

Estimation of Water Quality Parameters in the Sepidrood River by ANFIS, GEP and LS-SVM Models

A. Adib¹, H. Farajpanah², M. Mahmoudian Shoushtari³, I. Ahmadianfar⁴

1. Prof., Civil Engineering Dept., Engineering Faculty, Shahid Chamran University of Ahvaz, Ahvaz, Iran
(Corresponding Author) arashadib@yahoo.com
2. MSc Student, Civil Engineering Dept., Engineering Faculty, Shahid Chamran University of Ahvaz, Ahvaz, Iran
3. Prof., Civil Engineering Dept., Engineering Faculty, Shahid Chamran University of Ahvaz, Ahvaz, Iran
4. Assist. Prof., Civil Engineering Dept. Engineering Faculty, Behbahan Khatam Al-Anbia University of Technology, Behbahan, Iran

(Received May 24, 2019 Accepted Mar. 16, 2020)

To cite this article:

Adib, A., Farajpanah, H., Mahmoudian Shoushtari, M., Ahmadianfar, I. 2020. "Estimation of water quality parameters in the Sepidrood River by ANFIS, GEP and LS-SVM models" Journal of Water and Wastewater, 31(5), 1-10. Doi: 10.22093/wwj.2020.187271.2873. (In Persian)

Abstract

Rivers are the most important water supply resource for the drinkable, agricultural and industrial demands. Therefore, estimation of water quality parameters in rivers is an essential and necessary task. This research applies the Adaptive Neuro-Fuzzy Inference System (ANFIS), the Least Squares-Support Vector Machines (LS-SVM) and the Gene Expression Programming (GEP) for estimation of Total Dissolved Solids (TDS), Electrical Conductivity (EC) and Total Hardness (TH) in the Sepidrood River and a 40 year period. The applied performance criteria are the correlation coefficient (R), the Nash-Sutcliffe model Efficiency coefficient (NSE), the Normalized Mean Squared Error (NMSE) and the Mean Absolute Error (MAE). These methods have high ability for estimation of water quality parameters. The best method is LS-SVM method for estimation of TDS ($R_{Train}=0.95$ $R_{Test}=0.96$). The best method is GEP method for estimation of EC ($R_{Train}=0.94$ $R_{Test}=0.95$). The best method is ANFIS method for estimation of TH ($R_{Train}=0.92$ $R_{Test}=0.94$). This research shows that intelligence methods can estimate unmeasured concentration of qualitative parameters by concentration of other qualitative parameters.

Keywords: The Sepidrood River, Water Quality, ANFIS, LS-SVM, GEP.



مجله آب و فاضلاب، دوره ۳۱، شماره ۵، صفحه: ۱۰-۱

تخمین پارامترهای کیفیت آب رودخانه سفیدرود با استفاده از مدل‌های ANFIS، GEP و LS-SVM

آرش ادیب^۱، هیوا فرج‌پناه^۲، محمد محمودیان شوشتری^۳، ایمان احمدیان فر^۴

۱- استاده گروه مهندسی عمران، دانشکده مهندسی،

دانشگاه شهید چمران اهواز، اهواز، ایران

(نویسنده مسئول) arashadib@yahoo.com

۲- دانشجوی کارشناسی ارشد، گروه مهندسی عمران، دانشکده مهندسی،

دانشگاه شهید چمران اهواز، اهواز، ایران

۳- استاده گروه مهندسی عمران، دانشکده مهندسی،

دانشگاه شهید چمران اهواز، اهواز، ایران

۴- استادیار، گروه مهندسی عمران، دانشکده مهندسی،

دانشگاه صنعتی خاتم الانبیا بهبهان، بهبهان، ایران

(دریافت ۹۸/۳/۳ پذیرش ۹۸/۱۲/۲۶)

برای ارجاع به این مقاله به صورت زیر اقدام بفرمایید:

ادیب، آ.، فرج‌پناه، ه.، محمودیان شوشتری، م.، احمدیان فر، ا.، ۱۳۹۹، تخمین پارامترهای کیفیت آب رودخانه سفیدرود با استفاده از

مدل‌های ANFIS، GEP و LS-SVM "مجله آب و فاضلاب، ۳۱(۵)، ۱۰-۱. Doi: 10.22093/wwj.2020.187271.2873

چکیده

رودخانه‌ها مهم‌ترین منابع آب آشامیدنی، کشاورزی و صنعت به‌شمار می‌روند. از این رو بررسی و تخمین پارامترهای کیفی آب در طول یک رودخانه باید مورد توجه قرار گیرد. در این پژوهش از سه مدل سیستم سامانه استنتاجی عصبی- فازی تطبیقی (ANFIS)، حداقل مربعات ماشین بردار پشتیبان (LS-SVM) و برنامه‌ریزی بیان ژن (GEP) برای تخمین برخی از پارامترهای کیفی آب (کل جامدات محلول، هدایت الکتریکی و سختی کل) رودخانه سفیدرود طی یک دوره آماری ۴۰ ساله استفاده شد. ارزیابی این سه روش هوش مصنوعی توسط معیارهای آماری ضریب هم‌بستگی (R)، ضریب کارایی نش- ساتکلیف (NSE)، مربع میانگین خطای استاندارد نرمال شده (NMSE) و میانگین قدر مطلق خطا (MAE) انجام شد. نتایج حاصل از این پژوهش نشان داد که هر سه روش هوش مصنوعی ANFIS، LS-SVM و GEP قابلیت بسیار زیادی در تخمین پارامترهای کیفی TDS، TH، EC دارند. به‌صورتی که برای تخمین TDS روش LS-SVM (RTrain=0.95 RTest=0.96)، برای تخمین EC، روش GEP (RTrain=0.94 RTest=0.95) و برای تخمین TH روش ANFIS (RTrain=0.92 RTest=0.94) بهترین شبیه‌سازی را انجام دادند. این پژوهش نشان می‌دهد که با استفاده از روش‌های هوش مصنوعی می‌توان غلظت پارامترهای کیفی را در صورت عدم اندازه‌گیری آنها و بر اساس غلظت سایر پارامترهای کیفی محاسبه کرد.

واژه‌های کلیدی: سفیدرود، کیفیت آب، ANFIS، LS-SVM، GEP

۱- مقدمه

امروزه تعیین کیفیت آب به وسیله روش‌های هوش مصنوعی مناسب امری ضروری برای مدیریت منابع آب و حفظ محیط‌زیست به‌خصوص در مناطق فاقد وسایل اندازه‌گیری مناسب کیفیت آب است.



الگوریتم ژنتیک^{۱۶} و برنامه‌ریزی ژنتیک^{۱۷} هستند. ورودی‌ها با استفاده از تجزیه و تحلیل مولفه اصلی^{۱۸} تعیین شدند. نتایج نشان داد که الگوریتم GA-LSSVR^{۱۹} دارای دقت بیشتری برای مدل‌سازی پارامترهای کیفیت آب نسبت به GP است. این امر توسط به‌کارگیری ضریب تعیین و ضریب کارایی نش-ساتکلیف تأیید شد (Bozorg-Haddad, 2017).

آزاد و همکاران از ANFIS بهینه شده با روش‌های ازدحام ذرات^{۲۰} و کلونی مورچه‌ها^{۲۱} برای تخمین EC، کل جامدات محلول^{۲۲}، نسبت جذب سدیم^{۲۳}، سختی کربناتی^{۲۴} و سختی کل^{۲۵} در رودخانه زاینده‌رود استفاده کردند (Azad et al., 2019).

عاقل و همکاران از ANFIS-PSO و ANFIS برای تخمین EC، TDS، TH و قلیایی کل^{۲۶} در استان کرمانشاه استفاده کردند (Aghel et al. 2019).

حقیابی و همکاران از روش‌های ANN، ماشین بردار پشتیبان^{۲۷} و روش گروهی پردازش داده‌ها^{۲۸} برای تخمین دبی جریان، EC، pH، Ca²⁺، HCO₃⁻، SO₄²⁻، Mg²⁺، SAR، Na⁺ و TDS در رودخانه تیره از سرشاخه‌های رودخانه دز استفاده کردند (Haghiabi et al., 2018).

در این پژوهش از سه روش هوش مصنوعی GEP، ANFIS و LS-SVM برای تخمین پارامترهای کیفی آب در ایستگاه آستانه رودخانه سفیدرود استفاده شد و نوآوری این پژوهش استفاده از روش رگرسیون گام‌به‌گام برای معرفی داده‌های ورودی به هر مدل است و برای هر پارامتر کیفی سه مدل با ۱، ۳ و ۵ پارامتر ورودی انتخاب شدند و در نهایت مدلی انتخاب شد که علاوه بر داشتن بیشترین دقت، نیاز به پارامترهای ورودی کمتری داشت. این امر برای ایستگاه‌های هیدرومتری ایران بسیار مهم است زیرا در این

از جمله پژوهش‌های که در خصوص مدل‌سازی کیفیت آب رودخانه‌ها در کشورهای مختلف انجام شده به شرح زیر است:

کیسی و همکاران به بررسی دقت شبکه‌های عصبی مصنوعی^۱ و سیستم استنتاج فازی تطبیقی^۲ و برنامه‌ریزی بیان ژن^۳ در مدل‌سازی اکسیژن محلول^۴ در رودخانه‌ای در آمریکا پرداختند. آنها با استفاده از شاخص‌های آماری ضریب تعیین^۵، جزر میانگین مربع خطاها^۶، میانگین قدر مطلق خطا^۷ و میانگین قدر مطلق خطا نسبی^۸ به ارزیابی مدل‌ها پرداختند (Kisi et al., 2013).

روان‌سالار و رجایی به بررسی توانایی مدل ANN و مدل ترکیبی شبکه عصبی موجک^۹ برای پیش‌بینی ماهانه هدایت الکتریکی^{۱۰} در یک ایستگاه اندازه‌گیری واقع در ترکیه پرداختند (Ravansalar and Rajaei, 2015).

کیسی و پارمار دقت مدل‌های اسپلاین رگرسیونی تطبیقی چند متغیره^{۱۱}، حداقل مربعات-ماشین بردار پشتیبان^{۱۲} و مدل درختی M5 را در مدل‌سازی آلودگی آب رودخانه‌ای در هند را بررسی کردند (Kisi and Parmar, 2016). رجایی و شهابی به ارزیابی برنامه‌ریزی بیان ژن موجک^{۱۳} و WANN در پیش‌بینی مقدار کل نیتروژن در آب‌های ساحلی پرداختند و نتایج WANN و WGEF را با مدل‌های ANN، GEP، و روش رگرسیون چند خطی^{۱۴} مقایسه کردند. نتایج مقایسه مدل‌های ترکیبی موجک با دیگر مدل‌ها با استفاده از معیارهای آماری ضریب کارایی نش-ساتکلیف^{۱۵} و RMSE، برتری عملکرد مدل‌های ترکیبی موجک را برای پیش‌بینی کل نیتروژن در آب‌های ساحلی نشان داد (Rajaei and Shahabi, 2016).

بزرگ‌حداد و همکاران از دو روش داده کاوی برای مدل‌سازی پارامترهای کیفی آب رودخانه سفیدرود استفاده کردند. این روش‌ها شامل مدل حداقل مربعات رگرسیون بردار پشتیبان بهینه شده با

¹⁶ Genetic Algorithm (GA)

¹⁷ Genetic Programming (GP)

¹⁸ Principal Component Analysis (PCA)

¹⁹ Genetic Algorithm-Less Square Support Vector Regression (GA-LSSVM)

²⁰ Particle Swarm Optimization (PSO)

²¹ Ant Colony Optimization (ACO_R)

²² Total Dissolved Solids (TDS)

²³ Sodium Adsorption Ratio (Cheng et al., 2016) (SAR)

²⁴ Carbonate Hardness (CH)

²⁵ Total Hardness (TH)

²⁶ Total Alkalinity (TA)

²⁷ Support Vector Machines (SVMs)

²⁸ Group Method of Data Handling (GMDH)

¹ Artificial Neural Network (ANN)

² Adaptive Neuro-Fuzzy Inference System (ANFIS)

³ Gene Expression Programming (GEP)

⁴ Dissolved Oxygen (DO)

⁵ Coefficient of Determination (R)

⁶ Root Mean Square Error (RMSE)

⁷ Mean Absolute Error (MAE)

⁸ Mean Absolute Relative Error (MARE)

⁹ Wavelet Artificial Neural Network (WANN)

¹⁰ Electrical Conductivity (EC)

¹¹ Multivariate Adaptive Regression Spline (MARS)

¹² Least Square-Support Vector Machine (LS-SVM)

¹³ Wavelet Gene Expression Programming (WGEF)

¹⁴ Multiple Linear Regression (MLR)

¹⁵ Nash-Sutcliffe Efficiency Coefficient (NSE)



مدل‌سازی EC، TDS و TH رودخانه سفیدرود با روش ANFIS و LS-SVM در محیط نرم‌افزار MATLAB کدنویسی شد و GEP با استفاده از GeneXpro به‌کار گرفته شد.

۲-۲- مدل ANFIS

سامانه استنتاجی عصبی-فازی تطبیقی برای اولین بار توسط جانگ معرفی شد. این سامانه ابزاری کارآمد برای تخمین توابع پیوسته حقیقی در دامنه‌ای محدود است. این سیستم پیاده‌سازی یک سیستم فازی سوگنو به‌صورت ساختار شبکه‌ای پیش‌رونده است (Jang, 1993).

۲-۳- LS-SVM

سویکنز و اندوال در سال ۱۹۹۹، مدل حداقل مربعات ماشین‌بردار پشتیبان را معرفی کردند. این مدل محدودیت‌های یکسانی با روش SVM کلاسیک دارد اما نسبت به این روش از پیچیدگی محاسباتی کمتر و دقت و سرعت بیشتری دارد (Suykens and Vandewalle, 1999) در واقع الگوریتمی پیشرفته‌تر از الگوریتم SVM است (Mellit et al., 2013).

ایستگاه‌ها به‌دلیل نبودن تجهیزات و افراد خبره در بسیاری از سال‌ها فقط تعداد معدودی از پارامترهای کیفی اندازه‌گیری می‌شوند.

۲- مواد و روش‌ها

۲-۱- حوضه مورد مطالعه

حوضه آبریز سفیدرود بین مختصات جغرافیایی $37^{\circ} 37'$ تا $46^{\circ} 13'$ طول شرقی و $45^{\circ} 35'$ تا $55^{\circ} 37'$ عرض شمالی قرار گرفته است (Bozorg-Haddad et al., 2017). در این پژوهش از داده‌های کیفیت آب ایستگاه آستانه (با مختصات جغرافیایی $49^{\circ} 56' 2''$ طول شرقی و $37^{\circ} 16' 42''$ عرض شمالی و ارتفاع ۳- متر) واقع در رودخانه سفیدرود در طول دوره نمونه‌برداری ۴۰ ساله (۱۳۹۴-۱۳۵۰ به‌جز سال‌های ۱۳۵۷ تا ۱۳۶۰ و سال ۱۳۶۵ که داده‌ای برداشت نشده بود) استفاده شد. موقعیت جغرافیایی ایستگاه آستانه و مشخصات آماری داده‌های کیفیت آب و جریان رودخانه به‌ترتیب در شکل ۱ و جدول ۱ نشان داده شده است. همان‌طور که در جدول ۱ قابل مشاهده است، دبی رودخانه دارای بیشترین ضریب تغییرات و pH داری کمترین ضریب تغییرات است.

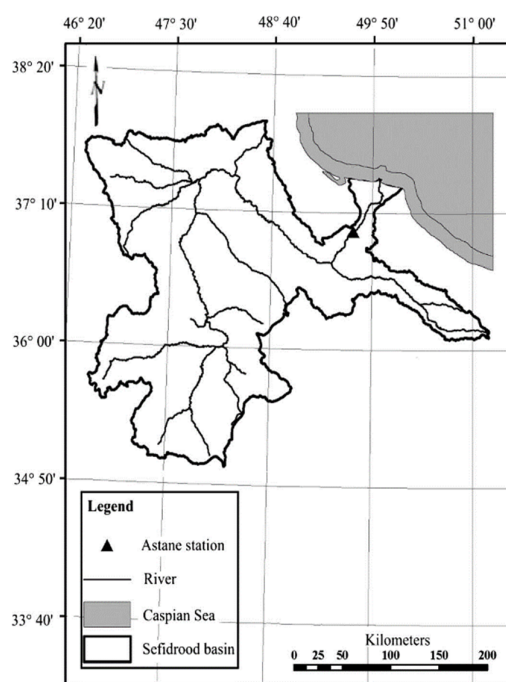


Fig. 1. The Sepidrood river watershed and the Astane hydrometric station (Bozorg-Haddad et al., 2017)
 شکل ۱- حوضه آبریز سفیدرود و ایستگاه هیدرومتری آستانه (Bozorg-Haddad et al., 2017)



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

Table 1. The stochastic characteristics of water quality parameters and flow discharge in the Astane hydrometric station

Parameter	Unit	Mean	Max	Min	SD	CV
Q	m ³ /s	105.4	1501	1.31	179.33	1.7
TDS	mg/L	680.8	1472	130	246.22	0.36
EC	μmhos/cm	1076	1236	185	383.8	0.36
CL	meq/L	4.48	15.6	0.17	2.83	0.63
HCO ₃	meq/L	3.57	9.25	0.15	1.5	0.42
SO ₄	meq/L	4.42	16.82	0.03	3.77	0.85
K	meq/L	0.08	0.68	0	0.05	0.61
Ca	meq/L	3.04	10.36	0.28	1.23	0.4
Mg	meq/L	2.71	8.7	0.05	1.72	0.64
Na	meq/L	3.82	15.65	0.01	3.37	0.86
SAR	meq/L	2.25	9.71	0	1.85	0.82
pH	-----	7.63	9.63	6.2	0.45	0.06
TH	meq/L	287.5	284.5	22.5	101.56	0.35

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

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

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

$$MAE = \frac{1}{N} \sum_{i=1}^N |O_i - E_i| \quad (4)$$

که در آنها O_i مقدار مشاهداتی، E_i مقدار برآوردی و \bar{O}_i متوسط مقدار مشاهداتی است.

۳- بحث و نتایج

برای انتخاب بهترین مدل برای شبیه‌سازی TH، EG و TDS رودخانه سفیدرود، دو فاکتور مهم هستند: ۱- شبکه طبق معیارهای ارزیابی دقت زیادی داشته باشد ۲- مدل انتخابی کمترین

۲-۴- GEP

برنامه‌ریزی بیان ژن، تعمیم‌یافته روش الگوریتم ژنتیک است که برای اولین بار توسط جان‌کوزا^۱ بر اساس تئوری داروین^۲ ارائه شد و روشی مناسب برای مدل‌سازی فرایندهای غیرخطی و پیچیده است که توسط (Ferreira, 2001) ابداع شد. این روش مانند GA و GP، جزو الگوریتم‌های تکاملی است که از جمعیتی از افراد استفاده کرده و آنها را مطابق برازندگی انتخاب می‌کند و تغییرات ژنتیکی را با استفاده از یک یا چند عملگر ژنتیکی اعمال می‌کند (Ferreira, 2001) خلاصه‌ای از پارامترهای به‌کار گرفته شده در GEP در جدول ۲ ارائه شده است.

۲-۵- معیارهای ارزیابی مدل‌ها

برای ارزیابی مدل‌های مورد استفاده، از ۴ شاخص آماری معتبر و رایج شامل NS_E ، R ، مربع میانگین خطای استاندارد نرمال شده^۳ و MAE استفاده شد. روابط شاخص‌های کمی چهارگانه مذکور در فرمول‌های ۱ تا ۴ نشان داده شده است.

بر اساس این شاخص‌های کمی، مدل برتر دارای مقدار حداکثر $|R|$ و NS_E (بیشترین مقدار NS_E برابر ۱ است) و حداقل NS_E و MAE خواهد بود

¹ John Koza

² Darwin's Theory

³ Normalized Mean Square Error (NMSE)



جدول ۲- پارامترهای GEP در این پژوهش

Table 2. The parameters of GEP in this research

General settings		Genetic operators	
Chromosome length	30	Mutation rate	0.044
Head length	7	IS transposition rate	0.1
Number of genes	3	RIS transposition rate	0.1
Link function	+	One-point recombination rate	0.2
Performance criterion	MSE	Two-point recombination rate	0.3
Population size	1000	Gene recombination rate	0.2
		Gene transposition rate	0.1

در این پژوهش برای مدل‌سازی EC، TDS و TH از ۷۰ درصد داده‌ها به صورت تصادفی برای آموزش و از ۳۰ درصد باقی‌مانده برای تست مدل استفاده شد.

جداول ۳ و ۴ نشان می‌دهند که برخلاف تصور، غلظت آلاینده‌های موردنظر در این پژوهش ارتباطی به دبی جریان در رودخانه ندارند. دلیل، این است که ایستگاه آستانه در منطقه‌ای قرار دارد که محل کشاورزی است و غلظت آلاینده‌ها، بیشتر تابع مقدار پسابی است که از مزارع وارد رودخانه می‌شود و این پساب‌ها حاوی سموم و کودهای شیمیایی است و غلظت آلاینده‌ها در رودخانه را افزایش می‌دهند.

همچنین این جداول نشان می‌دهند که EC و TDS بیشترین هم‌بستگی را با یکدیگر دارند و EC تقریباً ۰/۶ TDS است و به این دلیل هم‌بستگی آنها به ۰/۹۴ می‌رسد. در رتبه بعدی این دو پارامتر بیشترین هم‌بستگی را با سدیم و کلر دارند که اجزا تشکیل‌دهنده نمک طعام هستند. زیرا این دو پارامتر در واقع مقدار

پارامتر ورودی را داشته باشد. هر مدل و شبکه‌ای این دو فاکتور را به صورت هم‌زمان داشته باشد، به‌عنوان مدل و شبکه برتر برای پیش‌بینی پارامتر موردنظر انتخاب می‌شود.

۳-۱- انتخاب ورودی‌ها و تعیین ساختار مدل‌ها

برای تعیین پارامترهای ورودی مدل‌ها برای تخمین پارامترهای کیفی ابتدا ضرایب هم‌بستگی پیرسون در سطح معنی‌داری ۹۹ درصد با استفاده از نرم‌افزار SPSS 22، بین پارامترهای کمی و کیفی آب در ایستگاه آستانه محاسبه شد که در جدول ۳ نشان داده شده است. سپس پارامترهای که دارای هم‌بستگی معنی‌دار هستند بر اساس رگرسیون گام‌به‌گام به ترتیب اولویت تأثیرگذاری بر مقدار پارامتر خروجی مدل، به‌عنوان ورودی‌های مدل‌ها انتخاب شدند و برای هر پارامتر کیفی ۳ الگوی ورودی تحت عنوان a، b و c تعیین شدند که الگو a، ۱ پارامتر کیفی ورودی، الگو b، ۳ پارامتر کیفی ورودی و الگو c، ۳ پارامتر کیفی ورودی دارد که در جدول ۴ نشان داده شده است.

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

Table 3. The Pearson correlation coefficient between water quality parameters for confidence level of 99%

Q	TDS	EC	Cl	HCO ₃	SO ₄	K	Ca	Mg	Na	SAR	pH	TH	
TDS	-0.032	1	*0.94	*0.8	*0.45	*0.61	*0.34	*0.5	*0.5	*0.83	*0.66	*0.13	*0.73
EC	-0.026	*0.94	1	*0.79	*0.4	*0.63	*0.35	*0.46	*0.45	*0.85	*0.69	*0.12	*0.66
TH	-0.062	*0.73	*0.65	*0.51	*0.78	*0.43	*0.28	*0.54	*0.8	*0.54	*0.24	0.089	1

* Correlation is significant at the 0.99 level



جدول ۴- پارامترهای انتخاب شده توسط رگرسیون گام به گام

Table 4. The selected parameters by stepwise regression

Inputs of model	Model	Output
Ec	a	TDS
Ec, Cl, Na	b	
Ec, Cl, Na, SAR, TH	c	
Mg	a	TH
Mg, TDS, HCO ₃	b	
Mg, TDS, HCO ₃ , Ec, Na	c	
TDS	a	EC
TDS, Cl, Na	b	
TDS, Cl, Na, SAR, TH	c	

آزمون دقت کمتری دارد و روش ANFIS و LS-SVM هر دو در بخش آموزش تقریباً عملکرد مشابهی داشته‌اند ولی در بخش آموزش روش ANFIS دقت بهتری دارد. در نهایت برای تخمین مقدار TH آب رودخانه مدل b روش ANFIS پیشنهاد می‌شود. معادلات ارائه شده توسط روش GEP عبارت‌اند از

$$TDS = Cl + EC / 2.34 - (TH / 0.74) * \cos(9.42) + 2Na - 9.35 + SAR * Cl \quad (5)$$

$$EC = TDS + 9.24 + (8.91 + TDS * Tg^{-1}(6.72)) * \ln(5.87) + Tg^{-1}(7.65) - 2 * TDS - 6.34 \quad (6)$$

$$TH = \text{Exp}(\log(TDS * HCO_3 * 8.23 * 4.57)) + \text{Exp}(\log(TDS * Mg * 8.23 * 5.66)) + \ln(TDS * Mg) + 2HCO_3 \quad (7)$$

در شکل ۲ مقایسه بین مقدار مشاهداتی و مقدار محاسباتی توسط سه روش در فاصله زمانی مارس ۲۰۱۴ تا فوریه ۲۰۱۶ (بازه زمانی ۲ ساله) نشان داده شده است. در این شکل مشاهده می‌شود که هر سه روش به خوبی قادر به تخمین غلظت آلاینده‌ها هستند و نقاط حداقل و حداکثر غلظت آلاینده‌ها را به خوبی مدل‌سازی می‌کنند. بنابراین به جای اجرای مجدد این مدل‌ها در ایستگاه آستانه می‌توان از معادلات ۵ تا ۷ برای تخمین غلظت آلاینده‌ها استفاده کرد.

در شکل ۳ مقایسه بین نتایج سه روش محاسباتی توسط دیگرام تیلور نشان داده شده است. در این دیگرام از معیار کارایی انحراف ریشه میانگین مربعات^۱ استفاده شده است

نمک و شوری آب را نشان می‌دهند. پارامتر TH بیشترین هم‌بستگی را با یون‌های تشکیل‌دهنده خود دارد (در درجه اول، یون منیزیم و سپس یون‌های کربنات، سدیم و کلسیم).

۲-۳- مقایسه مدل‌های ANFIS, LS-SVM و GEP

با توجه به جدول ۵ برای پیش‌بینی TDS، هر سه روش هوش مصنوعی به خوبی توانستند عمل شبیه‌سازی را انجام دهند. در نهایت برای پیش‌بینی TDS رودخانه سفیدرود، مدل b روش LS-SVM پیشنهاد داده می‌شود. به دلیل اینکه با ۳ پارامتر ورودی در بخش آموزش عملکرد بهتری از مدل a روش ANFIS و مدل c GEP داشت و در بخش آزمون هم نسبت به دو روش ANFIS و GEP دقت بهتری داشت. برای مدل‌سازی EC هر سه روش هوش مصنوعی دقت خوب و قابل‌قبولی دارند. دو روش LS-SVM و GEP با یک پارامتر ورودی شبیه‌سازی را انجام دادند و دقت هر دو روش هم تقریباً مشابه است. با این تفاوت که روش GEP در بخش آموزش و آزمون مقدار MAE را نسبت به روش LS-SVM کاهش داده است. در کل، مدل c روش ANFIS نسبت به ۲ روش دیگر بعضی شاخص‌های آماری را به خصوص در بخش آزمون، بهبود داده است ولی اختلاف آن قدر محسوس نیست که مدل c روش ANFIS با ۵ پارامتر ورودی به عنوان مدل برتر انتخاب شود. در نتیجه مدل a روش GEP به عنوان مدل برتر برای تخمین مقدار EC رودخانه سفیدرود انتخاب می‌شود. علاوه بر این ارائه معادله‌ای برای مدل‌سازی در روش GEP یکی از مزیت‌های برتر این روش است. برای پیش‌بینی TH، طبق جدول ۵ مشاهده می‌شود که هر سه روش با ۳ پارامتر ورودی شبیه‌سازی را انجام داده‌اند. روش GEP نسبت به دو روش دیگر در بخش آموزش و

¹ Root-Mean-Square Deviation (RMSD)



جدول ۵- مقایسه نتایج روش‌های ANFIS, LS-SVM, GEP برای تخمین پارامترهای کیفیت آب TH, EC و TDS
Table 5. Comparison between results of ANFIS, LS-SVM and GEP for calculation of TH, EC and TDS

Parameter	Method	Model	Train				Test			
			R	NMSE	MAE	NS _E	R	NMSE	MAE	NS _E
TDS	ANFIS	a	0.94	0.12	38.86	0.88	0.95	0.1	42.4	0.9
	LS-SVM	b	0.95	0.09	41.57	0.91	0.96	0.09	39.11	0.91
	GEP	c	0.95	0.09	47.99	0.91	0.95	0.1	50.85	0.9
EC	ANFIS	c	0.94	0.11	72.24	0.89	0.96	0.07	70.42	0.92
	LS-SVM	a	0.94	0.12	67.53	0.88	0.95	0.1	70.43	0.9
	GEP	a	0.94	0.12	66.75	0.88	0.95	0.09	48.2	0.9
TH	ANFIS	b	0.92	0.16	27.6	0.84	0.94	0.11	24.2	0.89
	LS-SVM	b	0.92	0.15	26.3	0.85	0.92	0.16	25.9	0.84
	GEP	b	0.9	0.19	30.93	0.8	0.9	0.19	29.15	0.81

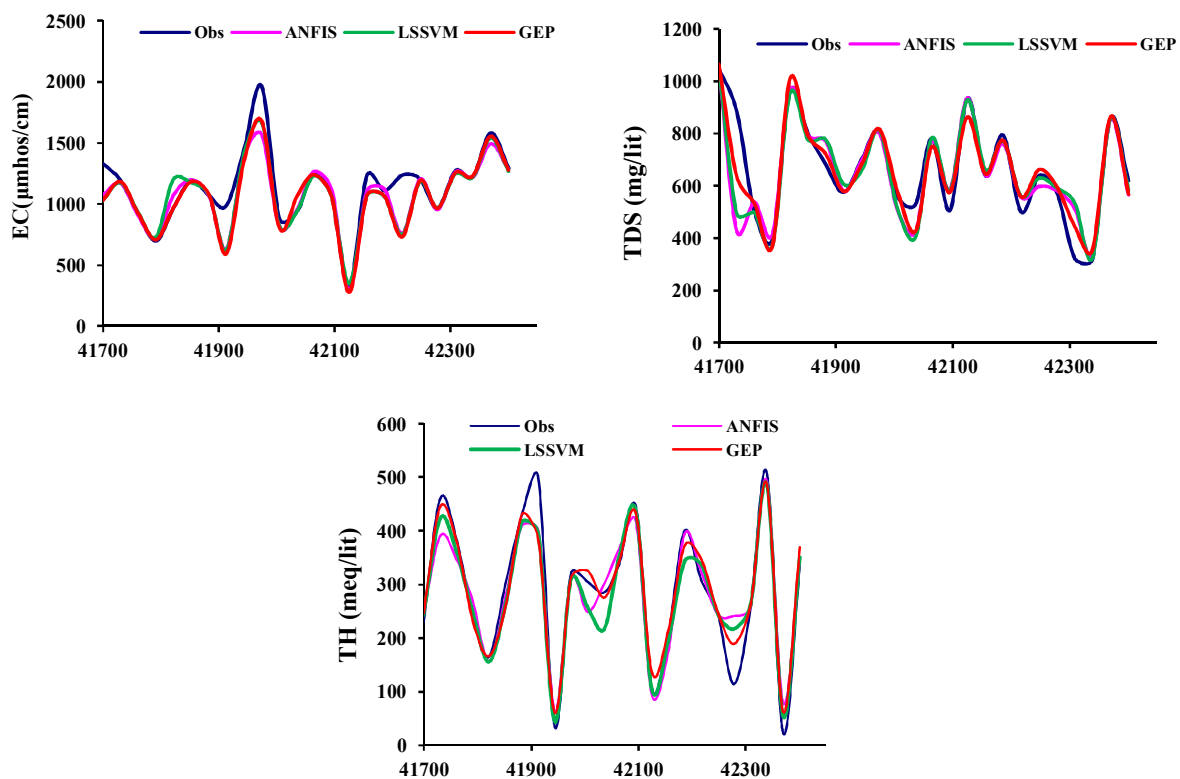


Fig. 2. Comparison between the observed data and estimated values by GEP, LS-SVM and ANFIS methods

شکل ۲- مقایسه بین داده‌های مشاهده‌ای و مقدار تخمین زده شده توسط روش‌های ANFIS و GEP, LS-SVM

مربوط به آن در دیگرام پایین تر قرار بگیرد) روش برتر است و در صورتی که در یک سطح قرار بگیرند روشی که RMSD کمتری داشته باشد و انحراف معیار آن با انحراف معیار داده‌های مشاهده‌ای نزدیک تر باشد (علامت آن در سمت راست علائم سایر روش‌ها

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

دیگرام تیلور نیز بیانگر صحت نتایج است. در دیگرام تیلور، روشی که بیشترین همبستگی را با مشاهدات نشان دهد (علامت



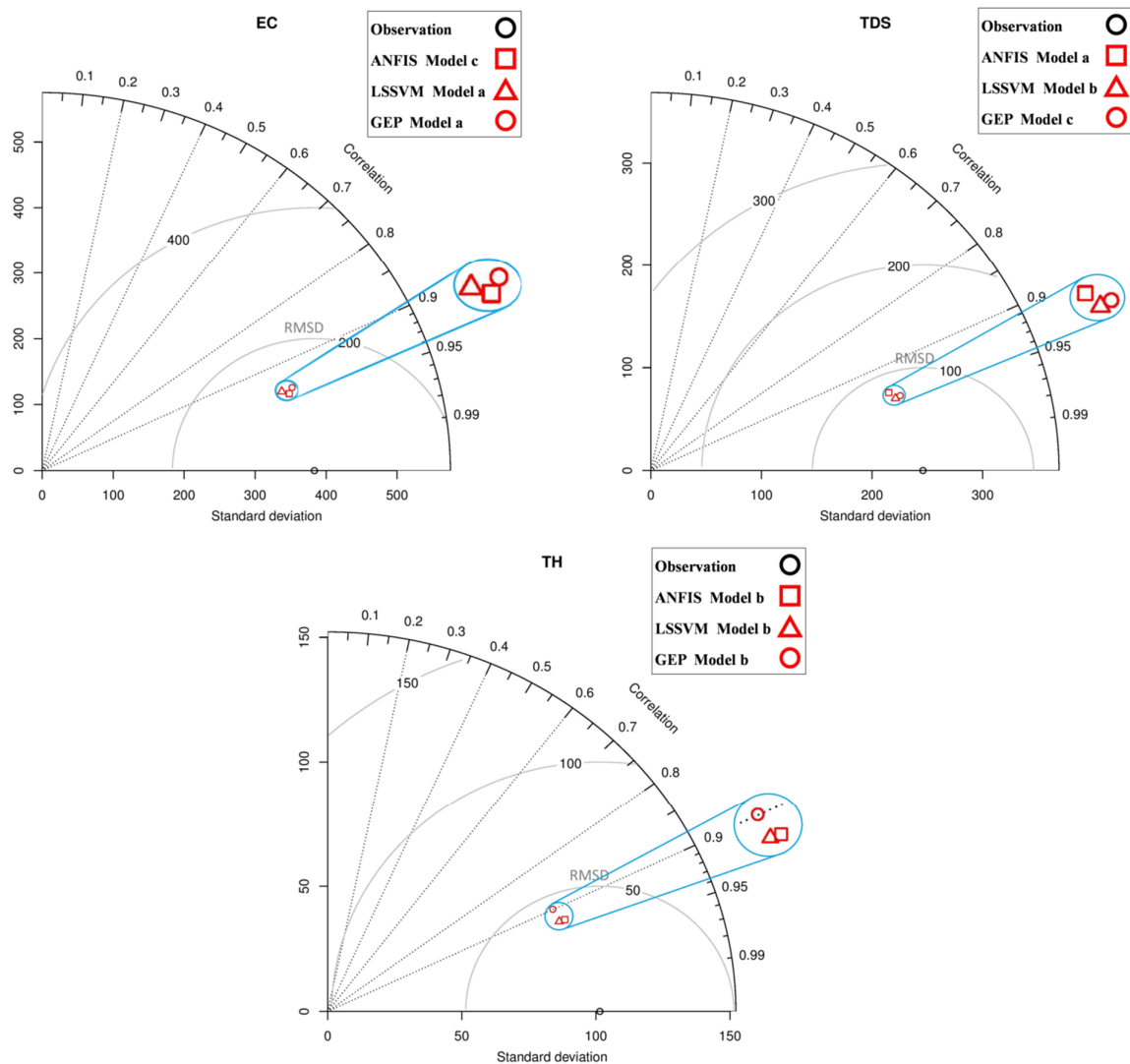


Fig. 3. Comparison between results of GEP, LS-SVM and ANFIS methods by the Taylor diagram

شکل ۳- مقایسه بین نتایج روش‌های GEP, LS-SVM و ANFIS به وسیله دیاگرام تیلور

اندازه‌گیری شده سایر پارامترهای کیفی است. این پارامترها مانند سدیم و کلسیم وابسته به سایر پارامترها نیستند و اندازه‌گیری آنها ساده‌تر از اندازه‌گیری این ۳ پارامتر است. این امر از لحاظ حرفه‌ای باعث کاهش هزینه‌ها می‌شود. نتایج این پژوهش بیانگر این است که این مدل‌ها قادر به تخمین پارامترهای کیفی TH, EC و TDS که اندازه‌گیری و کارهای آزمایشگاهی مربوط به آنها پرهزینه و زمان‌بر هستند، می‌باشند. هر سه روش هوش مصنوعی نام‌برده در پیش‌بینی پارامترهای کیفی TH, EC و TDS عملکرد قابل قبولی ارائه دادند.

قرار گیرد) به عنوان روش برتر انتخاب می‌شود.

۴- نتیجه‌گیری

در پژوهش انجام شده به مقایسه سه روش هوش مصنوعی ANFIS, LS-SVM و GEP برای تخمین پارامترهای کیفی TDS, EC و TH رودخانه سفیدرود پرداخته شد. برای تخمین این ۳ پارامتر که متشکل از چند پارامتر هستند و نمونه‌گیری و اندازه‌گیری آنها با مشکل انجام می‌شود، روش مرسوم، استفاده از مقدار



بهبود وضعیت کیفی آب شرب و کشاورزی این رودخانه باشد.

۵- قدردانی

نویسندگان این پژوهش، از ریاست و اعضای هیئت علمی دانشکده مهندسی عمران و معماری، دانشگاه شهید چمران اهواز که حمایت‌های معنوی برای انجام این پژوهش به عمل آوردند، تشکر می‌کنند.

ارائه معادله ریاضی بین متغیر وابسته و متغیرهای مستقل توسط مدل GEP از مهم‌ترین مزیت‌های این مدل است و برای پیش‌بینی ماه‌های آینده اهمیت بسزایی دارد. بنابراین مزیت ذکر شده، GEP را نسبت به دو روش دیگر متمایز می‌کند.

معادله ریاضی استخراج شده فقط برای پیش‌بینی پارامترهای کیفی EC، TDS و TH ایستگاه آستانه رودخانه سفیدرود کاربرد دارد و برای سایر رودخانه‌های کشور باید مجدداً محاسبه شود. این پژوهش می‌تواند گامی مؤثر در اتخاذ تصمیمات مدیریتی برای

References

- 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.
- 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.
- 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.
- Cheng, Q., Huang, Q., Khan, S., Liu, Y., Liao, Z., Li, G. et al. 2016. Adsorption of Cd by peanut husks and peanut husk biochar from aqueous solutions. *Ecological Engineering*, 87, 240-245.
- Ferreira, C. 2001. Gene expression programming: a new adaptive algorithm for solving problems. *Complex Systems*, 13(2), 87-129.
- Haghiabi, A. H., Nasrolahi, A. H. & Parsaie, A. 2018. Water quality prediction using machine learning methods. *Water Quality Research Journal*, 53(1), 3-13.
- Jang, J. S. 1993. ANFIS: adaptive-network-based fuzzy inference system. *IEEE Transactions on Systems, Man and Cybernetics*, 23(3), 665-685.
- 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.
- 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.
- Mellit, A., Pavan, A. M. & Benghanem, M. 2013. Least squares support vector machine for short-term prediction of meteorological time series. *Theoretical and Applied Climatology*, 111(1-2), 297-307.
- 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.
- Ravansalar, M. & Rajaei, T. 2015. Evaluation of wavelet performance via an ANN-based electrical conductivity prediction model. *Environmental Monitoring and Assessment*, 187(6), 366.
- Suykens, J. A. & Vandewalle, J. 1999. Least squares support vector machine classifiers. *Neural Processing Letters*, 9(3), 293-300.

