

مقاله علمی-پژوهشی

رویکرد تلفیقی در پایش کیفیت آب رودخانه سیاهرود با بهره‌گیری از روش‌های نوین تحلیل

داده

مجتبی خوش‌روش^{۱*}، لاله دیوبند هفشجانی^۲، مصطفی مرادزاده^۳، طاهره اسکندری^۴

تاریخ دریافت: ۱۴۰۴/۰۱/۳۱ تاریخ پذیرش: ۱۴۰۴/۰۳/۱۰

چکیده

شاخص کیفیت آب (WQI) ابزاری مؤثر برای ارزیابی منابع آبی است که با ترکیب پارامترهای مختلف در یک عدد ساده، امکان مقایسه و تصمیم‌گیری را فراهم می‌کند. با این حال، استفاده از وزن‌دهی‌های استاندارد بدون توجه به شرایط محلی ممکن است منجر به نتایج نادرست شود. این پژوهش با هدف بهینه‌سازی محاسبه WQI، از رویکردی تلفیقی شامل تحلیل مؤلفه‌های اصلی (PCA)، تصمیم‌گیری چندمعیاره (MERECE) و مدل‌های هوش مصنوعی بهره‌برده است. داده‌های ۲۰ ساله (۲۰۰۱-۲۰۲۱) مربوط به رودخانه سیاهرود در استان مازندران با ۱۱ پارامتر کیفی بررسی شد. کاهش ابعاد داده‌ها با PCA انجام گرفت و شاخص WQI با روش MERECE محاسبه شد. همچنین، پیش‌بینی WQI با دو الگوریتم رگرسیون خطی و جنگل تصادفی، در دو حالت داده‌های واقعی و کاهش‌یافته، انجام شد و با شاخص‌هایی نظیر MAPE، EVS و R² ارزیابی گردید. یافته‌ها نشان دادند کدورت به‌عنوان مؤلفه کلیدی در آلودگی فیزیکی با TSS، BOD و COD همبستگی بالایی دارد. مؤلفه‌های دیگر نمایانگر آلودگی‌های میکروبی، شیمیایی و حرارتی ناشی از فاضلاب و کشاورزی‌اند. میانگین WQI نشان‌دهنده کیفیت نامطلوب آب در بیشتر ایستگاه‌ها بود. مدل رگرسیون خطی عملکرد بهتری نسبت به جنگل تصادفی داشت، هرچند کاهش ابعاد کمی از دقت کاست. روش MERECE نشان داد کلیفرم کل و کدورت بیشترین وزن را دارند، که اهمیت آلودگی میکروبی و ذرات معلق را برجسته می‌سازد. نتایج بر ضرورت پایش مستمر و مدیریت منابع آلودگی تأکید دارند.

واژه‌های کلیدی: آلودگی آبی، شاخص کیفیت، کدورت، مؤلفه‌های اصلی

مقدمه

هستند. با رشد جمعیت، صنعتی شدن سریع و توسعه بدون برنامه‌ریزی، سطح آب‌ها از جمله رودخانه‌ها، دریاچه‌ها و برکه‌ها با خطراتی مواجه شده و کیفیت آب‌ها رو به وخامت می‌رود (et Prasad et al., 2022). ارزیابی کیفیت آب یک جزء اساسی از چارچوب‌های مدیریت و حفاظت محیط زیست در سراسر جهان را نشان می‌دهد. پیچیدگی سیستم‌های آبی که با پارامترهای فیزیکی، شیمیایی و بیولوژیکی متعدد مشخص می‌شود، نیازمند روش‌های ارزیابی جامع برای نظارت و طبقه‌بندی مؤثر آب‌ها است (Şahin Kiy and Arslan, 2021). شاخص WQI به عنوان ابزاری حیاتی برای ارزیابی و اطلاع رسانی کیفیت بدنه‌های آبی عمل می‌کند. این داده‌های پیچیده کیفیت آب را در یک مقدار عددی ساده می‌کند و امکان تفسیر و مقایسه آسان‌تر در منابع مختلف آب و در طول زمان را فراهم می‌کند. WQI معمولاً پارامترهای متعددی مانند دما، اکسیژن محلول، pH و آلاینده‌های مختلف را برای ارائه یک نمای کلی از شرایط کیفیت آب در خود جای می‌دهد (Jahan and Strezov, 2017);

بیش از ۹۷ درصد از سطح زمین را آب‌ها تشکیل می‌دهند که تنها ۱/۱ درصد آن آب شیرین است. از این مقدار، تنها یک درصد آن به آب‌های سطحی شیرین تعلق دارد (Mahbubul et al., 2023). آب‌های سطحی شیرین و کیفیت آن‌ها نقش حیاتی در توسعه اقتصادی ایفا می‌کنند و به سلامت انسان و محیط زیست مرتبط

- ۱- دانشیار، گروه مهندسی آب، دانشکده مهندسی زراعی، دانشگاه علوم کشاورزی و منابع طبیعی ساری، ساری، ایران
- ۲- استادیار گروه مهندسی محیط‌زیست، دانشکده مهندسی آب و محیط‌زیست، دانشگاه شهید چمران اهواز، اهواز، ایران
- ۳- آزمایشگاه زمین‌شناسی، دپارتمان علوم زمین، مدرسه عالی نرمال (ENS)، پاریس، فرانسه
- ۴- اداره کل حفاظت محیط زیست استان مازندران، پژوهشکده اکولوژی دریای خزر، ساری، ایران

(* - نویسنده مسئول: Email: m.khoshravesh@sanru.ac.ir)

و (۴) ایجاد یک شاخص کیفیت آب منطقه‌ای که با دقت بیشتری شرایط منحصر به فرد رودخانه سیاهرود را مشخص می‌کند. این رویکرد چند وجهی با مقابله هم‌زمان با چالش‌های بهینه‌سازی پارامتر و منطقه‌بندی وزن، به شکاف‌های تحقیقاتی شناسایی شده می‌پردازد.

مواد و روش‌ها

منطقه مورد مطالعه

رودخانه سیاهرود در استان مازندران و در محدوده مطالعاتی قائمشهر- جویبار جریان دارد. حوزه آبریز این رودخانه بین محدوده مطالعاتی ساری- نکا در شرق و بابل- آمل در غرب قرار گرفته است. نزدیک به ۹۳۵/۵ کیلومتر مربع این حوزه دشت و بقیه آن شامل ارتفاعات می‌باشد. بالاترین نقطه ارتفاعی این محدوده ۳۹۲ و پست‌ترین نقطه آن با ارتفاع ۲۶- متر از سطح دریای آزاد در خروجی حوزه قرار دارد. طول کل رودخانه ۵۹ کیلومتر بوده و منطقه مرتبط با آن دارای اقلیم معتدل فصلی است که در آن فصل‌های بارانی و خشک به‌طور مشخص از یکدیگر تفکیک می‌شوند. دمای هوا در این نواحی بین ۱۱ تا ۲۵ درجه سانتی‌گراد متغیر است و میانگین سالانه دما برابر با ۱۶ درجه سانتی‌گراد می‌باشد. استفاده‌های عمده انسانی از زمین در این حوضه شامل کشاورزی، صنعت، مناطق شهری و مختلط است (Charkhabi et al., 2005).

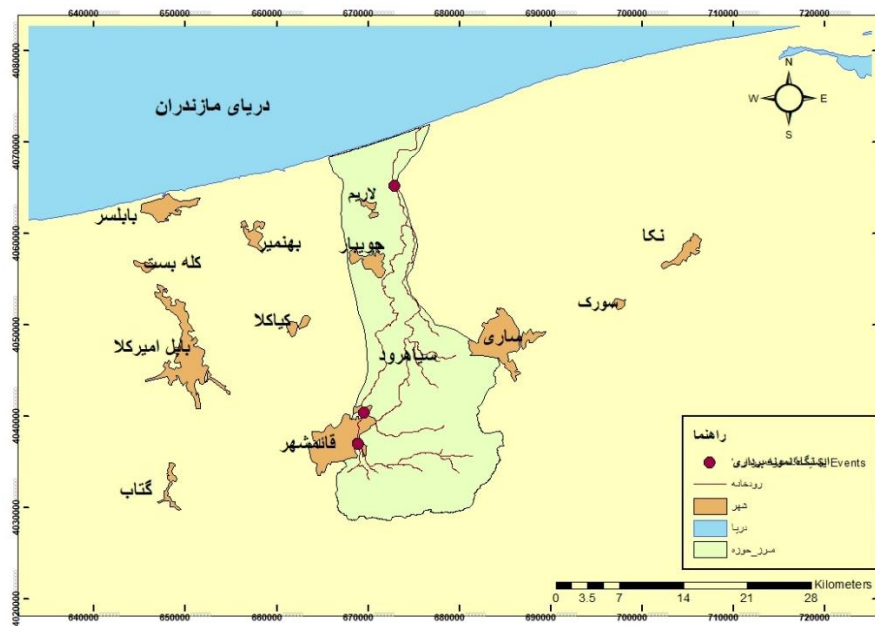
داده‌های مورد استفاده

داده‌های مورد استفاده در این پژوهش از بخش‌های مختلف (پایین دست، میان دست و بالا دست) رودخانه سیاهرود (شکل ۱) بین سال‌های ۲۰۰۱ تا ۲۰۲۱ جمع‌آوری شدند و شامل ۱۱ پارامتر مهم کیفیت آب pH، کل جامدات معلق (TSS)، کل جامدات محلول (TDS)، هدایت الکتریکی (EC)، کدورت (Turbidity)، فسفات (PO_4)، اکسیژن‌خواهی بیوشیمیایی (BOD)، اکسیژن‌خواهی شیمیایی (COD)، اکسیژن محلول (DO)، دما (T) و کلیفرم کل (Total coli) هستند. داده‌ها با بهره‌گیری از روش‌های آماری مختلف مورد تحلیل قرار گرفتند تا ویژگی‌های کلیدی آن‌ها شناسایی و ارزیابی شوند. ابتدا، به‌منظور درک توزیع و ویژگی‌های اصلی داده‌ها، میانگین و انحراف معیار محاسبه گردید. میانگین به عنوان نماینده‌ای از مرکز داده‌ها عمل کرده و انحراف معیار میزان پراکندگی داده‌ها نسبت به میانگین را اندازه‌گیری می‌کند. علاوه بر این، به‌منظور ارزیابی روابط بین ویژگی‌ها و شناسایی الگوهای موجود در داده‌ها، از ضریب همبستگی پیرسون برای تحلیل روابط خطی و از ضریب همبستگی اسپیرمن برای ارزیابی روابط ترتیبی استفاده شد. این تحلیل‌ها به‌منظور شناسایی ویژگی‌های برجسته و استخراج اطلاعات مفید، بهبود فرایند تصمیم‌گیری در مدل‌سازی و پیش‌بینی‌ها را فراهم ساخت.

(Bhutiani et al., 2016). توانایی این شاخص برای متراکم کردن پارامترهای متعدد کیفیت آب در یک امتیاز واحد نه تنها به ارتباطات عمومی کمک می‌کند، بلکه به سیاست‌گذاران در تصمیم‌گیری آگاهانه در مورد مدیریت منابع آب کمک می‌کند (Fashagba et al., 2024). با این حال، این رویکردهای مرسوم اغلب از طرح‌های وزن‌دهی پارامتر استاندارد استفاده می‌کنند که ویژگی‌های هیدرولوژیکی منطقه‌ای و تغییرات فصلی خاص منابع آبی را در نظر نمی‌گیرند. ارزیابی دقیق کیفیت آب پیامدهای گسترده‌ای برای سلامت اکوسیستم، ایمنی تامین آب عمومی و مدیریت منابع پایدار دارد. وزن‌دهی نامناسب پارامترها در محاسبات WQI می‌تواند منجر به طبقه‌بندی‌های نادرست قابل توجهی شود، به‌طور بالقوه نقاط حساس آلودگی را پنهان کند یا سطوح آلودگی را بیش از حد تخمین بزند، که متعاقباً تصمیمات خطمشی و استراتژی‌های کاهش را نادرست می‌سازد. با تشدید کمبود آب در سطح جهان و تغییرات آب و هوایی الگوهای هیدرولوژیکی را تغییر می‌دهد، نیاز به روش‌های ارزیابی کیفیت آب دقیق و خاص منطقه به‌طور فزاینده‌ای حیاتی می‌شود. توسعه رویکردهای بهینه که می‌تواند به‌طور موثر ویژگی‌های منحصر به فرد بدنه‌های آبی مانند رودخانه سیاهرود را به تصویر بکشد، برای مدیریت آگاهانه حوزه آبخیز و حفاظت از این منبع ارزشمند برای نسل‌های آینده ضروری است.

علی‌رغم پیشرفت‌های قابل توجه در روش‌های ارزیابی کیفیت آب، چندین شکاف حیاتی در ادبیات علمی وجود دارد. بهینه‌سازی و منطقه‌بندی وزن پارامترها در شاخص‌های کیفیت آب، با اجماع محدود در مورد روش‌های وزن‌دهی مناسب برای زمینه‌های مختلف جغرافیایی و اقلیمی، به‌طور کافی مورد توجه قرار نمی‌گیرد. علاوه بر این، در حالی که روش‌های کاهش ابعاد مانند تجزیه و تحلیل مؤلفه اصلی (PCA) و روش‌های تصمیم‌گیری چند معیاره (MCDM) به‌طور جداگانه در مطالعات کیفیت آب به کار گرفته شده‌اند، اثر هم‌افزایی بالقوه بهینه‌سازی هم‌زمان پارامترهای شاخص از طریق کاهش ابعاد در حالی که وزن پارامترها از طریق MCDM افزایش می‌یابد تا حد زیادی غیرقابل توصیف باقی می‌ماند. این شکاف دانش مانع از توسعه چارچوب‌های ارزیابی کیفیت آب قوی‌تر و سازگارتر منطقه‌ای شده است.

این پژوهش یک رویکرد یکپارچه جدید را با ترکیب کاهش ابعاد PCA با روش‌های MCDM و تکنیک‌های هوش مصنوعی برای بهینه‌سازی ارزیابی کیفیت آب برای رودخانه سیاهرود معرفی می‌کند. اهداف اولیه عبارتند از: (۱) شناسایی تاثیرگذارترین پارامترهای کیفیت آب از طریق PCA برای حذف افزونگی و افزایش کارایی محاسباتی. (۲) بهینه‌سازی وزن پارامترها با استفاده از تکنیک‌های متعدد MCDM برای انعکاس اهمیت منطقه‌ای. (۳) استفاده از مدل‌های هوش مصنوعی برای اعتبارسنجی و اصلاح چارچوب ارزیابی یکپارچه



شکل ۱- نقشه منطقه مورد مطالعه

همبستگی هستند، تبدیل می‌شوند. این مؤلفه‌های جدید ترکیبی خطی از متغیرهای اولیه هستند و تعدادی از آن‌ها، به‌عنوان جایگزین متغیرهای اولیه، در مدل‌سازی مورد استفاده قرار می‌گیرند (Bahrami et al., 2020). در PCA، ابتدا ماتریس همبستگی یا کواریانس متغیرها محاسبه شد. سپس، با تجزیه این ماتریس، بردارهای ویژه و مقادیر ویژه استخراج گردید. بردارهای ویژه جهت مؤلفه‌های اصلی را مشخص می‌کند و مقادیر ویژه میزان واریانس توضیح داده‌شده توسط هر مؤلفه را نشان می‌دهد. سپس انتخاب مؤلفه‌های اصلی بر اساس مقدار ویژه و درصد واریانس توضیح داده‌شده انجام شد. مؤلفه‌هایی که بیشترین واریانس را ثبت کردند، به‌عنوان مهم‌ترین مؤلفه‌ها در توصیف ساختار داده‌ها در نظر گرفته شد. در مرحله نهایی، امتیازات مؤلفه‌های اصلی با ضرب ماتریس داده‌های استاندارد شده در ماتریس بردارهای ویژه محاسبه گردید. این امتیازات، داده‌های اولیه را در فضای جدید مؤلفه‌های اصلی نمایش داد و در مدل‌سازی و تحلیل بیشتر مورد استفاده قرار گرفت (Karpagam and Christy, 2023).

الگوریتم‌های رگرسیون خطی و جنگل تصادفی

در این پژوهش، از دو الگوریتم پرکاربرد رگرسیون خطی و جنگل تصادفی برای مدل‌سازی داده‌ها استفاده شد. این الگوریتم‌ها به دلیل سادگی، کارایی و قابلیت‌های مختلفی که دارند، انتخاب شدند. رگرسیون خطی یک الگوریتم پایه در مسائل رگرسیونی است که برای مدل‌سازی روابط بین یک متغیر وابسته (هدف) و یک یا چند متغیر مستقل (ویژگی‌ها) استفاده می‌شود. اساس کار این الگوریتم،

محاسبه شاخص کیفیت آب

در پژوهش حاضر برای محاسبه شاخص کیفیت آب از رابطه (۱) استفاده شد.

$$WQI = \sum_{j=1}^n q_j \times w_j \quad (1)$$

$$q_j = \frac{C_j}{S_j} \quad (2)$$

در رابطه (۲)، C_j : غلظت پارامترهای شیمیایی در هر نمونه، S_j : غلظت مجاز آن پارامتر در آب براساس استانداردهای سازمان بهداشت جهانی، w_j : وزن هر پارامتر شیمیایی و WQI مقدار شاخص کیفیت آب می‌باشد. برای تعیین وزن پارامترها از روش MEREC به عنوان یکی از روش‌های تصمیم‌گیری چند معیاره استفاده شد (Keshavarz-Ghorabae et al., 2021). پس از محاسبه شاخص، کیفیت آب در یکی از چهار دسته زیر قرار می‌گیرد.

جدول ۱- طبقه‌بندی کیفیت آب بر اساس شاخص کیفیت آب

کیفیت آب	محدوده شاخص
عالی	۲۵-۰
خوب	۵۰-۲۶
ضعیف	۷۵-۵۱
بسیار ضعیف	۱۰۰-۷۶

تجزیه و تحلیل مؤلفه‌های اصلی (PCA)

در این روش، متغیرهای اولیه به مؤلفه‌هایی جدید که بدون

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{Y_i^{exp} - Y_i^{pred}}{Y_i^{exp}} \right| \quad (۴)$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (Y_i^{exp} - Y_i^{pred})^2}{\sum_{i=1}^n (Y_i^{exp} - Y_{ave}^{exp})^2} \quad (۵)$$

در معادله‌های بالا، Y_i^{pred} : مقدار \hat{Y} شاخص پیش بینی شده توسط مدل، Y_i^{exp} : مقدار Y شاخص محاسبه شده: Y_{ave}^{exp} ، مقدار متوسط شاخص محاسبه شده و n : تعداد داده‌ها می‌باشد.

نمره‌ی واریانس توضیح داده شده نشان دهنده‌ی نشان دهنده‌ی میزان واریانس داده‌های وابسته است که توسط مدل پیش بینی شده و مقدار آن بین ۰ و ۱ متغیر است. مقدار بالاتر این معیار نشان دهنده‌ی دقت بیشتر مدل در تبیین تغییرات داده‌های واقعی است. اندازه‌گیری درصد خطا یک معیار بسیار مفید برای ارزیابی عملکرد مدل‌ها به ویژه در مسائل پیش بینی است. به طور خاص، این معیار کمک می‌کند تا مشخص شود که پیش بینی‌های مدل چقدر از مقادیر واقعی انحراف دارند و این انحراف چند درصد است. ضریب تعیین، نسبت واریانس در متغیر وابسته را که توسط متغیرهای مستقل توضیح می‌شود را اندازه‌گیری می‌کند. مقدار بالاتر آن نشان دهنده‌ی تطابق بهتر مدل با داده‌ها است.

نتایج و بحث

پردازش و تحلیل آماری داده‌ها

تجزیه و تحلیل داده‌های پژوهش حاضر به طور خلاصه در جدول (۲) نشان داده شده است. مقدار اسیدیته در محدوده‌ی ۷/۱۱ تا ۸/۸۶ قرار دارد و میانگین آن ۷/۹۹ است که نشان دهنده‌ی قلیائیت ملایم آب است.

مدل سازی یک خط راست است که به بهترین شکل داده‌ها را برازش می‌کند. این مدل فرض می‌کند که رابطه‌ای خطی بین ورودی‌ها و خروجی وجود دارد. برای آموزش مدل، رگرسیون خطی به دنبال یافتن ضرایب خطی (وزن‌ها) است که کمترین خطا را در پیش بینی مقادیر هدف تولید کنند. ابتدا مدل رگرسیون خطی بر روی داده‌های واقعی آموزش داده شد و سپس برای داده‌های کاهش یافته نیز همین مدل اعمال گردید.

جنگل تصادفی یک الگوریتم پیشرفته و قدرتمند است که برای رگرسیون و طبقه بندی استفاده می‌شود. این الگوریتم از مجموعه‌ای از درخت‌های تصمیم گیری تشکیل شده است که به صورت تصادفی و مستقل از یکدیگر ساخته می‌شوند. درخت‌های تصمیم گیری به طور معمول داده‌ها را از طریق سوالات متوالی تقسیم بندی می‌کنند تا پیش بینی دقیق تری از مقدار هدف داشته باشند. در جنگل تصادفی، این درخت‌ها به طور مستقل از یکدیگر ساخته می‌شوند و پیش بینی نهایی با استفاده از میانگین پیش بینی‌های هر درخت به دست می‌آید. این الگوریتم قادر است با استفاده از برش‌های مختلف ویژگی‌ها و داده‌ها، عملکرد بسیار خوبی در مواجهه با داده‌های پیچیده و نویزدار داشته باشد. در این مطالعه نیز، مدل جنگل تصادفی ابتدا بر روی داده‌های واقعی آموزش داده شد و سپس همین فرایند برای داده‌های کاهش یافته نیز انجام شد.

ارزیابی عملکرد الگوریتم‌های مورد مطالعه

برای ارزیابی عملکرد الگوریتم‌های یادگیری ماشین از آماره‌های مختلفی مانند نمره‌ی واریانس توضیح داده شده (EVS)، اندازه‌گیری درصد خطا (MAPE) و ضریب تعیین (R^2) استفاده شد.

$$EVS = 1 - \frac{Var(Y_i^{exp} - Y_i^{pred})}{Var(Y_i^{exp})} \quad (۳)$$

جدول ۲- آمار توصیفی پارامترهای کیفی آب در ایستگاه‌های نمونه برداری

انحراف معیار Std	حداقل Min	حداکثر Max	متوسط Mean	پارامتر Parameter
۰/۲۹	۷/۱۱	۸/۸۶	۷/۹۹	اسیدیته
۲۶۰/۲۶	۴	۳۲۶۶	۸۸/۴۴	مواد جامد معلق
۴۰۰/۲	۲/۲۳	۴۴۰۸	۶۱۰/۶۲	مواد جامد محلول
۷۳۲/۰۲	۲۹۳	۷۷۸۰	۱۰۰۰/۳۲	هدایت الکتریکی
۲۳۵/۸۲	۳	۲۹۹۴	۷۰/۰۲	کدورت
۰/۳۱	۰/۰۵	۱/۹۵	۰/۴۷	فسفات
۱۴/۹۷	۰/۵	۱۳۶	۱۴/۱۲	نیاز اکسیژن خواهی بیوشیمیایی
۳۴/۷۷	۱	۳۱۳	۳۳/۰۷	نیاز اکسیژن خواهی بیوشیمیایی
۱/۹۱	۲/۱	۱۰/۸	۶/۱۸	اکسیژن محلول
۶/۲۷	۳	۳۲	۱۸/۴۰	دما
۴۶۷	۳۹	۲۴۰۰	۲۲۴۸/۹۱	کلیفرم کل
۱۳۷۰/۲۶	۹۱/۶۸	۱۷۴۷۰/۱۸	۵۱۳/۹۸	شاخص کیفیت آب

مقدار حداقلی ۲/۱ میلی‌گرم بر لیتر در برخی ایستگاه‌ها می‌تواند نشان‌دهنده‌ی شرایط کمبود اکسیژن باشد. داده‌های کلیفرم کل نشان می‌دهد که میانگین آن ۲۲۴۸ بوده و مقدار حداکثر آن به ۲۴۰۰ می‌رسد که نشان‌دهنده‌ی آلودگی میکروبی بالای آب است. در نهایت، مقدار شاخص کیفیت آب از ۹۱/۶۷ تا ۱۷۴۷۰/۱۷ متغیر است و مقدار میانگین آن ۵۱۳/۹۷ نشان‌دهنده‌ی کیفیت نامطلوب آب در اکثر ایستگاه‌ها است. به طور کلی، بررسی این داده‌ها نشان می‌دهد که کیفیت آب در برخی ایستگاه‌ها پایین بوده و تحت تأثیر آلاینده‌های فیزیکی، شیمیایی و میکروبی قرار دارد که می‌تواند ناشی از ورود فاضلاب، رواناب شهری، کشاورزی و صنعتی باشد. بنابراین، اقدامات مدیریتی برای کاهش آلاینده‌ها و بهبود کیفیت آب ضروری به نظر می‌رسد.

پردازش و تحلیل آماری داده‌ها

پیش از آغاز تحلیل‌های آماری، ضروری است اطمینان حاصل شود که داده‌های مورد استفاده دارای توزیع نرمال هستند. در این پژوهش، به منظور بررسی نرمال بودن توزیع ۱۱ پارامتر کیفی آب، از آزمون شاپیرو-ویلک بهره گرفته شد. نتایج این آزمون که در جدول ۳ ارائه شده‌اند، نشان می‌دهد که مقدار p برای تمام پارامترها بیشتر از ۰/۰۵ است؛ بنابراین، می‌توان فرض نرمال بودن داده‌ها را برای هر یک از پارامترها پذیرفت.

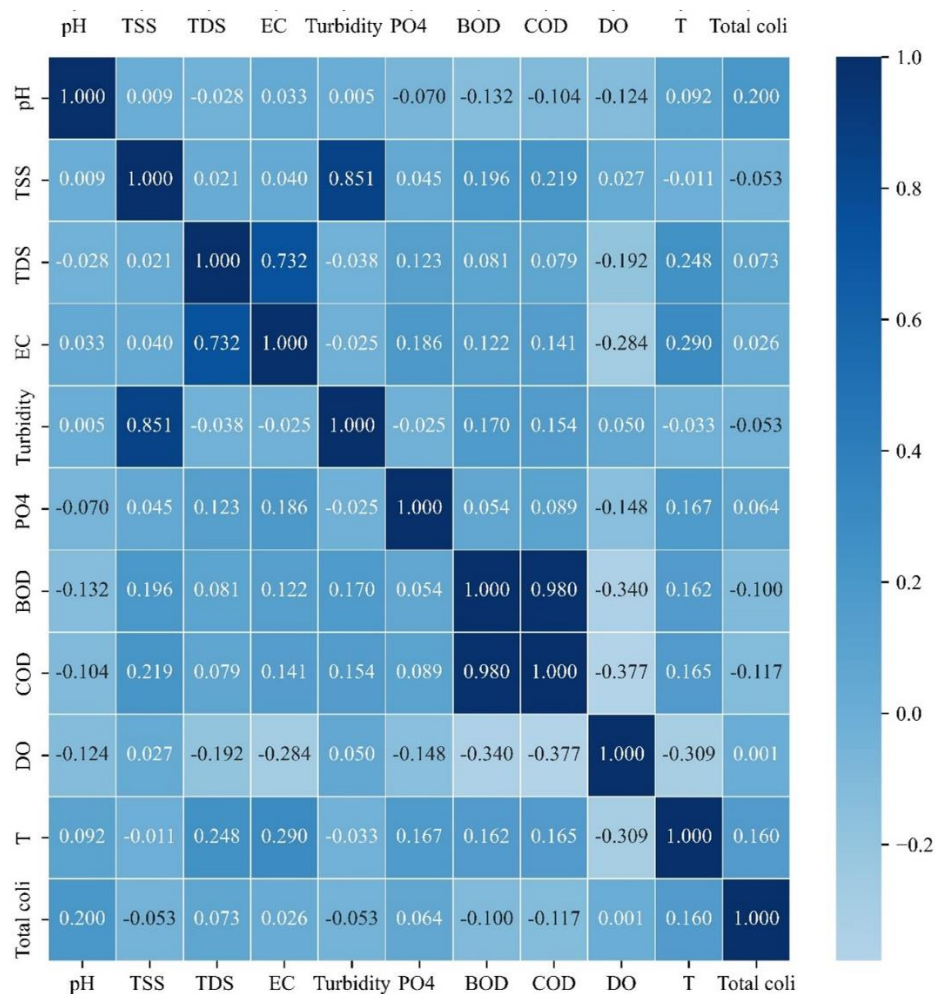
میزان کل جامدات معلق دارای پراکندگی بالایی بوده و مقدار آن از ۴ تا ۳۲۶۶ میلی‌گرم بر لیتر متغیر است. مقدار میانگین ۸۸/۴۴ میلی‌گرم بر لیتر نشان‌دهنده‌ی احتمال ورود رسوبات و آلاینده‌های ناشی از رواناب‌های سطحی است. میزان کل جامدات محلول در محدوده‌ی ۲/۲۳ تا ۴۴۰۸ میلی‌گرم بر لیتر قرار دارد و میانگین آن ۶۱۰/۶۲ میلی‌گرم بر لیتر است که نشان‌دهنده‌ی شوری متوسط آب است. هدایت الکتریکی نیز با مقدار میانگین ۱۰۰۰/۳۱ میکروزیمنس بر سانتی‌متر در برخی ایستگاه‌ها به مقدار حداکثری ۷۷۸۰ رسیده که احتمال ورود آلاینده‌های معدنی را نشان می‌دهد. کدورت در محدوده‌ی ۳ تا ۲۹۹۴ متغیر بوده و مقدار میانگین آن ۷۰/۰۲ نشان‌دهنده‌ی حضور ذرات معلق و مواد آلی در آب است. میزان فسفات نیز در محدوده‌ی ۰/۰۵ تا ۱/۹۵ میلی‌گرم بر لیتر قرار دارد و میانگین آن ۰/۴۷ میلی‌گرم بر لیتر است که می‌تواند ناشی از ورود فاضلاب شهری و کودهای کشاورزی باشد. نتایج مرتبط با اکسیژن‌خواهی بیوشیمیایی و اکسیژن‌خواهی شیمیایی نشان می‌دهد که میانگین نیاز اکسیژن‌خواهی بیوشیمیایی برابر ۱۴/۱۱ میلی‌گرم و میانگین نیاز اکسیژن‌خواهی شیمیایی برابر ۳۳/۰۷ میلی‌گرم بر لیتر است. مقدار حداکثری نیاز اکسیژن‌خواهی بیوشیمیایی برابر ۱۳۶ میلی‌گرم بر لیتر و نیاز اکسیژن‌خواهی شیمیایی برابر ۳۱۳ میلی‌گرم بر لیتر بوده که بیانگر وجود مقادیر بالای مواد آلی تجزیه‌پذیر و آلاینده‌های شیمیایی است. از سوی دیگر، مقدار اکسیژن محلول با میانگین ۶/۱۸ میلی‌گرم بر لیتر در محدوده‌ی طبیعی قرار داشته اما

جدول ۳- نتایج آزمون شاپیرو-ویلک برای بررسی نرمال بودن پارامترهای کیفیت آب مورد مطالعه

پارامتر	pH	TSS	TDS	EC	Turbidity	PO ₄	BOD	COD	DO	T	Total coli
مقدار p	۰/۹۷۴	۰/۹۶۶	۰/۹۹۱	۰/۹۸۷	۰/۹۷۷	۰/۹۷۵	۰/۹۷۸	۰/۸۱۴	۰/۷۱۲	۰/۸۴۵	۰/۸۵۲

در ادامه، تحلیل همبستگی پارامترهای کیفیت آب بر اساس نقشه حرارتی ماتریس همبستگی در شکل (۲) نشان داده شده است. کل مواد جامدات معلق همبستگی مثبت قوی با کدورت نشان می‌دهد. ارتباط کامل بین این دو پارامتر کاملاً منطقی است، زیرا افزایش مواد جامد معلق موجب افزایش کدورت آب می‌شود. رابطه نیاز اکسیژن‌خواهی بیوشیمیایی و شیمیایی منطقی است، زیرا هر دو پارامتر میزان آلودگی آلی را نشان می‌دهند و معمولاً همزمان افزایش یا کاهش می‌یابند.

در گام بعد، پیش از اجرای تحلیل مؤلفه‌های اصلی (PCA)، لازم است میزان کفایت نمونه و ارتباط میان متغیرها مورد ارزیابی قرار گیرد. بدین منظور، شاخص KMO و آزمون کرویت بارتلت به کار گرفته شد. مقدار شاخص KMO برابر با ۰/۷۳۱ بود که از حد توصیه‌شده ۰/۵ فراتر بوده و نشان‌دهنده کفایت مطلوب نمونه‌ها برای تحلیل مؤلفه‌های اصلی است. همچنین، نتایج آزمون بارتلت در سطح معناداری ۰/۵، نشان‌دهنده وجود همبستگی معنادار میان متغیرها است. این یافته‌ها مؤید آن هستند که داده‌های مورد بررسی برای تحلیل مؤلفه‌های اصلی مناسب هستند (Shrestha, 2021).



شکل ۲- ماتریس همبستگی پارامترهای کیفیت آب

شده است که حذف جامدات معلق از طریق فرایندهای تصفیه فاضلاب منجر به کاهش قابل توجهی در سطوح نیاز اکسیژن خواهی بیوشیمیایی و شیمیایی می‌شود که نشان‌دهنده ارتباط قوی بین جامدات معلق و شاخص‌های آلودگی آلی است (Jover-Smet et al., 2017; Spychała and Nguyen, 2019).

حضور جامدات معلق در سیستم‌های رودخانه‌ای اغلب نشان‌دهنده تغییرات اکولوژیکی گسترده، افزایش سطوح آلاینده‌ها و افزایش متناظر در سطوح نیاز اکسیژن خواهی بیوشیمیایی و شیمیایی مطابقت دارد که تحت‌تاثیر فعالیت‌های انسانی مانند تخلیه‌های صنعتی، رواناب شهری و شیوه‌های کشاورزی قرار دارند (Xu et al., 2021). علاوه بر این افزایش بار رسوب نه تنها مانع از نفوذ نور می‌شود، بلکه با ترویج شکوفه‌های جلبکی مضر که اکسیژن مورد نیاز موجودات آبی را مصرف می‌کند، سطح نیاز اکسیژن خواهی بیوشیمیایی را افزایش می‌دهد و بر زندگی آبزیان تداخل ایجاد می‌کنند (Yu, 2025; Liu and Chen, 2013). در نهایت، مطالعات متعددی بر ظرفیت

نتایج تحلیل همبستگی نشان می‌دهد که مواد جامد معلق دارای ارتباط قوی با شاخص‌های آلودگی آلی، شامل نیاز اکسیژن خواهی بیوشیمیایی و شیمیایی است. این ارتباط نشان می‌دهد که افزایش غلظت مواد جامد معلق معمولاً با افزایش بار آلی آب همراه است، زیرا مواد معلق می‌توانند حامل ترکیبات آلی قابل تجزیه باشند. همبستگی بین غلظت مواد جامد معلق و این آلاینده‌های آلی را می‌توان به مکانیسم‌های متعددی نسبت داد. اول و مهمتر از همه، مواد جامد معلق به عنوان حامل برای آلاینده‌های آلی و معدنی مختلف عمل می‌کنند. تحقیقات نشان می‌دهد که آلاینده‌هایی مانند فسفر و فلزات سنگین، تمایل به جذب روی ذرات معلق دارند، که انتقال آنها را در سیستم‌های آبی تسهیل می‌کند (Wakida et al., 2014; Sakson, 2023). این ارتباط به طور قابل توجهی بر کیفیت کلی آب تأثیر می‌گذارد، زیرا آلاینده‌های متصل به مواد جامد معلق می‌توانند بر فعالیت بیولوژیکی در آب تأثیر بگذارند و منجر به افزایش تقاضای اکسیژن بیوشیمیایی شوند (Dong et al., 2011). علاوه بر این، نشان داده

خواهی بیوشیمیایی و کاهش سطح اکسیژن محلول می‌شود (Gholizadeh et al. 2016). علاوه بر این، فعل و انفعالات بین دما و فعالیت میکروبی می‌تواند به طور قابل توجهی پویایی اکوسیستم را تغییر دهد، به ویژه از طریق تنفس و فرآیندهای جذب مواد آلی. با بررسی باکس پلات ارائه شده (شکل ۳) می‌توان به تحلیل جامعی از پارامترهای کیفیت آب پرداخت. پارامترهای مواد جامد محلول، مواد جامد معلق، هدایت الکتریکی و کدورت، نشان‌دهنده تغییرات اندک و مقادیر نزدیک به صفر هستند، که حاکی از کیفیت مطلوب آب از نظر مواد معلق و محلول است. در مقابل، پارامترهای فسفات، اکسیژن محلول، نیاز اکسیژن خواهی شیمیایی و بیوشیمیایی، دما و کلیرم کل، نوسانات قابل توجه و نقاط پرت متعددی را نشان می‌دهند. این نوسانات می‌تواند ناشی از آلودگی‌های کشاورزی، فاضلاب‌های شهری و صنعتی، و یا تغییرات فصلی باشد. به ویژه، وجود نقاط پرت در پارامترهای نیاز اکسیژن خواهی شیمیایی و بیوشیمیایی، نشان‌دهنده نوسانات زیاد در میزان مواد آلی قابل تجزیه بیولوژیکی و شیمیایی در آب است. همچنین، نوسانات در پارامتر اکسیژن محلول، نشان‌دهنده تغییرات در میزان اکسیژن محلول در آب است که برای بقای موجودات آبی ضروری است. در نهایت، نوسانات زیاد در کلیرم کل، نشان‌دهنده نوسانات زیاد در آلودگی میکروبی آب است. با وجود نوسانات در پارامترهای کیفیت آب، تغییرات شاخص کیفیت آب نسبتاً پایدار و نزدیک به صفر است. با این حال، نوسانات در پارامترهای فردی می‌تواند بر کیفیت کلی آب تأثیر بگذارد و نیاز به بررسی دقیق‌تر و اقدامات مدیریتی مناسب را ایجاد می‌کند.

نتایج تجزیه و تحلیل مولفه‌های اساسی و تعیین پارامترهای تاثیرگذار

ارزیابی کیفیت آب برای حفظ محیط زیست و سلامت عمومی ضروری است. برای ارزیابی دقیق عوامل موثر بر کیفیت آب، اغلب از آنالیز PCA استفاده می‌شود. PCA کاهش ابعاد داده را در عین حفظ تنوع، که برای شناسایی پارامترهای کلیدی بسیار مهم است، امکان‌پذیر می‌کند. PCA به مجموعه داده‌های پارامتر کیفیت آب منطقه مورد مطالعه برای شناسایی الگوهای زیربنایی و متغیرهای تأثیرگذار اعمال شد و نتایج در جدول‌های ۴ و ۵ نشان داده شده است. تحلیل مؤلفه‌های اصلی بر روی ۱۱ پارامتر انجام شد و ۷ مؤلفه اصلی با استفاده از معیار مقدار ویژه انتخاب شدند. این ۷ مؤلفه در مجموع ۷۷٪ از تغییرات کل را توضیح می‌دهند. مؤلفه اول و دوم مسئول بخش عمده‌ای از تغییرات (۴۵٪) هستند (جدول ۴). مؤلفه سوم تا هفتم حدود ۳۰٪ باقی مانده تغییرات را توضیح داده است. پارامترهای کلیدی با وزن بزرگتر از ۰/۵ برای مؤلفه‌های تأثیرگذار در تحلیل مؤلفه‌های اصلی (PCA) در جدول ۵ ارائه شده است. تحلیل مؤلفه‌های اصلی (PCA) بر اساس جدول ۵ نشان می‌دهد که مؤلفه

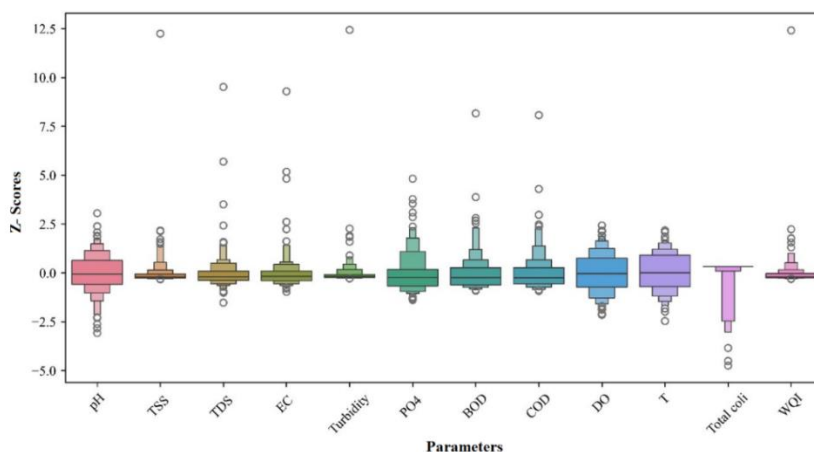
پیش‌بینی مواد جامد معلق به عنوان شاخصی از کیفیت کلی آب در رودخانه‌ها تأکید کرده‌اند. غلظت بالای جامدات معلق اغلب با سطوح بالای نیاز اکسیژن خواهی بیوشیمیایی و شیمیایی مرتبط است زیرا میزان آلودگی وارد شده به سیستم‌های رودخانه را منعکس می‌کند. به عنوان مثال، فریرا و همکاران اشاره کردند که نظارت بر جامدات معلق می‌تواند به عنوان یک نماینده مفید برای ارزیابی سلامت اکوسیستم‌های آبی در طول زمان باشد (Ferreira et al., 2023).

از آنجایی که تجزیه مواد آلی نیازمند مصرف اکسیژن است، افزایش نیاز اکسیژن خواهی بیوشیمیایی و شیمیایی می‌تواند منجر به کاهش اکسیژن محلول در آب شود و شرایط نامطلوبی را برای زیست‌بوم‌های آبی ایجاد کند. همچنین، مقادیر بالای مواد جامد معلق باعث افزایش کدورت آب و کاهش نفوذ نور شده که می‌تواند فتوسنتز گیاهان آبی را مختل کند. بر این اساس، همبستگی قوی بین مواد جامد معلق و شاخص‌های آلودگی آلی نشان می‌دهد که پایش این پارامترها برای ارزیابی کیفیت آب و مدیریت منابع آبی ضروری است. اسیدیته همبستگی مثبت ضعیفی با کلیرم کل دارد، که نشان می‌دهد افزایش اسیدیته ممکن است با افزایش فعالیت زیستی و حضور باکتری‌ها مرتبط باشد. همچنین کل جامدات محلول با هدایت الکتریکی و کدورت همبستگی مثبت متوسط دارد، که منطقی است، زیرا حضور جامدات محلول در آب، رسانایی الکتریکی و شفافیت را تحت تأثیر قرار می‌دهد.

اکسیژن محلول و دما همبستگی منفی متوسط با نیاز اکسیژن خواهی بیوشیمیایی و شیمیایی نشان می‌دهند. تجزیه مواد آلی در محیط‌های آبی منجر به تکثیر جمعیت‌های میکروبی و افزایش مصرف اکسیژن می‌شود، که نشان‌دهنده افزایش تقاضای اکسیژن بیوشیمیایی است (Nas and Nas, 2009). این پدیده به ویژه در آب‌های آلوده مشهود است، جایی که ورود مواد آلی متابولیسم میکروبی را تسریع و در نتیجه، اکسیژن محلول را کاهش می‌دهد. جومب و ناندینی دریافتند که سطوح بالای تقاضای اکسیژن بیوشیمیایی همبستگی معکوس با اکسیژن محلول دارد، که نشان می‌دهد کاهش غلظت اکسیژن محلول اغلب با افزایش مقادیر اکسیژن خواهی بیوشیمیایی همراه است و بیانگر کیفیت پایین‌تر آب و افزایش بار آلودگی آلی در سیستم‌های آبی است (Jumbe and Nandini, 2012). علاوه بر این، دما نقش مهمی در این رابطه ایفا می‌کند. دما بر حلالیت اکسیژن در آب تأثیر می‌گذارد، به طوری که آب‌های گرم‌تر معمولاً اکسیژن محلول کمتری دارند (Dong et al., 2019). دماهای بالا می‌توانند اثرات آلودگی را تشدید کنند، زیرا نرخ متابولیک میکروارگانیسم‌ها افزایش می‌یابد، نیاز اکسیژن خواهی بیوشیمیایی بیشتر افزایش می‌یابد و منجر به کاهش اکسیژن موجود می‌شود. تحقیقات قلی‌زاده و همکاران نشان داد که دمای بالای آب با افزایش تخریب ماده آلی مرتبط است که باعث افزایش نیاز اکسیژن

قرار دارد، که به وضعیت اکسیژن‌رسانی آب مرتبط است و می‌تواند نشان‌دهنده کیفیت زیستی یا فرایندهای هواهی در محیط باشد. مؤلفه اصلی پنجم (PC5) با فسفات و مواد جامد محلول (TDS) مشخص می‌شود، که به آلودگی‌های محلول و املاح آب اشاره دارد، احتمالاً ناشی از منابع کشاورزی یا صنعتی است. مؤلفه اصلی ششم (PC6) به دما (T) وابسته است، که شرایط حرارتی محیط را منعکس می‌کند و می‌تواند بر فرایندهای شیمیایی و بیولوژیکی تأثیر بگذارد. نهایتاً، مؤلفه اصلی هفتم (PC7) هیچ پارامتر کلیدی ندارد، که ممکن است به معنای تأثیر محدود این مؤلفه در توضیح واریانس باشد. این تحلیل بر اهمیت کدورت، فسفات، اکسیژن محلول و دما در ارزیابی کیفیت محیط، به‌ویژه در زمینه آلودگی و شرایط فیزیکی-شیمیایی، تأکید می‌کند و می‌تواند برای مدیریت منابع آب یا شناسایی منابع آلودگی مفید باشد.

اصلی اول (PC1) تحت تأثیر کدورت (Turbidity) است، که بیانگر نقش برجسته ذرات معلق در توضیح واریانس داده‌های محیطی است. کدورت به‌عنوان یک شاخص کلیدی کیفیت آب، احتمالاً به دلیل تأثیرات بصری، زیست‌محیطی و ارتباطش با آلاینده‌های فیزیکی یا آلی انتخاب شده، زیرا می‌تواند فرآیندهایی مانند نفوذ نور یا سلامت اکوسیستم‌های آبی را مختل کند. مؤلفه اصلی دوم (PC2) هیچ پارامتر کلیدی مشخصی ندارد، که ممکن است نشان‌دهنده پراکندگی تأثیر پارامترها یا اهمیت کمتر این مؤلفه باشد. مؤلفه اصلی سوم (PC3) با اسیدیته (pH)، فسفات (PO₄) و کلیفرم کل (Total coli) تعریف می‌شود، که به ویژگی‌های شیمیایی (اسیدیته و فسفات) و آلودگی میکروبی (کلیفرم) اشاره دارد و احتمالاً نشان‌دهنده تأثیر فاضلاب یا فعالیت‌های انسانی است. مؤلفه اصلی چهارم (PC4) تحت تأثیر اکسیژن محلول (DO)



شکل ۳ - نمودار جعبه‌ای پارامترهای کیفیت آب

جدول ۴- مقادیر ویژه مجموعه داده‌های پارامترهای کیفیت آب منطقه مورد مطالعه

مؤلفه اصلی	مقادیر ویژه	واریانس (%)	واریانس تجمعی (%)
۱	۳/۳۶	۳۰/۱۶	۳۰/۱۶
۲	۱/۶۹	۱۵/۲۸	۴۵/۴۳
۳	۱/۳۸	۱۲/۵۰	۵۷/۹۳
۴	۰/۰۳	۰/۰۲۵	۵۸/۱۸
۵	۰/۰۲	۰/۱۶	۵۸/۳۴
۶	۱/۱۲	۱۰/۱۱	۶۸/۴۵
۷	۰/۹۲	۸/۲۹	۷۶/۷۴
۸	۰/۵۲	۴/۶۸	۸۱/۴۳
۹	۰/۵۷	۱۵/۱۳	۹۶/۵۶
۱۰	۰/۷۸	۷/۰۴	۹۳/۶۰
۱۱	۰/۷۱	۶/۴۰	۱۰۰

جدول ۵- پارامترهای کلیدی برای مؤلفه‌های تأثیرگذار در تحلیل مؤلفه‌های اصلی

ردیف	پارامترهای کلیدی
۱	مؤلفه اصلی ۱ PC1 کدورت (Turbidity)
۲	مؤلفه اصلی ۲ PC2 -
۳	مؤلفه اصلی ۳ PC3 اسیدیته (pH) - فسفات (PO ₄) - کلیفرم کل (Total coli)
۴	مؤلفه اصلی ۴ PC4 اکسیژن محلول (DO)
۵	مؤلفه اصلی ۵ PC5 فسفات (PO ₄) - مواد جامد محلول (TDS)
۶	مؤلفه اصلی ۶ PC6 دما (T)
۷	مؤلفه اصلی ۷ PC7 -

برای ارزیابی داده‌های کیفیت آب تالاب و مباناد در مجاورت تأسیسات فرآوری غذاهای دریایی استفاده کردند. پارامترهای مورد بررسی شامل دمای اتمسفر و آب، TDS، pH، EC، شوری، DO، BOD، قلیابیت، COD، سختی و مواد مغذی مانند نیترات، فسفات، آمونیاک و سیلیس بودند. نتایج نشان دادند که PCA قادر است ۶ مؤلفه اصلی را استخراج کند که ۶۵/۷۹٪ از کل واریانس داده‌ها را تشکیل می‌دهند. پارامترهایی که در مؤلفه اول نقش دارند، شامل قلیابیت، BOD، COD و مواد مغذی هستند. آلودگی ارگانیک و تخلیه زباله از صنایع غذاهای دریایی تأثیر قابل توجهی بر کیفیت آب داشته است. این گونه مطالعه مزایای تکنیک‌های آماری چندمتغیره را در تجزیه و تحلیل و تفسیر داده‌های پیچیده نشان می‌دهد و به برنامه‌ریزی برای مطالعات آینده کمک می‌کند (Ecology and Prasad, 2024).

نتایج وزن‌دهی پارامترهای شاخص کیفیت آب با استفاده از روش MEREC در جدول ۶ نشان داده شده است. مقدار وزن اختصاص داده شده به کلیفرم کل در مدل MEREC برابر ۰/۱۹ است که نشان‌دهنده اهمیت بیشتر آن نسبت به سایر پارامترها مانند این یافته‌ها نشان می‌دهند که نظارت بر ورود انواع فاضلاب برای حفظ آب و کشاورزی پایدار ضروری است.

گربا و همکاران از تحلیل مؤلفه‌های اصلی (PCA) برای ارزیابی کیفیت آب استفاده کردند و پارامترهایی نظیر کدورت، دما، هدایت الکتریکی و pH به عنوان عوامل کلیدی در مدیریت کیفیت آب شناسایی شدند. در این میان، کدورت به عنوان یک عامل بنیادین در تعیین شفافیت آب شناخته شد، چرا که تأثیر قابل توجهی بر ویژگی‌های ظاهری و همچنین قابلیت زیستی آب دارد (Garba et al., 2022). ارزیابی کیفیت آب دریاچه جایی (نیجریه) در فصول تر و خشک با استفاده از ۲۱ پارامتر فیزیکوشیمیایی نشان داد که تغییرات فصلی قابل توجه در کیفیت آب وجود دارد. در فصل تر، مقادیر کل مواد جامد محلول، اکسیژن محلول و هدایت الکتریکی از جمله مقادیر کلیدی بودند. در فصل خشک، اکثر پارامترها افزایش قابل توجهی داشتند. تحلیل آماری همبستگی‌های قوی بین پارامترهای مختلف را شناسایی کرد، به‌ویژه بین فسفات و سختی کل در فصل تر و بین pH و دما در فصل خشک. این مطالعه همچنین چهار مؤلفه اصلی را شناسایی کرد که ۹۸/۵ - ۱۰۰٪ از واریانس کیفیت آب را توضیح می‌دهند، که نمایانگر منابع مختلف آلودگی از جمله ضایعات آلی، فاضلاب خانگی و عوامل طبیعی است (Adejuwon et al., 2025). ویدیا و پرساد در تحقیقی از تحلیل مؤلفه‌های اصلی (PCA)

جدول ۶- نتایج وزن‌دهی پارامترهای شاخص کیفیت آب منطقه مورد مطالعه با استفاده از روش MEREC

پارامتر	pH	TSS	TDS	EC	Turbidity	PO ₄	BOD	COD	DO	T	Total coli
مقدار p	۰/۰۱	۰/۱۳	۰/۰۴	۰/۰۶	۰/۱۵	۰/۰۳	۰/۰۶	۰/۰۶	۰/۰۲	۰/۰۵	۰/۱۹

ویژگی‌های مجموعه داده بسیار مهم است. رگرسیون خطی و جنگل تصادفی دو الگوریتم برجسته هستند که به طور گسترده برای وظایف رگرسیون مورد استفاده قرار می‌گیرند. مقایسه مقادیر واقعی در مقابل مقادیر پیش‌بینی شده این الگوریتم‌ها قبل و بعد از استفاده از

پیش‌بینی شاخص کیفیت آب با استفاده از الگوریتم‌های یادگیری ماشین

در حوزه مدل‌سازی پیش‌بینی‌کننده، درک عملکرد الگوریتم‌های مختلف یادگیری ماشین برای انتخاب مناسب‌ترین رویکرد بر اساس

اغلب عملکرد پیش‌بینی قابل قبولی مشابه فضای با ابعاد بالاتر را نشان می‌دهند (شکل ۶ و ۷). با این حال، هر دو تجسم داده‌های آموزشی و آزمایشی ممکن است افزایشی در پراکندگی نقاط در مقایسه با نمودارهای قبل از PCA نشان دهند. این نشان می‌دهد که کاهش ابعاد ممکن است منجر به از دست رفتن اطلاعات ظریفی شده باشد که می‌تواند برای پیش‌بینی‌های دقیق مرتبط باشد. مطابق با یافته‌های یک بررسی سیستماتیک، کاهش ابعاد می‌تواند مدل‌ها را ساده‌سازی کند و به‌طور بالقوه منجر به کاهش دقت پیش‌بینی به دلیل حذف ویژگی‌های مهم شود. با وجود این تغییرات، حفظ دقت پیش‌بینی قابل قبول پس از PCA، سودمندی کاهش ابعاد را در افزایش کارایی مدل بدون به خطر انداختن عملکرد قابل توجهی نشان می‌دهد. استفاده از چنین تکنیک‌هایی می‌تواند زمان‌های آموزش سریع‌تر و ساختارهای مدل ساده‌تر را تسهیل کند، که به‌ویژه در هنگام برخورد با داده‌های با ابعاد بالا یا زمانی که منابع محاسباتی محدود هستند مفید است.

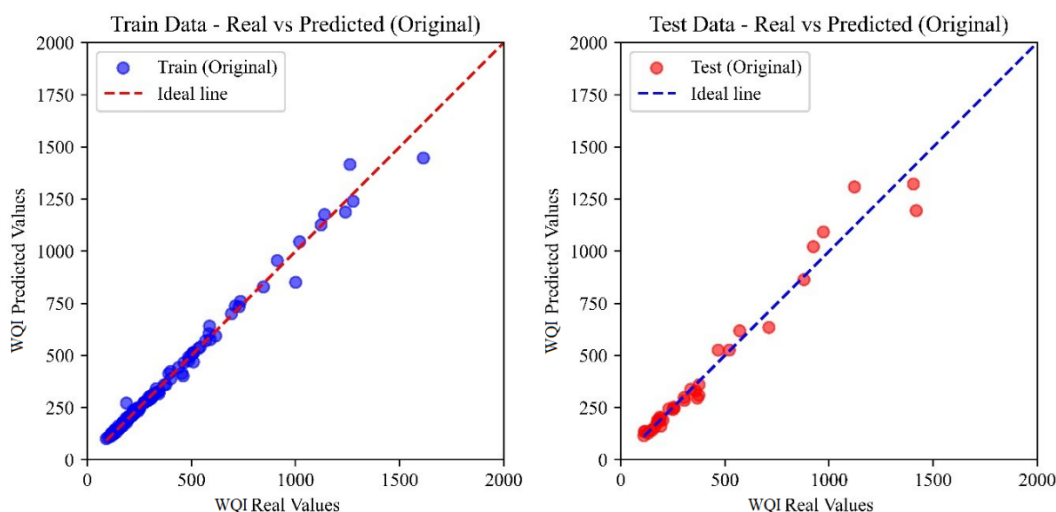
در جدول ۷، عملکرد مدل‌های مختلف پیش‌بینی در زمینه‌های یادگیری ماشین و آمار با تمرکز بر شاخص‌های ارزیابی بررسی شده است. مدل رگرسیون خطی نشان‌دهنده عملکرد بسیار خوبی در هر دو دسته داده (آزمون و آموزش) است. مقادیر EVS نزدیک به ۱ ($0/999$) و R^2 نزدیک به ۱ بیانگر این هستند که مدل تقریباً ۱۰۰ درصد از واریانس داده‌ها را توضیح می‌دهد. MAPE برای داده‌های آموزشی $0/021$ و برای داده‌های آزمایشی $0/012$ نشان‌دهنده دقت بالای پیش‌بینی‌هاست.

تکنیک‌های کاهش ابعاد مانند PCA به محققان این امکان را می‌دهد که بینش‌هایی در مورد عملکرد مدل، تعمیم و تطابق بیش از حد بالقوه به دست آورند.

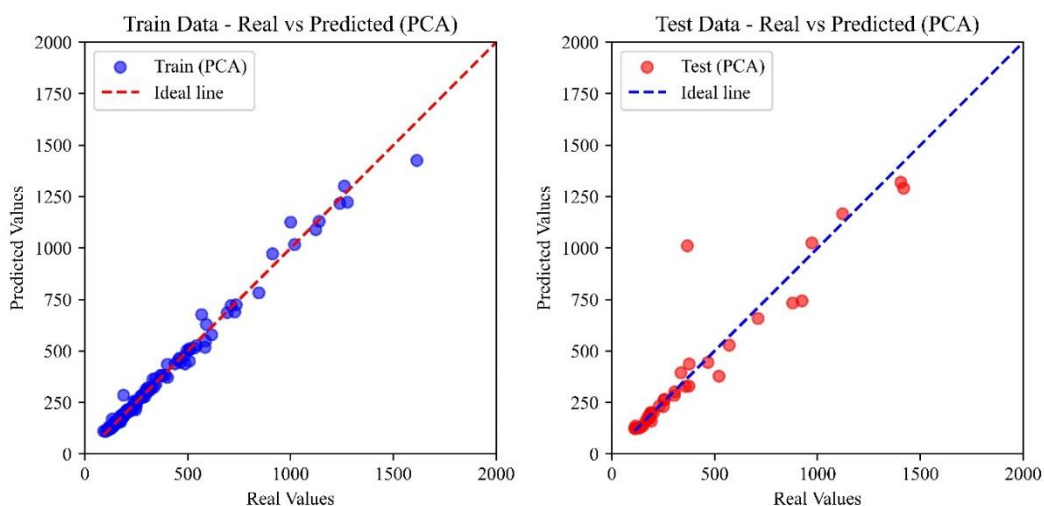
در بررسی عملکرد اولیه بدون کاهش ابعاد، مشاهده شده است که هر دو مدل رگرسیون خطی و جنگل تصادفی به‌طور موثر روابط زیربنایی در داده‌های آموزشی را ضبط می‌کنند (شکل ۴ و ۵). نزدیکی داده‌های آموزشی به خط ایده‌آل نشان‌دهنده دقت پیش‌بینی بالایی است. این مشاهدات با یافته‌های ارائه شده توسط روزنبوش و همکاران، که تأکید دارند روش‌های یادگیری نظارت شده، از جمله رگرسیون خطی و جنگل‌های تصادفی، می‌توانند در پیش‌بینی برتری داشته باشند، زمانی که روابط بین متغیرها به خوبی در مجموعه آموزشی تعریف شده باشد، همسو می‌شود (Rosenbusch et al., 2021).

علی‌رغم همسویی آشکار نقاط داده با خط ایده‌آل، افزایش پراکندگی بین نقاط داده‌های آزمون در مقایسه با داده‌های آموزشی نشان می‌دهد که مدل‌ها ممکن است چالش‌هایی را در تعمیم به داده‌های دیده نشده نشان دهند. افزایش واریانس نشان‌دهنده توسط گسترش گسترده‌تر پیش‌بینی‌ها در بین نمونه‌های آزمایشی، اغلب نشانه‌ی برازش بیش از حد است، جایی که یک مدل به جای سیگنال واقعی، نویز را در داده‌های آموزشی ثبت می‌کند (Yeom et al., 2018). پراکندگی بیشتر نشان می‌دهد که در حالی که مدل‌ها عملکرد مناسبی دارند، ممکن است در برابر نمونه‌های داده جدیدی که از شرایط آموزشی منحرف می‌شوند، مقاوم نباشند.

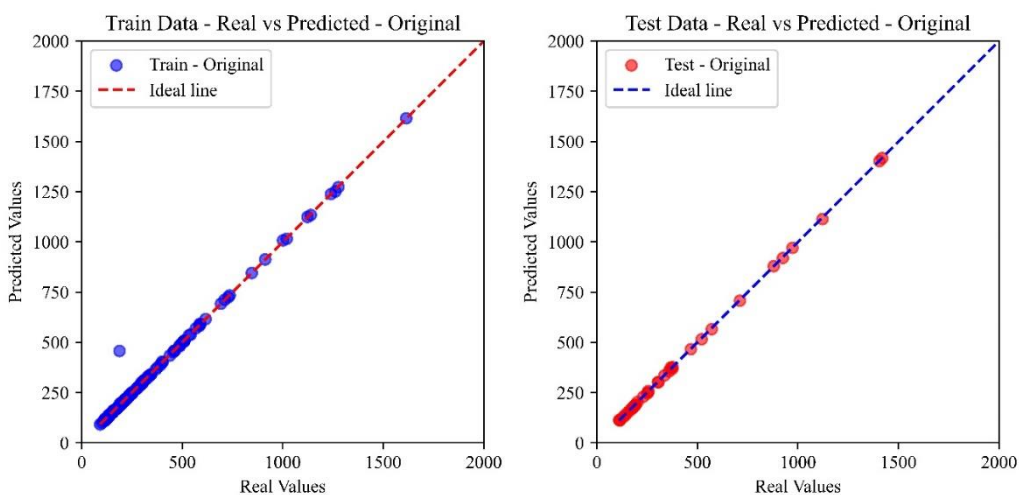
پس از اجرای PCA برای کاهش ابعاد مجموعه‌های داده، نتایج



شکل ۴- ارزیابی عملکرد مدل جنگل تصادفی در داده‌های آموزش و آزمون



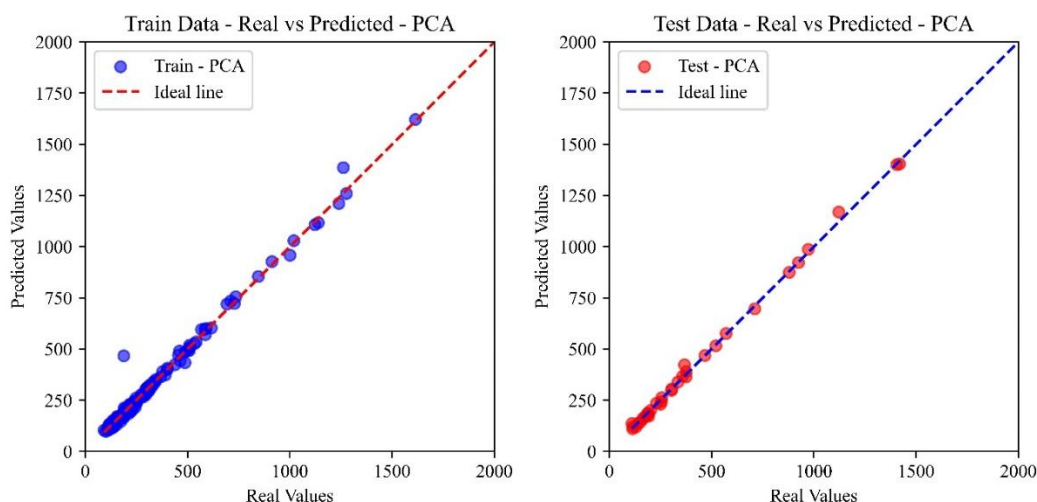
شکل ۵- ارزیابی عملکرد مدل جنگل تصادفی بعد از کاهش ابعاد در داده‌های آموزش و آزمون



شکل ۶- ارزیابی عملکرد مدل رگرسیون خطی در داده‌های آموزش و آزمون

جدول ۷- مقایسه عملکرد مدل‌های هوش مصنوعی در پیش‌بینی شاخص کیفیت آب رودخانه سیاهرود

داده‌های آزمایشی			داده‌های آموزشی			Model
EVS	MAPE	R ²	EVS	MAPE	R ²	
۰/۹۹۹	۰/۰۱۲	۰/۹۹۹	۰/۹۹۹	۰/۰۲۱	۰/۹۹۹	رگرسیون خطی
۰/۹۹۹	۰/۰۳۷	۰/۹۹۹	۰/۹۹۹	۰/۰۵۱	۰/۹۹۹	رگرسیون خطی + کاهش ابعاد
۰/۹۱۴	۰/۰۸۳	۰/۹۱۱	۰/۹۰۰	۰/۰۳۵	۰/۸۹۹	جنگل تصادفی
۰/۸۹۲	۰/۱۳۸	۰/۸۸۹	۰/۹۰۲	۰/۰۴۴	۰/۹۰۱	جنگل تصادفی + کاهش ابعاد

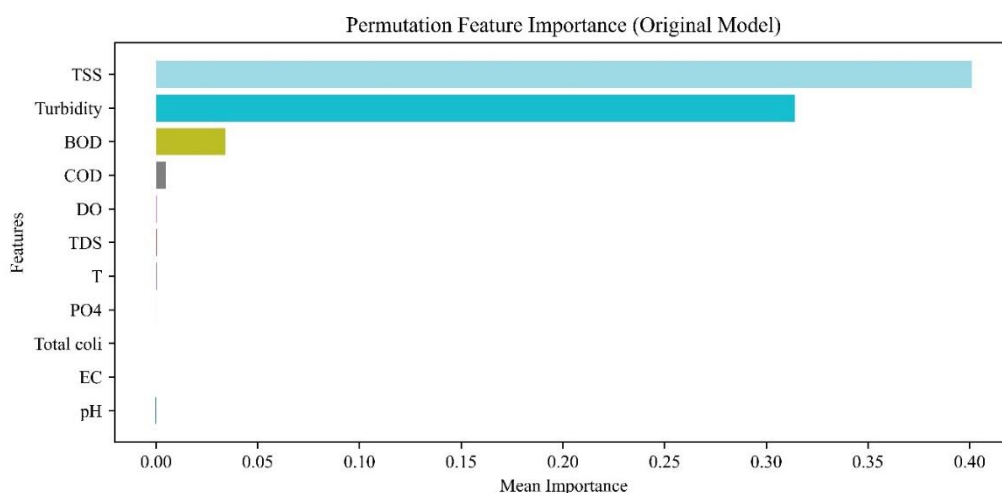


شکل ۷- ارزیابی عملکرد مدل رگرسیون خطی بعد از کاهش ابعاد در داده‌های آموزش و آزمون

کمک می‌کند. در زمینه ارزیابی کیفیت آب، درک این حساسیت‌ها می‌تواند قابلیت‌های پیش‌بینی را افزایش دهد و استراتژی‌های مدیریت منابع را آگاه کند. در شکل ۸ نتایج آنالیز حساسیت مدل رگرسیون خطی نشان داده شده است. مدل بیشترین حساسیت را به پارامترهای مرتبط با ذرات معلق و کدورت و سپس آلودگی آلی (نیاز اکسیژن خواهی بیوشیمیایی و شیمیایی) نشان می‌دهد. این نتیجه با انتظارات مطابقت دارد، زیرا این پارامترها به طور قابل توجهی بر شفافیت و کیفیت آب تأثیر می‌گذارند و مستقیماً بر اکوسیستم‌های آبی تأثیر می‌گذارند. پارامترهای مانند اکسیژن محلول، مواد جامد محلول، دما و pH تأثیر ناچیزی بر پیش‌بینی مدل دارند و این نشان می‌دهد که در حالی که آن‌ها در دینامیک کیفیت آب نقش ایفا می‌کنند، تغییرات آنها به طور قابل ملاحظه‌ای نتایج مدل رگرسیون خطی را در پیکربندی فعلی آن تغییر نمی‌دهد. قابل ذکر است که سایر پارامترها بدون تأثیرند.

عباس و همکاران (۲۰۲۴) نیز در مطالعه‌ای مدل‌های مختلف یادگیری ماشین برای پیش‌بینی کیفیت آب مورد مقایسه قرار دادند و اهمیت پارامترهایی مانند نیاز اکسیژن خواهی بیوشیمیایی و اکسیژن محلول برجسته شد. نتایج نشان داد که بین این پارامترها و سلامت آب همبستگی بالایی وجود دارد. این پارامترها به‌عنوان شاخص‌های جانسین برای سطح آلودگی آلی عمل کرده و تأثیر مستقیمی بر حیات آبیان دارند (Abbas et al., 2024).

دقت بالای مدل رگرسیون خطی، آن را به گزینه‌ای مناسب برای پیش‌بینی‌های خطی تبدیل می‌کند. بکارگیری تکنیک PCA در مدل رگرسیون خطی، عملکرد مناسب و قابل توجهی را به نمایش می‌گذارد. مقادیر مشابه EVS و R^2 (۰/۹۹۹) حاکی از این است که PCA تأثیری مخرب بر روی کیفیت مدل ندارد اما MAPE برای داده‌های آزمون (۰/۰۳۷) نشان می‌دهد که دقت مدل کاهش یافته است. ممکن است این موضوع به علت کاهش ابعاد و اتلاف بخشی از اطلاعات باشد که بر روی ظرفیت پیش‌بینی تأثیر می‌گذارد. مدل جنگل تصادفی با EVS برابر با ۰/۹۱۴ و R^2 برابر با ۰/۹۱۱ برای داده‌های آزمایشی عملکرد نسبتاً خوبی را نشان می‌دهد. MAPE برای مدل جنگل تصادفی نیز بالاتر از نتایج رگرسیون خطی است (۰/۰۸۳) که نشان‌دهنده عدم دقت کامل پیش‌بینی‌ها است. با این حال، جنگل تصادفی هنوز هم یک گزینه قدرتمند و جامع در پیش‌بینی داده‌ها محسوب می‌شود، به ویژه برای داده‌های غیرخطی. تکنیک PCA در جنگل تصادفی (مقدار EVS برابر با ۰/۹۰۲ و R^2 برابر با ۰/۹۰۱) منجر به بهبودی در EVS نسبت به مدل جنگل تصادفی بدون PCA نشده است و حتی MAPE برای داده‌های آزمایشی (۰/۰۴۴) نشان می‌دهد که کاربرد PCA به احتمال زیاد منجر به کاهش کیفیت پیش‌بینی در این مدل شده است. تجزیه و تحلیل حساسیت یک گام مهم در توسعه مدل است، زیرا به شناسایی پارامترهایی که به طور قابل توجهی بر نتایج مدل تأثیر می‌گذارد



شکل ۸- نتایج آنالیز حساسیت مدل رگرسیون خطی

uncertainty assessment in Mirpurkhas, Sindh, Pakistan. *Water*. 16(7): 941.

نتیجه‌گیری

Adejuwon, E. O., Ogwueleka, T. C., Ogungbemi, E. O., Prabhu, R., Rendon-Nava, A. and Yates, K. 2025. Assessment of surface water quality using chemometric tools: A case study of Jabi Lake, Abuja, Nigeria. *Iranian Journal of Science and Technology, Transactions of Civil Engineering*, 1-24.

این پژوهش با بهره‌گیری از رویکرد تلفیقی تحلیل مؤلفه‌های اصلی (PCA) تصمیم‌گیری چندمعیاره (MERECE) و مدل‌های هوش مصنوعی، کیفیت آب رودخانه سیاهرود را ارزیابی کرد و نشان داد که کدورت، کلیفرم کل، فسفات و آلاینده‌های آلی (COD و BOD) و عوامل اصلی کاهش کیفیت آب هستند. شاخص کیفیت آب (WQI) با میانگین بالا، کیفیت نامطلوب آب را در اکثر ایستگاه‌ها تأیید کرد. روش MERECE با اختصاص وزن‌های ۰/۱۹ به کلیفرم کل و ۰/۱۵ به کدورت، نقش برجسته آلودگی میکروبی و فیزیکی را برجسته ساخت. مدل رگرسیون خطی با دقت بالا بهترین عملکرد را در پیش‌بینی WQI داشت، هرچند کاهش ابعاد با PCA اندکی دقت را کاهش داد. این یافته‌ها بر ضرورت پایش مستمر پارامترهای کلیدی و اجرای اقدامات مدیریتی مانند تصفیه فاضلاب، کنترل رواناب‌های شهری، کشاورزی و صنعتی برای بهبود کیفیت آب و حفاظت از اکوسیستم رودخانه تأکید دارند.

Bahrami, M., Khaksar, E. and Khaksar, E. 2020. Spatial variation assessment of groundwater quality using multivariate statistical analysis (Case Study: Fasa Plain, Iran). *Journal of Groundwater Science and Engineering*. 8(3): 230-243.

Bhutiani, R., Khanna, D. R., Kulkarni, D. B. and Ruhela, M. 2016. Assessment of Ganga river ecosystem Haridwar, Uttarakhand, India with reference to water quality indices. *Applied Water Science*. 6: 107-113.

Charkhabi, A. H., Sakizadeh, M. and Rafiee, G. 2005. Seasonal fluctuation in heavy metal pollution in Iran s Siahroud River-a preliminary study (7 pp). *Environmental Science and Pollution Research*. 12: 264-270.

Dong, D., Guo, Z., Hua, X., Lan, Y., Zhou, J., Ding, X. and Qiao, Q. 2011. Sorption of DDTs on biofilms, suspended particles and river sediments: effects of heavy metals. *Environmental Chemistry Letters*. 9: 361-367.

Dong, S., Liu, L., Zhang, Y. and Jiang, F. 2019. Occurrence and succession of bacterial community in O3/BAC process of drinking water treatment. *International Journal of Environmental Research and Public Health*. 16(17): 3112.

Ecology, E. C. V. and Prasad, G. 2024. Interpretation of

سپاسگزاری

نویسندگان این مقاله از دانشگاه علوم کشاورزی و منابع طبیعی ساری بابت تأمین هزینه‌های طرح پژوهشی با کد قرارداد ۱۴۰۳-۰۴-۰۲ کمال تشکر را دارند.

منابع

Abbas, F., Cai, Z., Shoaib, M., Lqbal, J., Ismail, M., Arifullah, Alrefaei, A. F. and Albeshr, M. F. 2024. Machine learning models for water quality prediction: a comprehensive analysis and

- methods: A critical review. *Environmental and Sustainability Indicators*. 18: 100247.
- Nas, S. S. and Nas, E. 2009. Water quality modeling and dissolved oxygen balance in streams: a point source streeter-phelps application in the case of the harsit stream. *Clean-Soil, Air, Water*. 37(1): 67-74.
- Prasad, D. V. V., Venkataramana, L. Y., Kumar, P. S., Prasannamedha, G., Harshana, S., Srividya, S. J., Harrinei, K. and Indraganti, S. 2022. Analysis and prediction of water quality using deep learning and auto deep learning techniques. *Science of the Total Environment*. 821: 153311.
- Rosenbusch, H., Soldner, F., Evans, A. M. and Zeelenberg, M. 2021. Supervised machine learning methods in psychology: A practical introduction with annotated R code. *Social and Personality Psychology Compass*. 15(2): e12579.
- Şahin Kiy, M. and Arslan, H. 2021. Assessment of groundwater quality for irrigation and drinking using different quality indices and geostatistical methods in Çorum province (Turkey). *Irrigation and Drainage*. 70(4): 871-886.
- Sakson, G. 2023. Analysis of suspended solids emissions from a combined sewage system using the Stormwater Management Model (SWMM). *Geology, Geophysics and Environment*. 49(4): 389-399.
- Shrestha, N. 2021. Factor analysis as a tool for survey analysis. *American journal of Applied Mathematics and Statistics*. 9(1): 4-11.
- Spychała, M. and Nguyen, T. H. 2019. Preliminary study on greywater treatment using nonwoven textile filters. *Applied Sciences*. 9(15): 3205.
- Wakida, F. T., Huato, S. M., Flores, E. G., Colin, T. D. J. P., Gomez, H. E. and Lopez, A. A. 2014. Pollutant association with suspended solids in stormwater in Tijuana, Mexico. *International Journal of Environmental Science and Technology*. 11: 319-326.
- Xu, H., Xu, G., Wen, X., Hu, X. and Wang, Y. 2021. Lockdown effects on total suspended solids concentrations in the Lower Min River (China) during COVID-19 using time-series remote sensing images. *International Journal of Applied Earth Observation and Geoinformation*. 98: 102301.
- Yeom, S., Giacomelli, I., Fredrikson, M. and Jha, S. 2018. Privacy risk in machine learning: Analyzing the connection to overfitting. in 2018 IEEE 31st computer security foundations symposium (CSF). IEEE.
- Yu, Y. 2025. Key technologies for purification of suspended solids in turbid surface water under the influence of rainy seasons. *Polish Journal of Environmental Studies*. 34(2): 1445-1455.
- water quality using principal component analysis—A case study from a tropical Ramsar wetland site adjacent to the Seafood processing facilities. *Ecology, Environment & Conservation*. 30: 311-305.
- Fashagba, T. S., Bessedik, M., ElSayed, N. B., Abdelbaki, C. and Kumar, N. 2024. Evaluating the water quality of the Keddara dam (Algeria) using water quality indices. *Water*. 16(9): 1291.
- Ferreira, W. J., Targa, M. D. S. and Catelani, C. D. S. 2023. Temporal evolution of suspended solids in the management of medium-sized watersheds in the Metropolitan Region of Paraíba Valley. *Revista Ambiente & Água*. 18: e2945.
- Garba, N., Adamu, A. and Audu, I. A. 2022. Assessment of water quality treatment and costing: A case study of Sokoto state water board. *International Journal of Research Publication and Reviews*. 3(11): 1407-1413.
- Gholizadeh, M. H., Melesse, A. M. and Reddi, L. 2016. A comprehensive review on water quality parameters estimation using remote sensing techniques. *Sensors*. 16(8): 1298.
- Jahan, S. and Strezov, V. 2017. Water quality assessment of Australian ports using water quality evaluation indices. *PLoS One*. 12(12): e0189284.
- Jover-Smet, M., Martín-Pascual, J. and Trapote, A. 2017. Model of suspended solids removal in the primary sedimentation tanks for the treatment of urban wastewater. *Water*. 9(6): 448.
- Jumbe, A. S. and Nandini, N. 2012. Physicochemical investigation of surface water quality of urban wetlands of Bangalore, India. *Asian Journal of Water, Environment and Pollution*. 9(1): 59-69.
- Karpagam, V. and Christy, S. 2023. Classifying Indian rivers surface water quality using principal component analysis combined with principal component weighted arithmetic index techniques. In 2023 eighth international conference on science technology engineering and mathematics (ICONSTEM). IEEE.
- Keshavarz-Ghorabae, M. 2021. Assessment of distribution center locations using a multi-expert subjective-objective decision-making approach. *Scientific Reports*. 11(1): 19461.
- Liu, W. and Chen, W. 2013. Modeling hydrothermal, suspended solids transport and residence time in a deep reservoir. *International Journal of Environmental Science and Technology*. 10: 251-260.
- Mahbulbul, S. M. M., Shakhawat, H. M., Rajaul, K. M., Faisal, U. M., Mahady, H. and Hayat, K. R. 2023. Surface water quality profiling using the water quality index, pollution index and statistical

An Integrated Approach to Monitoring Water Quality in the Siahroud River using Modern Data Analysis Methods

M. Khoshravesh^{1*}, L. Divband Hafshejani², M. Moradzadeh³, T. Eskandari⁴

Received: Apr.20, 2025

Accepted: May.31, 2025

Abstract

The Water Quality Index (WQI) is an effective tool for assessing water resources, enabling comparison and decision-making by combining multiple parameters into a single simple numerical value. However, the use of standard weighting schemes without considering local conditions may lead to inaccurate results. This study aims to optimize WQI calculation by adopting an integrated approach involving Principal Component Analysis (PCA), Multi-Criteria Decision Making (MERC), and artificial intelligence models. Twenty years of data (2001–2021) from the Siahroud River in Mazandaran Province, encompassing 11 water quality parameters, were analyzed. Dimensionality reduction was performed using PCA, and WQI was calculated through the MERC method. Additionally, WQI prediction was conducted using two algorithms (Linear Regression and Random Forest) under both full and reduced datasets, and evaluated through indicators such as EVS, MAPE, and R^2 . The findings revealed that turbidity, as the key factor in physical pollution, exhibited strong correlations with TSS, BOD, and COD. Other principal components indicated microbial, chemical, and thermal pollution stemming from sources such as wastewater discharge and agricultural activities. The mean WQI values indicated poor water quality in most sampling stations. The Linear Regression model showed superior performance compared to the Random Forest model, although dimensionality reduction slightly reduced predictive accuracy. According to the MERC method, total coliform (weight: 0.19) and turbidity (weight: 0.15) were identified as the most influential parameters, highlighting the significance of microbial contamination and suspended solids. The results underscore the necessity for continuous monitoring and effective management of pollution sources to improve regional water quality.

Keywords: Organic pollution, Principal components, Quality index, Turbidity.

1- Associate Professor, Department of Water Engineering, Faculty of Agricultural Engineering, Sari Agricultural Sciences and Natural Resources University, Sari, Iran

2- Assistant Professor, Department of Environmental Engineering, Faculty of Water and Environmental Engineering, Shahid Chamran University of Ahvaz, Ahvaz, Iran

3- Laboratoire de Géologie, Département de Géosciences, Ecole Normale Supérieure (ENS), Paris, France

4- Department of Environment at Mazandaran Province (DOE), Caspian Sea Ecology Research Center (CSERC), Sari, Iran

(*- Corresponding Author Email: m.khoshravesh@sanru.ac.ir)