهداتی پیش‌بینی کرده است. بر این اساس که سختی کل دارای چندین روند متوالی ثابت افزایشی و کاهشی طی دوره ۲۰۲۱-۱۹۶۹ است. مقدار

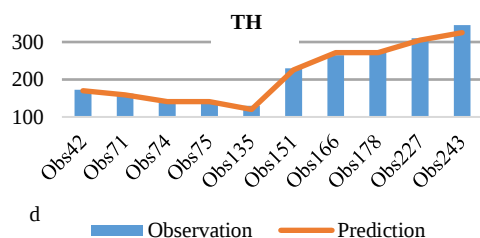
جدول ۸- قانون حاکم بر مدل‌سازی متغیر TH در الگوریتم رگرسیون و درخت طبقه‌بندی

Table 8. The rule governing the modeling of the TH element in the regression and the classification tree algorithm

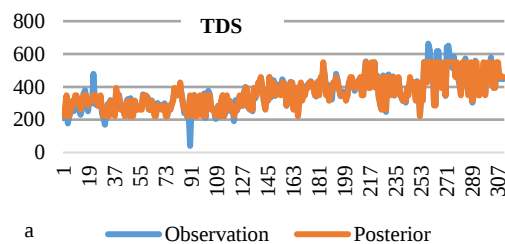
Nodes	TDS(Pred)	Objects	Rules
Node 1	237.2	311	
Node 2	135.3	32	If Cation ≤ 3.95 then TH = 135.3 in 10.3% of cases
Node 3	162.3	29	If Cation (3.95; 4.35] then TH = 162.3 in 9.3% of cases
Node ...	...	...	...
Node ...	...	...	...
Node ...	...	...	...
Node ...	...	...	...
Node 117	172.5	1	If Cation (4.35; 4.86] and %Na > 21.35 and $HCO_3^- \leq 3.15$ and PH ≤ 8.25 and Ec > 427 and Ca ≤ 1.60 then TH = 172.5 in 0/3% of cases
Node 118	170	3	If Cation (4.35; 4.86] and %Na > 21.35 and $HCO_3^- \leq 3.15$ and PH ≤ 8.25 and Ec > 427 and Ca (1/60; 2/20] then TH = 170 in 1/0% of cases
Node 119	173	1	If Cation (4.35; 4.86] and %Na > 21.35 and $HCO_3^- \leq 3.15$ and PH ≤ 8.25 and Ec > 427 and Ca > 2.20 then TH = 173 in 0.3% of cases



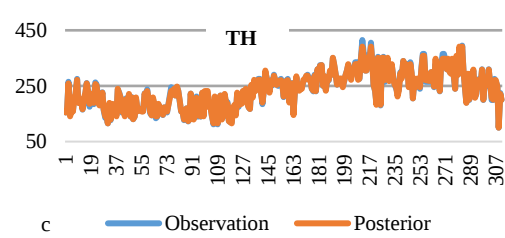
b Observation Prediction



d Observation Prediction



a Observation Posterior



c Observation Posterior

شکل ۸- تخمین متغیرهای شیمیایی کیفیت آب (TDS و TH) نمونه‌های آموزشی a-c و اعتبارسنجی b-d در الگوریتم رگرسیون و درخت طبقه‌بندی

Fig. 8. Prediction of water quality elements (TDS, TH) training examples a-c and validation b-d in the algorithm

میزان RMSE در مدل SVM ۲/۹۷ و در مدل RT ۵۷/۴ است. به‌طور کلی دقت الگوریتم SVM برای پیش‌بینی متغیر TDS بالاتر است.

در جدول ۹ میزان اعتبار و دقت دو مدل SVM و RT برای متغیر TH اندازه‌گیری و با یکدیگر مقایسه شده است. در مدل SVM دقت در ضریبها RMSE، MSE، r و MAE به ترتیب ۱/۶، ۲/۶، ۰/۹۹۷ و ۲۲/۴ و

صحت‌سنجی: در جدول ۸ میزان اعتبار و دقت دو مدل SVM و RT برای متغیر TDS با یکدیگر مقایسه شده است. در مدل SVM، میزان MAE ۵۳۰/۵ بوده که مقدار بالای این عدد به توجه به میانگین این متغیر است و در مدل RT برای ۱۰ نمونه مدل‌سازی شده، مقدار آن ۲۲/۶ است. در مدل SVM میزان همبستگی، مقدار آن ۰/۹۷۸ و در مدل RT این ضریب ۰/۹۸ است. در مدل SVM، مقدار MSE ۸/۸ و در مدل RT ۳۲۹۴/۲ است.

در مدل RT به‌ترتیب ۱۱/۵، ۱۳۲/۳، ۰/۹۹۶ و ۶/۲ عالی و بالاتری مدل‌سازی کرده و قابل قبول است. الگوریتم‌های SVM و RT متغیر TH را با دقت

جدول ۹- مقایسه صحت‌سنجی مدل‌های یادگیری ماشین مدل (RT و SVM) متغیر TDS

Table 9. Comparison of validation of model machine models (RT and SVM) TDS

TDS	SVM			Regression Tree
	Cross-validation results	Training	Validation	Obs-Pre
MAE	530.548	473.226	226.705	22.6
r	0.978	0.976	0.988	0.983
MSE	8.826	8.485	10.799	3294.2
RMSE	2.97	2.91	3.3	57.4

جدول ۱۰- مقایسه صحت‌سنجی مدل‌های یادگیری ماشین (مدل RT و SVM) متغیر TH

Table 10. Comparison of validation of model machine models (RT and SVM) TDS

TH	SVM			Regression Tree
	Cross-validation results	Training	Validation	Obs-Pre
MAE	22.384	8.228	12.378	6.2
r	0.997	0.999	0.999	0.996
MSE	2.613	2.386	2.885	132.3
RMSE	1.6	1.5	1.7	11.5

نتیجه‌گیری

دسترسی به آب سالم و با کیفیت از نظر تراکم متغیرهای شیمیایی شاخص مهمی در توسعه انسانی و بهداشت عمومی است. از این‌رو، پیش‌بینی و نظارت بر کیفیت آب برای جلوگیری از افت کیفیت آب بسیار مؤثرتر از بدتر شدن و آلوده شدن آن است. الگوریتم یادگیری ماشین برای تحقق این اهداف با صرف هزینه و زمان کمتر و دقت بالاتر در مدل‌سازی کیفیت آب ابزاری سودمند است.

در پژوهش حاضر به‌منظور تخمین کیفیت دبی آب‌خیز چم انجیر و انتخاب مدل بهینه در پیش‌بینی کیفیت آب، دو الگوریتم؛ ماشین بردار پشتیبان و طبقه رگرسیون درختی ارزیابی شدند. متوسط املاح محلول در آب‌خیز چم انجیر در دوره مورد مطالعه ۲۰۲۱-۱۹۹۶، ۳۶۳/۴۲ میلی‌گرم در لیتر بوده، متوسط سختی کل ۲۳۶/۶ میلی‌گرم در لیتر است. یافته‌ها نشان داد که غلظت متغیرهای شیمیایی آب در دوره ۲۰۲۱-۱۹۶۹ افزایشی بوده، نتایج Pettitt Test نشان داد که نقطه آغاز روند افزایشی و معنی‌دار این متغیرها از سال ۱۹۸۵ بوده است.

نتایج نمودارهای ترسیم شده، نشان داد الگوریتم‌های ماشین بردار ویژه و درخت رگرسیون در پیش‌بینی غلظت بار محلول و سختی کل آب عملکرد و دقت بالایی دارند. هر دو الگوریتم روند افزایشی کیفیت آب در آب‌خیز چم انجیر را پیش‌بینی کرده‌اند. طبق نتایج طبقه‌بندی

الگوریتم درخت رگرسیون، رابطه متغیر TDS با متغیر EC به‌خوبی مدل‌سازی شده است با افزایش EC مقدار TDS نیز افزایش می‌یابد. بر اساس آخرین مشاهدات درخت تصمیم (که معمولاً دقیق‌ترین ویژگی متغیر مورد پیش‌بینی را نشان می‌دهد)، افزایش ۰/۶ کلسیم باعث افزایش ۱۰۳ میلی‌گرم در لیتر میزان TH خواهد شد.

بر اساس یافته‌های صحت‌سنجی یادگیری ماشین، مقادیر اعتبارسنجی، ضرایب بالای ضریب تعیین و میانگین مربعات خطا، در پژوهش حاضر الگوریتم ماشین بردار ویژه عملکرد بهتری در مدل‌سازی متغیر شیمیایی غلظت رسوبات محلول و سختی کل آب‌خیز چم انجیر دارد. صحت‌سنجی دو الگوریتم SVM و CART داده‌های مشاهداتی و برآوردی نشان داد ضرایب MSE، MAE و r در هر دو مدل SVM و CART تفاوت ندارند. اما برآوردهای سختی آب و بار معلق میزان RMSE در الگوریتم SVM کمتر از ۱۰ درصد بوده است. همچنین، میزان Bias داده‌های آموزشی در مدل SVM به صفر نزدیک است؛ بنابراین، الگوریتم SVM نتایج بهتری در مقایسه با الگوریتم CART در تخمین داده‌ها حاصل می‌کند.

یافته‌ها نشان داد، در آب‌خیز چم انجیر به‌دلیل رخداد تغییر اقلیم و نوسان بارش، به‌دلیل عوامل زمین‌شناسی و فرسایش سازندهای کارستیک، عبور رودخانه از مناطق سکونتگاهی، تغییرات کاربری اراضی، اضافه

RMSE=2.97 و برای متغیر TH مقدار RMSE=1.6 است. صحت الگوریتم RT، برای این متغیرها به این ترتیب RMSE=57.2 و RMSE=11.5 است).

تشکر و قدردانی

نویسندگان از شرکت سهامی آب منطقه‌ای لرستان به دلیل همکاری در دسترسی به اطلاعات منابع آب آبخیز چم انجیر تشکر و قدردانی می‌کنند.

تعارض منافع

در این مقاله تضاد منافی وجود ندارد و این مسأله مورد تأیید همه نویسندگان است.

برداشت منابع آب در بخش کشاورزی و کاهش دبی رودخانه کیفیت آب طی سال‌های اخیر روند کاهش‌یافته داشته است. پایش و نظارت بر کیفیت آب به‌وسیله مدیران ذی‌ربط ضروری است. نتایج مطالعات مشابه نشان داد، الگوریتم یکسانی برای پیش‌بینی کیفیت آب وجود ندارد.

در این پژوهش نیز، نتایج حاصل از صحت‌سنجی الگوریتم‌های مختلف یادگیری ماشین در برآورد داده‌های کیفیت آب، نشان داد الگوریتم‌های درخت تصمیم و ماشین بردار ویژه در برآورد این داده‌ها دقیق تر هستند. همچنین، نتایج سنج‌های اعتبارسنجی نشان دادند، دقت الگوریتم ماشین بردار ویژه در برآورد کیفیت منابع آب سطحی آبخیز چم انجیر بالاتر است (صحت الگوریتم SVM، در برآورد متغیر TDS مقدار

منابع مورد استفاده

- Abuzir, S. Y., Abuzir, Y. S., 2022. Machine learning for water quality classification. *Water Qual. Res. J.*, 57(3), 152-164.
- Ahmed, U., Mumtaz, R., Anwar, H., Shah, A. A., Irfan, R., García-Nieto, J., 2019. Efficient water quality prediction using supervised machine learning. *Water*, 11(11), 2210.
- Al-Adhaileh, M. H., Aldhyani, T. H., Alsaade, F. W., Al-Yaari, M., Albaggar, A. K. A., 2022. Groundwater quality: The application of artificial intelligence. *J. Environ. Public Health*, 2022(1), 8425798.
- Arabgol, R., Sartaj, M., Asghari, K., 2016. Predicting nitrate concentration and its spatial distribution in groundwater resources using support vector machines (SVMs) model. *ENVIRON MODEL ASSESS*, 21, 71-82.
- Asadollah, S.B.H.S., Sharafati, A., Motta, D. and Yaseen, Z.M., 2021. River water quality index prediction and uncertainty analysis: A comparative study of machine learning models. *J. Environ. Chem. Eng.*, 9(1), p.104599.
- Barbieri, M., Barberio, M. D., Banzato, F., Billi, A., Boschetti, T., Franchini, S., Gori, F., Petitta, M., 2023. Climate change and its effect on groundwater quality. *J. Environ. Public Health*, 45(4), 1133-1144.
- Beiranvand, N., Sepahvand, A., Haghizadeh, A., 2023. Total Dissolved Solids modeling using machine learning algorithms in periods of low and high water (Case study: Khorramabad, Biranshahr and Alashtar watersheds, Lorestan province). *JRWM*, 76(3), 215-236. (in persian)
- Bui, D. T., Khosravi, K., Tiefenbacher, J., Nguyen, H., Kazakis, N., 2020. Improving prediction of water quality indices using novel hybrid machine-learning algorithms. *Sci. Total Environ.*, 721, 137612.
- Cojbasic, S., Dmitrasinovic, S., Kostic, M., Turk Sekulic, M., Radonic, J., Dodig, A., Stojkovic, M., 2023. Application of machine learning in river water quality management: A review. *Water Sci. Technol.*, 88(9), 2297-2308.
- Dezfooli, D., Hosseini-Moghari, S. M., Ebrahimi, K., Araghinejad, S., 2018. Classification of water quality status based on minimum quality parameters: application of machine learning techniques. *Model. Earth Syst. Environ.*, 4, 311-324.
- Delbaz, R., Ebrahimian, H., 2024. A Review of the Application of Data Science and Machine Learning in Agricultural Water Management. *Journal of Water and Sustainable Development*, 11(2), 39-56. doi: 10.22067/jwsd.v11i2.2402.1310 (in persian).
- Dritsas, E., Trigka, M., 2023. Efficient data-driven machine learning models for water quality prediction. *Computation J. (MDPI)*, 11(2), 16.
- 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(13), 2448.
- Eze, E., Kirby, S., Attridge, J., Ajmal, T., 2023. Aquaculture 4.0: hybrid neural network multivariate water quality parameters forecasting model. *Sci. Rep.*, 13(1), 16129.

- Farhadinejad, T., Vayskarami, I., Zand, M., 2024. Investigating the relationship between river flow changes caused by drought and the quality of surface water resources in the Tirah River Basin. *Watershed Engin. Manage.*, 16(1), 64-81. doi: 10.22092/ijwmse.2023.361325.2009.(in persian)
- Farid Giglou, B., Ghazavi, R., Dokhani, S., 2022. Evaluating the impact of climate change on Aras border river water quality using statistical methods, SWAT Model and WQISC Index. *Watershed Engin. Manage.*, 13(4), 718-731. doi: 10.22092/ijwmse.2021.351242.1822.(in persian)
- Fernández del Castillo, A., Yebra-Montes, C., Verduzco Garibay, M., de Anda, J., Garcia-Gonzalez, A., Gradilla-Hernández, M. S., 2022. Simple prediction of an ecosystem-specific water quality index and the water quality classification of a highly polluted river through supervised machine learning. *Water*, 14(8), 1235.
- Ghaemi, A., Azhdary Moghaddam, M., Keikha, S., 2024. Evaluation of integrated artificial intelligence models in estimating total dissolved solid concentrations in the upstream of Sari city. *Watershed Engin. Manage.*, 16(1), 50-63. doi: 10.22092/ijwmse.2023.358863.1975.(in persian)
- Goodarzi, M. R., Niknam, A. R. R., Barzkar, A., Niazkar, M., Zare Mehrjerdi, Y., Abedi, M. J., Heydari Pour, M., 2023. Water quality index estimations using machine learning algorithms: a case study of Yazd-Ardakan Plain, Iran. *Water*, 15(10), 1876.
- Granata, F., Papirio, S., Esposito, G., Gargano, R., De Marinis, G., 2017. Machine learning algorithms for the forecasting of wastewater quality indicators. *Water*, 9(2), 105.
- Hasheminasab, S., Rahimi, D., Zakerinejad, R., Kropáček, J., 2022. Assessment of climate change impact on surface water: a case study—Karoun River Basin, Iran. *Arab. J. Geosci.*, 15(9), 904.
- Haghiabi, A. H., Nasrolahi, A. H., Parsaie, A., 2018. Water quality prediction using machine learning methods. *Water Qual. Res. J.*, 53(1), 3-13.
- Hamada, M. S., Zaqoot, H. A., & Sethar, W. A. (2024). Using a supervised machine learning approach to predict water quality at the Gaza wastewater treatment plant. *Environ. Sci.: Adv.*, 3(1), 132-144.
- Hussein, E. A., Thron, C., Ghaziasgar, M., Bagula, A., Vaccari, M., 2020. Groundwater prediction using machine-learning tools. *Algorithms*, 13(11), 300.
- Kaur, I., Gulati, A., Lamba, P.S., Jain, A., Taneja, H. and Syal, J.S., 2024. Water Quality Assessment using Machine Learning: A Focus on Coliform Prediction in Water. *Asian J. Water Environ. Pollut.*, 21(5), pp.19-26.
- Khan, M. S. I., Islam, N., Uddin, J., Islam, S., Nasir, M. K., 2022. Water quality prediction and classification based on principal component regression and gradient boosting classifier approach. *J. King Saud Univ. - Comput. Inf. Sci.*, 34(8), 4773-4781.
- Khan, Y., See, C. S., 2016. Predicting and analyzing water quality using machine learning: a comprehensive model. In 2016 IEEE Long Island Systems, Applications and Technology Conference (LISAT) (pp. 1-6). IEEE.
- Khullar, S., Singh, N., 2021. Machine learning techniques in river water quality modelling: a research travelogue. *Water Supply*, 21(1), 1-13.
- Kouadri, S., Elbeltagi, A., Islam, A. R. M. T., Kateb, S., 2021. Performance of machine learning methods in predicting water quality index based on irregular data set: application on Illizi region (Algerian southeast). *Appl. Water Sci.*, 11(12), 190.
- Lakshmi, T. M., Martin, A., Begum, R. M., Venkatesan, V. P., 2013. An analysis on performance of decision tree algorithms using student's qualitative data. *Int. J. Mod. Educ. Comput. Sci.*, 5(5), 18-27.
- Li, D., Cui, B., Zuo, F., Zong, H. and Yu, W., 2023. Hydrological characteristics and water quality change in mountain river valley on Qinghai-Tibet Plateau. *Appl. Water Sci.*, 13(4), p.104.
- Makumbura, R.K., Mampitiya, L., Rathnayake, N., Meddage, D.P.P., Henna, S., Dang, T.L., Hoshino, Y. and Rathnayake, U., 2024. Advancing water quality assessment and prediction using machine learning models, coupled with explainable artificial intelligence (XAI) techniques like shapley additive explanations (SHAP) for interpreting the black-box nature. *Results Eng.*, 23, p.102831.
- Murdoch, P. S., Baron, J. S., Miller, T. L., 2000. Potential effects of climate change on surface-water quality in North America 1. *J. Am. Water Resour. Assoc.*, 36(2), 347-366.
- Murphy, J., Chanat, J., 2023. Leveraging machine learning to automate regression model evaluations for large multi-site water-quality trend studies. *Environ. Model. Softw.*, 170, 105864.
- Nallakaruppan, M.K., Gangadevi, E., Shri, M.L., Balusamy, B., Bhattacharya, S. and Selvarajan, S., 2024. Reliable water quality prediction and parametric analysis using explainable AI models. *Sci. Rep.*, 14(1), p.7520.
- Nuanmeesri, S., Poomhiran, L., Kadmateekarun, P., Chopvitayakun, S., 2023. Improving the Water Quality Classification Model for Various Farms Using Features Based on Artificial Neural Network. *TEM J.* 12(4), 2144.
- Oğuz, A., Ertuğrul, Ö. F., 2023. A survey on applications of machine learning algorithms in water quality assessment and water supply and management. *Water Supply*, 23(2), 895-922.

- Okafor, C.O., Ude, U.I., Okoh, F.N. and Eromonsele, B.O., 2024. Safe drinking water: The need and challenges in developing countries. In *Water quality-new perspectives*. IntechOpen.
- Omambia, A., Maake, B., Wambua, A., 2022. Water quality monitoring using IoT & machine learning. In 2022 IST-Africa Conference (IST-Africa) (pp. 1-8). IEEE.
- Qie, G., Zhang, Z., Getahun, E., Allen Mamer, E., 2023. Comparison of machine learning models performance on simulating reservoir outflow: A case study of two reservoirs in Illinois, USA. *J. Am. Water Resour. Assoc.*, 59(3), 554-570.
- Rahimi, D., Hasheminasab, S., 2017. Analysis water quality by artificial neural network in bazoft river (iran). *J Chem Pharm Res*, 9, 115-121.
- Sahoo, S., Russo, T. A., Elliott, J., Foster, I., 2017. Machine learning algorithms for modeling groundwater level changes in agricultural regions of the US. *Water Resour. Res.*, 53(5), 3878-3895.
- Sahour, S., Khanbeyki, M., Gholami, V., Sahour, H., Kahvazade, I., Karimi, H., 2023. Evaluation of machine learning algorithms for groundwater quality modeling. *Environ. Sci. Pollut. Res.*, 30(16), 46004-46021.
- Sayahi, F., Divband Hafshejani, L., Tishehzan, P. and Abdolabadi, H., 2024. The combination of dimensionality reduction methods and machine learning algorithms in the optimization of Maroon River water quality prediction. *IJSWR*, 55(9), pp.1601-1615.
- Sepahvand, A., Prelovšek, M., Nazari Samani, A. A., Wasson, R. J., 2021. SOLUTE TRANSPORT AND SOLUTIONAL DENUDATION RATE OF CARBONATE KARST IN THE SEMI-ARID ZAGROS REGION (SOUTHWESTERN IRAN). *J. Cave Karst Stud.*, 83(3).
- Shirazi, F., Zahiri, A., Piri, J., Dehghani, A. A., 2023. " Research Paper" Development a new hydraulic method for prediction of river flood discharge. *J. atershed Manag. Res.*, 14(28), 110-123. (in persian)
- Soleimani Motlagh, M., Talebi, A., Zareei, M., 2016. The study of drought on the quality of surface water resources in Kashkan watershed. *J. Watershed Manag. Res.*, 6(12), 154-165. (in persian)
- Yan, X., Zhang, T., Du, W., Meng, Q., Xu, X., Zhao, X., 2024. A Comprehensive Review of Machine Learning for Water Quality Prediction over the Past Five Years. *J. Mar. Sci. Eng.*, 12(1), 159.
- Zaresefat, M., Derakhshani, R., 2023. Revolutionizing groundwater management with hybrid AI models: A practical review. *Water*, 15(9), 1750.