

Estimating the Quality Index of Surface Waters Using Machine Learning Methods (Case Study: Zayandeh-Rood River)

Mohammad Shayannejad^{*}, Elham Fazel Najafabadi and Fahimeh Hatamian Jazi¹

(Received: March 13-2025 ; Revised: May 22-2025 ; Accepted: August 2-2025)

1. Department of Water Science and Engineering, College of Agriculture, Isfahan University of Technology, Isfahan, Iran.

*: Corresponding author, Email: shayannejad@iut.ac.ir

Abstract

Regarding the increasing need for water resources and the decline of surface water resources, awareness of these resources is a crucial need in planning, developing, and protecting them. This research was conducted to model the water quality index (the most widely used feature of determining water quality) using machine learning models (Random Forest and Support Vector Machine) in the Zayandehrood River. Regarding the large number of water quality indices, the NSFQI index was used in this study. First, this index was calculated, and then, input data, including water quality characteristics of 8 stations over 31 years, and the river water quality index were used. In this research, 80% of the data was used in the training stage, and the remaining 20% was used in the evaluation stage. The optimal model was selected based on the evaluation criteria, including R^2 , CRM, and NRMSE. The results showed that the Support Vector Machine algorithm ($0.931 < R^2 < 0.982$, $1.321 < \text{NRMSE} < 2.93$, $-0.00312 < \text{CRM} < 0.004576$) performed better in four stations, and the Random Forest algorithm ($0.993 < R^2 < 0.999$, $1.280 < \text{NRMSE} < 1.789$, $-0.00114 < \text{CRM} < 0.004698$) performed similarly in four stations. The results also showed that selecting series number 3, which uses three features — TDS, TH, and WQI — to estimate the water quality index (WQI) was the best combination for saving time and cost.

Keywords: Random Forest, Zayandeh-Rood, Support vector machine, Machine learning models, Water quality characteristics

How to cite this article: Shayannejad, M., Fazel Najafabadi, E. and Hatamian Jazi, F. 2025. Estimating the Quality Index of Surface Waters Using Machine Learning Methods (Case Study: Zayandeh-Rood River).. *J. Water and Soil Sci.* 29(3): 145-158. DOI: [10.47176/jwss.29.3.40413](https://doi.org/10.47176/jwss.29.3.40413)

Copyright © 2025 Shayannejad et al. Published by Isfahan University of Technology.



This work is licensed under a Creative Commons Attribution-Noncommercial 4.0 International license (<https://creativecommons.org/licenses/by-nc/4.0/>). Non-commercial uses of the work are permitted, provided the original work is properly cited.

تخمین شاخص کیفی آب‌های سطحی با استفاده از روش‌های یادگیری ماشین (مطالعه موردی: رودخانه زاینده‌رود)

محمد شایان‌نژاد*^{id}، الهام فاضل نجف‌آبادی و فهیمه حاتمیان جزی^۱

(تاریخ دریافت: ۱۴۰۳/۱۲/۲۳؛ تاریخ بازنگری: ۱۴۰۴/۳/۱؛ تاریخ پذیرش: ۱۴۰۴/۵/۱۱)

۱. گروه علوم و مهندسی آب، دانشکده کشاورزی، دانشگاه صنعتی اصفهان، اصفهان، ایران.

*: مسئول مکاتبات: پست الکترونیکی: shayannejad@iut.ac.ir

چکیده

باتوجه به نیاز روزافزون به منابع آب و کاهش منابع آب سطحی، آگاهی از کیفیت منابع آب سطحی از نیازهای مهم در برنامه‌ریزی، توسعه و حفاظت از منابع آب به شمار می‌رود. این پژوهش با هدف مدل‌سازی شاخص کیفیت آب (ویژگی پرکاربرد تعیین کیفیت آب) به کمک مدل‌های یادگیری ماشین جنگل تصادفی و ماشین بردار پشتیبان در رودخانه زاینده‌رود انجام شده است. باتوجه به تعدد شاخص‌های کیفیت آب، در این پژوهش از شاخص NSFQI استفاده شد. ابتدا این شاخص محاسبه و در ادامه، از داده‌های ورودی شامل ویژگی‌های کیفی آب ۸ ایستگاه در یک دوره ۳۱ ساله و شاخص کیفیت آب رودخانه استفاده شد. در این پژوهش در مرحله آموزش ۸۰ درصد داده‌ها و در مرحله ارزیابی ۲۰ درصد باقی‌مانده استفاده شده و بر اساس نتایج معیارهای ارزیابی R^2 ، CRM و NRMSE مدل بهینه انتخاب شد. نتایج نشان داد در چهار ایستگاه الگوریتم ماشین بردار پشتیبان ($0.982 < R^2 < 0.993$ ، $0.21 < NRMSE < 0.321$ ، $0.04576 < CRM < 0.0312$) و در چهار ایستگاه الگوریتم جنگل تصادفی ($0.999 < R^2 < 0.993$ ، $0.1789 < NRMSE < 0.280$ ، $0.04698 < CRM < 0.0114$) عملکرد بهتری نشان داد. همچنین نتایج نشان داد که برای صرفه‌جویی در زمان و هزینه، سری شماره ۳ (شامل سه ویژگی هدایت الکتریکی، کل مواد جامد محلول و سختی کل) به‌منظور برآورد شاخص کیفیت آب بهترین ترکیب بوده است.

واژه‌های کلیدی: جنگل تصادفی، زاینده‌رود، ماشین بردار پشتیبان، مدل‌های یادگیری ماشین، ویژگی‌های کیفیت آب

حق انتشار این مستند، متعلق به دانشگاه صنعتی اصفهان است. © ۱۴۰۴

این مقاله تحت گواهی زیر منتشر شده و هر نوع استفاده غیرتجاری از آن مشروط به استناد صحیح به مقاله و با رعایت شرایط مندرج در آدرس زیر مجاز است: (<https://creativecommons.org/licenses/by-nc/4.0/>)



مقدمه

بر هیچ کس پوشیده نیست که آب یک منبع ضروری برای همه موجودات زنده از جمله انسان بوده و نقش مهمی در شرب، صنعت، کشاورزی و محیط‌زیست ایفا می‌کند (۳۸). از سوی دیگر، با توجه به افزایش جمعیت و پدیده تغییر اقلیم و گرمایش جهانی، آلودگی آب یک چالش جهانی مهم به‌ویژه در کشورهای مناطق خشک و نیمه‌خشک مانند ایران است. طی دهه‌های اخیر، کاهش قابل توجهی در کیفیت آب وجود داشته است که می‌توان به رشد روزافزون جمعیت، صنعتی شدن و انواع مختلف آلاینده‌های ناشی از منابع شهری و صنعتی یا کشاورزی نسبت داد (۲). یک رویکرد ساده برای ارزیابی کیفیت آب شامل استفاده از شاخص کیفیت آب (Water Quality Index-WQI) است (۲۵). WQI مهم‌ترین شاخص در زمینه طبقه‌بندی کیفیت آبهای سطحی و زیرزمینی است. در محاسبه این شاخص از یک فرمول برای تبدیل ویژگی‌های کیفیت آب به مقادیر عددی استفاده شده و در نتیجه نمایشی از وضعیت کلی کیفیت آب ارائه می‌کند (۱۷، ۲۶ و ۳۱). این شاخص‌ها ابزارهای ساده اما قدرتمندی برای ارزیابی کیفیت آب هستند که با استفاده از مدل‌های ریاضی، طیفی از شرایط کیفیت آب، از ضعیف تا عالی را طبقه‌بندی می‌کنند (۱۶). شاخص کیفیت آب ایران، IRWQIsc، برای ویژگی‌های متعارف منابع آب سطحی یکی دیگر از شاخص‌های محلی است که بر اساس NSFQI توسعه یافته و معمولاً در ایران استفاده می‌شود (۱ و ۵). طبقه‌بندی شاخص کیفیت آب به سیاستمداران و تصمیم‌گیرندگان کمک می‌کند تا برای مدیریت منابع آب و انتخاب منابع مناسب، با توجه به استفاده مد نظر از آب، اقدامات لازم را انجام دهند (۱۱). برای محاسبه WQI روش‌های مختلفی توسط پژوهشگران ارائه شده که شاخص کیفیت آب بنیاد ملی بهداشت (National Sanitation Foundation Water Quality-NSFWQI) به دلیل دقت به نسبت بیشتر، سادگی و کاربرد وسیع شاخص پرکاربرد شناخته شده است (۱۸، ۲۳ و ۲۷). از جمله

پژوهش‌هایی که روی این پارامتر انجام شده است می‌توان به ماهرویان و همکاران (۲۲)، زمانی-احمدمحمودی و همکاران (۴۰)، مدیروس و همکاران (۲۴)، حسین‌زاده و همکاران (۱۵) و آکویونلو و آکینر (۳) اشاره کرد.

در سال‌های اخیر، مدل‌های هوش مصنوعی به‌ویژه روش‌های یادگیری ماشین (Machine Learning-ML)، به‌عنوان ابزارهایی مؤثر برای تحلیل پدیده‌های محیطی پیچیده و غیرخطی توسعه یافته‌اند. این رویکرد در مقایسه با روش‌های سنتی آماری، از دقت بیشتری برخوردار بوده و نیاز کمتری به فرضیات اولیه دارد. از جمله مزایای کلیدی استفاده از یادگیری ماشین در این زمینه می‌توان به افزایش دقت پیش‌بینی، کاهش هزینه و زمان پایش میدانی و امکان توسعه سامانه‌های پایش هوشمند و هشدار سریع اشاره کرد. این قابلیت‌ها یادگیری ماشین را به ابزاری کارآمد در مدیریت و پایش پایدار منابع آبی تبدیل کرده است. در این راستا چن و همکاران به تحلیل مقایسه‌ای عملکرد پیش‌بینی کیفیت آب سطحی در چین پرداخته و با استفاده از مدل‌های مختلف یادگیری ماشین به شناسایی ویژگی‌های کلیدی آب پرداختند. این پژوهشگران برای پایش کیفیت آب در آینده و ارائه هشدار به موقع کیفیت آب، از مدل‌های درخت تصمیم، جنگل تصادفی و جنگل آبشار عمیق استفاده کردند (۸). همچنین خوئی و همکاران به منظور پیش‌بینی WQI در رودخانه لابوونگ ویتنام از مدل‌های یادگیری ماشین استفاده کرده و نشان دادند که مدل XGBoost با ریشه میانگین مربعات خطا برابر با ۰/۱۰۷ عملکرد خوبی در پیش‌بینی WQI داشته است (۱۹). همچنین یوسری و همکاران به طبقه‌بندی کیفیت آب با استفاده از روش ماشین بردار پشتیبان و XGBoost پرداختند. این پژوهشگران یک الگوریتم طبقه‌بندی کیفیت آب ایجاد کرده و نتیجه گرفتند که مدل XGBoost با دقت ۹۴٪ در مقایسه با مدل SVM کارایی بهتری داشته است (۳۹).

رحمان و همکاران چهار الگوریتم یادگیری ماشین شامل DNN، GBM، GLM و RF را به‌منظور پیش‌بینی WQI به‌کار

مهم ایران به شمار می رود. این رودخانه از دامنه های شرقی زاگرس و کوه های مرتفع زردکوه بختیاری سرچشمه گرفته و پس از طی مسیر در استان های چهارمحال بختیاری و اصفهان به باتلاق گاوخونی در ۱۳۷ کیلومتری جنوب اصفهان می ریزد. در طول مسیر این رودخانه ارتفاع از ۴۵۵ متر تا ۸۶۱۱ متر متغیر بوده و به دلیل چنین تنوع ارتفاعی، آب و هوا به طور قابل توجهی در داخل حوزه تغییر می کند. دامنه بارندگی سالانه ۱۰۰-۲۰۰ میلی متر بوده و به دلیل تلاطم در عرض جغرافیایی میانی غربی رخ می دهد. منطقه مورد مطالعه در شکل ۱ نشان داده شده است.

رویکرد کلی مورد استفاده در این پژوهش شامل جمع آوری داده ها، تفسیر داده ها و کاربرد مدل های یادگیری ماشین در شکل ۲ ارائه شده است.

مجموعه داده های استفاده شده در این پژوهش شامل داده های کیفیت آب ۸ ایستگاه قلعه شاهرخ، سد تنظیمی زاینده رود، پل زمان خان، سد چم آسمان، پل کله، دیزیچه، لنج و موسیان در دوره ۳۱ ساله (۱۳۷۰ تا ۱۴۰۱) به صورت ماهانه اندازه گیری شده است. البته گفتنی است که در برخی ایستگاه ها تعداد سال های کمتری داده ها در دسترس بودند. این داده ها شامل هدایت الکتریکی (EC)، کل مواد جامد محلول (TDS)، pH، سختی کل (TH)، کلر (Cl⁻)، منیزیم (Mg²⁺)، سدیم (Na⁺)، کلسیم (Ca²⁺)، سولفات (SO₄²⁻) و بی کربنات (HCO₃⁻) به عنوان ویژگی های هدف برای برآورد شاخص کیفی آب (WQI) بودند.

محاسبه شاخص WQI

شاخص WQI یکی از شاخص های پرکاربرد برای تعیین کیفیت شیمیایی آب های سطحی و زیرزمینی است که روش های متفاوتی برای محاسبه آن در پژوهش ها ارائه شده است. در این پژوهش شاخص NSFQI استفاده شده که یکی از اولین و پرکاربردترین شاخص ها برای ارزیابی کلی وضعیت کیفیت آب بوده (۴ و ۳۴) و بر اساس جدول ۱ در پنج کلاس خیلی بد،

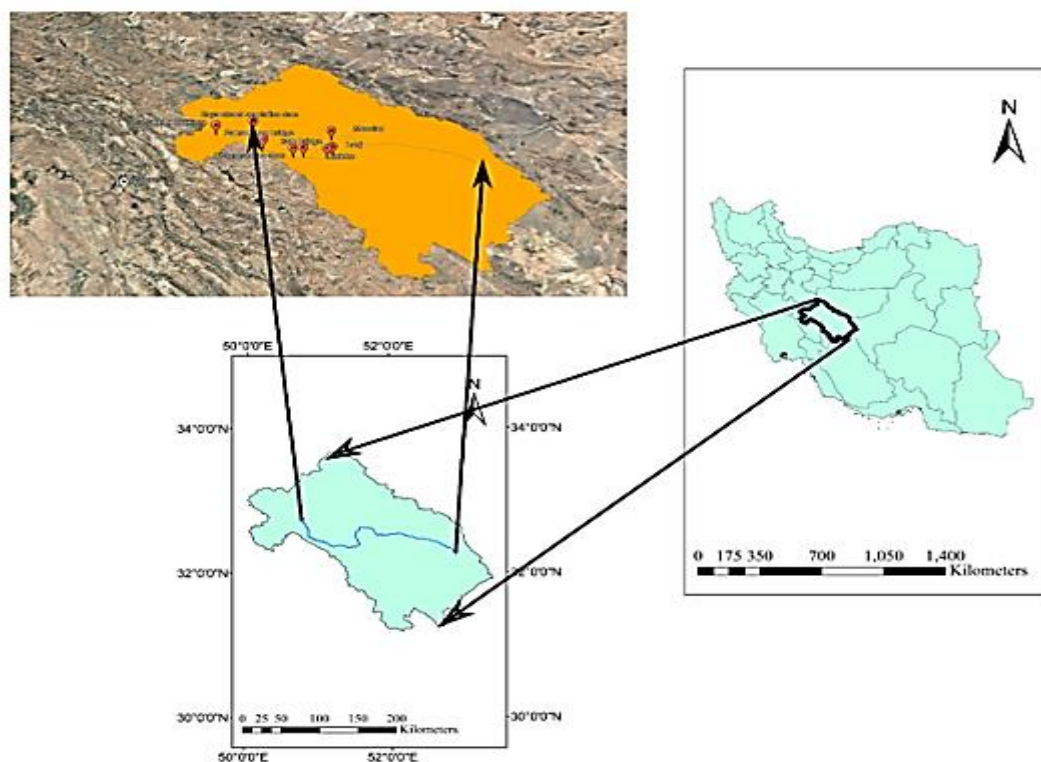
بردند. نتایج نشان داد که الگوریتم GLM مؤثرترین الگوریتم برای پیش بینی WQI بود و BOD، COD، EC، کدورت و سختی کل پارامترهای مؤثر در پیش بینی کیفیت آب سطحی بودند (۲۹). سینگ و همکاران پژوهش خود را روی رودخانه های هندوستان و با کاربرد ۵ الگوریتم یادگیری ماشین انجام دادند. این پژوهشگران نشان دادند که WQI را می توان با داشتن تنها سه ورودی به مدل GBT با شاخص ارزیابی بیشتر از ۸۰٪ تخمین زد. این در حالی است که نتایج این پژوهش نشان داد با داشتن دو ویژگی ورودی به مدل های GBT، RF و ANN می توان به دقت قابل قبولی در تخمین WQI دست یافت (۳۲). لین و همکاران با کاربرد الگوریتم های ترکیبی یادگیری ماشین LGB-Pearson و LGB-Spearman در منطقه رودخانه ای کره جنوبی، نشان دادند که الگوریتم ترکیبی LGB دقت بیشتری در برآورد و پیش بینی WQI دارد (۲۱). پالابیک و آکان با برداشت ۱۲ ویژگی کیفی آب در ۵ ایستگاه روی رودخانه Aksu و کاربرد ۲ روش نوظهور MLR و MLP-ANN و سه الگوریتم یادگیری ماشین NN، SVR و GPN به پیش بینی WQI پرداختند. نتایج این پژوهش نشان داد که روش های نوظهور MLR و MLP-ANN الگوریتم های قدرتمندی در این برآورد هستند (۲۸).

با توجه به اینکه رودخانه زاینده رود رودخانه ای حیاتی و مهم در فلات مرکزی ایران بوده و از چندین شهر و استان عبور کرده، آب شرب، کشاورزی و صنعت را فراهم می کند، اهمیت تشخیص و برآورد شاخص کیفی آب این رودخانه بسیار ضروری است. همچنین با توجه به اینکه در برخی ایستگاه های هیدرومتری موجود روی این رودخانه داده های بلندمدت در دسترس نیستند، از این رو هدف مطالعه حاضر بررسی توانایی دو الگوریتم ماشین بردار پشتیبان و جنگل تصادفی در برآورد WQI است.

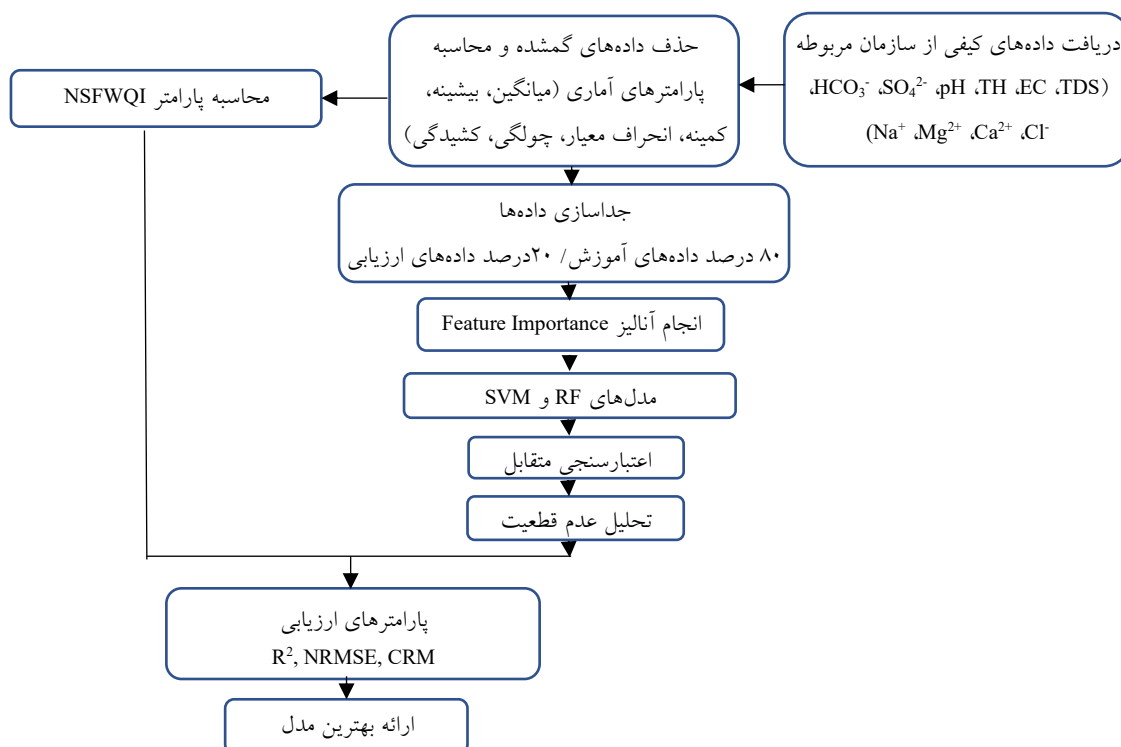
مواد و روش ها

منطقه و داده های مورد مطالعه

رودخانه زاینده رود ۲۸۸۰ کیلومتر طول داشته و یکی از رودهای



شکل ۱. حوزه زاینده‌رود و ایستگاه‌های مورد مطالعه



شکل ۲. نمودار مفهومی روش انجام پژوهش

انجام شده ویژگی‌ها اولویت‌بندی شده و ۶ ترکیب تعریف شد. سپس مدل‌های یادگیری ماشین برای هر ترکیب و برای همه ایستگاه‌ها اجرا و مقادیر WQI پیش‌بینی شده به دست آمدند. در این مطالعه، از دو الگوریتم یادگیری ماشین نظارت‌شده جنگل تصادفی و ماشین بردار پشتیبان برای پیش‌بینی شاخص کیفی آب (WQI) بر اساس ویژگی‌های انتخاب‌شده استفاده شد که در محیط برنامه‌نویسی پایتون کدنویسی شده است. این دو روش هم در مسائل رگرسیونی (خروجی پیوسته) و هم در مسائل طبقه‌بندی (خروجی گسسته) استفاده می‌شود که در این پژوهش با توجه به ماهیت داده‌ها، مسئله از نوع رگرسیون است. با توجه به توصیه‌های پژوهش‌های پیشین، مجموعه آزمون (اعتبارسنجی) متشکل از ۲۰ درصد داده‌های تصادفی بوده؛ درحالی‌که ۸۰ درصد باقی‌مانده برای آموزش استفاده شد. به‌منظور اعتبارسنجی مقادیر پیش‌بینی شده WQI با مقادیر اندازه‌گیری شده، از معیارهای ارزیابی استفاده شد.

جنگل تصادفی (RF)

Random Forest یک الگوریتم یادگیری گروهی نظارت‌شده است که توسط بریمن (۲۰۰۱) ارائه شده است. این الگوریتم در طول آموزش چندین درخت تصمیم را ساخته و میانگین خروجی‌های آنها را برای کاربرد در رگرسیون محاسبه می‌کند (۶ و ۷). در این الگوریتم مشکل بیش‌برازش کردن درخت تصمیم به ویژه زمانی که درخت دارای شاخه‌های زیادی باشد، حل می‌شود. این الگوریتم روش درخت تصمیم‌گیری را از طریق کیسه‌گذاری (Bagging) بهبود می‌بخشد. کیسه‌گذاری، ترکیبی از دو مرحله زیر است:

۱- Bootstrapping: یک تکنیک آماری نمونه‌برداری مجدد از داده‌های مرحله آموزش است. بدین معنا که چندین بار به طور تصادفی داده‌های آموزش از کل داده‌ها انتخاب شده و هر بار یک درخت تصمیم‌گیری ساخته می‌شود. مجموع این درخت‌ها جنگل تصمیم‌گیری را تشکیل می‌دهد.

۲- Aggrigation: در این مرحله نتایج هر درخت تصمیم‌گیری که در مرحله قبل ساخته شده، جمع‌آوری شده و اگر مسئله از

بد، متوسط، خوب و عالی طبقه‌بندی می‌شود (۱۲، ۳۵ و ۳۷). شاخص NSFQI از طریق معادله ۱ محاسبه شده است (۳۵):

$$NSFWQI = \sum_{i=1}^n q_i w_i \quad (1)$$

که در این معادله، w_i وزن نسبی هر ویژگی کیفیت آب به دلیل اهمیت آن برای آب آشامیدنی است (مجموع w_i ها برابر با ۱ است). وزن هر ویژگی با استفاده از دستورالعمل ارائه شده توسط سازمان محیط زیست ایران و اهمیت هر ویژگی برای آب آشامیدنی در جدول ۲ ارائه شده است (۱۰).

همچنین q_i رتبه‌بندی کیفیت هر ویژگی مشخص و نسبت مقدار واقعی به مقدار استاندارد آب آشامیدنی است (معادله ۲) (۳۵):

$$q_i = \frac{c_i}{s_i} \times 100 \quad (2)$$

که c_i مقدار اندازه‌گیری شده ویژگی و s_i مقدار استاندارد ویژگی مد نظر است.

مدل‌های یادگیری ماشین

مدل‌های یادگیری ماشین به دلیل توانایی آنها در ایجاد روابط پیچیده بین ویژگی‌های ورودی و متغیرهای هدف، به‌طور گسترده در مدل‌سازی استفاده شده‌اند. مدل‌های یادگیری ماشین بر اساس محاسبات داخلی، امتیازها یا وزن‌های اهمیت به متغیرهای ورودی اختصاص داده و آنها را ارزیابی می‌کنند (۳۰). این امتیازها نشان‌دهنده میزان تأثیر هر متغیر ورودی بر پیش‌بینی‌های ویژگی هدف است. انتخاب ویژگی‌های ورودی نقشی کلیدی در پایداری و انعطاف‌پذیری مدل پیش‌بینی‌کننده دارد. با بررسی و انتخاب مناسب متغیرهای ورودی مرتبط، می‌توان ثبات مدل‌ها را افزایش داده و از پیش‌بینی‌های قابل‌اعتماد در سری‌های مختلف اطمینان حاصل کرد (۱۳، ۱۴ و ۳۳). در این پژوهش آنالیز اهمیت ویژگی توسط روش SHAP (SHapley Additive exPlanations) انجام شد. آنالیز SHAP مبتنی بر تئوری بازی‌ها و برای توضیح خروجی هر مدل یادگیری ماشین بوده و همچنین اثرگذاری هر یک از ویژگی‌ها را روی ویژگی هدف مشخص می‌کند. بر اساس آنالیز

جدول ۱. طبقه‌بندی NSFQI

کیفیت آب	مقدار NSFQI
خیلی بد	۰-۲۵
بد	۲۶-۵۰
متوسط	۵۱-۷۵
خوب	۷۶-۹۰
عالی	۹۱-۱۰۰

جدول ۲. ویژگی‌های کیفیت آب و وزن ویژگی‌ها برای NSFQI

وزن‌های اصلی			وزن‌های اصلاح‌شده		
شماره	ویژگی	وزن	شماره	ویژگی	وزن
۱	اکسیژن محلول	۰/۱۷	۱	pH	۰/۰۵
۲	pH	۰/۱۱	۲	هدایت الکتریکی	۰/۰۹۶
۳	اکسیژن‌خواهی بیولوژیکی	۰/۱۱	۳	کل جامدات	۰/۱
۴	دما	۰/۱۰	۴	سولفات	۰/۱
۵	فسفات	۰/۱۰	۵	بیکربنات	۰/۱
۶	نیتрат	۰/۱۰	۶	کلر	۰/۱
۷	کدورت	۰/۰۸	۷	کلسیم	۰/۱۴۵
۸	کل جامدات	۰/۰۷	۸	منیزیم	۰/۱۵
۹	کلیفرم مدفوعی	۰/۱۶	۹	سدیم	۰/۱
			۱۰	سختی	۰/۰۵۹
	کل	۱		کل	۱

نوع رگرسیون باشد، میانگین آن به عنوان خروجی و اگر از نوع طبقه‌بندی باشد، نتایج شمارش و رأی‌گیری آن به عنوان خروجی در نظر گرفته می‌شود.

ماشین بردار پشتیبان (SVM)

ماشین‌های بردار پشتیبان (SVM) که در ابتدا توسط واپنیک (۱۹۹۵) معرفی شد، یک کلاس همه‌کاره از الگوریتم‌های یادگیری نظارت‌شده هستند که در اصل برای دسته‌بندی طراحی شده‌اند، اما به روش رگرسیون به عنوان رگرسیون بردار پشتیبان (SVR) گسترش یافته‌اند (۳۶). در این روش یک مرز جداکننده (مرز تصمیم) بین داده‌ها ترسیم می‌شود. اگر تعداد ویژگی‌ها دو

عدد باشد این مرز یک خط، اگر این تعداد ۳ عدد باشد مرز یک صفحه و اگر تعداد بیش از ۳ عدد باشد، مرز یک آبر صفحه خواهد بود. مرز تصمیم باید بیشترین فاصله را با نزدیک‌ترین نمونه‌های آموزشی داشته باشد. نزدیک‌ترین نمونه‌های آموزشی بردار پشتیبان نام دارد. به فاصله بین مرز تصمیم و نقاط بردار پشتیبان حاشیه گفته می‌شود. در این روش هدف این است که حاشیه و احتمال خطا کمینه شود. در این پژوهش برای مرز تصمیم از تابع پایه شعاعی (Radial basis function) استفاده شده که به آن Kernel می‌گویند. همچنین ضریب جریمه (C) برابر با ۲/۲ در نظر گرفته شده است. هرچه ضریب جریمه بیشتر باشد، به این معناست که حاشیه کوچک و خطا کاهش می‌یابد.

معیارهای ارزیابی

۱- ضریب تعیین (R^2)

معیار R^2 نسبت واریانس در متغیر هدف را که توسط مدل توضیح داده شده، اندازه‌گیری می‌کند. این معیار به صورت معادله ۳ تعریف می‌شود:

$$R^2 = 1 - \left[\frac{\sum_{i=1}^n (TP_{actual,i} - TP_{pred,i})^2}{\sum_{i=1}^n (TP_{actual,i} - \text{mean}(TP_{actual}))^2} \right] \quad (3)$$

که در این معادله مقدار R^2 از ۰ تا ۱ متغیر است. مقدار بیشتر نشان‌دهنده تناسب بهتر مدل با داده‌هاست. مقدار برابر با ۱ نشان‌دهنده یک تناسب کامل است؛ درحالی‌که مقدار نزدیک به صفر نشان می‌دهد که مدل در گرفتن واریانس در داده‌ها ناموفق بوده است.

۲- ضریب جرم باقی‌مانده (CRM)

ضریب جرم باقی‌مانده یا CRM (معادله ۴) بیان‌کننده آن است که مدل مدنظر مقدار پارامتر هدف را بیشتر (مقادیر منفی) یا کمتر (مقادیر مثبت) از مقادیر واقعی برآورد می‌کند. مقادیر نزدیک به صفر برای این شاخص نشان‌دهنده عملکرد بهتر مدل است.

$$CRM = \frac{\sum_{i=1}^n TP_{actual,i} - \sum_{i=1}^n TP_{pred,i}}{\sum_{i=1}^n TP_{actual,i}} \quad (4)$$

۳- جذر میانگین مربعات خطای نرمال‌شده (NRMSE)

این معیار (معادله ۵)، حاصل تقسیم جذر میانگین مربعات خطا بر بازه ویژگی هدف است. به همین دلیل عددی بدون بُعد و واحد بوده و برای گزارش بسیار مناسب است.

$$NRMSE = \frac{\sqrt{\sum_{i=1}^n (TP_{actual,i} - TP_{pred,i})^2}}{\text{mean}(TP_{actual})} \quad (5)$$

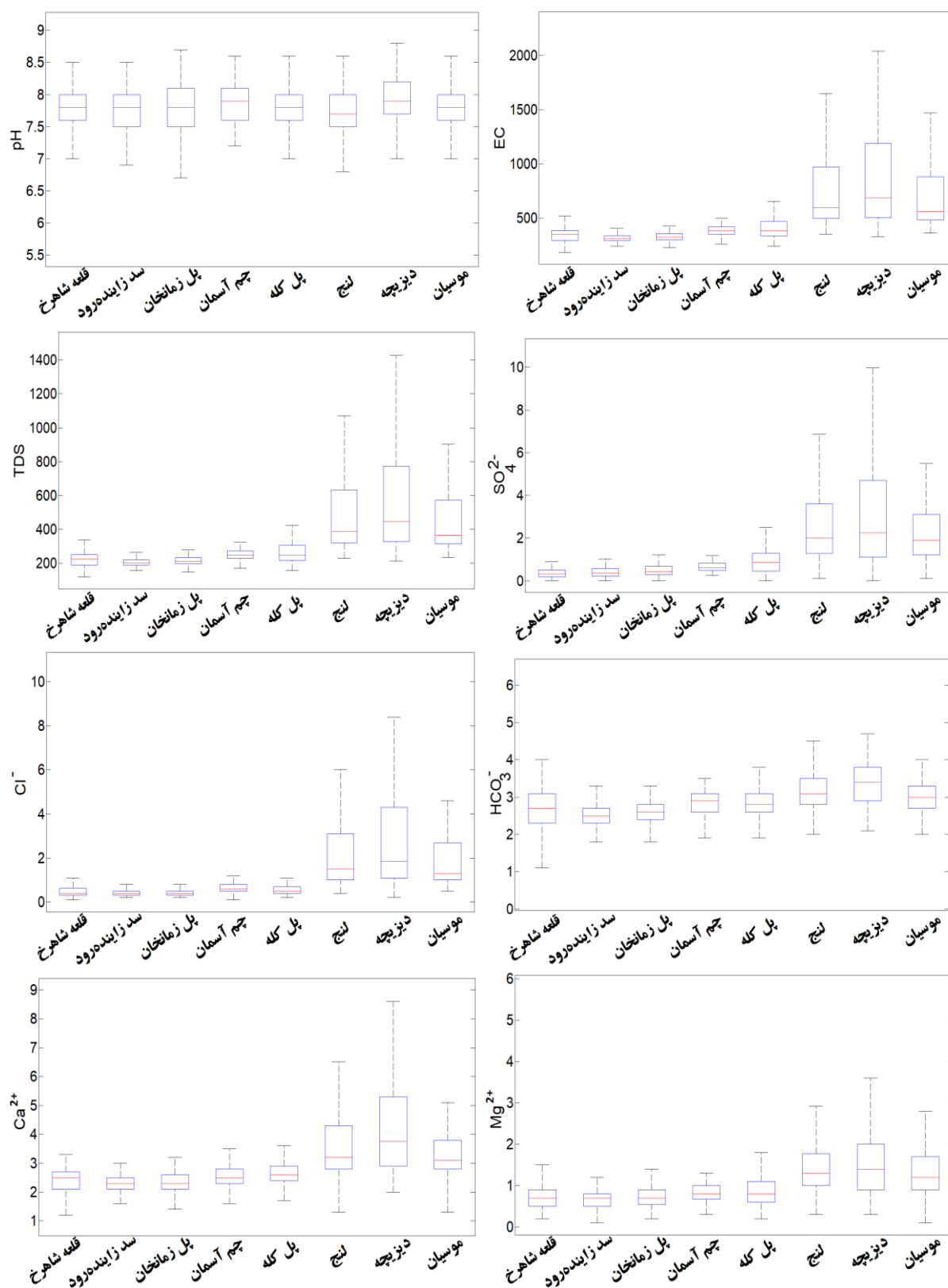
نتایج و بحث

در شکل ۲ نمودارهای جعبه‌ای ویژگی‌های استفاده شده در این پژوهش ارائه شده است.

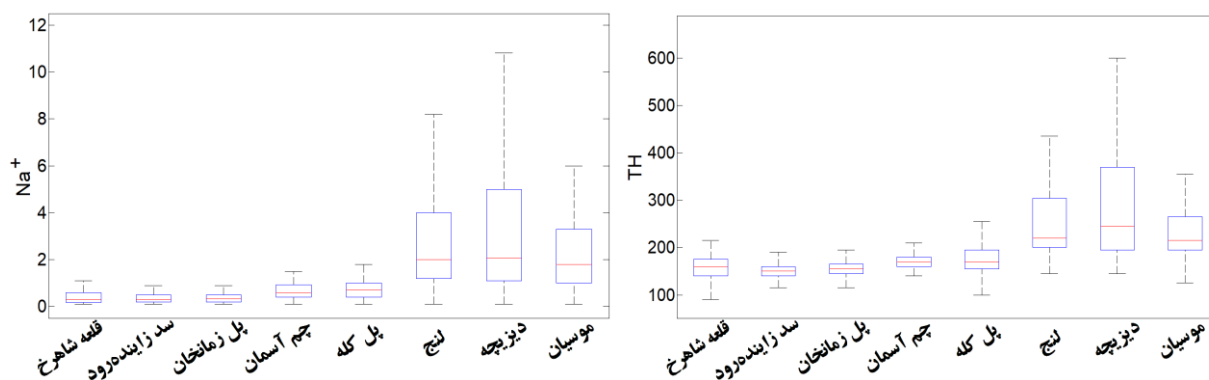
نتایج نشان داد که میانگین pH، استانداردهای کیفیت آب سطحی را برآورده کرده و انحراف معیار در بازه ۰/۳۲۴-۰/۳۸۱ میانگین نشان‌دهنده خواص فیزیکی به نسبت پایدار است. میانگین هدایت الکتریکی در تمام ایستگاه‌ها در محدوده استاندارد WHO قرار گرفت؛ اگرچه بیشینه آن در ایستگاه‌های دیزپچه، لنج و موسیان از مقدار مجاز بیشتر بود. در ایستگاه‌های نامبرده انحراف معیار نیز مقادیر زیادی نشان داد. گفتنی است که EC شاخص کیفیت فیزیکی آب است که بیشتر تحت تأثیر شرایط هواشناسی و املاح آب قرار می‌گیرد (۹ و ۲۰). از سوی دیگر میانگین ویژگی TDS در ایستگاه‌های دیزپچه و لنج، و میانگین سولفات در ایستگاه دیزپچه بیشتر از حد استاندارد WHO قرار گرفت که نشان از زیاد بودن میزان بار رسوب معلق در این ایستگاه‌ها بوده و به منظور تأمین آب شرب در آنها نیاز به تمهیدات جدی است. دو آنیون بیکربنات و کلر، و همچنین سه کاتیون کلسیم، منیزیم و سدیم مقادیر میانگین بسیار کمتر از حد استاندارد نشان دادند. ویژگی TH هم اگر چه از حد استاندارد کیفیت آبهای سطحی گذر نکرده، اما مقادیر بیشتر از ۱۰۰ mg/lit نشان داد.

از آنجایی که متغیرهای مختلف کیفیت آب به‌طور یکسان در پدیدآمدن مقادیر WQI مشارکت نمی‌کنند، می‌توان چندین متغیر تعیین‌کننده که وضعیت کیفیت آب را به‌طور معقول و بدون از دست‌دادن اطلاعات کلیدی منعکس کنند، شناسایی کرد. در این پژوهش تلاش شد تا با استفاده از شاخص‌های حیاتی کمتری که کیفیت آب را به‌طور مناسب منعکس می‌کند، یک ارزیابی جامع پیش‌بینی‌کننده کیفیت آب انجام شود. این آنالیز با استفاده از روش SHAP انجام شد که نتیجه آن در شکل ۳ ارائه شده است.

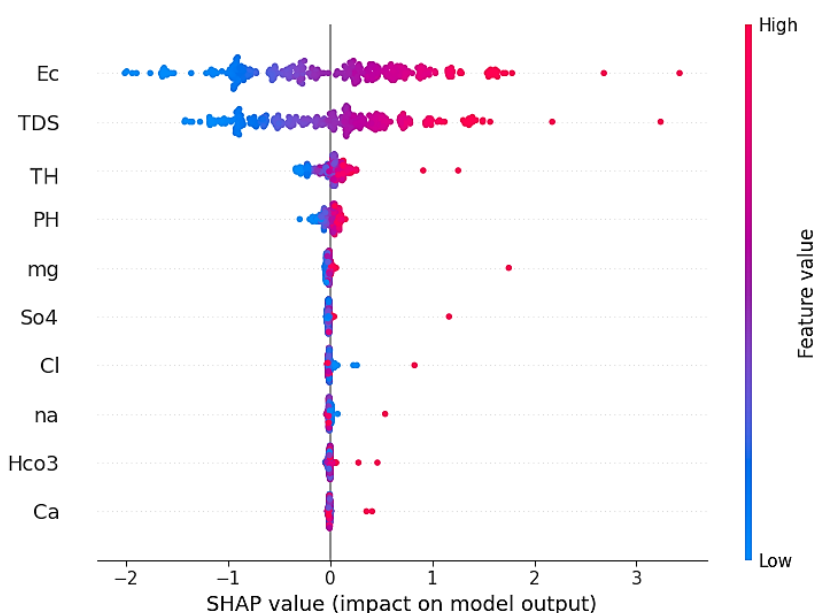
همان‌طور که در شکل ۳ نشان داده شده، بر اساس رتبه‌بندی اهمیت مدل جنگل تصادفی، EC تأثیرگذارترین عامل در پیش‌بینی WQI بوده و پس از آن TDS، TH، pH، Mg^{2+} ، SO_4^{2-} ، Cl^- ، Na^+ و HCO_3^- قرار گرفتند. شاخص‌های رتبه برتر آنهایی بودند که به درجات مختلف از حد مطلوب و یا استاندارد فراتر رفتند. بدین‌صورت که میانگین مقدار EC اگر چه از حد استاندارد فراتر نبوده، اما در حد مطلوب آب شرب



شکل ۲. مشخصات آماری ویژگی‌های مورد مطالعه



شکل ۲. مشخصات آماری ویژگی‌های مورد مطالعه (ادامه شکل ۲)



شکل ۳. رتبه‌بندی اهمیت ۱۰ ویژگی در الگوریتم جنگل تصادفی

شکل ۳، تعداد ۶ سری تعریف شد که در جدول ۳ ارائه شده است. در هر سری بسته به درجه اهمیت به‌دست‌آمده ترکیب‌های مختلفی از ویژگی‌های ورودی به مدل در نظر گرفته شد.

تمامی ترکیب‌های تعریف‌شده در جدول ۳ برای همه ۸ ایستگاه اجرا شد. نتایج نشان داد که برای تمامی ایستگاه‌ها سری ۳ که شامل پارامترهای EC، TDS و TH بوده، بهترین نتایج را ارائه کرده است. مقایسه نتایج سه معیار ارزیابی در پیش‌بینی WQI در تمامی ایستگاه‌ها و برای سری ۳ در جدول ۴ ارائه شده‌اند. نتایج نشان داد که در ایستگاه‌های

همچنین میانگین مقدار نیز قرار نداشته است. همچنین میانگین مقدار TDS در دو ایستگاه از حد استاندارد فراتر بوده و در بقیه ایستگاه‌ها مقادیر میانگین به بیشینه استاندارد نزدیک‌تر بوده است. از سوی دیگر اگرچه میانگین TH از حد استاندارد 500 mg/lit که توسط WHO ارائه شده فراتر نرفته است، اما از حد مطلوب و متعادل 100 mg/lit بیشتر بوده است. کمترین اهمیت HCO_3^- و Ca^{2+} احتمالاً به دلیل مقادیر نسبتاً کم آنها در کل سری زمانی بود، با بازه میانگین غلظت‌های $279/694 \text{ meq/lit}$ برای کلسیم و $519/345 \text{ meq/lit}$ برای بیکربنات، که تأثیر کمی بر کیفیت آب داشت. بر اساس درجات اهمیت به‌دست‌آمده از

قلعه‌شاهرخ، سد تنظیمی زاینده‌رود و سد چم‌آسمان،

جدول ۳. نتایج انتخاب ویژگی‌ها برای پیش‌بینی شاخص کیفیت آب (WQI)

شماره سری	ویژگی	درجه اهمیت ویژگی	شماره سری	ویژگی	درجه اهمیت ویژگی
۱	EC, TDS, TH, pH, Mg ²⁺	۱/۵۰۳	۴	EC, TDS	۱/۳۰۲
۲	EC, TDS, TH, pH	۱/۴۷۶	۵	TDS	۰/۵۷۹
۳	EC, TDS, TH	۱/۴۱۷	۶	EC	۰/۷۲۳

جدول ۴. معیارهای ارزیابی برای سری شماره ۳، مدل‌ها و ایستگاه‌های مختلف

نام ایستگاه	مدل یادگیری ماشین	R ²	CRM	NRMSE
قلعه‌شاهرخ	RF	۰/۹۷۱	۰/۰۰۰۳۶	۱/۹۳۷
	SVM	۰/۹۸۲	-۰/۰۰۱۹۹	۱/۵۱۶
سد تنظیمی زاینده‌رود	RF	۰/۹۵۰	-۰/۰۰۱۱۳	۱/۶۸۵
	SVM	۰/۹۵۹	-۰/۰۰۳۱۲	۱/۵۱۶
پل زمان‌خان	RF	۰/۹۱۰	-۰/۰۰۰۲۷	۲/۳۸۳
	SVM	۰/۹۳۱	۰/۰۰۳۷۷۶	۲/۰۹۳
سد چم‌آسمان	RF	۰/۹۵۶	۰/۰۰۴۵۰	۱/۶۳۶
	SVM	۰/۹۷۱	۰/۰۰۴۵۷۶	۱/۳۲۱
پل کله	RF	۰/۹۹۴	۰/۰۰۴۶۹۸	۱/۷۶۰
	SVM	۰/۹۳۱	-۰/۰۰۰۲۰	۵/۸۹۸
دیزیچه	RF	۰/۹۹۹	۰/۰۰۰۱۳۸	۱/۲۸۰
	SVM	۰/۹۹۷	۰/۰۱۴۲۹۳	۲/۶۳۳
لنج	RF	۰/۹۹۸	۰/۰۰۳۰۷۱	۱/۷۸۹
	SVM	۰/۹۹۶	-۰/۰۱۸۰۲	۲/۵۴۸
موسیان	RF	۰/۹۹۳	-۰/۰۰۱۱۴	۲/۴۴۰
	SVM	۰/۹۶۴	۰/۰۰۰۶۷	۵/۵۱۶

الگوریتم SVM نسبت به الگوریتم جنگل تصادفی نتایج بهتری داشته است. این در حالی است که در ایستگاه‌های پل زمان‌خان، پل کله، دیزیچه، لنج و موسیان مقایسه معیارهای ارزیابی نشان داد که الگوریتم جنگل تصادفی نسبت به الگوریتم SVM نتایج بهتری ارائه کرده است.

شایان ذکر است که در انتخاب بهترین سری از بین شش

سری تعریف شده، به دلیل نزدیک بودن نتایج معیارهای ارزیابی می‌توان سری‌های دیگری به جای سری ۳ در نظر گرفت. برای نمونه در ایستگاه پل کله نتایج نشان داد که مقادیر پارامترهای ارزیابی برای سری شماره ۶ بسیار به مقادیر مشابه در سری ۳ نزدیک بوده است؛ از این‌رو چنانچه دقت کمتری در برآورد را بتوان قابل قبول دانست، می‌توان سری ۶ (استفاده از ویژگی

TDS به منظور برآورد ویژگی کیفیت آب) را نیز به عنوان سری مناسب در این ایستگاه پیشنهاد داد. همچنین در ایستگاه موسیان با در نظر گرفتن اندک دقت کمتر می توان سری ۵ را نیز که تنها نیاز به اندازه گیری EC به منظور برآورد ویژگی WQI دارد، معرفی کرد. اما در ایستگاه های نامبرده همچنان سری ۳ بهترین نتایج را نشان داد.

مقایسه با پژوهش های پیشین انجام شده نشان داد که چن و همکاران نیز در پژوهش خود مدل RF را به منظور برآورد شاخص کیفی آب به عنوان بهترین مدل معرفی کردند. در این پژوهش ویژگی هایی همچون DO، COD، NH₃-N و pH استفاده شده که در این بین تنها پارامتر ورودی مشترک با پژوهش حاضر pH بوده است. این در حالی است که مدل پیشنهادی چن و همکاران مشابه با پژوهش حاضر بوده است. همچنین رحمان و همکاران در پژوهش خود کارایی روش های یادگیری ماشین را برای پیش بینی WQI بررسی کرده و ۱۲ پارامتر را به عنوان پارامترهای ورودی مدل ها استفاده کرده که از این بین ۶ پارامتر با پژوهش حاضر مشترک بوده است. در پژوهش یاد شده نیز مدل RF به عنوان مدل مناسب به منظور پیش بینی WQI معرفی شده است. همچنین یوسری و همکاران با کاربرد مدل های SVM و XGBoost اظهار داشتند که در پیش بینی شاخص کیفیت آب با استفاده از پارامترهای ورودی pH، EC، BOD، نترات، کلیفرم مدفوعی و کلیفرم کلی، XGBoost دقت بیشتری نشان داده است. با توجه به تفاوت داده های ورودی مدل ها در این پژوهش با مطالعه حاضر، می توان اظهار داشت که با تغییر پارامترهای ورودی و اضافه کردن پارامترهای شیمیایی مدل XGBoost نیز می تواند به عنوان یک مدل پیشنهادی برای ایستگاه های استفاده شده در این پژوهش ارائه شود.

نتیجه گیری

یافته های این پژوهش نشان داد که در چهار ایستگاه الگوریتم SVM و به طور مشابه در چهار ایستگاه الگوریتم جنگل تصادفی

برای پیش بینی شاخص کیفیت آب رودخانه زاینده رود در ایستگاه های متعدد بهترین عملکرد را نشان داد. ضریب تعیین در برآوردها در بازه ۰/۹۹۹-۰/۸۱۸ و شاخص NRMSE در محدوده ۷/۱۹۴-۱/۲۸۰ قرار گرفت. از سوی دیگر بررسی نتایج به دست آمده برای ضریب CRM نشان داد که در ایستگاه های پل زمان خان، سد چم آسمان، پل کله، دیزیچه و لنج پیش بینی ها کم برآورد بوده و در ایستگاه های قلعه شاهرخ، سد تنظیمی زاینده رود و موسیان بیش برآورد رخ داده است. با وجود این نتایج، به دلیل اینکه تمامی ضرایب نزدیک به صفر بوده است، برآوردها قابل قبول بوده اند. بر اساس نتایج به دست آمده تفاوت مقادیر R² بسیار کم بوده، با این وجود در بین ۸ ایستگاه موجود بیشترین مقدار R² مربوط به ایستگاه دیزیچه و کمترین R² را ایستگاه سد تنظیمی زاینده رود به خود اختصاص داد. نتایج حاصل از تحلیل اهمیت ویژگی ها نشان داد که ویژگی های EC، TDS، TH و Mg به ترتیب بیشترین اهمیت را در برآورد شاخص کیفیت آب (WQI) داشته است. از طرفی Ca²⁺ کمترین اهمیت را در بین ویژگی ها نشان داد. نتایج تحلیل عاملی نشان داد که سه ویژگی EC، TDS و TH نقش کلیدی در پیش بینی شاخص کیفیت آب داشته و بنابراین اندازه گیری پیوسته و دقیق آن باید در برنامه ریزی منابع آب مورد توجه قرار گیرد؛ بنابراین از آنجا که هدف هوش مصنوعی و به عبارت دیگر یادگیری ماشین کاهش زمان و هزینه در مطالعات است، با جایگزینی ۳ ویژگی به جای ۱۰ ویژگی می توان به پیش بینی مناسبی از شاخص کیفیت آب (WQI) رسید.

نتایج این پژوهش به مدیران و سیاست گذاران آب در محاسبه دقیق و هدفمند شاخص کیفیت آب برای رودخانه ها و نهرها کمک می کند. این امر با کاهش زمان محاسبات، هزینه های کمتر و نظارت بر ایستگاه های آلوده به دست می آید. چنین رویکردی می تواند استراتژی های مدیریت منابع آب را افزایش داده و به مدیریت پایدار رودخانه ها کمک کند. با این حال، اذعان به چندین محدودیت در این مطالعه مهم است که در درجه اول مربوط به انتخاب متغیرهای فیزیکی و شیمیایی و مکان های نمونه برداری محدود است.

پیشنهاد می‌شود پژوهش‌های آتی با تعداد بیشتری از ایستگاه‌های نمونه‌برداری و طیف وسیع‌تری از متغیرهای کیفیت آب برای پیش‌بینی شاخص کیفیت آب رودخانه زاینده‌رود با استفاده از رویکردهای مدل‌سازی ترکیبی توسعه یابد. این رویکردها نقاط قوت چندین الگوریتم یادگیری ماشین را ترکیب کرده و می‌توانند برای بهبود عملکرد در ایستگاه‌های مختلف بررسی شوند. همچنین توصیه می‌شود مدل پیشنهادی در رودخانه‌های مختلف با شرایط آبی - اقلیمی متفاوت اعتبارسنجی شود.

References

منابع مورد استفاده

1. Aazami, J., A. Esmaili-Sari, A. Abdoli, H. Sohrabi and P.J. Van Den Brink. 2015. Monitoring and assessment of water health quality in the Tajan River, Iran using physicochemical, fish and macroinvertebrates indices. *Journal of Environmental Health Science and Engineering* 13(1): 29.
2. Akhtar, N., M. I. Syakir Ishak, S. A. Bhawani and K. Umar. 2021. Various natural and anthropogenic factors responsible for water quality degradation: A review. *Water* 13(19): 2660.
3. Akkoyunlu, A. and M.E. Akiner. 2012. Pollution evaluation in streams using water quality indices: a case study from Turkey's Sapanca Lake Basin. *Ecological Indicators* 18: 501–511.
4. Alexakis, D., V. A. Tsihrintzis, G. Tsakiris and G. D. Gikas. 2016. Suitability of water quality indices for application in lakes in the Mediterranean. *Water Resources Management* 30(5): 1621–1633.
5. Baghapour, M. A. and M. R. Shooshtarian. 2017. Assessment of groundwater quality for drinking purposes using water quality index (WQI) in Shiraz, Iran (2011 to 2015). *Iranian Journal of Health, Safety and Environment* 5(1): 888–893.
6. Breiman, L. 2001. Random forests. *Machine Learning* 45(1): 5–32.
7. Bui, D.T., K. Khosravi, J. Tiefenbacher, H. Nguyen and N. Kazakis. 2020. Improving prediction of water quality indices using novel hybrid machine-learning algorithms. *Science of the Total Environment* 721: 137612. <https://doi.org/10.1016/j.scitotenv.2020.137612>.
8. Chen, K., H. Chen, C. Zhou, Y. Huang, X. Qi, R. Shen, F. Liu, M. Zuo, J. Wang, Y. Zhang, D. Chen, X. Chen, Y. Deng and H. Ren. 2020. Comparative analysis of surface water quality prediction performance and identification of key water parameters using different machine learning models based on big data. *Water Research* 171: 115454. doi:10.1016/j.watres.2019.115454
9. Cheng, G., M. Wang, Y. Chen and W. Gao. 2020. Source apportionment of water pollutants in the upstream of Yangtze River using APCS-MLR. *Environmental Geochemistry and Health* 42(11): 3795–3810.
10. DOE (Department of Environment, Islamic Republic of Iran), Guideline for calculating water quality index in Iran (in Farsi). <https://www.doe.ir/portal/file/?958505/5->.
11. Ebraheim, G., M.H. Zonoozi and M. Saeedi. 2020. A comparative study on the performance of NSFQI m and IRWQI_{sc} in water quality assessment of Sefidroud River in northern Iran. *Environmental Monitoring and Assessment* 192(11): 1–13.
12. Effendi, H. and Y. Wardiatno. 2015. Water quality status of Ciambulawung River, Banten Province, based on pollution index and NSF-WQI. *Procedia Environmental Sciences* 24: 228–237.
13. Gültekin, B. and Sakar, B.E. 2018. Variable importance analysis in default prediction using machine learning techniques. In: Proceeding of 7th International Conference on Data Science, Technology and Applications. Institute for Systems and Technologies of Information, Control and Communication (INSTICC), Porto, Portugal.
14. Hameed, M., S.S. Sharqi, Z.M. Yaseen, H.A. Afan, A. Hussain and A. Elshafie. 2017. Application of artificial intelligence (AI) techniques in water quality index prediction: a case study in tropical region, Malaysia. *Neural Computing and Applications* 28(Suppl 1): 893–905.
15. Hoseinzadeh, E., H. Khorsandi, C. Wei and M. Alipour. 2015. Evaluation of Aydughmush river water quality using the National Sanitation Foundation Water Quality Index (NSFWQI), River Pollution Index (RPI), and Forestry Water Quality Index (FWQI). *Desalination and Water Treatment* 54(11): 2994–3002.
16. Horton, R.K. 1965. An index number system for rating water quality. *Journal of the Water Pollution Control Federation* 37(3): 300–306.
17. Içaga, Y. 2007. Fuzzy evaluation of water quality classification. *Ecological Indicators* 7(3): 710–718.
18. Javid, A., K. Yaghmaeian, E. Abbasi and A. Roudbari. 2014. An evaluation of water quality from Mojen River, by NSFQI index. *Journal of Ecological Engineering* 15(4): 1–6.
19. Khoi, D. N., N. T. Quan, D. Q. Linh, P. T. T. Nhi and N. T. D. Thuy. 2022. Using machine learning models for predicting the water quality index in the La Buong River, Vietnam. *Water* 14(10): 1552. doi: 10.3390/w14101552.

20. Li, X., P. Li, D. Wang and Y. Wang. 2014. Assessment of temporal and spatial variations in water quality using multivariate statistical methods: a case study of the Xin'anjiang River, China. *Frontiers of Environmental Science and Engineering* 8(6): 895–904.
21. Lin, Z., J. Y. Lim and J. M. Oh. 2024. Innovative interpretable AI-guided water quality evaluation with risk adversarial analysis in river streams considering spatial-temporal effects. *Environmental Pollution* 350: 124015.
22. Mahrooyan, F., L. Taghavi, M. Sarai Tabrizi and H. Babazadeh. 2021. Water quality assessment of Qazvin River using NSFQI Index for water quality classification. *Journal of Wetland Ecobiology* 12(1): 99-112.
23. Matta, G., A. Nayak, A. Kumar and P. Kumar. 2020. Water quality assessment using NSFQI, OIP and multivariate techniques of Ganga River system, Uttarakhand, India. *Applied Water Science* 10(9): 206.
24. Medeiros, A. C., K. R. F. Faial, K. do Carmo Freitas Faial, I. D. da Silva Lopes, M. de Oliveira Lima, R. M. Guimarães and N. M. Mendonça. 2017. Quality index of the surface water of Amazonian rivers in industrial areas in Pará. Brazil. *Marine Pollution Bulletin* 123(1-2): 156–164.
25. Misaghi, F., F. Delgosha, M. Razzaghamanesh and B. Myers. 2017. Introducing a water quality index for assessing water for irrigation purposes: A case study of the Ghezel Ozan River. *Science of the Total Environment* 589: 107-116.
26. Ocampo-Duque, W., N. Ferre-Huguet, J. L. Domingo and M. Schuhmacher. 2006. Assessing water quality in rivers with fuzzy inference systems: A case study. *Environment International* 32(6): 733-742.
27. Omid, A. and F. Shariati. 2021. Evaluation of Pasikhan River, north of Iran using water quality index (NSFWQI). *Caspian Journal of Environmental Sciences* 19(2): 219–230.
28. Palabiyik, S. and T. Akkan. 2024. Evaluation of water quality based on artificial intelligence: performance of multilayer perceptron neural networks and multiple linear regression versus water quality indexes. *Environment, Development and Sustainability* 1-24.
29. Rahaman, M. H., H. Sajjad, S. Hussain, M. Masroor and A. Sharma. 2024. Surface water quality prediction in the lower Thoubal river watershed, India: A hyper-tuned machine learning approach and DNN-based sensitivity analysis. *Journal of Environmental Chemical Engineering* 12(3): 112915.
30. Siham, A., S. Sara and A. Abdellah. 2021. Feature selection based on machine learning for credit scoring: an evaluation of filter and embedded methods. In: *Proceeding of 2021 International Conference on Innovations in Intelligent Systems and Applications (INISTA)*. Institute of Electrical and Electronics Engineers (IEEE), Kocaeli, Turkey.
31. Silvert, W. 2000. Fuzzy indices of environmental conditions. *Ecological Modelling* 130(1–3): 111–119.
32. Singh, S., A. Das and P. Sharma. 2024. Predictive modeling of water quality index (WQI) classes in Indian rivers: Insights from the application of multiple Machine Learning (ML) models on a decennial dataset. *Stochastic Environmental Research and Risk Assessment* 38(8): 3221-3238.
33. Singha, S., S. Pasupuleti, S.S. Singha, R. Singh and S. Kumar. 2021. Prediction of groundwater quality using efficient machine learning technique. *Chemosphere* 276: 130265. <https://doi.org/10.1016/j.chemosphere.2021.130265>
34. Sutadian, A. D., N. Muttill, A. G. Yilmaz and B. J. C. Perera. 2016. Development of river water quality indices-a review. *Environmental Monitoring and Assessment* 188(1): 58.
35. Tiwari, N. K., S. Das Gupta, H. S. Swain, D. N. Jha, S. Samanta, R. K. Manna, A. K. Das and B. K. Das. 2022. Water quality assessment in the ecologically stressed lower and estuarine stretches of river Ganga using multivariate statistical tool. *Environmental Monitoring and Assessment* 194(7): 469.
36. Vapnik, V.N. 1995. *The Nature of Statistical Learning Theory*. Springer-Verlag, New York.
37. Yadav, A. K., P. Khan and S. K. Sharma. 2010. Water quality index assessment of groundwater in todaraisingh tehsil of Rajasthan State, India—A greener approach. *E-Journal of Chemistry* 7(SUPPL. 1): 428–433. <https://doi.org/10.1155/2010/419432>
38. Yousefi, A. and S. Mahdian. 2015. The economic and social necessity of water reuse in Iran. *Water Reuse* 1(1): 1–7.
39. Yusri, H., A. Ab Rahim, S. Hassan, I. Halim and N. Abdullah. 2022. Water quality classification using SVM and XGBoost method. In: *IEEE 13th Control and System Graduate Research Colloquium (ICSGRC)*. Shah Alam, Malaysia. [doi: 10.1109/ICSGRC55096.2022.9845143](https://doi.org/10.1109/ICSGRC55096.2022.9845143).
40. Zamani-Ahmadmohammadi, R., F. Ghaedamini and M. Najafi. 2021. Spatial distribution pattern of water quality of Pireghar river with using water quality index (WQI) and geographic information system (GIS). *Journal of Environmental Science and Technology* 23(2): 73–86.