



## Evaluation of integrated artificial intelligence models in estimating total dissolved solid concentrations in the upstream of Sari city

Alireza Ghaemi\*, Mahdi Azhdary Moghaddam\*<sup>2</sup> and Sarina Keikha<sup>3</sup>

<sup>1</sup> Ph.D Student, Department of Civil Engineering, University of Sistan and Baluchestan, Zahedan, Iran

<sup>2</sup> Professor, Department of Civil Engineering, University of Sistan and Baluchestan, Zahedan, Iran

<sup>3</sup> Msc Student, Department of Civil Engineering, University of Sistan and Baluchestan, Zahedan, Iran

Received: 28 March 2023

Accepted: 24 June 2023

### Extended abstract

#### Introduction

Rivers are known as the vital resources of nature and the main foundations of sustainable development. Therefore, the quantity and quality of river water are considered valuable parameters. The increase in agricultural and industrial activities has reduced the quality of water resources in many places. The discharge of sewage, garbage and chemical fertilizers in the villages along the rivers is one of the most important sources of water quality pollution. The amount of urban and agricultural wastewater entering this surface has caused an increase in the amount of pollutants, so that in the period of 1993 to the end of 2017, the average amounts the three pollutants of total dissolved solid, chlorine and sodium in Varand Station are respectively 507.49, 2.16 and 2.47. Therefore, accurate estimation of water quality parameters is a basic requirement for water quality management, human health, public consumption and domestic use.

#### Materials and methods

Tajan River basin with an area of about 4147.22 square kilometers has an average river discharge and annual rainfall of 20 cubic meters per second and 539 mm respectively. The highest and lowest elevations of this River basin have been reported as 3728 and 26 meters, respectively. Various human activities such as agriculture and dam construction are carried out in this river. Therefore, evaluating the water quality of this river basin is required. In this research, the combination of two Gene Expression Programming Models (GEP) and Artificial Neural Network (ANN) with a data preprocessing algorithm called Empirical Mode Decomposition (EMD) was used to estimate one of the important parameters of water quality called Total Dissolved Solids (TDS). For this purpose, in this research, some of qualitative parameters including sodium, calcium, magnesium, sulfate bicarbonate, sulfuric acid and chlorine, which were measured in the period of 1993 to the end of 2017 at Varand station, were used to estimate the concentration of total dissolved solids.

#### Results and discussion

At first, the results of the observation data during the sampling period indicated that the TDS values in about 80% of the samples were in the range of 300 to 600 mg.liter<sup>-1</sup>, which represented the good quality of the water of this river. In order to compare the performance of independent and integrated approaches in estimating the quality parameters of the Tajan River in the training and testing stages, the evaluation benchmarks including Correlation Coefficient (R), Root Mean Square Error (RMSE), Mean Deviation of Error (MBE), Nash Coefficient (NSE), Objective Function (OBJ) and RSD ratio were applied. The results of this study demonstrated that the integrated model of Gene Expression Programming and Empirical Mode Decomposition (EMD-GEP) with the lowest error (RSD=0.23 and RMSE=24.41) was the most accurate model in TDS estimating compared to other models such as GEP (RSD=0.44 and RMSE=47.27). In addition, the integrated

\* Corresponding author: mazhdary@eng.usb.ac.ir

model of Artificial Neural Network and Empirical Mode Decomposition (EMD-ANN) with RMSE=36.64 and R=0.95 was stood at the second rank. Additionally, the outcomes of the Objective Function (OBJ) represented that EMD-GEP model could achieved the lowest OBJ value (15.92) than other techniques in the TDS modeling. While, the highest value of the OBJ=29.34 belonged to the GEP model.

### Conclusion

ANN and GEP methods were applied in this research to estimate TDS concentration in the Tajan River. After that, to increase the accuracy of the models, EMD technique was recruited to decompose the time series dataset. The results obtained from the integrated models were evaluated using some error statistical benchmarks such as correlation coefficient, root mean square error. The results showed that the EMD method could play an essential role in increasing the ANN and GEP performance so as to estimate this water quality parameter in Varand station. So that EMD-GEP and EMD-ANN could reduce the RMSE error by 48.35% and 14.02%, respectively, compared to the two independent models of GEP and ANN.

**Keywords:** Artificial neural network, Data pre-processing algorithms, Gene expression programming, Tajan river

Cite this article: Ghaemi, A., Azhdary Moghaddam, M., Keikha, S., 2024. Evaluation of integrated artificial intelligence models in estimating total dissolved solid concentrations in the upstream of Sari city. *Watershed Engineering and Management* 16(1), 50-63.

ش 2024, The Author(s). Published by Soil Conservation and Watershed Management Research Institute (SCWMRI). This is an open-access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0>).



## ارزیابی مدل‌های تلفیقی هوش مصنوعی در تخمین غلظت مواد جامد محلول در بالادست شهر ساری

علیرضا قائمی<sup>۱</sup>، مهدی اژدری مقدم<sup>۲\*</sup> و سارینا کیخا<sup>۳</sup>

<sup>۱</sup> دانشجوی دکتری مهندسی عمران، دانشکده مهندسی شهید نیکبخت، دانشگاه سیستان و بلوچستان، زاهدان، ایران

<sup>۲</sup> استاد گروه مهندسی عمران، دانشکده مهندسی شهید نیکبخت، دانشگاه سیستان و بلوچستان، زاهدان، ایران

<sup>۳</sup> دانشجوی کارشناسی ارشد مهندسی عمران، دانشکده مهندسی شهید نیکبخت، دانشگاه سیستان و بلوچستان، زاهدان، ایران

تاریخ پذیرش: ۱۴۰۲/۰۴/۰۳

تاریخ دریافت: ۱۴۰۲/۰۱/۰۸

### چکیده مبسوط

#### مقدمه

رودخانه‌ها، به‌عنوان منابع حیاتی طبیعت و پایه‌های اصلی توسعه پایدار شناخته می‌شوند. لذا، کمیت و کیفیت آب رودخانه‌ها جز پارامترهای ارزشمند محسوب می‌شوند. افزایش فعالیت‌های کشاورزی و صنعتی، باعث کاهش کیفیت منابع آب در بسیاری از نقاط شده است. تخلیه فاضلاب، زباله و کودهای شیمیایی در روستاهای حاشیه رودخانه‌ها از مهم‌ترین منابع آلودگی کیفیت آب به شمار می‌رود. حجم ورود پساب‌های شهری و کشاورزی به این رودخانه، باعث افزایش مقادیر آلاینده‌ها شده، به‌طوری‌که در بازه سال‌های ۱۳۷۲ تا انتهای سال ۱۳۹۶، میانگین آلاینده‌های موجود در ایستگاه ورنه برای سه آلاینده مواد جامد محلول، کلر و سدیم به ترتیب برابر ۵۰۷/۴۹، ۲/۱۶ و ۲/۴۷ بوده است. بنابراین، برآورد دقیق پارامترهای کیفیت آب، یک نیاز اساسی برای مدیریت کیفیت آب، سلامت انسان، مصرف عمومی و مصارف خانگی است.

#### مواد و روش‌ها

حوضه تجن، با مساحتی در حدود ۴۱۴۷/۲۲ کیلومتر مربع دارای میانگین دبی رودخانه و بارندگی سالانه به ترتیب ۲۰ متر مکعب بر ثانیه و ۵۳۹ میلی‌متر است. بیشترین و کمترین ارتفاع حوضه، به ترتیب ۳۷۲۸ و ۲۶ متر گزارش شده است. فعالیت‌های مختلفی نظیر کشاورزی و سدسازی در این رودخانه صورت می‌گیرد. در این پژوهش، از تلفیق دو مدل برنامه نویسی بیان ژن (GEP) و شبکه عصبی مصنوعی (ANN)، با یکی از الگوریتم‌های پیش‌پردازش داده به نام EMD برای تخمین یکی از پارامتر مهم کیفیت آب به نام مواد جامد محلول (TDS) استفاده شد. بدین منظور، در این پژوهش، برای مدل‌سازی و برآورد مقادیر مواد جامد محلول از چندین داده کیفی شامل سدیم، کلسیم، منیزیم، سولفات بی‌کربنات، سولفوریک اسید و کلر که در بازه سال‌های ۱۳۷۲ تا انتهای سال ۱۳۹۶ در ایستگاه ورنه اندازه‌گیری شده، استفاده شد.

#### نتایج و بحث

در ابتدا، نتایج حاصل از داده‌های مشاهداتی در طول مدت نمونه برداری نشان دادند که مقادیر TDS در حدود ۸۰ درصد از نمونه‌ها در بازه ۳۰۰ تا ۶۰۰ میلی گرم بر لیتر بوده که بیانگر کیفیت خوب آب این رودخانه بوده است. برای مقایسه عملکرد مدل‌های مستقل و ترکیبی برای تخمین پارامترهای کیفی رودخانه در مراحل آموزش و آزمون، شاخص‌های

آماري ضريب همبستگي (R)، ريشه ميانگين مربعات خطا (RMSE)، ميانگين انحراف خطا (MBE)، ضريب نش (NSE)، تابع هدف (OBJ) و نسبت RSD، استفاده شدند. نتايج حاصل از اين مطالعه نشان داد که مدل ترکيبی حاصل از برنامه نویسی بيان زن و تجزيه مود تجربي (EMD-GEP) توانست با کمترین خطا ( $RSD=0/23$  و  $RMSE=24/41$ ) عملکرد بالاتری را در مقایسه با ساير مدل‌ها مانند مدل GEP ( $RSD=0/44$  و  $RMSE=47/27$ ) در تخمین این پارامتر کیفی آب داشته باشد. علاوه بر این، مدل ترکيبی حاصل از شبکه عصبی مصنوعی و تجزيه مود تجربي (EMD-ANN) با  $36/64$   $RMSE=$  و  $R=0/95$  بعد از EMD-GEP، در رتبه دوم قرار گرفت. مطابق نتايج حاصل از تابع هدف برای هریک از مدل‌ها، مدل EMD-GEP دارای مقادير تابع هدف کمتری ( $OBJ=15/92$ ) نسبت به ساير مدل‌ها در مدل‌سازی مواد جامد محلول است. درحالی که، بیشترین مقدار تابع هدف ( $OBJ=29/34$ ) متعلق به مدل GEP است.

### نتیجه‌گیری

به‌منظور، تخمین مواد جامد محلول در رودخانه تجن، از روش‌های ANN و GEP در این پژوهش استفاده شد. سپس، برای تقویت برآوردها از الگوریتم‌های تجزيه مود تجربي برای پیش‌پردازش داده‌های سری زمانی استفاده شد. نتايج به‌دست آمده از مدل‌های هوشمند تلفیقی با استفاده از شاخص‌های آماری خطا، مانند ضريب همبستگي، ريشه ميانگين مربعات خطا، مورد ارزیابی قرار گرفت. نتايج نشان داد که روش EMD، باعث تقویت مدل‌های ANN و GEP، در راستای تخمین این پارامتر کیفی آب در ایستگاه ورنده شده‌اند. به‌طوری که مدل‌های ترکيبی حاصل از GEP و ANN توانست  $48/35$  درصد و  $14/02$  درصد خطای RMSE را به‌ترتیب در مقایسه با دو مدل مستقل GEP و ANN کاهش دهد.

**واژه‌های کلیدی:** الگوریتم‌های پیش‌پردازش داده‌ها، برنامه نویسی بيان زن، شبکه عصبی مصنوعی، رودخانه تجن

### مقدمه

امروزه، رودخانه‌ها یکی از مهم‌ترین منابع حیاتی طبیعت و پایه‌های اصلی توسعه پایدار به شمار می‌روند. در مواردی که رودخانه‌ها به‌عنوان یکی از منابع تامین کننده نیاز جوامع بشری بیان می‌شوند، علاوه بر کمیت و میزان آورد رودخانه، کیفیت آب نیز جز پارامترهای مهم مورد توجه قرار می‌گیرد. افزایش فعالیت‌های کشاورزی و صنعتی باعث افزایش قابل توجه حجم فاضلاب‌های شهری و کاهش کیفیت منابع آب در بسیاری از نقاط شده است.

افزایش مواد آلی و معدنی و همچنین، ترکیبات خطرناک فلزات سنگین ناشی از ورود پساب‌های صنعتی در آب، علاوه بر آلودگی محیط زیست آبیان، سبب برهم خوردن تعادل ترکیبات موجود در آب نیز شده است (Mohammadpour Rezaie-Balf et al., 2020). از (Mohammadpour et al., 2012; et al., 2015). از سوی، با توجه به اقلیم خشک و نیمه‌خشک ایران، تدوین برنامه‌های مدیریت کیفی منابع آبی، به‌منظور

رویارویی با بحران‌های کم آبی، یکی از راهکارهای ضروری در راستای حفاظت و بهره‌برداری پایدار از منابع آبی است. از این رو، بررسی نوسانات کمی و کیفی پارامترهای مختلف آب در طول یک رودخانه، از بخش‌های مهم در مدیریت جامع منابع آب محسوب می‌شود. رودخانه تجن نیز به‌عنوان یکی از رودخانه‌های مهم و پر آب کشور در معرض تهدیدات بسیاری قرار گرفته است. ورود رواناب‌های کشاورزی، فاضلاب‌های مناطق مسکونی و صنعتی، سبب افزایش غلظت برخی آلاینده‌ها و مواد سمی مانند کدورت و مواد جامد محلول شده است. مطالعات انجام شده در این رودخانه، نشان‌دهنده وجود رسوبات و آلاینده‌های مختلف مانند آفت‌کش‌ها و فلزات سنگین ناشی از فعالیت‌های انسانی در آب است. به همین دلیل، بررسی کیفیت این رودخانه را می‌توان به‌عنوان یکی از مهم‌ترین مشکلات کشور در دو دهه اخیر دانست (Saeedi et al., 2005; Ebadi et al., 2006; Ahmadi Momqani et al., 2010).

مواد جامد محلول<sup>۱</sup>، به‌عنوان نمک‌های معدنی یا مواد آلی محلول در آب، یکی از مهم‌ترین پارامترهای

<sup>1</sup> Total Dissolved Solid (TDS)

Peipolzadeh et al., Banadkooki et al., 2020 (2022).

Fethian and Hormozinejad (2011) با استفاده از شبکه‌های عصبی مصنوعی به مدل‌سازی پارامترهای کیفی و کمی جریان رودخانه شامل دبی، اشل، TDS و هدایت الکتریکی<sup>۷</sup> در رودخانه کارون پرداختند. به منظور تعیین تغییرات مکانی پارامترهای کیفی و کمی جریان، فاصله ایستگاه‌های پایین‌دست و بالادست نیز به عنوان یک متغیر ورودی در یک مدل شبکه عصبی شبکه پیش رونده<sup>۸</sup>، پرسپترون چند لایه<sup>۹</sup> و شبکه پایه شعاعی<sup>۱۰</sup> در نظر گرفته شد. مقایسه شبکه‌ها حاکی از دقت بالاتر شبکه پیش‌رونده نسبت به دو شبکه دیگر در برآورد مکانی پارامترهای کیفی و کمی جریان بود. Ghavidel and Montaseri (2014) با استفاده از برنامه نویسی بیان ژن<sup>۱۱</sup> رابطه‌ای با دقت قابل قبول را برای تخمین مقادیر TDS در زاینده‌رود ارائه کردند. Amiri and Amiri (2015) با استفاده از روش شبکه عصبی مصنوعی به برآورد پارامترهای کیفی رودخانه شامل EC، نسبت جذب سدیم<sup>۱۲</sup> و TDS پرداختند. مشاهده شد که مدل در نظر گرفته شده پارامترهای کیفی جریان آب رودخانه را با دقت قابل قبول برآورد می‌کند. Sharafati et al., (2020) با استفاده از روش‌های یادگیری ماشین به مدل‌سازی سه پارامتر اکسیژن مورد نیاز شیمیایی<sup>۱۳</sup>، اکسیژن مورد نیاز بیولوژیکی<sup>۱۴</sup> و TDS در تصفیه‌خانه‌های استان قم پرداختند. نتایج آن‌ها بیان گر دقت قابل قبول مدل‌ها در تخمین پارامترهای کیفی آب است. Sun et al., (2021) عملکرد روش ترکیبی ITD-MARS-CSA را در تخمین TDS دو ایستگاه ریگ چشمه و سلیمان تنگه بررسی کردند. نتایج آن‌ها حاکی از عملکرد قابل قبول روش در نظر گرفته شده در مقایسه با مدل MARS<sup>۱۵</sup> بود. Peipolzadeh et al., (2022) با استفاده از تلفیق الگوریتم پیش‌پردازش

تاثیرگذار در سطح کیفی آب آشامیدنی، نظر بسیاری از محققین را به خود جلب کرده است (Mohammadpour et al., 2021). به عنوان مثال، Saeedi et al., (2006) پتانسیل خود پالایی رودخانه تجن را بررسی کردند. نتایج آن‌ها نشان داد که با توجه به TDS، رودخانه مذکور در فصول سرد کیفیت قابل قبولی را دارد. این در حالی است که افزایش مقدار این پارامتر در بهار و تابستان باعث ایجاد برخی محدودیت‌ها در استفاده از آب این رودخانه برای آبیاری گیاهان حساس می‌شود. کیفیت و میزان آلودگی آب رودخانه‌های پروز و سبزکوه در استان چهارمحال و بختیاری، به وسیله Bagheri et al., (2018) بررسی شد. نتایج نشان داد که مقدار کل مواد جامد محلول در آب رودخانه پروز با مقدار استاندارد تفاوت آماری معنی‌داری ندارد.

تا کنون، مدل‌های گوناگونی به منظور تخمین میزان آبدی رودخانه‌ها همچون مدل‌های سری زمانی، مدل‌های ترکیبی و انواع مدل‌های مفهومی بارش-رواناب، ارائه شده است. لیکن روابط ارائه شده به دلیل عدم شناخت دقیق و پیچیدگی عوامل موثر در آبدی رودخانه‌ها، در بسیاری از موارد با مقادیر مشاهده شده مطابقت لازم را نداشته، حتی تفاوت معنی‌داری با یکدیگر دارند. از سویی، در دهه‌های گذشته، مدل‌های مختلف یادگیری ماشین<sup>۱</sup> نظیر شبکه عصبی مصنوعی<sup>۲</sup>، مدل درخت<sup>۳</sup>، برنامه‌نویسی ژنتیک<sup>۴</sup> و ماشین بردار پشتیبان<sup>۵</sup> به طور گسترده‌ای برای حل مشکلات مختلف مهندسی محیط‌زیست و کیفیت آب توسعه یافته‌اند.

همچنین، مطالعات گسترده‌ای در زمینه تخمین مواد جامد محلول (TDS) انجام شده که نشان دهنده عملکرد خوب این مدل‌ها در تخمین پارامترهای کیفی آب<sup>۶</sup> است (Safar Talory and Dehghani, 2012)؛ Jamei et al., 2020؛ Banadkooki et al., 2020

<sup>9</sup> Multilayer Perceptron (MLP)

<sup>10</sup> Radial Basis Function (RBF)

<sup>11</sup> Gene Expression Programming (GEP)

<sup>12</sup> Sodium Adsorption Ratio (SAR)

<sup>13</sup> Chemical Oxygen Demand (COD)

<sup>14</sup> Biochemical Oxygen Demand (BOD)

<sup>15</sup> Multivariate Adaptive Regression Spline

<sup>1</sup> Machine Learning (ML)

<sup>2</sup> Artificial Neural Network (ANN)

<sup>3</sup> Model Tree (MT)

<sup>4</sup> Genetic Programming (GP)

<sup>5</sup> Support Vector Machine (SVM)

<sup>6</sup> Water Quality Parameter (WQPs)

<sup>7</sup> Electrical Conductivity (EC)

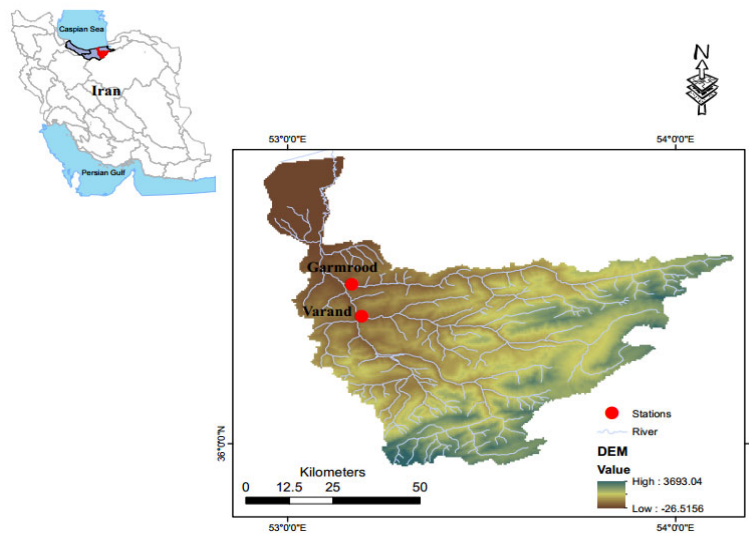
<sup>8</sup> Feed-Forward Neural Network (FFN)

با دو روش هوشمند به نام برنامه نویسی بیان ژن و شبکه عصبی مصنوعی به منظور بررسی عملکرد مدل‌های ترکیبی حاصل از آن‌ها در تخمین مواد جامد محلول استفاده شده است.

### مواد و روش‌ها

**منطقه مورد مطالعه:** حوضه تجن (۵۳ درجه و ۵۶ دقیقه تا ۳۶ درجه و ۱۷ دقیقه عرض شمالی و ۵۳ درجه و ۷ دقیقه تا ۵۳ درجه و ۴۲ دقیقه طول شرقی) مساحتی در حدود ۴۱۴۷/۲۲ کیلومتر مربع دارد. میانگین دبی رودخانه و بارندگی سالانه به ترتیب ۲۰ متر مکعب بر ثانیه و ۵۳۹ میلی‌متر است. کمترین و بیشترین ارتفاع حوضه به ترتیب ۲۶ و ۳۷۲۸ متر است. این رودخانه میزبان فعالیت‌ها و عملیات‌های مختلف کشاورزی، آبی‌پروری و صنعتی مانند سدسازی و استخراج شن و ماسه است. لذا، ایستگاه سنجش هیدرومتری فعال متعددی در حوضه رودخانه تجن (شکل ۱) برای ارزیابی مدلسازی TDS استفاده شده‌اند (Sun et al., 2021; Peipolzadeh et al., 2022).

تجزیه مد تجربی و ماشین بردار پشتیبان مقادیر TDS را در دو ایستگاه گرمرو و ورنند تخمین زدند. بررسی‌ها حاکی از تاثیر قابل توجه EEMD<sup>۱</sup> بر روی عملکرد SVM در تخمین این پارامتر کیفی مهم بود. مطالعات گسترده‌ای جهت تخمین شاخص‌های کیفی آب با استفاده از مدل‌های مختلف یادگیری ماشین (مانند شبکه عصبی مصنوعی و ماشین بردار پشتیبان) انجام شده که نشان‌دهنده عدم قطعیت آن‌ها بوده است. به عبارت دیگر، توانایی‌های منحصر به فرد هر مدل باعث تاثیر چشمگیری در تخمین پارامترهای گوناگون در ایستگاه‌های مختلف شده است. علاوه بر این، به دلیل داده‌های فصلی و غیرخطی بودن رکورد‌های سری زمانی، تغذیه فراداده خام به طور مستقیم به مدل ممکن است بیش از حد قابل توجهی برای تخمین پارامتر کیفیت آب ارائه نکند. لذا، به منظور افزایش عملکرد مدل‌ها، روش‌های پیش‌پردازش مختلفی برای استخراج ویژگی‌های تعبیه شده در سیگنال‌های سری زمانی دینامیکی و غیر ثابت پیشنهاد شده است (Wu and Huang, 2009; Sun et al., 2021). از این رو، در این پژوهش از ترکیب یکی از روش‌های پیش‌پردازش به نام تجزیه مد تجربی



شکل ۱- حوزه آبخیز رودخانه تجن، استان مازندران (Sun et al., 2021)  
Fig. 1. Tajan River watershed, Mazandaran Province (Sun et al., 2021)

کلسیم، منیزیم، سولفات بی‌کربنات، سولفوریک اسید و کلر برای ایستگاه ورنند در بازه سال‌های ۱۳۷۲ تا انتهای

در این پژوهش، برای برآورد و مدلسازی مقادیر مواد جامد محلول از چندین داده کیفی شامل سدیم،

<sup>۱</sup> Ensemble empirical Mode Decomposition

ورند نشان داده شده است (Peipolzadeh et al., 2022)؛  
(Sun et al., 2021).

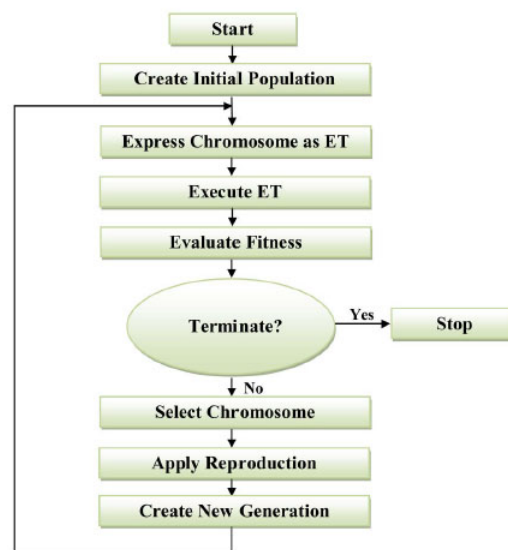
سال ۱۳۹۶ استفاده شد. داده‌های مورد استفاده از سازمان آب منطقه‌ای استان مازندران اخذ شده است. در جدول ۱، مشخصات آماری داده‌های کیفی ایستگاه

جدول ۱- شاخص‌های آماری داده‌های کیفی آب  
Table 1. Statistical indicators of water quality parameters

Station	WQP	Statistical indicators					
		Minimum	Average	Maximum	Skewness	Standard deviation	Coefficient of variation
Varand	HCO <sub>3</sub> <sup>-</sup>	1.5	4.36	22	5.54	1.38	0.31
	Cl <sup>-</sup>	0.2	2.16	29.5	4.13	2.81	1.30
	So <sub>4</sub> <sup>2-</sup>	0.03	1.42	9.4	2.85	0.87	0.61
	Ca <sup>2+</sup>	1	3.16	6	0.62	0.86	0.27
	Mg <sup>2+</sup>	0.5	2.37	8.2	2.09	0.86	0.36
	Na <sup>+</sup>	0.2	2.47	26	3.93	3.21	1.29
	K <sup>+</sup>	0.01	0.05	0.2	1.37	0.024	0.48
	TDS	31	507.49	2430	3.45	254.05	0.50

احتمال انتخاب شدنش افزایش می‌یابد. سپس، افراد انتخاب شده، با استفاده از یکی از عملگرهای ژنتیکی جهش یا همبری، جمعیت جدیدی را ایجاد می‌کنند و دوباره مورد بررسی قرار می‌گیرند. این فرایند با پیش رفتن نسل به جلو باعث بهبود نسبی کیفیت خواهد شد. همچنین، در صورت دست یابی به نتیجه مطلوب و یا رسیدن نسل‌ها به مقدار معین، فرایند تکامل متوقف شده و بهترین نتیجه ارائه می‌شود. شمای کلی فرایند اجرای برنامه‌نویسی بیان ژن به صورت زیر است (Esmaeili-Gisavandani et al., 2021).

برنامه نویسی بیان ژن: برنامه‌نویسی بیان ژن توسط فریرا در سال ۱۹۹۹ از توسعه GA و GP ارائه شد. در GEP، ابتدا یک جمعیت اولیه تصادفی حاصل از مجموعه توابع (عملگرهای ریاضی مورد استفاده در فرمول‌ها) و ترمینال‌ها (متغیرهای مساله و اعداد ثابت) ایجاد می‌شود. سپس، هر یک از افراد با استفاده از توابع برازش مورد ارزیابی قرار می‌گیرند. در این قسمت، از روش معروف چرخ رولت برای انتخاب بهترین افراد استفاده می‌شود. در چرخ رولت، هر یک از کروموزوم‌ها بسته به میزان مناسب بودنشان (بر اساس تابع برازش)،



شکل ۲- فلوچارت گام‌های اجرایی برنامه‌ریزی بیان ژن (Gandomi et al., 2013)

Fig. 2. Flowchart of the executive steps of gene expression programming (Gandomi et al., 2013)

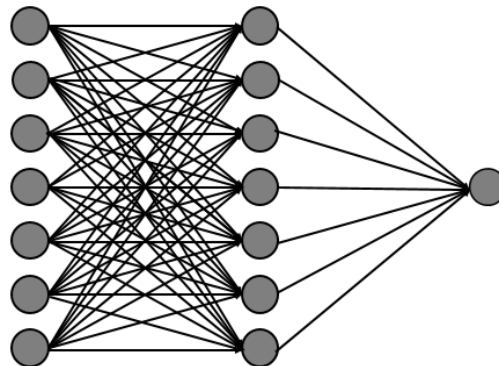
ساختار شبکه‌ای متشکل از تعدادی واحد (نرون‌های مصنوعی) دارند که در داخل شبکه به هم متصل شده اند و هر واحد دارای یک مشخصه ورودی و خروجی هستند که وظیفه محاسبات جزئی را بر عهده دارند. خروجی هر واحد، با توجه به مشخصه ورودی و خروجی آن، تعیین می‌شود. شکل ۳، ساختار مدل شبکه عصبی را نشان می‌دهد (Arya؛Mokhtarzad et al., 2017)؛ (Azar et al., 2021).

از نرم‌افزار GeneXproTools برای مدلسازی TDS در ایستگاه با استفاده از مدل GEP استفاده شد. پارامترهای اولیه مدل GEP بعد از انجام سعی و خطاهای متعدد، در جدول ۲، مشخص شده‌اند. همچنین تعداد نسل معین (۳۰۰۰) به‌عنوان معیار توقف مشخص شد.

شبکه عصبی مصنوعی: شبکه عصبی مصنوعی،

جدول ۲- پارامترهای تنظیم مدل GEP جهت تخمین مقادیر TDS  
Table 2. Setting parameters for GEP development in TDS estimation

Number	Description of parameter	Setting of parameter
1	Function set	$+, -, \times, \div, x^2, x^3, \frac{1}{x}, \sqrt{x}, \sqrt[3]{x}, \log x, \ln x$
2	Inversion	0.00546
3	Mutation	0.00138
4	Gene recombination	0.00277
5	Gene transportation	0.00277
6	Number of gene	3
7	Number of chromosomes	30
8	Maximum tree depth	5



شکل ۳- شماتیک شبکه عصبی مصنوعی (Mokhtarzad et al., 2017)

Fig. 3. Schematic of artificial neural network (Mokhtarzad et al., 2017)

بسط می‌دهد. این IMFها تمایل شدیدی به باندهای متعامد و محدود دارند، به طوری که هر کدام به یک مقیاس زمانی خاص در اطلاعات اختصاص داده می‌شوند. IMFها از طریق EMD یک ابزار چند وضوحی را فراهم کرده و تحلیل طیفی، اطلاعات محلی با فاز و دامنه متغیر با زمان بر طبق مقیاس‌ها را ارائه می‌دهد.

تجزیه مود تجربی: تجزیه مود تجربی<sup>۱</sup>، یکی از روش‌های تجزیه اطلاعات است که با عملیات غیرخطی و غیرایستا در راستا استخراج سیگنال‌ها در اطلاعات تولید شده، انجام می‌شود. این تکنیک شامل جداسازی‌های کامل و محلی برگرفته شده از اطلاعات یک سیگنال در نوسان‌های آهسته و سریع است. این روش توسط نوردن‌ای-هوانگ در سال ۱۹۹۸ ابداع شده که یک سری زمانی از اعداد را به تعداد محدودی از حالت‌های نوسانی به نام توابع مد ذاتی (IMF)

<sup>1</sup> Empirical Mode Decomposition (EMD)



$$MBE = \frac{1}{N} \sum_{i=1}^N (P_i - O_i) \quad (5)$$

که در آن‌ها،  $O_i$  مقادیرهای مشاهده‌ای،  $P_i$  مقادیرهای پیش‌بینی،  $\bar{O}$  میانگین داده‌های مشاهده‌ای،  $\bar{P}$  میانگین داده‌های پیش‌بینی و  $M$  تعداد داده‌ها است. از سویی، محاسبه هر آماره خطا به صورت جداگانه، نمی‌تواند معیار مناسبی برای بررسی عملکرد مدل مورد نظر باشد. از این رو، از شاخص آماری دیگری به نام تابع هدف<sup>۵</sup> استفاده شده که، به‌طور همزمان پارامترهای R، RMSE و میانگین خطای مطلق را در نظر می‌گیرد. مقادیر کمتر OBJ، نشان‌دهنده عملکرد قابل قبول مدل در تخمین مقادیر TDS بوده که از رابطه زیر به دست می‌آید (Najafzadeh and Ghaemi, 2019؛ Gandomi et al., 2013).

$$OBJ = \left( \frac{N_{tr} - N_{te}}{N_{total}} \right) \frac{MAE_{tr}}{R_{tr}^2} + \frac{2N_{te} \times MAE_{te}}{N_{total} \times R_{te}^2} \quad (6)$$

### نتایج و بحث

همانطور که ذکر شد، بررسی مقادیر TDS به‌عنوان یکی از مهمترین پارامترهای کیفیت آب رودخانه را می‌توان یکی از مهم‌ترین چالش کشور دانست. سازمان جهانی بهداشت کیفیت آب را بر اساس TDS به صورت جدول ۳، تقسیم‌بندی نموده است (Nowzari and Nematollahi, 2021). نتایج حاصل از داده‌های مشاهداتی در طول مدت نمونه‌برداری نشان دادند که مقادیر TDS در حدود ۸۰ درصد از نمونه‌ها در بازه ۳۰۰ تا ۶۰۰ میلی گرم بر لیتر بوده که بیانگر کیفیت خوب آب این رودخانه بوده است.

برخلاف روش‌های تجزیه سنتی مانند تبدیل موجک، این روش جدید داده-وقتی است. به بیان دیگر، این روش از تبدیل یا فیلتر از پیش تعریف شده‌ای استفاده نمی‌کند. همچنین، برخلاف تبدیل فوری، این روش نیاز به فرض خطی و ایستا بودن داده‌ها ندارد. بنابراین، این روش برای داده‌های غیرخطی و غیرایستا نیز عملکرد خوبی دارد (Zhang et al., 2010؛ Yang et al., 2019).

در این پژوهش، ابتدا داده‌های مورد استفاده به دو بخش آموزش و صحت‌سنجی تقسیم شدند. برای بخش آموزش مدل، ۷۵ درصد داده‌ها و برای صحت‌سنجی مدل در بخش آزمون، ۲۵ درصد داده‌ها اختصاص داده شد. سپس، تمامی سری‌های زمانی با استفاده از یکی از روش‌های پیش‌پردازش بنام الگوریتم تجزیه مود تجربی به چندین زیر سری (سیگنال) تجزیه شدند.

**معیار ارزیابی:** برای مقایسه عملکرد مدل‌های مستقل و ترکیبی ارائه شده برای تخمین پارامترهای کیفی رودخانه در مراحل آموزش و آزمون، شاخص‌های آماری ضریب همبستگی (R)، ریشه میانگین مربعات خطا<sup>۱</sup>، میانگین انحراف خطا<sup>۲</sup>، ضریب نش<sup>۳</sup> و نسبت RSD<sup>۴</sup> استفاده شدند.

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

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

$$NSE = 1 - \frac{\sum_{i=1}^N (P_i - O_i)^2}{\sum_{i=1}^N (O_i - \bar{O})^2} \quad (3)$$

$$RSD = \frac{RMSE}{STDEV_{obs}} = \frac{\sqrt{\sum_{i=1}^N (O_i - P_i)^2}}{\sqrt{\sum_{i=1}^N (O_i - \bar{O})^2}} \quad (4)$$

جدول ۳- تقسیم بندی کیفیت آب بر اساس TDS (Nowzari and Nematollahi, 2021)

Table 3. Classification of water quality based on TDS (Nowzari and Nematollahi, 2021)

Water quality status	WQP level (mg/litr)
Excellent	Less than 300
Good	300-600
Average	600-900
Poor	900-1200
Unfit quality	More than 1200

<sup>4</sup> Ratio of RMSE to Standard Deviation

<sup>5</sup> Objective Function (OBJ)

<sup>1</sup> Root Mean Square Error (RMSE)

<sup>2</sup> Mean Bias Error (MBE)

<sup>3</sup> Nash-Sutcliffe Efficiency Coefficient (NSE)

EMD-GEP در مقایسه با سایر مدل‌ها بود. در حالی که مدل مستقل GEP با  $RMSE=47/270$  و  $RSD=0/447$  ضعیف‌ترین عملکرد را از خود نشان داد. علاوه بر این، نتایج حاصل از آماره MBE در جدول ۴، نشان داد که مدل‌های مستقل و ترکیبی در مرحله آموزش به ترتیب بیش برآورد و کم برآورد بودند. در حالی که، مقادیر TDS تخمین زده شده توسط مدل‌های مستقل و ترکیبی مورد نظر در مرحله آزمون کمتر از مقادیر مشاهده‌ای بودند. به علاوه، نتایج حاصل از تابع هدف برای هر یک از مدل‌ها در جدول ۵، بیان شده است. مطابق جدول ۵، مدل EMD-GEP دارای مقادیر تابع هدف کمتری ( $OBJ=15/922$ ) نسبت به سایر مدل‌ها در مدلسازی مواد جامد محلول است. در حالی که، بیشترین مقدار تابع هدف ( $OBJ=29/340$ ) متعلق به مدل GEP است.

بر اساس جدول ۴، برای تخمین مواد جامد محلول در ایستگاه ورنه با تعداد ۷ ورودی، بالاترین دقت در مرحله آموزش ( $NSE=0/981$  و  $R=0/99$ )، متعلق به روش حاصل از ترکیب تجزیه مود تجربی و برنامه نویسی بیان ژن (EMD-GEP) بوده است. علاوه بر این، نتایج نشان داد که فرایند پیش‌پردازش داده‌ها می‌تواند تاثیر مطلوبی در تخمین مواد جامد محلول داشته باشد. روش مستقل ANN با داشتن ضریب همبستگی  $0/948$  کمترین دقت در مدلسازی مقادیر TDS را دارا بوده است. در حالی که، مدل EMD-ANN توانست مقادیر TDS را با  $46/15$  درصد خطای  $RMSE$  کمتر در مقایسه با مدل مستقل ANN ( $RMSE=50/086$ ) برآورد کند. معیارهای ارزیابی برای بررسی نحوه عملکرد مدل‌ها در ایستگاه ورنه در مرحله آزمون نشان داد که شاخص های  $RMSE$  و  $RSD$  بیانگر برتری روش ترکیبی

جدول ۴- نتایج ارزیابی مدل‌های هوشمند مصنوعی تلفیقی برای مدلسازی مواد جامد محلول در ایستگاه ورنه

Table 4. Evaluation of ML models for TDS estimation at the Varand station in training and testing periods

Model	Evaluation metrics				
	R	RMSE	MBE	NSE	RSD
Training period					
ANN	0.948	50.086	5.350	0.868	0.363
GEP	0.974	33.863	1.241	0.940	0.245
EMD-ANN	0.983	26.967	-9.461	0.962	0.195
EMD-GEP	0.991	18.739	-0.11	0.981	0.136
Testing period					
ANN	0.921	42.625	-11.958	0.836	0.403
GEP	0.933	47.270	-28.122	0.798	0.447
EMD-ANN	0.956	36.646	-18.270	0.878	0.347
EMD-GEP	0.984	24.413	-13.13	0.946	0.231

جدول ۵- ارزیابی عملکرد مدل‌های پیشنهادی با استفاده از مقادیر تابع هدف

Table 5. Performance evaluation of the proposed models using the objective function values

Model	OBJ
ANN	24/763
GEP	29.340
EMD-ANN	23.409
EMD-GEP	15.922

مطلوب برای مقادیر TDS ماهانه در ایستگاه ورنه را ارائه کرده‌اند. علاوه بر این، نزدیک‌تر بودن مقادیر TDS تخمین زده با EMD-GEP به خط نیم‌ساز اول و سوم در مقایسه با سایر روش‌ها نشان‌دهنده همبستگی بیشتر

شکل ۴، نمودار پراکنندگی روش‌های روش‌های مستقل و تلفیقی GEP، ANN، EMD-GEP و EMD-ANN را برای تخمین TDS ایستگاه ورنه نشان می‌دهد. بر اساس این شکل، تمامی روش‌ها تا حدودی تخمینی

در جدول ۶، نشان داده شده است. باتوجه به جدول یاد شده می‌توان بیان نمود که بیشترین مقدار آماره F و کمترین سطح معنی‌داری مربوط به مدل GEP است. همچنین، مدل ترکیبی EMD-GEP دارای کمترین آماره F و بیشترین سطح معنی‌داری است.

مقایسه مدل برتر این پژوهش (EMD-GEP) با پژوهش‌های قبل، مانند پژوهش Peipolzadeh et al., (2022) نشان داد که مدل ترکیبی EMD-GEP با ۱/۴۴ درصد ضریب همبستگی بالاتر و ۱۰/۵۴ درصد میانگین مربعات خطای کمتر در مقایسه با مدل EEMD-SVM ( $RMSE=27/29$  و  $R=0/97$ ) توانست مقادیر TDS را با دقت قابل‌قبول‌تری در مرحله آزمون تخمین بزند. علاوه بر این، خطای  $RMSE$  محاسبه شده با مدل EMD-GEP در تخمین مقادیر TDS در حدود ۶۸ درصد کمتر از مقدار  $RMSE$  ایجاد شده با روش Ghavidel and Montaseri (2014) بود.

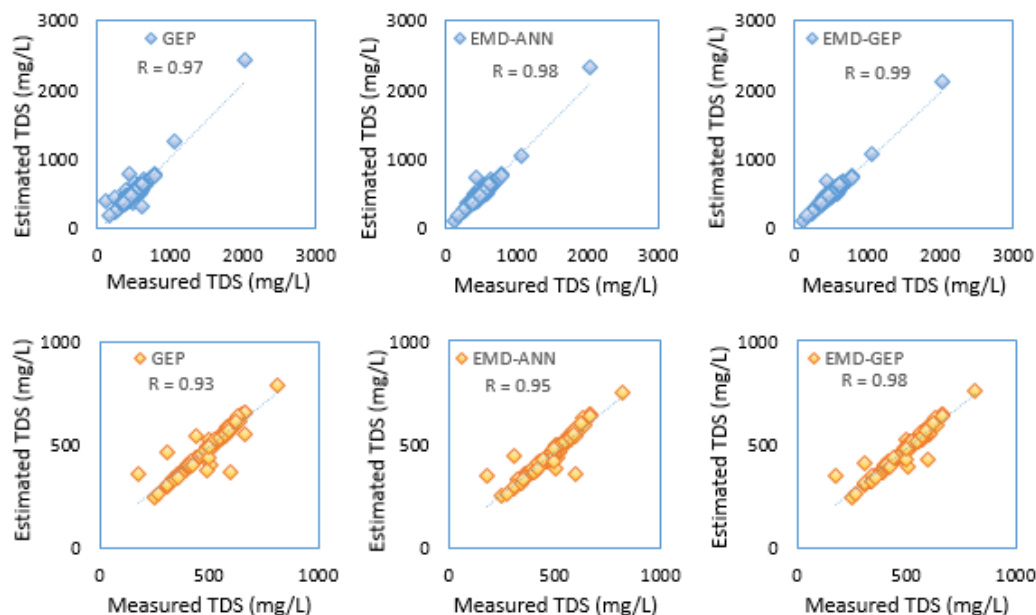
مقادیر مشاهده‌ای و تخمین زده شده با مدل EMD-GEP است. در خصوص ترکیب الگوریتم تجزیه مود تجربی این چنین می‌توان استنتاج کرد که روش مستقل GEP به تنهایی قادر به برآورد مقادیر حدی دبی جریان نبوده است. درحالی‌که ادغام آن با الگوریتم EMD موجب عملکرد بهتر مدل GEP شده است.

نمودار سری زمانی (شکل ۵) نیز نشان می‌دهد که مقادیر TDS تخمین زده شده با مدل EMD-GEP همپوشانی بهتری با مقادیر مشاهداتی دارد. علیرغم ضعف مدل‌ها در تخمین TDS، دقت مدل یاد شده نسبت به سایر مدل‌ها بیشتر و تخمین بهتری از مقادیر حدی به همراه داشته است. اغلب مدل‌های یادگیری ماشین توانایی بالاتری در درون‌یابی مقادیر نسبت به برون‌یابی دارند. بنابراین، آن‌ها توانایی بهتری در برآورد مقادیر میانگین در مقایسه با مقادیر حدی دارند. علاوه بر این، نتایج آزمون ANOVA برای بررسی عملکرد مدل‌های در نظر گرفته شده باتوجه به داده‌های مشاهداتی

جدول ۶- نتایج آزمون ANOVA برای مدل‌های هوشمند مصنوعی به ازای سری داده‌های مرحله آزمون

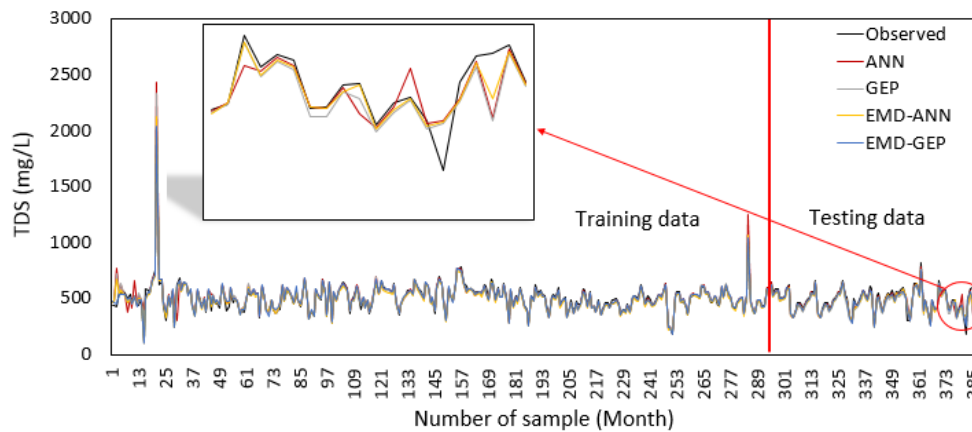
Table 6. ANOVA test results for ML models in testing period

Model	Significance level	F
ANN	0.20	1.63
GEP	0.05	3.83
EMD-ANN	0.36	0.82
EMD-GEP	0.41	0.67



شکل ۴- نمودارهای پراکندگی روش‌های هوشمند برای مدلسازی TDS در ایستگاه ورنند

Fig. 4. Scatter plots of ML models for TDS estimation at the Verand station



شکل ۵- نمودار سری زمانی برای دو مرحله آموزش و آزمون برای تخمین TDS در ایستگاه ورنند

Fig. 5. Monthly TDS predictions for training and testing periods at Varand station

قبولی برخوردار بوده اما می‌توان از سایر الگوریتم‌های یادگیری ماشین تکاملی و ادغام آن‌ها با روش‌های پیش پردازش دیگری مانند تجزیه حالت ارتعاشی (VMD) استفاده نمود. این فرایند باعث ایجاد مدل‌های دقیق تری در تخمین پارامترهای کیفی آب می‌شود. همچنین، با هدف افزایش دقت برآورد TDS، استفاده از متغیرهای ورودی مختلف بر اساس مقیاس‌های زمانی روزانه یا ساعتی توصیه می‌شود.

به‌عنوان راه‌حلی بالقوه برای مطالعات آینده، عدم قطعیت مرتبط با متغیرها و مدل‌های ورودی/خروجی را می‌توان برای ارائه مدل‌های برآورد قابل اعتمادتر بررسی کرد. همچنین، می‌توان چگونگی تاثیر ورودی‌های مدل و عدم قطعیت پارامترها بر نتایج تخمین TDS را بررسی نمود. به‌عنوان پیشنهاد نهایی و محدودیت پژوهش حاضر، سایر پارامترهای کیفی آب و پارامترهای هیدرولوژیکی مانند بارندگی، دما و دبی رودخانه را می‌توان به‌عنوان لایه‌های ورودی برای محاسبه بهتر تغییرپذیری و الگوهای TDS، به‌ویژه در رویدادهای جریان کم و زیاد به مدل وارد کرد. مزیت استفاده از مدل‌های یادگیری ماشین کمک موثری در حل مسایل مختلف به‌خصوص مدیریت منابع آب داشته است. به طور مثال، می‌توان مقادیر پارامترهای کیفی آب را تابعی از پارامترهای دیگر تصور کرده و با استفاده از این مدل‌ها مقادیر پارامترهای کیفی آب را در بازه‌های زمانی مختلف تخمین زد.

### نتیجه‌گیری

به‌منظور تخمین مواد جامد محلول در رودخانه تجن از روش‌های ANN و GEP در این پژوهش استفاده شد. سپس، برای تقویت برآوردها از الگوریتم‌های تجزیه مود تجربی برای پیش‌پردازش داده‌های سری زمانی استفاده شد. نتایج به‌دست آمده از مدل‌های هوشمند تلفیقی با استفاده از شاخص‌های آماری خطا نظیر ضریب همبستگی، ریشه میانگین مربعات خطا مورد ارزیابی قرار گرفت. نتایج نشان داد که روش EMD باعث تقویت مدل‌های ANN و GEP در راستای تخمین پارامترهای کیفی آب در ایستگاه ورنند واقع در رودخانه تجن شده اند. به‌طوری‌که مدل‌های ترکیبی حاصل از GEP و ANN توانست ۴۸/۳۵ و ۱۴/۰۲ درصد خطای RMSE را به‌ترتیب در مقایسه با دو مدل مستقل ANN و GEP کاهش دهد. علاوه بر این، مقایسه دو مدل EMD-GEP و EEMD-SVM نشان داد که مدل ترکیبی EMD-GEP مقادیر TDS را با ۱۰/۵۴ درصد خطای RMSE کمتر در مقایسه با مدل برتر پژوهش Peipolzadeh et al., (2022) تخمین زد. علاوه بر این، مقایسه خطای RMSE محاسبه شده با مدل EMD-GEP و فرمول Ghavidel and Montaseri (2014) در تخمین مقادیر TDS بیانگر دقت بالاتر EMD-GEP بود.

اصولاً، هر مدل دارای پارامترهای درونی خاصی بوده که تاثیر قابل توجهی بر عملکرد آن دارد. این پارامترهای مدل را می‌توان با سعی و خطا محاسبه نموده که می‌تواند باعث عدم قطعیت بالایی در نتایج حاصل از مدل شود. اگرچه مدل پیشنهادی از دقت قابل

## تشکر و قدردانی

بدینوسیله از کلیه افرادی که در ارتقای کیفیت مقاله نقش موثری ایفا کردند، صمیمانه قدردانی می‌شود.

## تعارض منافع

در این مقاله تضاد منافی وجود ندارد و این مساله مورد تایید همه نویسندگان است.

## منابع مورد استفاده

- Ahmadi Momqani, Y., Khorasani, N.A., Rafiei, Gh. 2010. Investigating polluting sources and water quality of Tajen river. *J. Nat. Environ.* 63(4), 317-327 (in Persian).
- Amiri, A., Amiri, S. 2015. Prediction of river quality parameters using artificial neural network. International Conference on New Achievements in Civil Engineering, Architecture, Environment and Urban Management, Tehran, 13 pages (in Persian).
- Arya, A.N., Kardan, N., Ghordoyee, S. 2021. Developing the artificial neural network evolutionary algorithms hybrid models (ANN-EA) to predict the daily evaporation from dam reservoirs. *Eng. Comput.* 1-19.
- Bagheri, M., Mansouri, P., Talebi, M.A., Karmi, M., Farzan, M. 2018. Comparison of water quality parameters of Paruz and Sabzkoh Rivers (Chaharmahal and Bakhtiari Provinces) with cold water fish farming standards. *Aquac. Dev. J.* 12(3), 1-14 (in Persian).
- Banadkooki, F.B., Ehteram, M., Panahi, F., Sammen, S.S., Othman, F.B., Ahmed, E.S. 2020. Estimation of Total Dissolved Solids (TDS) using new hybrid machine learning models. *J. Hydrol.* 587, 124989.
- Esmaili-Gisavandani, H., Lotfirad, M., Sofla, M.S.D., Ashrafzadeh, A. 2021. Improving the performance of rainfall-runoff models using the gene expression programming approach. *J. Water Climate Change* 12(7), 3308-3329.
- Ebadi, A.G., Zare, S. 2005. Measurement of organophosphorus pesticide in fish from the Tajan River. *Pakistan J. Biolog. Sci.* 8(10), 1463-1465
- Fethian H., Hormozinejad A. 2011. Prediction of quantitative and qualitative flow parameters in the Karun River (Malathani-Farsiat Basin) using artificial neural networks. *J. Wet. Ecobio.* 2(8), 29-43 (in Persian).
- Gandomi, A.H., Alavi, A.H., Ting, T.O., Yang, X.S. 2013. Intelligent modeling and prediction of elastic modulus of concrete strength via gene expression programming. In *Advances in Swarm Intelligence: Proceedings of 4th International Conference, ICSI 2013, Harbin, China, Part I4* Springer Berlin Heidelberg.
- Gazzaz, N.M., Yusoff, M.K., Aris, A.Z., Juahir, H., Ramli, M.F. 2012. Artificial neural network modeling of the water quality index for Kinta River (Malaysia) using water quality variables as predictors. *Marine Pollu. Bulletin* 64(11), 2409-2420
- Ghavidel, S., Montaseri, M. 2014. Application of different data-driven methods for the prediction of total dissolved solids in the Zarinehroud Basin. *Stoch. Environ. Res. Risk Assess.* 28, 2101-2118.
- Jamei, M., Ahmadianfar, I., Chu, X., Yaseen, Z.M. 2020. Prediction of surface water total dissolved solids using hybridized wavelet-multigene genetic programming: new approach. *J. Hydrol.* 589, 125335.
- Mohammadpour, R., Shaharuddin, S., Chang, C.K., Zakaria, N.A., Ab-Ghani, A., Chan, N.W. 2015. Prediction of water quality index in constructed wetlands using support vector machine. *Environ. Sci. Pollut. Res.* 22(8), 6208-6219.
- Najafzadeh, M., Ghaemi, A. 2019. Prediction of the five-day biochemical oxygen demand and chemical oxygen demand in natural streams using machine learning methods. *Environ. Monit. Assess.* 191, 1-21.
- Nowzari, H., Nematollahi, L. 2021. Investigating the qualitative characteristics of underground water sources that supply drinking water in Abade City. *Hum. Environ.* 19(1), 23-33 (in Persian).
- Peipolzadeh, S., Mastouri, R., Shahkarmi, N. 2022. Modeling and prediction of dissolved solids in Tajen River using an artificial intelligence-based approach based on pre-processing algorithms. *Iran Water Res.* 16(1), 25-37 (in Persian).
- Rezaie-Balf, M., Attar, N.F., Mohammadzadeh, A., Murti, M.A., Ahmed, A.N., Fai, C.M., El-Shafie, A. 2020. Physicochemical parameters data assimilation for efficient improvement of water quality index prediction: Comparative assessment of a noise suppression hybridization approach. *J. Clean. Prod.* 271, 122576.
- Saeedi M., Karbassi, A.R., Nabi Bidhendi, Gh.R., Mehrdadi, N. 2006. Impact of anthropogenic activities on heavy metals pollution in Tajan River water, Mazandaran Province. *J. Environ. Stu.* 32(40), 41-50.

- Safar Talory, M., Dehghani, A. 2012. Predicting changes in water quality of Shast Kalateh River using artificial neural network. Proceedings of National Conference on Water Flow and Pollution, Tehran (in Persian).
- Sharafati, A., Asadollah, S.B.H.S., Hosseinzadeh, M. 2020. The potential of new ensemble machine learning models for effluent quality parameters prediction and related uncertainty. *Process Saf. Environ. Prot.* 140, 68-78.
- Shiri, N., Shiri, J., Yaseen, Z.M., Kim, S., Chung, I.M., Nourani, V., Zounemat-Kermani, M. 2021. Development of artificial intelligence models for well groundwater quality simulation: Different modeling scenarios. *Public Library of Science (PLOS) One*, 16(5), e0251510.
- Sun, K., Rajabtabar, M., Samadi, S., Rezaie-Balf, M., Ghaemi, A., ShamshirBand, S., Mosavi, A. 2021. An integrated machine learning, noise suppression, and population-based algorithm to improve total dissolved solids prediction. *Eng. Appl. Comput. Fluid Mech.* 15(1), 251-271.
- Wu, Z., Huang, N.E. 2009. Ensemble empirical mode decomposition: a noise-assisted data analysis method. *Advan. Adap. Ddata Analy.* 1(01), 1-41.
- Mokhtarzad, M., Eskandari, F., Vanjani, N.J., Arabasadi, A. 2017. Drought forecasting by ANN, ANFIS, and SVM and comparison of the models. *Environ. Earth Sci.* 76(21), 1-10.
- Yang, H.F., Chen, Y.P.P. 2019. Hybrid deep learning and empirical mode decomposition model for time series applications. *Expert Syst Appl.* 120, 128-138.
- Zhang, J., Yan, R., Gao, R.X., Feng, Z. 2010. Performance enhancement of ensemble empirical mode decomposition. *Mech. Syst. Signal Process.* 24(7), 2104-2123.