



Application of Gene Expression Programming Method in Estimation of Shannon Entropy Water Quality Index

Amir hosein Shakarami¹, Laleh Divband Hafshejani^{*2}, Parvaneh Tishehzan³, Hamid Abdolabadi⁴

1. Department of Environmental Engineering, Faculty of Water and Environmental Engineering, Shahid Chamran University of Ahvaz, Ahvaz, Iran
2. Department of Environmental Engineering, Faculty of Water and Environmental Engineering, Shahid Chamran University of Ahvaz, Ahvaz, Iran
3. Department of Environmental Engineering, Faculty of Water and Environmental Engineering, Shahid Chamran University of Ahvaz, Ahvaz, Iran
4. Department of Environmental Engineering, Faculty of Water and Environmental Engineering, Shahid Chamran University of Ahvaz, Ahvaz, Iran

Received: 21 October 2024/ Revised: 12 November 2024/ Accepted: 14 November 2024

<https://doi.org/10.22034/arwe.2024.2043938.1031>

Abstract

Water quality, due to its vital role in sustaining life and supporting various aspects of human life, is a key issue in water resource management. With industrial advancements and increasing water demands, water pollution has become a serious concern. This research investigates the water quality of the Zohreh River using the Shannon Entropy Water Quality Index (SEWQI) and Gene Expression Programming (GEP) model. Data on water quality parameters, including pH, electrical conductivity, phosphate, nitrate, ammonium, total hardness, fecal coliform, dissolved oxygen, BOD5, COD, and turbidity, were collected from the Firoozabad and 720-meter Suire stations over a ten-year period (2011-2020). Firstly, the entropy of each parameter was calculated, and their weights were determined. Then, using the calculated weights, the Shannon Entropy Water Quality Index was obtained. The results obtained from the GEP model showed that this method has a high accuracy in predicting the water quality index with a coefficient of determination (R^2) of 0.803 and root mean square error (RMSE) of 0.07. The results of this study highlight the use of GEP as an effective tool for assessing and managing water quality in the Zohreh River and can contribute to improving water resource management policies and preventing risks to public health and ecosystems.

Keywords: Artificial intelligence, Firozabad, Optimization, Pearson correlation coefficient.



کاربرد روش برنامه‌ریزی بیان ژن در تخمین شاخص کیفیت آب آنتروپی شانون

امیرحسین شاکرمی^۱، لاله دیوبند هفشجانی^{۲*}، پروانه تیشه‌زن^۳، حمید عبدالآبادی^۴

۱. گروه مهندسی محیط زیست، دانشکده مهندسی آب و محیط‌زیست، دانشگاه شهید چمران اهواز، اهواز، ایران
۲. گروه مهندسی محیط زیست، دانشکده مهندسی آب و محیط‌زیست، دانشگاه شهید چمران اهواز، اهواز، ایران
۳. گروه مهندسی محیط زیست، دانشکده مهندسی آب و محیط‌زیست، دانشگاه شهید چمران اهواز، اهواز، ایران
۴. گروه مهندسی محیط زیست، دانشکده مهندسی آب و محیط‌زیست، دانشگاه شهید چمران اهواز، اهواز، ایران

دریافت: ۳۰ مهر ماه ۱۴۰۳ / اصلاحات: ۲۲ آبان ماه ۱۴۰۳ / پذیرش: ۲۴ آبان ماه ۱۴۰۳

<https://doi.org/10.22034/arwe.2024.2043938.1031>

چکیده

کیفیت آب به دلیل اهمیت حیاتی آن در حفظ زندگی و پشتیبانی از جنبه‌های مختلف، یکی از موضوعات کلیدی در مدیریت منابع آبی محسوب می‌شود. با پیشرفت صنعت و افزایش نیاز به آب، آلودگی منابع آبی به یک نگرانی جدی تبدیل شده است. این تحقیق به بررسی کیفیت آب رودخانه زهره با استفاده از شاخص کیفیت آب آنتروپی شانون و مدل برنامه‌ریزی بیان ژن (GEP) پرداخته است. داده‌های مربوط به پارامترهای کیفیت آب، شامل pH، هدایت الکتریکی، فسفات، نترات، آمونیوم، سختی کل، کلیفرم مدفوعی، اکسیژن محلول، COD، BOD5 و کدورت، طی دوره ده ساله (۱۳۹۰-۱۴۰۰) از ایستگاه‌های فیروزآباد و ۷۲۰ متری سویره تهیه شدند. ابتدا، آنتروپی هر پارامتر محاسبه و وزن‌های آن‌ها تعیین شد. سپس، با استفاده از وزن‌های محاسبه‌شده، شاخص کیفیت آب آنتروپی شانون به دست آمد. نتایج حاصل از مدل GEP نشان داد که این روش با ضریب تعیین (R^2) برابر ۰/۸۰۳ و ریشه میانگین مربعات خطا (RMSE) برابر ۰/۰۷، دقت بالایی در تخمین شاخص کیفیت آب دارد. نتایج این مطالعه به استفاده از GEP به عنوان ابزاری مؤثر برای ارزیابی و مدیریت کیفیت آب در رودخانه زهره اشاره دارد و می‌تواند به بهبود سیاست‌های مدیریتی منابع آبی و پیشگیری از خطرات برای سلامت عمومی و اکوسیستم‌ها کمک کند.

کلمات کلیدی: بهینه، ضریب پیرسون، فیروزآباد، هوش مصنوعی.

مقدمه

منابع آبی به دلیل نقش حیاتی در حفظ زندگی و پشتیبانی از جنبه‌های مختلف اهمیت فراوانی دارند. با پیشرفت صنعت و توسعه تمدن، نیاز به آب برای مصارف شرب، کشاورزی، پزشکی و صنعت به‌طور چشمگیری افزایش یافته است. این وضعیت به انتشار آلاینده‌های خطرناک مختلفی، از جمله فلزات سنگین، آلاینده‌های آلی و آلودگی‌های مرتبط با مراکز درمانی در منابع آب انجامیده و به یک نگرانی جدی تبدیل شده است (Hojjati-Najafabadi et al., 2022). یکی از راه‌های ارزیابی کیفیت آب استفاده از شاخص‌های ارزیابی می‌باشد. در این میان استفاده از شاخص کیفیت آب آنتروپی شانون برای بهینه‌سازی نظارت و مدیریت کیفیت آب از اهمیت ویژه‌ای برخوردار است. این شاخص به شناسایی متغیرهای حیاتی کمک می‌کند. به عنوان مثال، در شبکه‌های پایش، آنتروپی شانون به اولویت‌بندی متغیرهایی که خطرات زیست‌محیطی بیشتری دارند، کمک می‌کند. همانطور که در مطالعه حوضه رودخانه دوس نشان داده شد، که در آن ۳۲ تا ۵۰ درصد از متغیرهای نظارت‌شده غیر بحرانی تلقی شدند (Cruz and Silva., 2024). علاوه بر این، چارچوب‌های مبتنی بر آنتروپی، ارزیابی فرآیندهای انتقال آلاینده‌ها را تسهیل می‌کنند و بینش‌های ارزشمندی را برای مدیریت رویدادهای حیاتی ارائه می‌دهند. علاوه بر این، شاخص‌های کیفیت آب با وزن آنتروپی (EWQI) قابلیت اطمینان ارزیابی‌های آلودگی را با به حداقل رساندن سوگیری‌های ذهنی در وزن‌دهی پارامترها بهبود می‌بخشد (Singh et al., 2022). به طور کلی، شاخص آنتروپی شانون به عنوان یک ابزار قوی برای مدیریت مؤثر کیفیت آب و نظارت بر بهینه‌سازی در سناریوهای مختلف عمل می‌کند (Pang et al., 2023). از طرفی پیش‌بینی کیفیت آب به دلایل گوناگون برای بقای تمام موجودات زنده، از جمله انسان، اهمیت ویژه‌ای دارد، چرا که آب اساسی‌ترین منبع حیات به شمار می‌رود. کیفیت آب تأثیر مستقیمی بر سلامت عمومی و اکوسیستم‌ها دارد، از این رو ارزیابی خلوص آن پیش از هرگونه استفاده ضروری است. پیش‌بینی کیفیت آب می‌تواند به شناسایی و پیشگیری از خطرات احتمالی که ممکن است در آینده بر سلامت انسان تأثیر بگذارد کمک کند و تأثیر مثبتی بر جامعه به جا بگذارد (Sarma., 2023). به دلیل پیچیدگی فراوان سیستم‌های آبی و تأثیرگذاری طیف گسترده‌ای از عوامل محیطی و انسانی پیش‌بینی کیفیت آب از جمله چالش‌های مهم در مدیریت منابع آب است. مدل‌های توسعه‌یافته با استفاده از الگوریتم‌های هوش مصنوعی، دقت و استحکام را در پیش‌بینی پارامترهای کیفیت آب و طبقه‌بندی کیفیت آب نشان داده‌اند. استفاده از هوش مصنوعی در پیش‌بینی کیفیت آب می‌تواند به طور قابل توجهی هزینه‌ها را کاهش دهد، کارایی را بهبود بخشد و به مدیریت آب و کنترل آلودگی کمک کند (Nordin et al., 2021). برنامه‌ریزی بیان ژن (GEP) یکی از روش‌های قدرتمند در حوزه هوش مصنوعی و بهینه‌سازی است که مبتنی بر الگوریتم‌های تکاملی عمل می‌کند. این روش می‌تواند به طور خودکار مدل‌های غیرخطی را ایجاد کند که قادر به شناسایی روابط پیچیده و غیرمستقیم میان پارامترهای کیفیت آب هستند. این قابلیت برای مسائل محیطی که پارامترها معمولاً تأثیرات پیچیده‌ای بر هم دارند، بسیار حیاتی است. علاوه بر این به دلیل ساختار ژنتیکی، برنامه‌ریزی بیان ژن انعطاف‌پذیری بالایی در تطبیق با داده‌های مختلف دارد. این ابزار به راحتی می‌تواند با انواع مختلف داده‌های کیفی و کمی کار کند و مدل‌های مناسب برای پیش‌بینی کیفیت آب تولید کند. در راستای استفاده از مدل‌های هوش مصنوعی برای پیش‌بینی شاخص کیفیت آب تحقیقاتی انجام شده است. Goodarzi et al., 2023، در پژوهشی از شاخص کیفیت آب و فرآیند تحلیل سلسله مراتبی فازی شاخص کیفیت آب برای بررسی وضعیت کیفی آب ۹۶ حلقه چاه عمیق کشاورزی در دشت یزد-اردکان، ایران استفاده کردند. برای این منظور از سه الگوریتم یادگیری ماشین شامل الگوریتم برنامه‌ریزی بیان ژن، درخت مدل M_5P و رگرسیون تطبیقی چند متغیره (MARS) برای پیش‌بینی شاخص کیفیت آب استفاده کردند. برای بررسی کیفیت آب در اهداف آب آشامیدنی و کشاورزی از نمودارهای Wilcox و Schoeller استفاده کردند. نتایج نشان داد بر اساس روش WQI و FAHP-WQI به ترتیب ۷۵٪ و ۳۳٪ از منطقه مورد مطالعه از کیفیت خوبی برخوردار هستند. طبق نتایج نمودار Wilcox، حدود ۳۷/۲۵٪ از چاه‌ها کیفیت آب پایینی داشتند. شاخص کیفیت آب پیش‌بینی شده با استفاده از مدل‌های ML، بر اساس چندین معیار آماری مقایسه شد و این نتیجه حاصل شد که MARS برای تخمین شاخص کیفیت آب دقیق‌تر از مدل M_5P عمل می‌کند. Mohammadpour et al., 2016، در مطالعه‌ای با استفاده از مدل‌های برنامه‌ریزی بیان ژن و شبکه عصبی مصنوعی به پیش‌بینی شاخص کیفیت آب در تالاب‌های مصنوعی پرداختند. به طوری که ۱۱ متغیر کیفیت آب در ۱۷ نقطه تالاب، ۲ بار در ماه و طی یک دوره ۱۴ ماهه مورد پایش قرار گرفت. نتایج تحلیل عاملی نشان داد که شاخص کیفیت آب به شدت تحت تأثیر pH و مواد جامد معلق قرار دارد، در حالی که دما تأثیر قابل توجهی بر شاخص کیفیت آب در مناطق گرمسیری ندارد.

در نهایت پنج پارامتر مهم pH، جامد معلق، نیتروژن آمونیاکی، اکسیژن محلول و تقاضای اکسیژن شیمیایی برای توسعه مدل‌های مورد مطالعه انتخاب شدند. نتایج نشان داد که شبکه عصبی مصنوعی با MAE برابر ۰/۱۳ و R^2 برابر ۰/۹۹، نتایج بهتری در مقایسه با برنامه ریزی بیان ژن ارائه کرد. Shahinejad & Dehgani., 2018 در پژوهشی جهت تخمین اکسیژن محلول در آب رودخانه کامبرلند واقع در ایالت تنسی عملکرد مدل‌های ماشین بردار پشتیبان، شبکه عصبی موجک و برنامه ریزی بیان ژن را مورد بررسی قرار دادند. نتایج بدست آمده آن‌ها حاکی از آن بود که ساختارهای ترکیبی در هر سه مدل عملکرد بهتری نسبت به سایر ساختارها ارائه می‌دهد. نتایج نهایی آن‌ها نشان داد که مدل شبکه عصبی موجک بیشترین ضریب همبستگی (۰/۹۶۰)، کمترین جذر میانگین مربعات خطا (۰/۶۶۸) و نیز کمترین میانگین قدرمطلق خطا (۰/۵۱۹) را در مرحله صحت سنجی دارا می‌باشد و در نهایت به این نتیجه رسیدند که کمک گرفتن از شبکه عصبی موجک و حذف نویزهای سری‌های زمانی در تخمین پارامترهای کیفی آب رودخانه، می‌تواند روشی مناسب و سریع در مدیریت کیفیت منابع آب باشد.

با توجه به اهمیت موضوع و نبود تحقیقات پیشین در زمینه استفاده از شاخص کیفیت آب آنتروپی شانون برای ارزیابی کیفیت آب رودخانه زهره و تخمین شاخص با مدل برنامه‌ریزی بیان ژن، تحقیق حاضر انجام شد.

مواد و روش‌ها

معرفی منطقه مطالعاتی و داده‌های مورد استفاده

رودخانه زهره، که به نام رودخانه هندیجان نیز شناخته می‌شود، یکی از طولانی‌ترین و پرآب‌ترین رودخانه‌های ایران است که در جنوب غربی این کشور جریان دارد. این رودخانه با طی مسافتی حدود ۴۷۵ کیلومتر از میان سه استان فارس، کهگیلویه و بویراحمد و خوزستان عبور کرده و در نهایت به خلیج فارس می‌ریزد. سرچشمه رودخانه زهره در ارتفاعات کوه سفید در استان فارس قرار دارد. در این پژوهش پارامترهای اسیدیته، هدایت الکتریکی، فسفات، نترات، آمونیوم، سختی کل، کلیرم مدفوعی، اکسیژن محلول، نیاز اکسیژن‌خواهی بیوشیمیایی، نیاز اکسیژن‌خواهی شیمیایی و کدورت به صورت ماهانه برای ۱۰ سال آبی (۱۴۰۰-۱۳۹۰) و ایستگاه‌های فیروزآباد و ۷۲۰ متری سویره در رودخانه زهره از سازمان آب و برق خوزستان دریافت شد و سپس شاخص کیفیت آب آنتروپی شانون در هر ماه و هر ایستگاه محاسبه گردید.

نرمال‌سازی داده‌ها

برای جلوگیری از تأثیر نامطلوب مقیاس متفاوت متغیرها بر مدل‌ها، ضروری است داده‌ها از طریق پیش‌پردازش تصحیح شوند. هدف از نرمال‌سازی آن است که داده‌ها به اعدادی بین صفر و یک تبدیل شوند بنابراین ابتدا تمامی داده‌ها (ورودی و خروجی) با استفاده از رابطه (۱) در محدوده صفر و ۱ نرمال‌سازی شدند (Divband Hafshejani et al., 2022).

$$x_i^* = \frac{(x_i - x_{min})}{(x_{max} - x_{min})} \quad (1)$$

که، x ، x_{min} ، x_{max} به ترتیب مقدار هر پارامتر، حداقل و حداکثر آن در مجموعه داده مورد نظر هستند.

شاخص کیفیت آب آنتروپی شانون

استانداردسازی داده‌ها: برای جلوگیری از تأثیر مقیاس‌های مختلف پارامترها، داده‌ها باید نرمال‌سازی یا استاندارد شوند. روش‌های مختلفی برای نرمال‌سازی وجود دارد که در این تحقیق با استفاده از رابطه (۲) داده‌ها نرمال شدند.

$$n_{ij} = \frac{x_{ij}}{\sum_{j=1}^m x_{ij}} \quad (2)$$

که در آن:

n_{ij} : مقدار نرمال‌سازی شده برای پارامتر j در ایستگاه i

x_{ij} : مقدار اندازه‌گیری شده برای پارامتر j در ایستگاه i

m : تعداد نمونه‌ها

محاسبه آنتروپی هر پارامتر: آنتروپی هر پارامتر با استفاده از فرمول زیر محاسبه گردید:

$$E_j = - \frac{\sum_{j=1}^m n_{ij} \ln n_{ij}}{\ln m} \quad (3)$$

که در آن:

E_j : مقدار آنتروپی پارامتر j

m : تعداد نمونه

محاسبه وزن هر پارامتر: پس از محاسبه آنتروپی هر پارامتر، وزن مربوط به آن محاسبه شد. وزن پارامترها نشان می‌دهد که هر کدام تا چه اندازه بر روی کیفیت آب تأثیرگذار هستند. وزن هر پارامتر با استفاده از آنتروپی آن پارامتر محاسبه گردید (رابطه ۴):

$$W_j = \frac{1-E_j}{\sum_{j=1}^n (1-E_j)} \quad (4)$$

که در آن:

W_j : وزن پارامتر j

n : تعداد پارامترهای کیفیت آب

محاسبه شاخص کیفیت آب آنتروپی (EWQI): برای هر ایستگاه، شاخص کیفیت آب آنتروپی با استفاده از وزن پارامترها و مقادیر نرمال‌سازی شده محاسبه می‌شود.

$$q_i = \frac{c_i}{s_i} \times 100 \quad (5)$$

$$EWQI_i = \sum_{j=1}^n w_j q_i \quad (6)$$

که در آن:

$EWQI_i$: شاخص کیفیت آب آنتروپی در ایستگاه i

q_i : مقدار واقعی هر پارامتر مطابق با مقدار استاندارد WHO

براساس EWQI، کیفیت آب به پنج درجه طبقه بندی می‌شوند، از کیفیت آب بسیار خوب تا کیفیت آب بسیار ضعیف. استانداردهای طبقه بندی کیفیت آب‌های سطحی بر اساس EWQI در جدول (۱) ارائه شده است.

جدول ۱. طبقه بندی کیفیت آب‌های سطحی بر اساس مقادیر EWQI
Table 1. Surface water quality classification based on EWQI values

کیفیت آب	رتبه	EWQI
کیفیت آب عالی	1	<50
کیفیت آب خوب	2	50-100
کیفیت آب متوسط	3	100-150
کیفیت آب ضعیف	4	150-200
کیفیت آب بسیار ضعیف	5	>200

کاربرد مدل برنامه ریزی بیان ژن در تخمین شاخص کیفیت آب آنتروپی شانون

پیش‌بینی کیفیت آب یکی از موضوعات مهم در مدیریت منابع آب است که به تحلیل پیچیده‌ای از داده‌های محیطی نیاز دارد. پارامترهای مختلفی مانند pH، نیترات، فسفات، دما، اکسیژن محلول، کدورت، BOD5 و COD بر کیفیت آب تأثیر می‌گذارند و اغلب این پارامترها به صورت غیرخطی با هم در تعامل هستند. مدل‌های سنتی مانند رگرسیون خطی به دلیل سادگی در مدل‌سازی این روابط غیرخطی ناتوان هستند. GEP به عنوان یک ابزار قدرتمند برای حل این مشکل وارد عمل می‌شود و می‌تواند روابط غیرخطی و پیچیده بین پارامترهای کیفیت آب را کشف کند و مدل‌های پیش‌بینی دقیقی ارائه دهد. با استفاده از الگوریتم تکاملی خود، قادر است از داده‌های محیطی گذشته برای یادگیری الگوهای پنهان و پیش‌بینی شاخص‌های کیفیت آب استفاده کند (Shah et al., 2021). بنابراین ابتدا یک جمعیت اولیه از کروموزوم‌ها که به صورت رشته‌های خطی از ژن‌ها نمایش داده می‌شوند، ایجاد می‌شود. هر کروموزوم نشان‌دهنده یک مدل بالقوه است که به پیش‌بینی کیفیت آب کمک می‌کند. هر کروموزوم خطی به یک درخت بیان ریاضی تبدیل می‌شود که می‌تواند معادلات یا مدل‌های غیرخطی را برای داده‌های ورودی که پارامترهای کیفیت هستند تولید کند. برای هر مدل تولید شده، مقدار خطا یا دقت پیش‌بینی محاسبه می‌شود. این فرآیند به مدل‌ها امتیازی می‌دهد تا بهترین آن‌ها برای نسل‌های بعدی انتخاب شوند. از روی کروموزوم‌های با امتیاز بالا، با استفاده از عملیات‌هایی نظیر جهش ژنتیکی، تلاقی و انتخاب طبیعی، کروموزوم‌های جدید تولید می‌شوند. این فرآیند در چندین نسل تکرار می‌شود تا مدلی با دقت بالا برای پیش‌بینی کیفیت آب به دست آید (Ferreira., 2001). در نهایت، مدلی که بهترین عملکرد را بر اساس معیارهای انتخاب شده داشته باشد، به عنوان مدل نهایی GEP انتخاب می‌شود.

ارزیابی دقت مدل برنامه ریزی بیان ژن در تخمین شاخص کیفیت آب آنتروپی شانون

در این پژوهش، مجموعه داده‌های مورد استفاده به منظور آموزش و ارزیابی مدل برنامه ریزی بیان ژن برای پیش‌بینی شاخص کیفیت آب آنتروپی شانون، به دو زیرمجموعه تقسیم شدند. ۸۰ درصد از داده‌ها برای آموزش مدل‌ها و ۲۰ درصد برای سنجش عملکرد آن در پیش‌بینی شاخص کیفیت آب اختصاص یافت (Aldrees et al., 2022). این روش، که به عنوان "تقسیم تصادفی داده‌ها" شناخته می‌شود، رویکردی متداول در علم داده برای اطمینان از صحت و تعمیم‌پذیری نتایج به شمار می‌رود. کارایی مدل برنامه‌ریزی بیان ژن در پیش‌بینی شاخص کیفیت آب آنتروپی شانون با استفاده از معیارهای آماری متداول مانند ریشه میانگین مربعات خطا، (RMSE) و ضریب تعیین (R^2) ارزیابی گردید.

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (P_i - O_i)^2}{n}} \quad (7)$$

$$R^2 = \frac{\sum_{i=1}^n (O_i - \bar{O})(P_i - \bar{P})}{\sqrt{\sum_{i=1}^n (O_i - \bar{O})^2 \sum_{i=1}^n (P_i - \bar{P})^2}} \quad (8)$$

در روابط فوق n تعداد داده‌ها، P_i و O_i ، به ترتیب مقادیر اندازه‌گیری شده و پیش‌بینی شده عملکرد هستند. هر چه شاخص RMSE به صفر نزدیکتر باشد، نشان‌دهنده خطای کمتر و کارایی بهتر مدل می‌باشد. مقدار شاخص ضریب تبیین در دامنه منفی یک تا یک تغییر می‌کند. مقدار یک و منفی یک به ترتیب نشان‌دهنده همبستگی کامل و معکوس مقادیر مشاهداتی و شبیه‌سازی شده می‌باشد. هر چه مقدار این شاخص به یک نزدیکتر باشد، کارایی مدل بهتر است.

نتایج و بحث

ارزیابی کیفیت آب با استفاده از شاخص کیفیت آب آنتروپی شانون

شکل ۱، نتایج بررسی کیفیت آب در ایستگاه فیروز آباد و پل ۷۲۰ متری سوپره توسط شاخص کیفیت آب آنتروپی را نشان می‌دهد. مقدار بدست آمده از میانگین شاخص کیفیت آب آنتروپی در ایستگاه فیروز آباد معادل ۲۰۹/۱۰ محاسبه شد که نشان‌دهنده کیفیت آب بسیار ضعیف این ایستگاه می‌باشد. همچنین میانگین ۱۰ ساله شاخص کیفیت آب آنتروپی شانون در ایستگاه پل ۷۲۰ متری سوپره معادل ۱۷۶/۴۴ بدست آمد، که مطابق با جدول ۱ نشان‌دهنده کیفیت ضعیف آب در این ایستگاه می‌باشد.

از مقایسه شاخص کیفیت آب آنتروپی در بازه زمانی ۱۰ ساله در دو ایستگاه فیروز آباد و پل ۷۲۰ متری سویره این نتیجه حاصل می‌شود که کیفیت آب ایستگاه ۷۲۰ متری سویره نسبت به ایستگاه فیروز آباد در وضعیت بهتری قرار دارد.



شکل ۱. مقایسه بین مقدار شاخص کیفیت آب آنتروپی شانون برای ایستگاه‌های مورد مطالعه
 Fig 1. Comparison of Shannon Entropy Water Quality Index values for the studied stations

تعیین پارامترهای تاثیرگذار با استفاده از وزن دهی آنتروپی شانون

خلاصه ضریب وزنی پارامترهای کیفیت آب حاصل از روش آنتروپی در جدول ۱ آورده شده است.

جدول ۲. وزن پارامترهای کیفیت آب در شاخص کیفیت آب آنتروپی شانون

Table 2. Weight of Water Quality Parameters in Shannon Entropy Water Quality Index

پارامتر	BOD	COD	DO	EC	NH ₄	NO ₃	PO ₄	Total Hardness	Turbidity	pH
وزن	0.1235	0.1177	0.1261	0.1217	0.1238	0.1332	0.0518	0.1025	0.0005	0.0992

وزن بالاتر به معنای تاثیر بیشتر آن پارامتر بر کیفیت آب است. به عنوان مثال، پارامتری که وزن بالاتری دارد، تغییرات در آن پارامتر تاثیر بیشتری بر کیفیت کلی آب خواهد گذاشت. وزن پایین نشان می‌دهد که تغییرات در آن پارامتر تاثیر کمتری بر کیفیت کلی آب دارد. در ادامه تحلیل کاملی از وزن‌های اختصاص داده شده به هر پارامتر ارائه شده است. وزن نسبتاً بالای BOD (۰/۱۲۳۵) نشان‌دهنده تاثیر قابل توجه آن بر کیفیت آب است. مقادیر بالای BOD معمولاً نشان‌دهنده آلودگی‌های آلی بیشتر است که می‌تواند سطح اکسیژن محلول را کاهش داده و به حیات آبیان آسیب برساند. وزن COD (۰/۱۱۷۷) نزدیک به BOD است که اهمیت اندازه‌گیری آلاینده‌های آلی و معدنی که می‌توانند به صورت شیمیایی سطح اکسیژن را کاهش دهند، نشان می‌دهد. وزن DO نیز بالا است که نقش مهم آن در ارزیابی کیفیت آب را نشان می‌دهد. سطح پایین DO می‌تواند نشانه‌ای از آلودگی به‌ویژه ناشی از ضایعات آلی باشد و بر موجودات آبی تاثیر منفی بگذارد (Maddah., 2022). وزن اختصاص داده شده به هدایت الکتریکی (۰/۱۲۱۷) اهمیت آن را در تشخیص غلظت یون‌ها در آب که می‌تواند نشان‌دهنده وجود آلاینده‌های مختلف باشد، منعکس می‌کند. آمونیوم با وزن ۰/۱۲۳۸ یک پارامتر مهم برای ارزیابی کیفیت آب است. سطوح بالای آمونیوم ممکن است نشان‌دهنده روان‌آب‌های کشاورزی یا تخلیه فاضلاب باشد. نترات بالاترین وزن را در میان پارامترهای ذکر شده دارد که تاثیر قابل توجه آن بر کیفیت آب را نشان می‌دهد. سطوح بالای نترات می‌تواند ناشی از روان‌آب‌های کشاورزی باشد و منجر به یوتروفیکاسیون شود (Aldrees et al., 2020). فسفات نسبت به سایر پارامترها وزن کمتری دارد (۰/۰۵۱۸) که نشان‌دهنده تاثیر نسبتاً کمتر اما همچنان مهم آن بر کیفیت آب است که اغلب با آلودگی‌های غذایی و یوتروفیکاسیون مرتبط است. سختی کل با وزن ۰/۱۰۲۵ نشان می‌دهد که در حالی که سختی بر کیفیت آب تاثیر دارد، تاثیر آن نسبت به پارامترهایی مانند

نیترات یا BOD کمتر است. وزن pH نشان دهنده نقش آن در تأثیرگذاری بر شیمی آب و حیات آبیان است، هرچند به اندازه پارامترهایی مانند نیترات یا اکسیژن محلول تأثیرگذار نیست.

به طور کلی وزن‌ها نشان می‌دهند که پارامترهای مرتبط با مواد مغذی مانند NO_3 و BOD تأثیر بیشتری بر شاخص کیفیت آب در این مدل دارند و نیاز به کنترل آلودگی‌های آلی و روان‌آب‌های کشاورزی را برجسته می‌سازند. از طرف دیگر، پارامترهایی مانند کدورت و فسفات وزن کمتری دارند که نشان می‌دهد تأثیر آن‌ها در این زمینه کمتر است. این روش وزن‌دهی به اولویت‌بندی اقدامات مدیریتی برای بهبود کیفیت آب بر اساس اهمیت نسبی هر پارامتر کمک می‌کند.

تعیین پارامترهای تأثیرگذار با استفاده از ضریب همبستگی پیرسون

همبستگی بین متغیرهای ورودی و متغیر هدف با استفاده از ضریب همبستگی پیرسون (رابطه ۷) اندازه‌گیری شد و نتایج آن در شکل ۲ نشان داده شده است (Divband Hafshejani et al., 2020).

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

در رابطه بالا \bar{x} و \bar{y} میانگین متغیرهای x و y می‌باشد.

شکل ۲ ماتریس همبستگی، رابطه بین پارامترهای مختلف کیفیت آب و شاخص کیفیت آب آنتروپی شانون (EWQI) را نشان می‌دهد. در این ماتریس، اعداد بزرگتر و نزدیک به ۱ نشان‌دهنده همبستگی قوی‌تر مثبت و اعداد نزدیک به -۱ نشان‌دهنده همبستگی منفی قوی‌تر هستند. اعداد نزدیک به ۰ به معنی همبستگی ضعیف یا عدم همبستگی بین دو پارامتر است. همبستگی مثبت نسبتاً بالایی (۰/۹) بین هدایت الکتریکی (EC) و شاخص کیفیت آب آنتروپی شانون دیده می‌شود. این نشان می‌دهد که با افزایش EC، شاخص کیفیت آب نیز افزایش می‌یابد. این موضوع می‌تواند نشان‌دهنده نقش هدایت الکتریکی به عنوان یک عامل مهم در تعیین کیفیت آب باشد. سختی کل با شاخص کیفیت آب آنتروپی شانون همبستگی متوسط (۰/۶) دارد، که نشان می‌دهد سختی آب می‌تواند در کیفیت آب تأثیرگذار باشد.



شکل ۲. ضرایب همبستگی پیرسون بین پارامترهای کیفیت آب و شاخص کیفیت آب آنتروپی شانون در منطقه مورد مطالعه
 Fig 2. Pearson correlation coefficients between water quality parameters and Shannon entropy-based water quality index in the study area

تخمین شاخص کیفیت آنتروپی شانون با مدل برنامه ریزی بیان ژن

به منظور تخمین شاخص کیفیت آب آنتروپی شانون با استفاده از مدل برنامه‌ریزی بیان ژن، از سه نوع داده ورودی مختلف استفاده شد. این تنوع در انتخاب داده‌های ورودی به دلایل مختلفی صورت گرفت که هر یک به منظور بهبود دقت مدل و شناسایی بهتر پارامترهای موثر بر شاخص کیفیت آب بود. هدف اصلی این مرحله، یافتن ترکیب بهینه‌ای از داده‌های ورودی بود که منجر به بهترین عملکرد مدل در پیش‌بینی شاخص مورد نظر شود (Khair & Dhanalakshmi, 2022).

حالت اول: استفاده از همه داده‌های کیفیت آب موجود به عنوان ورودی

در این حالت، تمام پارامترهای اندازه‌گیری شده کیفیت آب به مدل داده شد. هدف از این روش، ایجاد یک مبنای کلی برای بررسی دقت مدل بود که تمامی اطلاعات در دسترس را در بر می‌گرفت. این روش اجازه می‌داد تا تأثیر همه پارامترها بر شاخص کیفیت آب به صورت همزمان بررسی شود و در صورت وجود همبستگی‌های پنهان یا پیچیدگی‌های غیرمنتظره بین داده‌ها، آن‌ها در مدل لحاظ شوند.

حالت دوم: استفاده از داده‌های موثر استخراج شده از وزن‌دهی آنتروپی شانون به عنوان ورودی

در این حالت، پارامترهای کیفیت آب بر اساس وزن‌دهی آنتروپی شانون فیلتر شدند. پارامترهایی که وزن بیشتری داشتند، به عنوان ورودی انتخاب شدند زیرا نقش مهم‌تری در تعیین شاخص کیفیت آب ایفا می‌کردند. هدف از این روش، کاهش تعداد ورودی‌ها و تمرکز بر پارامترهایی بود که تأثیر بیشتری بر شاخص آنتروپی شانون داشتند. این می‌توانست منجر به کاهش پیچیدگی مدل و افزایش دقت در تخمین شود، زیرا پارامترهای غیرموثر و نویز احتمالی حذف شده بودند.

حالت سوم: استفاده از داده‌های مؤثر استخراج‌شده از همبستگی پیرسون به عنوان ورودی

در این حالت، پارامترهای ورودی بر اساس ضریب همبستگی پیرسون با شاخص آنتروپی شانون انتخاب شدند. پارامترهایی که همبستگی قوی‌تری با شاخص داشتند، به عنوان ورودی انتخاب شدند. این روش کمک می‌کند تا پارامترهایی که بیشترین تأثیر مستقیم بر شاخص کیفیت آب داشتند، شناسایی شوند. هدف از این رویکرد، شناسایی روابط خطی بین پارامترها و شاخص هدف و بهبود عملکرد مدل از طریق انتخاب ویژگی‌های مناسب بود.

نتایج تأثیر انتخاب داده‌های ورودی مختلف بر دقت مدل برنامه‌ریزی بیان ژن در تخمین شاخص کیفیت آب آنتروپی شانون در جدول می‌باشد. این جدول به وضوح تفاوت عملکرد مدل را در سه حالت مختلف بیان می‌کند: استفاده از همه داده‌های موجود، داده‌های مؤثر استخراج‌شده از وزن‌دهی آنتروپی شانون، و داده‌های مؤثر بر اساس همبستگی پیرسون. بررسی نتایج موجود در جدول به ما کمک می‌کند تا درک بهتری از پارامترهای مؤثر و روش‌های بهینه برای بهبود دقت مدل در پیش‌بینی شاخص کیفیت آب به دست آوریم.

جدول ۳. مقایسه عملکرد مدل برنامه‌ریزی بیان ژن در پیش‌بینی شاخص کیفیت آب آنتروپی شانون با استفاده از داده‌های ورودی مختلف
Table 3. Comparison of the Performance of Gene Expression Programming Model in Predicting Shannon Entropy Water Quality Index Using Different Input Data

مدل	R ²	RMSE
برنامه‌ریزی بیان ژن	0.803	0.070
برنامه‌ریزی بیان ژن و آنتروپی شانون	0.851	0.067
برنامه‌ریزی بیان ژن و همبستگی پیرسون	0.723	0.073

نتیجه‌گیری

در این پژوهش، مدل برنامه‌ریزی بیان ژن (GEP) برای پیش‌بینی شاخص کیفیت آب آنتروپی شانون به کار گرفته شد و دقت آن با استفاده از معیارهای آماری مانند ریشه میانگین مربعات خطا (RMSE) و ضریب تعیین (R²) ارزیابی گردید. نتایج نشان دادند که این مدل توانایی خوبی در پیش‌بینی شاخص کیفیت آب دارد. استفاده از مدل GEP امکان کشف روابط پیچیده و غیرخطی بین پارامترهای مختلف کیفیت آب را فراهم کرد و منجر به بهبود دقت پیش‌بینی شد.

تحلیل نتایج نشان داد که انتخاب داده‌های ورودی مختلف تأثیر قابل توجهی بر عملکرد مدل دارد. استفاده از پارامترهای مؤثر استخراج‌شده از وزن‌دهی آنتروپی شانون یا همبستگی پیرسون منجر به بهبود دقت مدل و کاهش خطا شد. به‌طور کلی، پارامترهایی مانند نیترات، BOD، و آمونیوم تأثیر بیشتری بر کیفیت آب نشان دادند و مدل بر مبنای این پارامترها عملکرد بهتری داشت. این یافته‌ها اهمیت انتخاب ویژگی‌ها و استفاده از مدل‌های پیشرفته را برای بهبود دقت در پیش‌بینی شاخص‌های کیفیت آب تأیید می‌کنند و می‌توانند به مدیران منابع آب در تصمیم‌گیری‌های بهتر کمک کنند.

منابع

- Aldrees et al., 2022. (, L. N. T., Minh, D. N., Quang, T. P., Mong, T. V. T., Le Xuan, T., Son, H. P., & Van, P. N. (2020). Assessment of nitrogen nutrient sources in aquatic environment of Tuyen Lam sub-catchment based on its stable isotopes ratio ($\delta^{15}\text{N-NO}_3$) combined with geochemical parameters. *Nuclear Science and Technology*, 10(2), 32-43.
- Aldrees, A., Khan, M. A., Tariq, M. A. U. R., Mustafa Mohamed, A., Ng, A. W. M., & Bakheit Taha, A. T. (2022). Multi-expression programming (MEP): water quality assessment using water quality indices. *Water*, 14(6), 947.
- Cruz, F. M., & Silva, T. F. D. G. (2024). Water quality emergency monitoring networks: A method for identifying non-critical variables based on Shannon's entropy. *Journal of Hydroinformatics*, 26(3), 658-682.
- Divband Hafshejani, L., Naseri, A. A., Moradzadeh, M., Daneshvar, E., & Bhatnagar, A. (2022). Applications of soft computing techniques for prediction of pollutant removal by environmentally friendly adsorbents (case study: the nitrate adsorption on modified hydrochar). *Water Science & Technology*, 86(5), 1066-1082.
- Ferreira, C. (2001). Gene expression programming: a new adaptive algorithm for solving problems. *arXiv preprint cs/0102027*.
- 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.

- Hojjati-Najafabadi, A., Mansoorianfar, M., Liang, T., Shahin, K., & Karimi-Maleh, H. (2022). A review on magnetic sensors for monitoring of hazardous pollutants in water resources. *Science of the Total Environment*, 824, 153844.
- Khaire, U. M., & Dhanalakshmi, R. (2022). Stability of feature selection algorithm: A review. *Journal of King Saud University-Computer and Information Sciences*, 34(4), 1060-1073.
- Maddah, H. A. (2022). Predicting optimum dilution factors for BOD sampling and desired dissolved oxygen for controlling organic contamination in various wastewaters. *International Journal of Chemical Engineering*, 2022(1), 8637064.
- Mohammadpour, R., Shaharuddin, S., Zakaria, N. A., Ghani, A. A., Vakili, M., & Chan, N. W. (2016). Prediction of water quality index in free surface constructed wetlands. *Environmental Earth Sciences*, 75, 1-12.
- Nordin, N. F. C., Mohd, N. S., Koting, S., Ismail, Z., Sherif, M., & El-Shafie, A. (2021). Groundwater quality forecasting modelling using artificial intelligence: A review. *Groundwater for Sustainable Development*, 14, 100643.
- Pang, T., Jiang, J., Alfonso, L., Yang, R., Zheng, Y., Wang, P., & Zheng, T. (2023). Deriving analytical expressions of the spatial information entropy index on riverine water quality dynamics. *Journal of Hydrology*, 623, 129806.
- Sarma, A. (2023, February). An Analysis on the Techniques for Water Quality Prediction from Remotely Sensed data. In *2023 International Conference on Recent Trends in Electronics and Communication (ICRTEC)* (pp. 1-6). IEEE.
- Sarma, A. (2023, February). An Analysis on the Techniques for Water Quality Prediction from Remotely Sensed data. In *2023 International Conference on Recent Trends in Electronics and Communication (ICRTEC)* (pp. 1-6). IEEE.
- Shah, M. I., Javed, M. F., & Abunama, T. (2021). Proposed formulation of surface water quality and modelling using gene expression, machine learning, and regression techniques. *Environmental Science and Pollution Research*, 28, 13202-13220.
- Shahinejad, B., & Dehgani, R. (2018). Comparison of wavelet neural network models, support vector machine and gene expression programming in estimating the amount of oxygen dissolved in rivers. *Iran Water Resources Research*, 14(3), 226-238.
- Singh, K. R., Dutta, R., Kalamdhad, A. S., & Kumar, B. (2022). Study of physicochemical parameters and wetland water quality assessment by using Shannon's entropy. *Applied Water Science*, 12(11), 247.
- Zheng, K. G. Q. L. J. Z. J. X. W. M. (2017). Applications of support vector machine and improved k-Nearest Neighbor algorithm in fault diagnosis and fault degree evaluation of gas insulated switchgear. 1st International Conference on Electrical Materials and Power Equipment, Xian, Pppler R China.