



Simulation of Water Quality Parameters Using Intelligent Machine Learning Methods in the Halil River

Ghasem Shokrizadeh¹, Mohammad Reza Rezaei^{2*}, Mohammad Hossein Sayadi³, Saeid Akbarifard⁴

¹PhD Student, Department of Environment, Faculty of Natural Resources and Environment, University of Birjand, Iran

²Associate Professor, Department of Environment, Faculty of Natural Resources and Environment, University of Birjand, Iran

³Professor, National Institute of Oceanography and Atmospheric Science, Tehran, Iran

⁴Assistant Professor, Department of Water Engineering, Faculty of Civil and Surveying Engineering, Graduate University of Advanced Technology, Kerman, Iran.

Abstract

In this study, three intelligent modeling approaches, including Adaptive Neuro-Fuzzy Inference System, Artificial Neural Network–Multilayer Perceptron, and Artificial Neural Network–Radial Basis Function, were employed to simulate 14 water quality parameters, in the Halil River over a long-term monthly dataset spanning 52 years. Streamflow discharge was used as an input variable in some models. The performance of the developed models was evaluated using statistical indicators including the coefficient of determination (R^2), root mean square error (RMSE), mean absolute error (MAE), and Willmott's index of agreement (d). The results indicated that although all three artificial intelligence models demonstrated high capability in simulating the investigated water quality parameters, the ANFIS model outperformed the other models in most cases. Specifically, for SAR, values of $R^2 = 0.99$ and $RMSE = 1.04$ were obtained; for %Na, $R^2 = 0.99$ and $RMSE = 3.02$; and for SO_4 , $R^2 = 0.93$ and $RMSE = 0.61$. Among the input scenarios, model (b), which incorporates five easily measurable parameters, is recommended as a practical and efficient model, as it reduced RMSE by approximately 65% compared to model (a) (based solely on discharge). The findings of this study demonstrate that intelligent machine learning techniques can be effectively used to estimate water quality parameters in cases where direct measurements are unavailable, based on other measured variables, thereby reducing laboratory analysis costs.

Keywords: Halil River, Water quality parameters, Water quality simulation, Machine learning; ANFIS, ANN-MLP, ANN-RBF.

*Corresponding Author: Mohammad Reza Rezaei

Email: mrrezaei@birjand.ac.ir

Phone: 09151615414

Doi: 10.48306/juem.2026.579245.1150



شبیه‌سازی پارامترهای کیفی آب با استفاده از روش‌های هوشمند یادگیری ماشین در رودخانه هلیل‌رود

قاسم شکری زاده^۱، محمدرضا رضایی^{۲*}، محمدحسین صیادی^۳، سعید اکبری فرد^۴

^۱ دانشجوی دکتری، گروه محیط زیست، دانشکده منابع طبیعی و محیط زیست دانشگاه بیرجند، ایران

^۲ دانشیار، گروه محیط زیست، دانشکده منابع طبیعی و محیط زیست دانشگاه بیرجند، ایران

^۳ استاد، پژوهشگاه ملی اقیانوس شناسی و علوم جوی، تهران، ایران

^۴ استادیار، گروه مهندسی آب، دانشکده مهندسی عمران و نقشه برداری، دانشگاه تحصیلات تکمیلی صنعتی و فناوری پیشرفته، کرمان، ایران

چکیده

در این پژوهش از سه مدل سیستم استنتاجی عصبی-فازی تطبیقی، شبکه عصبی مصنوعی چند لایه پرسپترون و شبکه عصبی مصنوعی پایه شعاعی برای شبیه‌سازی ۱۴ پارامتر کیفی آب در رودخانه هلیل‌رود طی یک دوره آماری ماهانه بلند مدت ۵۲ ساله استفاده شد. دبی جریان (Q) به عنوان ورودی برخی مدل‌ها به کار گرفته شد. ارزیابی این سه روش هوش مصنوعی توسط معیارهای آماری ضریب تعیین (R^2)، جذر میانگین مربعات خطا (RMSE)، میانگین قدر مطلق خطا (MAE) و شاخص توافق ویلموت (d) انجام شد. نتایج حاصل از این پژوهش نشان داد، علی‌رغم اینکه هر سه روش هوش مصنوعی ANFIS، ANN-MLP و ANN-RBF قابلیت بسیار بالایی در شبیه‌سازی ۱۴ پارامتر کیفی مورد بررسی داشته‌اند، روش ANFIS در اکثر مدل‌ها برتر بوده است؛ به طوری که برای پارامتر SAR مقدار $R^2=0.99$ و $RMSE=1.04$ ، برای پارامتر Na^+ مقدار $R^2=0.99$ و $RMSE=3.02$ و برای پارامتر SO_4 مقدار $R^2=0.93$ و $RMSE=0.61$ حاصل گردید. مدل b (با پنج پارامتر آسان‌اندازه‌گیری) به عنوان مدل عملیاتی پیشنهادی معرفی می‌شود که خطای RMSE را نسبت به مدل a (تنها با دبی) به طور متوسط ۶۵٪ کاهش داد. این پژوهش نشان می‌دهد که با استفاده از روش‌های هوشمند یادگیری ماشین می‌توان غلظت پارامترهای کیفی را در صورت عدم اندازه‌گیری آنها و بر اساس غلظت سایر پارامترهای کیفی محاسبه کرد، که این امر هزینه‌های آزمایشگاهی را کاهش می‌دهد.

کلمات کلیدی: رودخانه هلیل‌رود، پارامترهای کیفی آب، شبیه‌سازی کیفیت آب، ANFIS، ANN-MLP، ANN-RBF

۱- مقدمه

در عصر حاضر، به کارگیری رویکردهای مبتنی بر یادگیری ماشین در پایش کیفیت آب، به عنوان یک ضرورت اساسی برای مدیریت یکپارچه منابع آبی و حفاظت از اکوسیستم‌های طبیعی مطرح است. این امر به ویژه در مناطقی که با کمبود زیرساخت‌های دقیق سنجش کیفیت آب مواجه هستند، اهمیتی مضاعف می‌یابد. روش‌های هوش مصنوعی مانند ANN، ANN-MLP و ANN-RBF با ترکیب قابلیت‌های یادگیری ماشین و مدل‌سازی غیرخطی، امکان شبیه‌سازی دقیق پارامترهای پیچیده کیفی آب را فراهم می‌کنند. این روش‌ها می‌توانند روابط پیچیده بین پارامترهای ورودی و خروجی را بدون نیاز به فرضیات ساده‌کننده، مدل‌کنند. پژوهش‌های گسترده‌ای در زمینه کاربرد مدل‌های محاسباتی نرم و روش‌های یادگیری ماشین در پایش و پیش‌بینی پارامترهای کیفیت آب انجام شده است. این مطالعات عمدتاً بر بهبود دقت پیش‌بینی از طریق بهینه‌سازی الگوریتم‌ها و یا ترکیب مدل‌های مختلف با یکدیگر متمرکز بوده‌اند.

در یکی از تحقیقات اولیه، به مقایسه عملکرد شبکه عصبی مصنوعی (ANN)، سیستم استنتاج فازی-عصبی تطبیقی (ANFIS) و برنامه‌ریزی بیان ژن (GEP) در مدل‌سازی غلظت اکسیژن محلول (DO) در یک رودخانه در آمریکا پرداخته شد. آن‌ها با استفاده از معیارهای ارزیابی نظیر ضریب تعیین (R^2)، ریشه میانگین مربعات خطا (RMSE)، میانگین قدر مطلق خطا (MAE) و میانگین قدر مطلق خطای نسبی (MAPE)، به تحلیل عملکرد مدل‌ها پرداختند [۱]. در [۲] قابلیت مدل ANN و مدل ترکیبی شبکه عصبی-موجک (WANN) در پیش‌بینی ماهانه هدایت الکتریکی (EC) آب یک ایستگاه هیدرومتری در ترکیه مورد بررسی و ارزیابی قرار گرفت [۲]. همچنین در پژوهشی در هند، دقت سه روش رگرسیون اسپلاین تطبیقی چندمتغیره (MARS)، حداقل مربعات ماشین بردار پشتیبان (LSSVM) و مدل درختی M5 در شبیه‌سازی آلودگی رودخانه با یکدیگر مقایسه شدند [۳]. در مطالعه دیگری، به ارزیابی مدل‌های ترکیبی موجک با برنامه‌ریزی بیان ژن (WGEP) و شبکه عصبی (WANN) در پیش‌بینی نیتروژن کل (TN) در آب‌های ساحلی پرداخته شد. عملکرد این مدل‌ها با نسخه‌های ساده ANN، GEP و نیز رگرسیون چندمتغیره خطی (MLR) مقایسه شد. نتایج، بر اساس معیارهای نش-ساتکلیف (NSE) و RMSE، نشان‌دهنده برتری قابل توجه مدل‌های ترکیبی مبتنی بر تبدیل موجک در پیش‌بینی این پارامتر بود [۴]. بزرگ‌حداد و همکاران با بهره‌گیری از دو رویکرد داده‌کاوی، به مدل‌سازی پارامترهای کیفی آب رودخانه سفیدرود پرداختند. مدل‌های مورد استفاده شامل حداقل مربعات رگرسیون بردار پشتیبان بهینه‌شده با الگوریتم ژنتیک (GA-LSSVR) و برنامه‌ریزی ژنتیک (GP) بود. در این پژوهش، متغیرهای ورودی بر اساس تحلیل مؤلفه‌های اصلی (PCA) انتخاب شدند. ارزیابی‌ها با استفاده از شاخص‌های R^2 و NSE نشان داد که مدل هیبریدی GA-LSSVR از دقت بالاتری نسبت به GP در تخمین پارامترهای کیفیت آب برخوردار است [۵].

در تحقیقات داخلی، در [۶] از مدل ANFIS بهینه‌شده با الگوریتم‌های ازدحام ذرات (PSO) و کلونی مورچگان (ACO) برای تخمین پارامترهای EC، کل مواد جامد محلول (TDS)، نسبت جذب سدیم (SAR)، سختی کربناته و سختی کل (TH) در رودخانه زاینده‌رود استفاده شد. در منطقه‌ای مشابه، نیز کارایی مدل‌های ANFIS و ANFIS-PSO در تخمین پارامترهای EC، TDS، TH و قلیائیت کل در منابع آبی استان کرمانشاه مورد بررسی قرار گرفت [۷]. همچنین، حقی‌آبی و همکاران در مطالعه‌ای بر روی رودخانه تیر، توانایی روش‌های مختلف هوش محاسباتی شامل ANN، ماشین بردار پشتیبان (SVM) و روش گروهی پردازش داده‌ها (GMDH) را در تخمین غلظت پارامترهایی نظیر یون‌های سدیم (Na^+)، کلسیم (Ca^{2+})، منیزیم (Mg^{2+})، کلر (Cl^-)، سولفات (SO_4^{2-})، بیکربنات (HCO_3^-)، اسیدپته (PH)، EC، TDS و SAR مورد ارزیابی و مقایسه قرار دادند [۸].

در سال‌های اخیر، با پیشرفت مدل‌های یادگیری عمیق، استفاده از این مدل‌ها در پیش‌بینی پارامترهای کیفیت آب به‌طور قابل توجهی افزایش یافته است. به‌عنوان نمونه، وو و وانگ یک مدل هیبریدی مبتنی بر شبکه عصبی، تبدیل موجک و LSTM ارائه دادند که توانست دقت پیش‌بینی کیفیت آب را نسبت به مدل‌های منفرد به‌طور معنی‌داری بهبود دهد [۹]. همچنین تیان و همکاران با ترکیب تحلیل ارتباط خاکستری و الگوریتم LSTM نشان دادند که انتخاب ویژگی‌های مؤثر می‌تواند عملکرد پیش‌بینی پارامترهای کیفیت آب را بهبود دهد [۱۰]. در ادامه، گائو و همکاران با توسعه یک مدل چندمتغیره مبتنی بر LSTM برای پیش‌بینی پارامترهایی نظیر DO و TN، به دقت بالایی در مدل‌سازی روابط غیرخطی داده‌های کیفی آب دست یافتند [۱۱]. همچنین کای و همکاران با ارائه مدل KF-LSTM مبتنی بر ساختار encoder-decoder، مشکل داده‌های گمشده در پایش کیفیت آب را برطرف کرده و عملکرد مدل را بهبود دادند [۱۲]. علاوه بر این، روآن و همکاران یک مدل هیبریدی RF-CEEMD-LSTM را برای پیش‌بینی

آلودگی آب توسعه دادند که با ترکیب جنگل تصادفی، تجزیه سیگنال و LSTM توانست دقت پیش‌بینی را در داده‌های غیرایستا و غیرخطی افزایش دهد [۱۳]. در همین راستا، چن و همکاران با ترکیب تحلیل موجک و انتقال یادگیری در قالب مدل WA-LSTM-TL، نشان دادند که استفاده از تکنیک‌های هیبریدی می‌تواند محدودیت‌های داده‌ای و ناپایداری سری‌های زمانی کیفیت آب را کاهش دهد [۱۴]. همچنین استفاده از معماری‌های پیشرفته‌تر مانند BiLSTM و مدل‌های مبتنی بر مکانیزم توجه (Attention) نیز مورد توجه قرار گرفته است، به‌طوری‌که مطالعات اخیر نشان داده‌اند این مدل‌ها توانایی بالایی در استخراج وابستگی‌های زمانی پیچیده در داده‌های کیفیت آب دارند [۱۵].

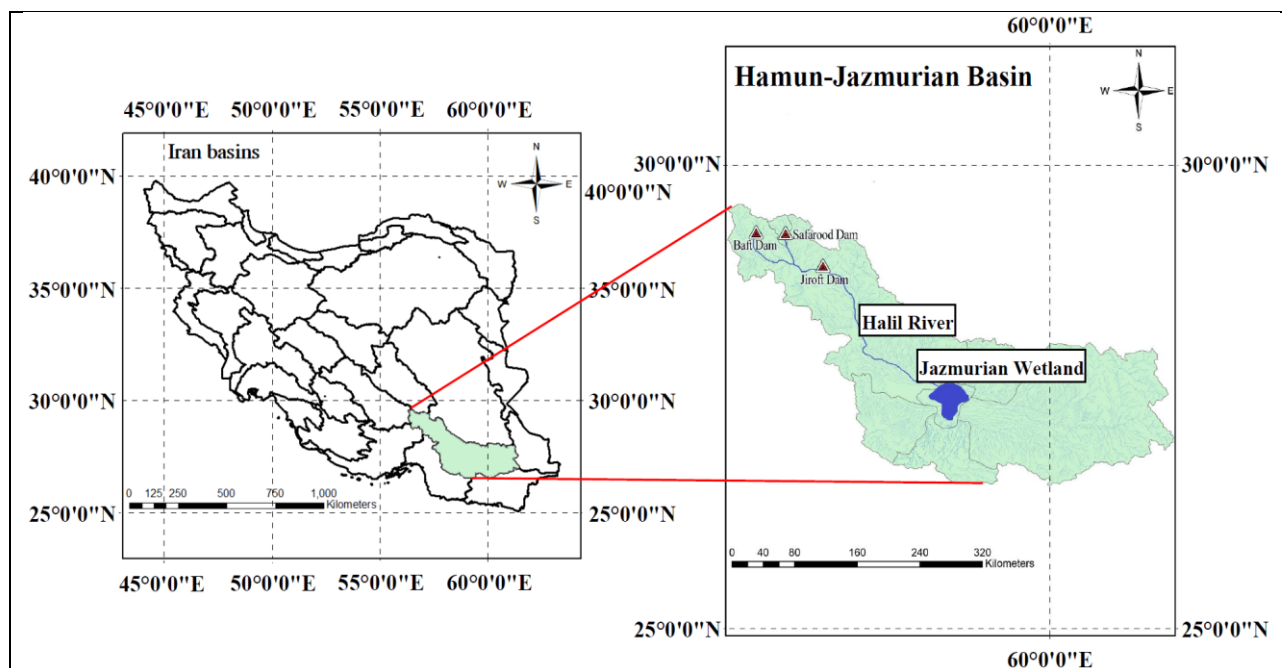
نوآوری اصلی این پژوهش عبارت است از: (۱) استفاده همزمان از سه روش هوش مصنوعی ANFIS، ANN-MLP و ANN- RBF برای شبیه‌سازی ۱۴ پارامتر کیفی مختلف در یک بازه آماری بلندمدت ۵۲ ساله در حوضه نیمه‌خشک هلیل‌رود که تاکنون مشابه آن برای این منطقه انجام نشده است؛ (۲) طراحی سه الگوی ورودی متفاوت (a، b و c) برای هر پارامتر کیفی و مقایسه سیستماتیک آنها، به‌طوری‌که مدل b با پنج پارامتر آسان‌اندازه‌گیری (Q، TDS، EC، PH و TH) برای اولین بار به عنوان «مدل عملیاتی پیشنهادی» برای ایستگاه‌های با داده‌های محدود معرفی می‌شود؛ (۳) استفاده از روش رگرسیون گام به گام به همراه ضریب همبستگی پیرسون برای انتخاب بهینه ورودی‌ها و کاهش وابستگی به اندازه‌گیری‌های پرهزینه آزمایشگاهی. این نوآوری برای ایستگاه‌های هیدرومتری ایران بسیار حائز اهمیت است، زیرا در بسیاری از سال‌ها فقط تعداد محدودی از پارامترهای کیفی اندازه‌گیری می‌شوند.

۲- مواد و روش‌ها

در این بخش به بررسی حوضه آبریز مورد مطالعه و مدل‌های به کار رفته برای شبیه‌سازی پارامترهای کیفی حوضه مورد مطالعه پرداخته شده است.

۲-۱- حوضه مورد مطالعه

حوضه آبریز هلیل رود بین مختصات جغرافیایی ۵۶°۵۴' تا ۵۹°۱۵' طول شرقی و ۲۷°۱۵' تا ۲۹°۳۰' عرض شمالی قرار گرفته است. این حوضه یکی از مهم‌ترین منابع آب در جنوب شرقی ایران است و رودخانه هلیل رود نقش کلیدی در تامین آب کشاورزی و شرب منطقه دارد. در این پژوهش از داده‌های کیفیت آب ایستگاه هیدرومتری شور (با مختصات جغرافیایی ۵۷°۴۵' طول شرقی و ۲۸°۲۰' عرض شمالی و ارتفاع تقریبی ۱۵۰۰ متر) واقع در رودخانه هلیل رود در طول دوره نمونه‌برداری ماهانه ۵۲ ساله (از ابتدای ۱۳۴۶ تا ابتدای ۱۳۹۸) استفاده شد. این دوره شامل نوسانات اقلیمی مختلف مانند خشکسالی‌ها و ترسالی‌ها است که تنوع داده‌ها را افزایش می‌دهد. موقعیت جغرافیایی حوضه و مشخصات آماری داده‌های کیفیت آب و جریان رودخانه به ترتیب در شکل (۱) و جدول (۱) نشان داده شده است. همان‌طور که در جدول (۱) قابل مشاهده است، دبی رودخانه دارای بیشترین ضریب تغییرات ($CV=2.44$) به دلیل نوسانات فصلی و سالانه جریان است، در حالی که PH دارای کمترین ضریب تغییرات ($CV=0.04$) است که نشان دهنده پایداری نسبی این پارامتر است.



شکل ۱- موقعیت رودخانه هلیل بر روی حوضه جازموریان

جدول ۱- مشخصات آماری پارامترهای کیفیت آب و دبی جریان در ایستگاه شور

پارامتر	واحد	میانگین	Min	Max	انحراف معیار	واریانس	چولگی	ضریب تغییرات
Q	m ³ /s	36.12	0	390	87.96938	7738.612	2.803379	2.44
TDS	mg/L	628.63	180	2640	370.9736	137621.4	2.128558	0.59
EC	µmhos/cm	902.21	7.6	4000	632.8589	400510.3	1.564317	0.70
PH	بدون بعد	7.8528	6.5	8.8	0.277803	0.077174	-0.37918	0.04
Co3	mg/L	0.5369	0	3.9	0.859673	0.739037	1.907967	1.60
Hco3	mg/L	3.2637	0.5	6	0.90732	0.823229	-0.64888	0.28
Cl	mg/L	4.0042	0.13	28	4.069447	16.5604	2.427716	1.02
So4	mg/L	2.402	0	32	2.259799	5.106691	5.971697	0.94
Ca	mg/L	1.9488	0.6	5.5	0.550407	0.302948	1.594448	0.28
Mg	mg/L	1.6961	0.12	7.7	0.78434	0.615189	2.218471	0.46
Na	mg/L	6.2365	0.4	35.8	5.398446	29.14322	2.120319	0.87
K	mg/L	0.4549	0	3.873	0.449792	0.202313	2.470454	0.99
SAR	بدون بعد	7.8874	0.344	62.5	9.982599	99.65228	2.599421	1.27
%Na	درصد	64.471	11.54	195	29.13159	848.6498	1.652405	0.45
TH	mg/L	354.18	160	700	65.7854	4327.718	1.011311	0.19

مدلسازی پارامترهای کیفی رودخانه هلیل رود با روش‌های ANFIS، ANN-MLP و ANN-RBF در محیط نرم‌افزار MATLAB نسخه R2018b کدنویسی شد.

۲-۲-۲ مدل ANFIS

سامانه استنتاجی عصبی-فازی تطبیقی (ANFIS) برای اولین بار توسط جانگ در سال ۱۹۹۳ معرفی شد. این سامانه ابزاری کارآمد برای تخمین توابع پیوسته حقیقی در دامنه‌ای محدود است. ANFIS پیاده‌سازی یک سیستم فازی تاکاگی-سوگنو (Takagi-Sugeno) به صورت ساختار شبکه‌ای پیش‌رونده است [۱۶]. این روش، ترکیبی از منطق فازی و شبکه‌های عصبی است که مزایای هر دو را ترکیب می‌کند: قابلیت تفسیرپذیری فازی و توانایی یادگیری شبکه‌های عصبی. ساختار ANFIS شامل پنج لایه است:

۱. لایه فازی سازی (Fuzzification Layer): ورودی‌ها به عضویت در مجموعه‌های فازی تبدیل می‌شوند. برای هر ورودی x ، درجه عضویت $\mu_A(x)$ محاسبه می‌شود، معمولاً با تابع عضویت گاوسی:

$$\mu_A(x) = \exp\left(-\frac{(x-c)^2}{2\sigma^2}\right) \quad (۱)$$

که c مرکز و σ عرض تابع است.

۲. لایه قوانین (Rule Layer): درجه شلیک هر قانون محاسبه می‌شود. برای دو ورودی x و y و قوانین تاکاگی-سوگنو:

قانون ۱: اگر x در A_1 و y در B_1 ، سپس $f_1 = p_1x + q_1y + r_1$

قانون ۲: اگر x در A_2 و y در B_2 ، سپس $f_2 = p_2x + q_2y + r_2$

درجه شلیک: $w_i = \mu_{\{A_i\}}(x) * \mu_{\{B_i\}}(y)$

۳. لایه نرمال سازی (Normalization Layer): درجه شلیک نرمال می‌شود:

$$\bar{w}_i = \frac{w_i}{\sum w_j} \quad (۲)$$

۴. لایه نتیجه گیری (Consequent Layer): خروجی هر قانون محاسبه می‌شود:

$$\bar{w}_i f_i = \bar{w}_i (p_i x + q_i y + r_i) \quad (۳)$$

۵. لایه خروجی (Output Layer): خروجی نهایی جمع خروجی‌های قوانین است:

$$f = \sum \bar{w}_i f_i \quad (۴)$$

آموزش ANFIS با الگوریتم ترکیبی حداقل مربعات و گرادیان نزولی انجام می‌شود تا پارامترهای مقدماتی (c , σ) و نتیجه‌ای (p , q , r) تنظیم شوند.

۲-۳-۲ مدل ANN-MLP

شبکه عصبی مصنوعی چند لایه پرسپترون (ANN-MLP) یک مدل شبکه عصبی خوراک پیش‌رو است که از لایه‌های ورودی، پنهان و خروجی تشکیل شده است. این مدل برای مسائل رگرسیون و طبقه‌بندی مناسب است و از الگوریتم پس‌انتشار (Backpropagation) برای آموزش استفاده می‌کند. MLP می‌تواند روابط غیرخطی پیچیده را تقریب بزند. ساختار MLP شامل:

- لایه ورودی: ورودی‌ها مستقیماً به لایه پنهان منتقل می‌شوند.
- لایه‌های پنهان: هر نورون جمع وزنی ورودی‌ها را محاسبه کرده و تابع فعال‌سازی اعمال می‌کند. برای نورون j در لایه پنهان:

$$z_j = \sum_i w_{ji} x_i + b_j \quad (5)$$

سپس خروجی $a_j = \sigma(z_j)$ جایی که σ تابع فعال سازی است، مانند $\sigma(z) = \max(0, z)$ یا ReLU یا سیگموئید: $\sigma(z) = 1 / (1 + \exp(-z))$.

- **لایه خروجی:** مشابه لایه پنهان، اما اغلب با تابع خطی برای رگرسیون. آموزش با الگوریتم پس انتشار:

۱. **گذر رو به جلو (Forward Pass):** خروجی محاسبه می شود.

۲. **محاسبه خطا:** برای رگرسیون، $MSE: L = (1/2) \sum (y - \hat{y})^2$

۳. **گذر رو به عقب (Backward Pass):** گرادیان ها محاسبه و وزنه ها به روز می شوند:

$$w_{new} = w_{old} - \eta \frac{\partial L}{\partial w} \quad (6)$$

که η نرخ یادگیری است.

در این پژوهش، از دو لایه پنهان با ۱۰ نورون و تابع فعال سازی tansig استفاده شد.

۲-۴-مدل ANN-RBF

شبکه عصبی پایه شعاعی (ANN-RBF) شامل لایه ورودی، لایه پنهان با توابع پایه شعاعی و لایه خروجی خطی است. این مدل برای تقریب توابع پیچیده و مدل سازی غیرخطی مناسب است و سرعت آموزش بالایی دارد. ساختار RBF:

- **لایه پنهان:** هر نورون یک تابع پایه شعاعی (معمولاً گاوسی) دارد:

$$\phi_i(x) = \exp\left(-\frac{\|x - c_i\|^2}{2\sigma_i^2}\right) \quad (7)$$

که c_i مرکز و σ_i عرض است. مراکز با الگوریتم K-means تعیین می شوند.

- **لایه خروجی:** خروجی خطی است:

$$y(x) = \sum_{i=1}^M w_i \phi_i(x) + b \quad (8)$$

آموزش دو مرحله ای:

۱. تعیین مراکز و عرض ها (بدون نظارت).

۲. آموزش وزنه های خروجی با حداقل مربعات.

در این پژوهش، تعداد مراکز ۱۰ انتخاب شد و عرض $\sigma = d_{\max} / \sqrt{2M}$ ، که $d_{\max}$ حداکثر فاصله بین مراکز است.

۲-۵-پیش پردازش داده ها

به منظور بهبود عملکرد مدل ها و جلوگیری از تأثیر دامنه اندازه گیری متفاوت پارامترها، تمام داده های ورودی و خروجی قبل از مدلسازی استانداردسازی (نرمال سازی) شدند. از روش استانداردسازی Z-Score استفاده گردید که بر اساس رابطه (۹) داده ها را به توزیعی با میانگین صفر و انحراف معیار یک تبدیل می کند:

$$Z = \frac{X - \mu}{\sigma} \quad (9)$$

که در آن X مقدار اصلی، μ میانگین و σ انحراف معیار هر پارامتر است. پس از اجرای مدل‌ها، خروجی‌های شبیه‌سازی شده مجدداً به مقیاس اصلی بازگردانده شدند تا محاسبه شاخص‌های خطا (مانند RMSE و MAE) بر حسب واحد اصلی هر پارامتر انجام شود. **تقسیم داده‌ها:** مجموعه داده‌های ماهانه ۵۲ ساله (۶۲۴ داده) به دو دسته تقسیم گردید. ۷۰ درصد از داده‌ها برای آموزش و ۳۰ درصد باقیمانده برای آزمون مدل‌ها در نظر گرفته شد. انتخاب داده‌های آموزش و آزمون به صورت تصادفی انجام گرفت تا از سوگیری فصلی یا دوره‌ای جلوگیری شود. لازم به تأکید است که داده‌های آزمون در هیچ مرحله‌ای از آموزش مدل‌ها استفاده نشدند و صرفاً برای ارزیابی نهایی قابلیت تعمیم‌پذیری (Generalization) مدل‌ها به کار رفتند.

تنظیم پارامترهای مدل ANFIS: پس از اجرای متعدد و تحلیل حساسیت، بهترین مقادیر پارامترها به صورت زیر تعیین گردید:

- نوع توابع عضویت: gaussmf

- تعداد توابع عضویت برای هر ورودی: ۳

- الگوریتم آموزشی: ترکیبی (hybrid) شامل حداقل مربعات (برای پارامترهای نتیجه‌ای) و گرادیان نزولی (برای پارامترهای مقدماتی)

- حداکثر تعداد دوره‌های آموزش: ۲۰۰

- خطای هدف: ۰/۰۱

تنظیم پارامترهای مدل ANN-MLP: ساختار شبکه شامل یک لایه ورودی، دو لایه پنهان و یک لایه خروجی بود. مقادیر بهینه عبارت بودند از:

- تعداد نرون‌ها در لایه‌های پنهان: ۱۰ نرون در لایه اول و ۱۰ نرون در لایه دوم

- تابع فعال‌سازی لایه‌های پنهان: tansig, logsig

- تابع فعال‌سازی لایه خروجی: purelin

- الگوریتم آموزش: پس‌انتشار با لونیبرگ-مارکوارت (Levenberg-Marquardt)

- نرخ یادگیری اولیه: ۰/۰۱

- حداکثر تعداد دوره‌های آموزش: ۳۰۰

- خطای هدف: 10^{-6}

تنظیم پارامترهای مدل ANN-RBF: مقادیر بهینه به صورت زیر انتخاب شد:

- تعداد نرون‌ها (مراکز) در لایه پنهان: ۱۰

- نوع تابع پایه شعاعی: گاوسی

- تعیین مراکز: الگوریتم K-means

- محاسبه عرض توابع پایه شعاعی: $\sigma = \frac{d_{max}}{\sqrt{2M}}$ که در آن d_{max} حداکثر فاصله بین مراکز و M تعداد مراکز است.

- آموزش وزنه‌های لایه خروجی: روش حداقل مربعات خطی (Linear Least Squares)

اجرای چندباره مدل‌ها: به دلیل وابستگی نتایج برخی الگوریتم‌ها (به ویژه K-means در RBF و مقداردهی اولیه تصادفی وزنه‌ها در MLP) به شرایط اولیه، هر سه مدل برای هر پارامتر کیفی ۱۰ بار به طور مستقل اجرا شدند و در پایان میانگین شاخص‌های ارزیابی (R^2 , RMSE, MAE و d) به عنوان نتایج نهایی گزارش گردید. این کار از گزارش نتایج تصادفی و غیرقابل تکرار جلوگیری می‌کند.

۲-۶- معیارهای ارزیابی مدل‌ها

برای ارزیابی مدل‌های مورد استفاده، از ۴ شاخص آماری معتبر و رایج شامل ضریب تعیین (R)، جذر میانگین مربعات خطا (RMSE)، میانگین قدر مطلق خطا (MAE) و شاخص توافق ویلموت (d) استفاده شد. روابط شاخص‌های کمی در فرمول‌های زیر نشان داده شده است.

$$R = \frac{\sum_{i=1}^N (O_i - \bar{O})(E_i - \bar{E})}{\sqrt{\sum_{i=1}^N (O_i - \bar{O})^2 \sum_{i=1}^N (E_i - \bar{E})^2}} \quad (10)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (O_i - E_i)^2}{N}} \quad (11)$$

$$MAE = \frac{1}{N} \sum_{i=1}^N |O_i - E_i| \quad (12)$$

$$d = 1 - \frac{\sum_{i=1}^N (O_i - E_i)^2}{\sum_{i=1}^N (|E_i - \bar{O}| + |O_i - \bar{O}|)^2} \quad (13)$$

که در آنها O_i مقادیر مشاهده‌ای، E_i مقادیر برآوردی، $\bar{O}$ و $\bar{E}$ متوسط مقادیر مشاهده‌ای و برآوردی و n تعداد مشاهدات است. بر اساس این شاخص‌های کمی، مدل برتر دارای مقدار حداکثر R و d (بیشترین مقدار d برابر ۱ است) و حداقل $RMSE$ و MAE خواهد بود. شاخص d (Willmott's index) حساسیت کمتری به مقادیر بزرگ خطا دارد و توافق کلی مدل را نشان می‌دهد.

۳- بحث و نتایج

انتخاب مدل بهینه برای شبیه‌سازی پارامترهای کیفی رودخانه هلیل‌رود مستلزم در نظر گرفتن همزمان دو معیار اساسی است: نخست، برخورداری از دقت پیش‌بینی بالا که با استفاده از شاخص‌های آماری نظیر ضریب تعیین ($\hat{R}$) و معیارهای خطا (مانند $RMSE$ و MAE) قابل ارزیابی است؛ و دوم، سادگی ساختاری و استفاده حداقلی از متغیرهای ورودی به منظور افزایش قابلیت کاربرد عملی مدل. بر این اساس، مدلی که به‌طور همزمان هر دو معیار مذکور را تأمین نماید، به عنوان گزینه برتر برای پیش‌بینی پارامتر مورد نظر انتخاب می‌گردد. یافته‌های این پژوهش حاکی از آن است که در میان روش‌های مورد بررسی، سیستم استنتاج فازی-عصبی تطبیقی (ANFIS) و در میان مدل‌های توسعه‌یافته، مدل (c) از عملکرد بهتری برخوردار بوده‌اند که این برتری را می‌توان به استفاده از پارامترهای با همبستگی بالا در ساختار این مدل‌ها نسبت داد.

۳-۱- انتخاب ورودی‌ها و تعیین ساختار مدل‌ها

به منظور تعیین متغیرهای ورودی به مدل‌های شبیه‌سازی پارامترهای کیفی، ابتدا ماتریس همبستگی بین پارامترهای هیدرولوژیکی و کیفی آب در ایستگاه شور با استفاده از ضریب همبستگی پیرسون در سطح اطمینان ۹۹ درصد (سطح معنی‌داری ۰/۰۱) محاسبه گردید. این تحلیل با بهره‌گیری از نرم‌افزار SPSS نسخه ۲۲ انجام شد و نتایج آن در جدول (۳) ارائه شده است. در گام بعدی، متغیرهایی که دارای همبستگی معنی‌دار با پارامتر هدف بودند، بر اساس روش رگرسیون گام‌به‌گام (Stepwise Regression) و به ترتیب اهمیت آن‌ها در تبیین تغییرات متغیر وابسته، به عنوان ورودی مدل‌ها انتخاب شدند. بر این مبنا، برای شبیه‌سازی هر یک از پارامترهای کیفی، سه الگوی ورودی متفاوت تحت عناوین a، b و c تعریف گردید: الگوی a تنها شامل دبی جریان (Q) به عنوان ساده‌ترین متغیر ورودی (شبیه‌سازی کمی-کیفی)؛ الگوی b شامل متغیرهای قابل اندازه‌گیری با سهولت بیشتر، یعنی Q، TDS، EC، PH و TH؛ و الگوی c شامل سه پارامتری که بالاترین ضریب همبستگی را با متغیر هدف داشتند. ساختار دقیق ورودی‌ها و خروجی مدل‌ها در جدول (۴) به تفصیل ارائه شده است.

جدول ۳- ضریب همبستگی پیرسون بین پارامترهای کیفیت آب در سطح معنی داری ۹۹ درصد

	Q	TDS	EC	PH	Co3	Hco3	Cl	So4	Ca	Mg	Na	K	SAR	%Na	TH
Q	1	-0.12	-0.49	0.00	0.85	-0.73	-0.25	-0.13	-0.11	0.16	-0.24	0.84	0.90	0.78	-0.13
TDS	-0.12	1	0.92	0.11	-0.16	0.40	0.94	0.77	0.40	0.61	0.97	-0.04	0.24	0.32	0.76
EC	-0.49	0.92	1	0.11	-0.47	0.63	0.93	0.72	0.38	0.47	0.95	-0.36	-0.14	-0.02	0.71
PH	0.00	0.11	0.11	1	-0.18	0.05	0.13	0.04	-0.20	0.08	0.15	-0.11	0.08	0.12	0.00
Co3	0.85	-0.16	-0.47	-0.18	1	-0.67	-0.27	-0.15	-0.12	0.07	-0.26	0.84	0.76	0.69	-0.16
Hco3	-0.73	0.40	0.63	0.05	-0.67	1	0.47	0.21	0.22	0.11	0.47	-0.56	-0.53	-0.43	0.58
Cl	-0.25	0.94	0.93	0.13	-0.27	0.47	1	0.57	0.37	0.46	0.96	-0.15	0.10	0.19	0.66
So4	-0.13	0.77	0.72	0.04	-0.15	0.21	0.57	1	0.33	0.61	0.75	-0.09	0.12	0.19	0.62
Ca	-0.11	0.40	0.38	-0.20	-0.12	0.22	0.37	0.33	1	0.09	0.33	-0.03	-0.01	-0.06	0.53
Mg	0.16	0.61	0.47	0.08	0.07	0.11	0.46	0.61	0.09	1	0.51	0.16	0.36	0.29	0.70
Na	-0.24	0.97	0.95	0.15	-0.26	0.47	0.96	0.75	0.33	0.51	1	-0.14	0.13	0.24	0.69
K	0.84	-0.04	-0.36	-0.11	0.84	-0.56	-0.15	-0.09	-0.03	0.16	-0.14	1	0.84	0.70	-0.11
SAR	0.90	0.24	-0.14	0.08	0.76	-0.53	0.10	0.12	-0.01	0.36	0.13	0.84	1	0.92	0.11
%Na	0.78	0.32	-0.02	0.12	0.69	-0.43	0.19	0.19	-0.06	0.29	0.24	0.70	0.92	1	0.13
TH	-0.13	0.76	0.71	0.00	-0.16	0.58	0.66	0.62	0.53	0.70	0.69	-0.11	0.11	0.13	1

- Correlation is significant at the 0.01 level

جدول ۴- پارامترهای منتخب توسط رگرسیون گام به گام برای تخمین ۱۴ پارامتر کیفی در مدل‌های مورد استفاده

پارامتر خروجی	مدل	ورودی‌های مدل	پارامتر خروجی	مدل	ورودی‌های مدل
SO4	a	Q	Na	a	Q
	b	Q, TDS, EC, pH, TH		b	Q, TDS, EC, pH, TH
	c	TDS, EC, Cl		c	TDS, EC, Na
HCO3	a	Q	K	a	Q
	b	Q, TDS, EC, pH, TH		b	Q, TDS, EC, pH, TH
	c	Q, CO3, SAR		c	EC, Na, TH
CO3	a	Q	Cl	a	Q
	b	Q, TDS, EC, pH, TH		b	Q, TDS, EC, pH, TH
	c	TDS, EC, Na		c	Q, K, SAR
SAR	a	Q	TH	a	Q
	b	Q, TDS, EC, pH		b	Q, TDS, EC, pH, TH
	c	TDS, EC, Mg		c	Q, K, %Na
%Na	a	Q	PH	a	Q
	b	Q, TDS, EC, TH		b	Q, TDS, EC, pH, TH
	c	Cl, Na, %Na		c	Q, K, SAR
Ca	a	Q	EC	a	Q
	b	Q, TDS, pH, TH		b	Q, TDS, EC, pH, TH
	c	TDS, Cl, Na		c	TH, TDS, EC
Mg	a	Q	TDS	a	Q
	b	Q, EC, pH, TH		b	Q, TDS, EC, pH, TH
	c	EC, Cl, Na		c	TH, SO4, TDS

پیش از ارائه نتایج عددی مدل‌ها، لازم است به تحلیل هیدروشیمیایی روابط بین پارامترها پرداخته شود. همان طور که در جدول (۳) مشاهده می‌شود، همبستگی منفی بین دبی جریان (Q) با TDS (۰/۱۲-) و EC (۰/۴۹-) نشان‌دهنده رقیق‌شدگی املاح در زمان‌های پربابی است. در مقابل، همبستگی مثبت قوی Q با K (۰/۸۴) و SAR (۰/۹۰) پدیده‌ای غیرمنتظره به نظر می‌رسد که می‌تواند ناشی از ورود پساب‌های کشاورزی و زهکش‌های حاوی این یون‌ها در بالادست ایستگاه شور باشد. همچنین، همبستگی بسیار

بالای TDS با EC (۰/۹۲)، Cl (۰/۹۴) و Na (۰/۹۷) تأیید می‌کند که شوری آب عمدتاً توسط کلرید سدیم کنترل می‌شود که از هوازدگی سازندهای تبخیری و همچنین برگشت آب آبیاری ناشی می‌شود. ضعیف‌ترین همبستگی‌ها مربوط به PH با سایر پارامترها است (کمتر از ۰/۲۰ در همه موارد) که دلیل اصلی عملکرد ضعیف مدل‌ها در شبیه‌سازی PH خواهد بود؛ زیرا PH در آب‌های طبیعی بیشتر تابع تعادل دی‌اکسید کربن و سیستم کربناته است تا غلظت یون‌های اصلی.

۲-۳- مقایسه مدل‌های ANN-RBF و ANN-MLP، ANFIS

به منظور ارزیابی دقیق عملکرد مدل‌های توسعه‌یافته ANN-RBF، ANN-MLP، ANFIS، نتایج شبیه‌سازی هر سه روش برای ۱۴ پارامتر کیفی آب با استفاده از چهار معیار آماری کلیدی یعنی d ، R^2 و MAE و RMSE مقایسه شد. جدول (۵) خلاصه‌ای جامع از این نتایج را برای مدل‌های a، b و c هر روش ارائه می‌دهد. مقادیر بهینه (بالاترین R^2 و d و پایین‌ترین MAE و RMSE) در هر ستون با برجسته‌سازی آبی نشان داده شده است. نتایج بر اساس بهترین ساختارهای تنظیم‌شده هر مدل (بهینه‌سازی‌شده با تحلیل حساسیت و تکرار بیش از ۱۰ باره برای جلوگیری از overfitting) به دست آمده است.

جدول ۵- مقایسه نتایج روش‌های ANN-RBF و ANN-MLP، ANFIS برای تخمین پارامترهای کیفیت آب

پارامتر	روش	RMSE			MAE			d			R2		
		c	b	a	c	b	a	c	b	a	c	b	a
SO4	ANFIS	0.61	0.82	1.88	0.40	0.38	0.89	0.98	0.97	0.66	0.93	0.88	0.31
	MLP	0.74	1.22	1.87	0.47	0.40	0.86	0.97	0.93	0.66	0.89	0.77	0.32
	RBF	0.82	0.88	1.86	0.49	0.54	0.81	0.97	0.96	0.68	0.87	0.86	0.32
HCO3	ANFIS	0.24	0.32	0.57	0.18	0.16	0.40	0.98	0.97	0.87	0.93	0.88	0.60
	MLP	0.34	0.39	0.57	0.21	0.21	0.39	0.97	0.95	0.87	0.87	0.82	0.61
	RBF	0.29	0.33	0.57	0.19	0.21	0.39	0.97	0.97	0.87	0.90	0.87	0.61
CO3	ANFIS	0.25	0.36	0.42	0.11	0.14	0.32	0.98	0.96	0.93	0.92	0.85	0.76
	MLP	0.25	0.31	0.40	0.11	0.14	0.27	0.98	0.97	0.94	0.92	0.88	0.79
	RBF	0.24	0.20	0.40	0.12	0.13	0.27	0.98	0.99	0.94	0.93	0.95	0.79
SAR	ANFIS	1.04	1.32	3.94	0.45	0.55	2.92	1.00	1.00	0.96	0.99	0.98	0.85
	MLP	1.78	2.31	3.06	0.55	0.78	2.03	0.99	0.99	0.98	0.97	0.95	0.91
	RBF	1.55	2.16	2.64	0.58	0.71	1.72	0.99	0.99	0.98	0.98	0.96	0.93
%Na	ANFIS	3.02	3.89	14.93	1.84	2.54	12.18	1.00	1.00	0.92	0.99	0.98	0.74
	MLP	4.36	6.61	11.39	2.22	3.59	7.92	0.99	0.99	0.96	0.98	0.95	0.85
	RBF	5.90	8.56	11.46	2.12	4.24	8.13	0.99	0.98	0.96	0.96	0.92	0.85
Ca	ANFIS	0.35	0.32	0.54	0.23	0.20	0.37	0.86	0.88	0.26	0.61	0.65	0.07
	MLP	0.35	0.32	0.53	0.24	0.22	0.35	0.87	0.89	0.33	0.61	0.66	0.10
	RBF	0.33	0.37	0.52	0.21	0.21	0.34	0.88	0.85	0.39	0.64	0.55	0.13
Mg	ANFIS	0.40	0.35	0.73	0.29	0.23	0.47	0.92	0.94	0.41	0.74	0.80	0.15
	MLP	0.41	0.41	0.66	0.30	0.26	0.40	0.92	0.92	0.65	0.73	0.73	0.29
	RBF	0.38	0.37	0.66	0.26	0.25	0.40	0.93	0.94	0.66	0.77	0.78	0.30
Na	ANFIS	0.96	1.31	3.50	0.48	0.40	2.23	0.99	0.99	0.85	0.97	0.95	0.58
	MLP	1.28	1.71	3.79	0.61	0.60	2.16	0.98	0.97	0.78	0.95	0.90	0.59
	RBF	1.66	1.35	3.78	0.49	0.43	2.16	0.98	0.98	0.78	0.92	0.94	0.59
K	ANFIS	0.14	0.14	0.24	0.10	0.10	0.21	0.97	0.97	0.92	0.90	0.91	0.74
	MLP	0.21	0.22	0.22	0.12	0.14	0.19	0.95	0.94	0.94	0.84	0.81	0.79
	RBF	0.18	0.19	0.21	0.14	0.16	0.18	0.96	0.95	0.94	0.87	0.85	0.80
Cl	ANFIS	0.54	1.03	2.63	0.32	0.28	1.62	1.00	0.98	0.85	0.98	0.94	0.58
	MLP	2.15	1.55	2.86	0.55	0.46	1.58	0.92	0.96	0.77	0.74	0.86	0.59
	RBF	1.45	1.21	2.85	0.46	0.36	1.57	0.97	0.98	0.78	0.88	0.91	0.59
TH	ANFIS	31.68	36.87	57.04	21.49	26.42	40.13	0.93	0.90	0.59	0.77	0.69	0.26
	MLP	34.18	36.45	56.96	22.98	24.92	38.33	0.92	0.91	0.57	0.73	0.69	0.29
	RBF	31.46	35.78	56.46	21.19	22.89	38.19	0.93	0.91	0.59	0.77	0.70	0.30
PH	ANFIS	0.24	0.25	0.27	0.17	0.18	0.19	0.61	0.51	0.30	0.26	0.16	0.05
	MLP	0.24	0.25	0.27	0.17	0.17	0.19	0.64	0.67	0.31	0.25	0.24	0.05
	RBF	0.23	0.24	0.27	0.16	0.16	0.18	0.71	0.66	0.36	0.34	0.27	0.08

0.98	0.99	0.67	1.00	1.00	0.89	49.47	41.56	234.02	86.30	83.11	364.98	ANFIS	EC
0.97	0.98	0.67	0.99	0.99	0.84	57.98	47.51	225.14	113.00	92.80	396.34	MLP	
0.98	0.97	0.66	1.00	0.99	0.84	42.46	32.00	228.32	82.19	111.05	397.22	RBF	
0.99	0.99	0.56	1.00	1.00	0.84	32.94	25.68	160.14	44.23	42.79	244.77	ANFIS	TDS
0.97	0.94	0.57	0.99	0.98	0.75	36.92	42.14	152.18	68.20	105.59	268.99	MLP	
0.98	0.94	0.59	0.99	0.99	0.86	27.53	31.47	143.40	52.59	87.37	238.86	RBF	

نتایج مندرج در جدول (۵) حاکی از عملکرد بسیار قوی و قابل اعتماد هر سه روش هوش مصنوعی در شبیه‌سازی پارامترهای کیفی پیچیده آب رودخانه هلیل‌رود است. به طور کلی، مدل c (سه ورودی با بیشترین همبستگی) در ۱۰ پارامتر از ۱۴ پارامتر عملکرد برتری نسبت به مدل‌های a و b نشان داد. این امر به طور قاطع بر اهمیت انتخاب هوشمندانه ورودی‌ها بر پایه تحلیل همبستگی پیرسون و رگرسیون گام‌به‌گام تأکید می‌کند و نشان می‌دهد که روابط غیرخطی میان پارامترهای کیفی (مانند TDS-EC-Na با SO_4 و Cl) بسیار قوی‌تر از تأثیر مستقیم دبی جریان است. در مقابل، مدل a (تنها ورودی Q) ضعیف‌ترین عملکرد را داشت (R^2 غالباً بین ۰.۲۶-۰.۷۹ و RMSE بسیار بالا) که کاملاً منطبق با انتظارات هیدرو شیمیایی است؛ زیرا دبی جریان به تنهایی نمی‌تواند فرآیندهای پیچیده انحلال، رسوب‌گذاری، تبادل یونی و آلودگی‌های کشاورزی را مدل کند.

مدل b (با پنج پارامتر آسان-اندازه‌گیری) عملکردی از متوسط تا عالی ارائه داد و از منظر کاربردی بسیار ارزشمند است. این مدل امکان شبیه‌سازی پارامترهای پرهزینه و زمان‌بر (مانند SAR، EC، TDS، $Na\%$ ، SO_4 و Cl) را تنها با اندازه‌گیری پنج پارامتر ساده فراهم می‌کند که در اکثر ایستگاه‌های هیدرومتری ایران به راحتی قابل دسترس هستند. بنابراین، مدل b به عنوان «مدل عملیاتی پیشنهادی» برای نظارت در زمان واقعی کیفیت آب حوضه هلیل‌رود معرفی می‌شود.

از منظر مقایسه روش‌ها، ANFIS در اکثر پارامترهای کلیدی (TDS، EC، SAR، $Na\%$ ، Cl، TH، HCO_3) برتری مطلق داشت. بالاترین $R^2=0.99$ و $d=1.00$ برای EC، TDS، SAR و $Na\%$ در مدل‌های b و c توسط ANFIS به دست آمد. دلیل این برتری، معماری هیبریدی ANFIS است که ترکیبی از منطق فازی تاکاگی-سوگنو (قابلیت مدیریت عدم قطعیت و روابط زبانی) و شبکه عصبی (قابلیت یادگیری الگوهای پیچیده) می‌باشد. ANFIS به ویژه در پارامترهایی که عدم قطعیت اندازه‌گیری بالا (مانند شوری و نسبت‌های یونی) وجود دارد، عملکرد بهتری نشان داد.

ANN-RBF در پارامترهایی مانند CO_3 ($R^2=0.95$)، SAR ($R^2=0.98$)، $Na\%$ ($R^2=0.96$)، TH ($R^2=0.77$)، EC ($R^2=0.98$) و TDS ($R^2=0.98$) عملکرد بسیار درخشانی داشت. سرعت همگرایی فوق‌العاده بالا و قابلیت تقریب محلی قوی توابع شعاعی، این روش را برای داده‌های با پراکندگی بالا و روابط غیرخطی محلی مناسب ساخته است. ANN-RBF همچنین در مدل b برای چندین پارامتر (مانند CO_3 ، TH، PH و EC) بهترین RMSE را ثبت کرد که نشان‌دهنده تعادل عالی بین دقت و سرعت محاسباتی است.

ANN-MLP در پارامترهایی مانند CO ، SAR، $Na\%$ ، Na، EC، TDS و برخی موارد Ca عملکرد قابل قبولی داشت، اما به طور کلی پس از ANFIS و RBF قرار گرفت. حساسیت بیشتر MLP به انتخاب اولیه وزن‌ها و احتمال گیر کردن در نقاط محلی بهینه، دلیل اصلی این اختلاف است.

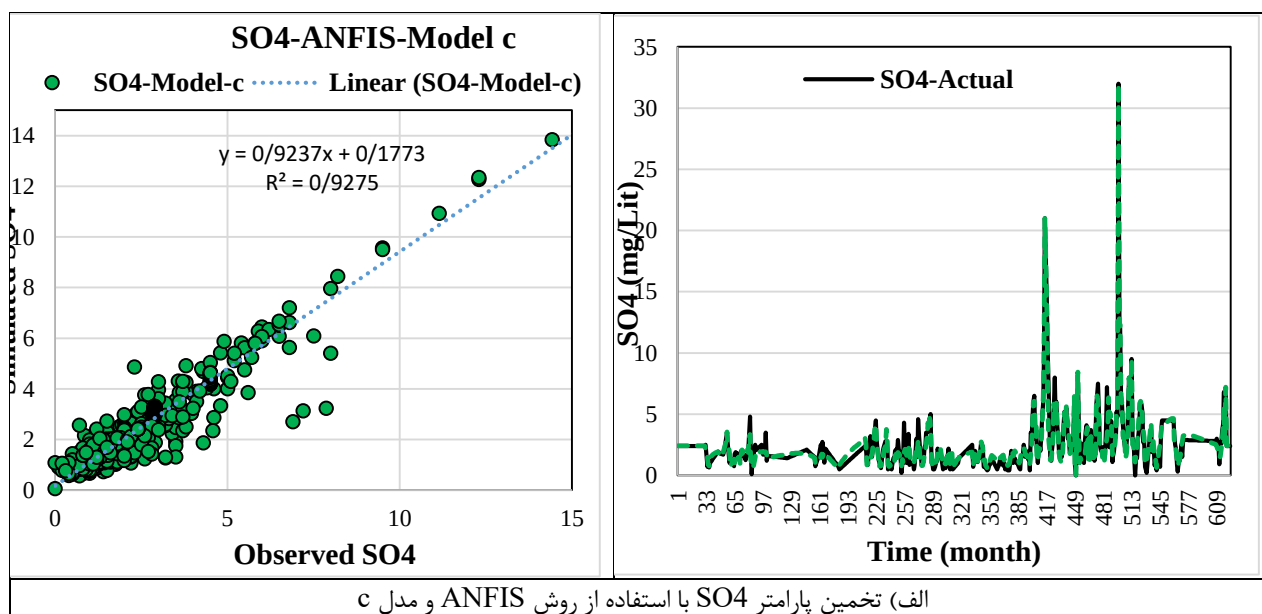
مقایسه بین پارامترها نیز نکات جالبی را آشکار می‌کند:

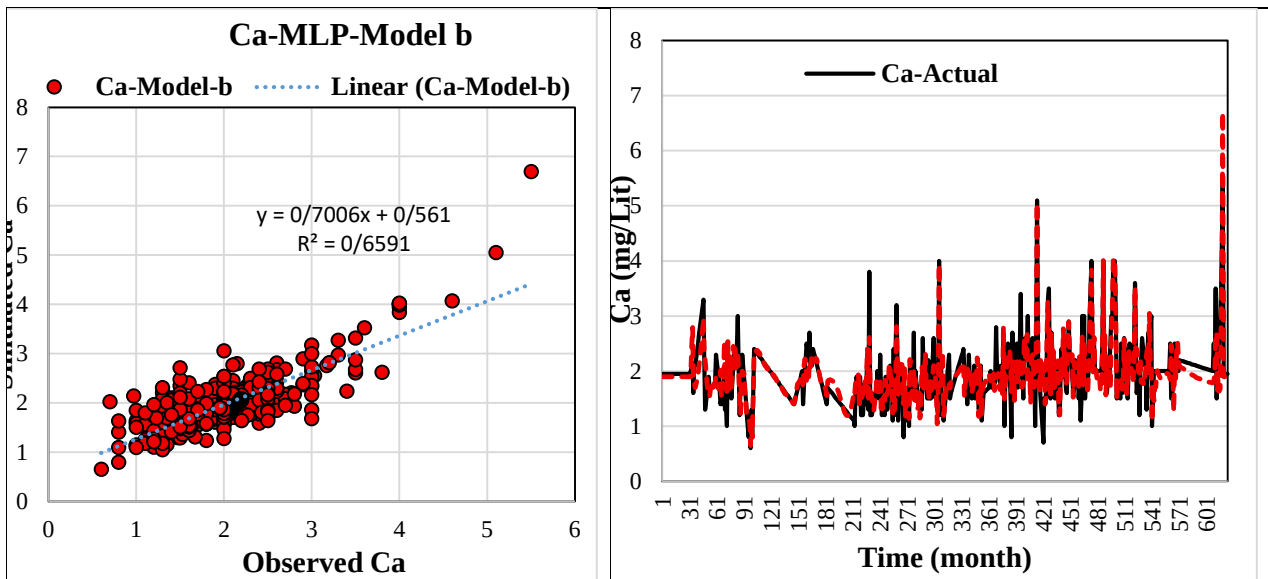
- پارامترهای شوری و رسانایی (EC و TDS) با R^2 حدود ۰/۹۹ توسط ANFIS بهترین شبیه‌سازی شدند که دقیقاً با همبستگی بسیار بالای این دو پارامتر (۰/۹۲) در جدول همبستگی همخوانی دارد.
- پارامتر سختی کل (TH) توسط مدل RBF و به خوبی مدل شد، زیرا این پارامتر عمدتاً تحت تأثیر کاتیون‌های Ca^{2+} و Mg^{2+} است که در مدل c پارامتر Mg^{2+} به عنوان ورودی وارد شده است.
- نسبت‌های سدیمی (SAR و $Na\%$) توسط ANFIS تقریباً به طور کامل ($d=1.00$ ، $R^2=0.99$) شبیه‌سازی شدند که اهمیت بالای این پارامترها در ارزیابی آب کشاورزی را برجسته می‌کند.
- یون‌های عمده (Ca ، Na، SO_4 ، Cl) در مدل c عمدتاً با ورودی‌های EC، TDS و Na بهترین عملکرد را داشتند که منطق شیمیایی قوی دارد (این یون‌ها عمدتاً نمک‌های محلول را تشکیل می‌دهند).

• مقادیر شاخص d نزدیک به ۱/۰۰ در مدل‌های برتر (۰/۹۸ - ۱/۰۰) نشان‌دهنده توافق عالی و عدم وجود bias سیستماتیک بین مقادیر مشاهده‌شده و شبیه‌سازی‌شده است. همچنین کاهش چشمگیر RMSE و MAE در مدل‌های b و c نسبت به مدل a (گاهی تا ۸۰٪ کاهش) تأیید می‌کند که انتخاب ورودی‌های همبسته، خطای مدل را به طور معناداری کاهش می‌دهد. این نتایج به طور قابل توجهی بهتر از مطالعات پیشین در رودخانه‌های ایران (مانند سفیدرود با $R=0.94-0.96$ توسط روش‌های GEP و LS-SVM) و مطالعات مختلف نظیر [۶] در زاینده‌رود است. نوآوری اصلی پژوهش حاضر، ترکیب سه روش پیشرفته با سه الگوی ورودی استراتژیک و تمرکز بر مدل b به عنوان مدل عملیاتی است که امکان پیاده‌سازی در شرایط واقعی با حداقل هزینه را فراهم می‌کند.

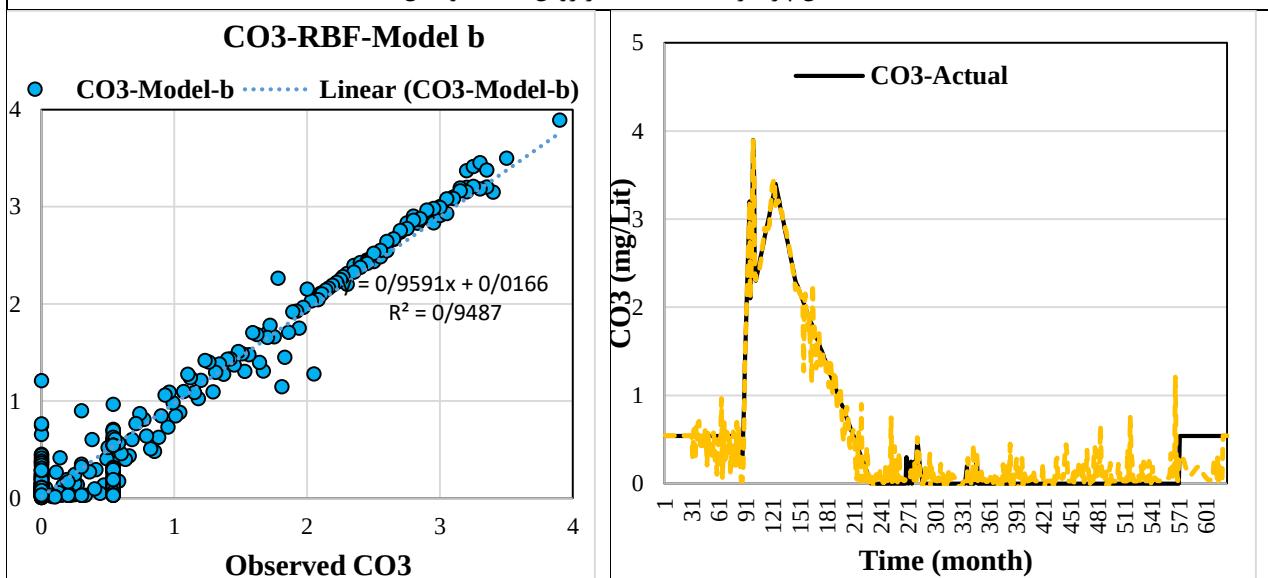
در نهایت، یافته‌های این پژوهش ثابت می‌کند که روش‌های هوشمند یادگیری ماشین نه تنها دقت بسیار بالایی در شبیه‌سازی پارامترهای کیفی پیچیده ارائه می‌دهند، بلکه با کاهش نیاز به اندازه‌گیری‌های آزمایشگاهی پرهزینه، ابزاری قدرتمند برای مدیریت یکپارچه منابع آب در حوضه هلیل‌رود محسوب می‌شوند. این رویکرد می‌تواند مبنایی برای توسعه سیستم‌های هشدار زود هنگام کیفیت آب و پر کردن داده‌های گمشده در ایستگاه‌های هیدرومتری باشد.

به منظور ارزیابی بصری دقت و قابلیت تعمیم‌پذیری مدل‌های پیشنهادی، از نمودارهای سری زمانی (Time Series) و پراکندگی (Scatter Plot) برای مقایسه مقادیر مشاهده‌شده (Observed) و شبیه‌سازی‌شده (Simulated) استفاده شد. این نمودارها امکان بررسی الگوهای زمانی نوسانات پارامترهای کیفی (مانند پیک‌ها، دره‌ها و روندهای کلی) و همخوانی خطی بین داده‌های واقعی و پیش‌بینی‌شده را فراهم می‌کنند. در این پژوهش، بهترین مدل‌ها (بر اساس معیارهای آماری جدول ۵) برای پارامترهای منتخب SO_4 (مدل c با ANFIS)، Ca (مدل b با ANN-MLP)، CO_3 (مدل b با ANN-RBF) و TDS (مدل b با ANFIS) انتخاب گردیدند. این پارامترها به عنوان نماینده‌های کلیدی شوری، سختی و یون‌های عمده انتخاب شدند، زیرا نوسانات آن‌ها تحت تأثیر عوامل هیدروژئوشیمیایی مانند بارندگی، تبخیر، کشاورزی و آلودگی صنعتی قرار دارد. نمودارهای سمت راست سری زمانی را نشان می‌دهند که نوسانات ماهانه را در طول ۵۲ سال دوره آماری (۱۳۴۶-۱۳۹۸) پوشش می‌دهند، در حالی که نمودارهای سمت چپ پراکندگی را با خط رگرسیون خطی (به همراه معادله و R^2) نمایش می‌دهند. این تحلیل گرافیکی مکمل ارزیابی عددی است و به شناسایی نقاط قوت (مانند دقت در پیک‌ها) و ضعف (مانند انحراف در مقادیر پایین) کمک می‌کند.

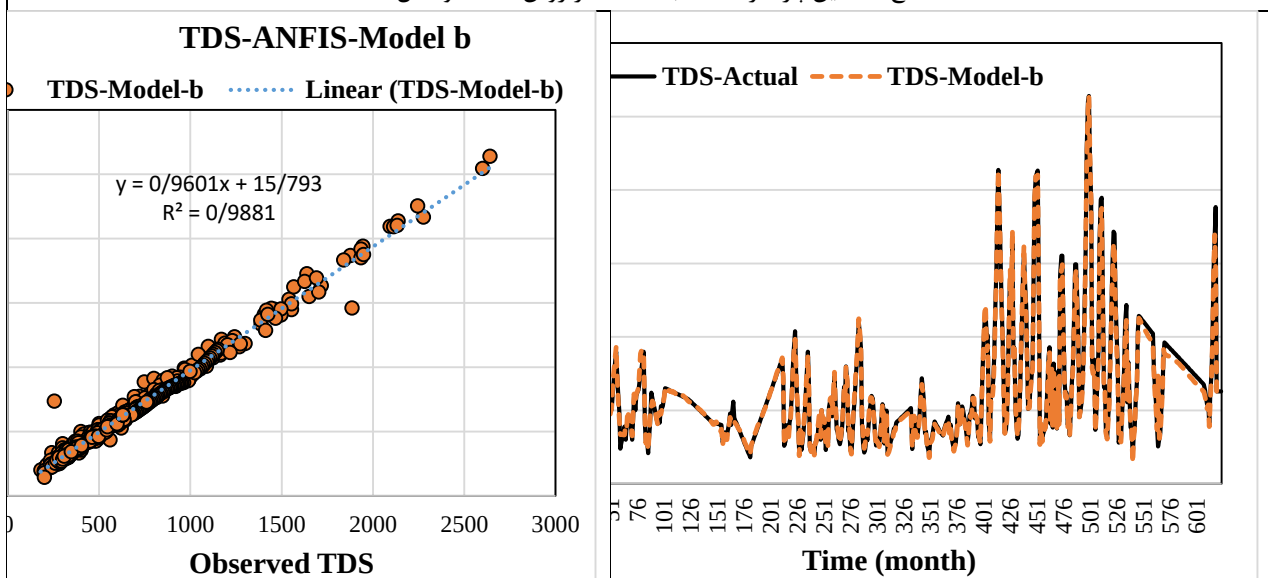




ب) تخمین پارامتر Ca با استفاده از روش MLP و مدل b



ج) تخمین پارامتر CO3 با استفاده از روش RBF و مدل b



د) تخمین پارامتر TDS با استفاده از روش ANFIS و مدل b

شکل ۲- مقایسه بین داده‌های مشاهده‌ای و مقدار شبیه‌سازی شده برخی از پارامترهای کیفی در مدل‌های برتر

نتایج گرافیکی مندرج در شکل (۲)، عملکرد برتر مدل‌های هوش مصنوعی را در شبیه‌سازی دینامیک‌های پیچیده کیفی آب رودخانه هلیل‌رود تأیید می‌کند و جزئیات زیر را برجسته می‌سازد:

• شکل (۲-الف) پارامتر SO_4 (مدل c با ANFIS): در نمودار سری زمانی، مقادیر شبیه‌سازی‌شده (خط سبز) به طور قابل توجهی با داده‌های واقعی (خط سیاه) هم‌خوانی دارد، به ویژه در ثبت پیک‌های ناگهانی (مانند ماه‌های ۶۰۰-۴۰۰) که احتمالاً ناشی از افزایش ورودی‌های کشاورزی (کودهای سولفاتی) یا فرآیندهای ژئوشیمیایی مانند اکسیداسیون پیریت است. انحرافات جزئی در مقادیر پایین (کمتر از ۵ mg/L) مشاهده می‌شود که ممکن است به دلیل حساسیت ANFIS به عدم قطعیت در داده‌های اندازه‌گیری در غلظت‌های پایین باشد. در نمودار پراکندگی، شیب خط رگرسیون ($y = 0.9237x + 0.1773$) نزدیک به ۱ است و $R^2 = 0.9275$ نشان‌دهنده همبستگی قوی و عدم وجود bias سیستماتیک است. این عملکرد عالی ANFIS را می‌توان به قابلیت ترکیبی آن در مدیریت قوانین فازی و یادگیری عصبی نسبت داد، که روابط غیرخطی بین ورودی‌ها (Na, EC, TDS) و SO_4 را به خوبی تقریب می‌زند. مقایسه با مطالعات مشابه مانند [۸] در رودخانه تیر نشان می‌دهد که این دقت ($R^2 > 0.92$) برتر از مدل‌های سنتی رگرسیون چندخطی است.

• شکل (۲-ب) پارامتر Ca (مدل b با ANN-MLP): سری زمانی (خط قرمز) نوسانات واقعی (خط سیاه) را با دقت بالا دنبال می‌کند، به ویژه در کاهش‌های ناگهانی (مانند ماه‌های ۱۰۰-۵۰ و ۵۰۰-۳۵۰) که می‌تواند به دلیل رسوب‌گذاری کربنات کلسیم در دوره‌های خشک باشد. MLP با ساختار چندلایه و الگوریتم پس‌انتشار، قادر به یادگیری الگوهای فصلی (مانند افزایش Ca در فصل‌های پر باران به دلیل شستشوی خاک‌های آهکی) است، اما در برخی نقاط (مانند ماه ۶۰۰) اندکی overestimation مشاهده می‌شود که ممکن است ناشی از overfitting جزئی در فاز آموزش باشد. نمودار پراکندگی با معادله $y = 0.7006x + 0.561$ و $R^2 = 0.6591$ همبستگی متوسط تا خوب را نشان می‌دهد، که با توجه به ورودی‌های آسان مدل b (TH, EC, TDS, Q و PH)، عملکرد قابل قبولی است. این نتیجه تأکید می‌کند که ANN-MLP برای پارامترهای وابسته به سختی (مانند Ca) که تحت تأثیر PH و TH هستند، گزینه مناسبی است، و دقت آن ($R^2 \approx 0.66$) با پژوهش [۱] در مدل‌سازی یون‌ها همخوانی دارد.

• شکل (۲-ج) پارامتر CO_3 (مدل b با ANN-RBF): در سری زمانی، مقادیر شبیه‌سازی‌شده (خط آبی) الگوی نوسانی واقعی را به خوبی بازسازی کرده، به ویژه در پیک‌های بالا (مانند ماه‌های ۲۰۰-۱۰۰) که مرتبط با افزایش قلیائیت در دوره‌های کم‌آب است. RBF با توابع پایه شعاعی، قابلیت تقریب محلی قوی دارد و انحرافات را در مقادیر پایین (کمتر از ۱ mg/L) به حداقل می‌رساند، که این امر به دلیل سرعت همگرایی بالا و حساسیت کمتر به نویز داده‌هاست. نمودار پراکندگی با شیب $y = 0.9581x + 0.0166$ و $R^2 = 0.9487$ همبستگی عالی را نشان می‌دهد، که نزدیک به خط ۱:۱ است و عدم پراکندگی زیاد در نقاط بالا را تأیید می‌کند. این عملکرد برتر RBF را می‌توان به ماهیت محلی آن نسبت داد، که روابط غیرخطی بین CO_3 و ورودی‌های مدل b (مانند PH و TH) را بهتر مدل می‌کند. مقایسه با [۶] در زاینده‌رود نشان می‌دهد که این دقت ($R^2 > 0.94$) بهبود قابل توجهی نسبت به ANFIS ساده است.

• شکل (۲-د) پارامتر TDS (مدل b با ANFIS): سری زمانی (خط نارنجی) نوسانات واقعی (خط سیاه) را با دقت فوق‌العاده‌ای دنبال می‌کند، به ویژه در روندهای افزایشی (مانند ماه‌های ۵۵۰-۴۰۰) که احتمالاً ناشی از تبخیر و آلودگی کشاورزی است. ANFIS با قوانین فازی، قادر به مدیریت عدم قطعیت‌های ذاتی در TDS (مانند تأثیرات فصلی Q) است و تنها انحرافات جزئی در پیک‌های شدید مشاهده می‌شود. نمودار پراکندگی با معادله $y = 0.9601x + 15.793$ و $R^2 = 0.9881$ همبستگی نزدیک به کامل را نشان می‌دهد، که شیب نزدیک به ۱، دقت بالا و عدم bias را تأیید می‌کند. این نتیجه برجسته، قابلیت ANFIS در مدل‌سازی پارامترهای جامع مانند TDS را اثبات می‌کند، که با یافته‌های [۵] در سفیدرود ($R^2 \approx 0.95$) مقایسه می‌شود و بهبود ۳-۵ درصدی را نشان می‌دهد.

به طور کلی، این تحلیل گرافیکی تأیید می‌کند که مدل‌های پیشنهادی نه تنها دقت عددی بالایی دارند ($R^2 > 0.9$ در اکثر موارد)، بلکه الگوهای زمانی و روابط خطی (روندهای فصلی و سالانه) را به طور مؤثر بازسازی می‌کنند. انحرافات جزئی عمدتاً در مقادیر حدی (به ویژه حین خشکسالی‌های شدید) رخ می‌دهد که می‌تواند با ادغام داده‌های بیشتر یا بهینه‌سازی‌های هیبریدی (مانند الگوریتم‌های فراکاوشی) کاهش یابد. این یافته‌ها مبنایی محکم برای کاربرد عملی این مدل‌ها در پیش‌بینی کیفیت آب و مدیریت حوضه هلیل‌رود فراهم می‌کند؛ به ویژه در شرایط تغییرات اقلیمی که نوسانات پارامترهای کیفی افزایش می‌یابد.

۳-۳- محدودیت‌های مطالعه

علیرغم عملکرد عالی مدل‌ها در اکثر پارامترها، چند محدودیت اساسی وجود دارد: (۱) مدل‌های توسعه‌یافته برای پارامتر PH عملکرد ضعیفی داشتند (بیشینه $R^2=0.34$) که نشان می‌دهد داده‌های یونی به تنهایی برای شبیه‌سازی PH کافی نیستند و افزودن پارامترهایی مانند دما، فشار جزئی CO_2 و قلیائیت کل ضروری است. (۲) در مقادیر حدی (به ویژه TDS بالای ۲۰۰۰ میلی‌گرم بر لیتر در سال‌های خشک)، خطای نسبی مدل‌ها به بیش از ۱۰ درصد می‌رسد که لزوم استفاده از روش‌های ترکیبی با الگوریتم‌های فراکاوشی را برای بهبود تخمین نقاط پرت نشان می‌دهد. (۳) دوره آماری بلندمدت ۵۲ ساله علیرغم مزیت آن، شامل تغییرات اقلیمی و انسانی بوده که مدل‌های خطی-غیرخطی ساده قادر به تفکیک این اثرات نیستند.

۴- نتیجه‌گیری

در پژوهش حاضر، سه روش قوی هوش مصنوعی شامل ANN-RBF، ANN-MLP، ANFIS برای شبیه‌سازی دقیق ۱۴ پارامتر کیفی آب ($Na, Mg, Ca, SO_4, Cl, HCO_3, CO_3, PH, EC, TDS$) در ایستگاه هیدرومتری شور رودخانه هلیل‌رود با استفاده از داده‌های ماهانه ۵۲ ساله (۱۳۴۶-۱۳۹۸) به کار گرفته شد. نتایج نشان داد که هر سه روش قابلیت بسیار بالایی در تخمین پارامترهای کیفی پیچیده دارند؛ به طوری که روش ANFIS در پارامترهای کلیدی شوری و نسبت‌های سدیمی (Na, EC, TDS, SAR, Cl) با R^2 نزدیک به ۰/۹۹ و $d \approx 1.00$ عملکرد برتر و تفسیرپذیری ارائه داد، ANN-RBF سرعت آموزش بالا و دقت عالی در پارامترهای محلی (مانند CO_3 و TH) نشان داد و ANN-MLP نیز در برخی یون‌های سختی (مانند Ca) قابل اعتماد بود. نوآوری اصلی پژوهش، طراحی سه الگوی ورودی استراتژیک بود که مدل b (با پنج پارامتر آسان اندازه‌گیری Q, EC, PH, TH) به عنوان مدل عملیاتی پیشنهادی برجسته شد؛ زیرا امکان شبیه‌سازی پارامترهای پرهزینه و زمان‌بر را بدون نیاز به تجهیزات آزمایشگاهی پیشرفته فراهم می‌کند و در اکثر ایستگاه‌های هیدرومتری ایران قابل اجرا است. تحلیل گرافیکی (سری زمانی و پراکندگی) نیز تأیید کرد که مدل‌های پیشنهادی نه تنها دقت عددی بالایی ($R^2 > 0.90$ در اکثر موارد) دارند، بلکه الگوهای فصلی، پیک‌های ناگهانی و روابط غیرخطی را به طور مؤثر بازسازی می‌کنند. این یافته‌ها در شرایطی که حوضه هلیل‌رود با چالش‌های جدی تغییرات اقلیمی (کاهش بارش، افزایش تبخیر و تشدید خشکسالی)، کاهش جریان سطحی، افزایش شوری و کمبود آب روبرو است، اهمیت دوچندانی می‌یابد؛ زیرا مدل‌های توسعه‌یافته می‌توانند به عنوان ابزاری کم‌هزینه برای پر کردن داده‌های گمشده، پیش‌بینی کیفیت آب، هشدار زودهنگام آلودگی و مدیریت یکپارچه منابع آب در حوضه‌های نیمه‌خشک جنوب شرقی ایران مورد استفاده قرار گیرند. در نهایت، پژوهش حاضر نه تنها عملکرد خوب روش‌های یادگیری ماشین را اثبات کرد، بلکه چارچوبی عملی برای کاهش وابستگی به اندازه‌گیری‌های آزمایشگاهی گران‌قیمت ارائه داد که می‌تواند در سیاست‌گذاری‌های پایدار منابع آب، به ویژه تحت سناریوهای آینده تغییرات اقلیمی و فشارهای انسانی، نقش مؤثری ایفا کند.

۵- منابع و مآخذ

- [1]. Kisi, O., Akbari, N., Sanatipour, M., Hashemi, A., Teimourzadeh, K. & Shiri, J. 2013. Modeling of dissolved oxygen in river water using artificial intelligence techniques. *Journal of Environmental Informatics*, 22(2), 92-101. <https://doi.org/10.3808/jei.201300248>
- [2]. Ravansalar, M. & Rajaei, T. 2015. Evaluation of wavelet performance via an ANN-based electrical conductivity prediction model. *Environmental Monitoring and Assessment*, 187(6), 366. <https://doi.org/10.1007/s10661-015-4600-9>
- [3]. Kisi, O. & Parmar, K. S. 2016. Application of least square support vector machine and multivariate adaptive regression spline models in long term prediction of river water pollution. *Journal of Hydrology*, 534, 104-112. <https://doi.org/10.1016/j.jhydrol.2015.12.014>

- [4]. Rajaei, T. & Shahabi, A. 2016. Evaluation of wavelet-GEP and wavelet-ANN hybrid models for prediction of total nitrogen concentration in coastal marine waters. *Arabian Journal of Geosciences*, 9(3), 176. <https://doi.org/10.1007/s12517-015-2236-0>
- [5]. Bozorg-Haddad, O., Soleimani, S. & Loáiciga, H. A. 2017. Modeling water-quality parameters using genetic algorithm-least squares support vector regression and genetic programming. *Journal of Environmental Engineering*, 143(7), 04017021. [https://doi.org/10.1061/\(ASCE\)EE.1943-7870.0001216](https://doi.org/10.1061/(ASCE)EE.1943-7870.0001216)
- [6]. Azad, A., Karami, H., Farzin, S., Mousavi, S. F. & Kisi, O. 2019. Modeling river water quality parameters using modified adaptive neuro fuzzy inference system. *Water Science and Engineering*, 12(1), 45-54. <https://doi.org/10.1016/j.wse.2018.11.001>
- [7]. Aghel, B., Rezaei, A. & Mohadesi, M. 2019. Modeling and prediction of water quality parameters using a hybrid particle swarm optimization-neural fuzzy approach. *International Journal of Environmental Science and Technology*, 16(8), 4823-4832. <https://doi.org/10.1007/s13762-019-02331-6>
- [8]. Haghiabi, A. H., Nasrolahi, A. H. & Parsaie, A. 2018. Water quality prediction using machine learning methods. *Water Quality Research Journal*, 53(1), 3-13. <https://doi.org/10.2166/wqrj.2018.025>
- [9]. Wu, J. and Wang, Z., 2022. A hybrid model for water quality prediction based on an artificial neural network, wavelet transform, and long short-term memory. *Water*, 14(4), p.610. <https://doi.org/10.3390/w14040610>
- [10]. Tian, X., Wang, Z., Taalab, E., Zhang, B., Li, X., Wang, J., Ong, M.C. and Zhu, Z., 2022. Water quality predictions based on grey relation analysis enhanced LSTM algorithms. *Water*, 14(23), p.3851. <https://doi.org/10.3390/w14233851>
- [11]. Gao, Z., Chen, J., Wang, G., Ren, S., Fang, L., Yinglan, A. and Wang, Q., 2023. A novel multivariate time series prediction of crucial water quality parameters with Long Short-Term Memory (LSTM) networks. *Journal of Contaminant Hydrology*, 259, p.104262. <https://doi.org/10.1016/j.jconhyd.2023.104262>
- [12]. Cai, H., Zhang, C., Xu, J., Wang, F., Xiao, L., Huang, S. and Zhang, Y., 2023. Water quality prediction based on the KF-LSTM encoder-decoder network: A case study with missing data collection. *Water*, 15(14), p.2542. <https://doi.org/10.3390/w15142542>
- [13]. Ruan, J., Cui, Y., Song, Y. and Mao, Y., 2023. A novel RF-CEEMD-LSTM model for predicting water pollution. *Scientific Reports*, 13(1), p.20901. <https://doi.org/10.1038/s41598-023-48409-6>
- [14]. Chen, S., Huang, J., Wang, P., Tang, X. and Zhang, Z., 2024. A coupled model to improve river water quality prediction towards addressing non-stationarity and data limitation. *Water Research*, 248, p.120895. <https://doi.org/10.1016/j.watres.2023.120895>
- [15]. Bi, J., Zhang, L., Yuan, H. and Zhang, J., 2023. Multi-indicator water quality prediction with attention-assisted bidirectional LSTM and encoder-decoder. *Information Sciences*, 625, pp.65-80. <https://doi.org/10.1016/j.ins.2022.12.091>

- [16]. Jang, J. S. 1993. ANFIS: adaptive-network-based fuzzy inference system. *IEEE Transactions on Systems, Man and Cybernetics*, 23(3), 665-685. <https://doi.org/10.1109/21.256541>