



Estimation of the Qualitative Parameters of the Tajan River in Mazandaran Province Using Supporting Vector Machines and Expression Gene Programming Models

Document Type
Research Paper

Received
2022/05/01

Accepted
2023/04/15

Mehdi Yousefi¹, Samad Imamqolizadeh², Hadi Ghorbani³, Mehboobe Vanak⁴

- 1 Master's student in Soil Science Engineering, Department of Water and Soil, Faculty of Agriculture, Shahrood University of Technology, Iran
- 2 Professor, Department of Water and Environment, School of Civil Engineering, Shahrood University of Technology, Iran
- 3 Associate Professor, Department of Water and Soil, Faculty of Agriculture, Shahrood University of Technology, Iran
- 4 Students of the Department of Water and Environment, Faculty of Civil Engineering, Shahrood University of Technology, Iran

DOI: 10.22034/eiap.2023.169995

Abstract

Due to the importance of predicting and monitoring river water quality parameters, in the present study, the capability of two models of support vector machines (SVM) and gene expression planning methods (GEP) used to estimate two water quality parameters namely the TDS and SAR in the Tajan River. To estimate the TDS and SAR, parameters which the measurement of these parameters are easy and has low cost, were used as input data to the model. These parameters included the electrical conductivity of water, acidity, sodium, calcium, potassium, the adsorption ratio of sodium, magnesium, chlorine, sulfate, bicarbonate, and river flow. SPSS was used to determine the correlation between independent and dependent variables. Based on data analysis using step-by-step method, different scenarios of combining input data for TDS and SAR prediction were considered. Comparison of the results using statistical criteria showed that the use of three variables, EC, SO₄ and SAR in estimating TDS in Tajan river, had the highest correlation coefficient and the lowest error rate. The Na and EC variables also give the best results in estimating the river SAR parameter. In general, the obtained results show the high performance of gene expression planning method in estimating the quality parameters of the Tajan river and therefore it can be used to predict the qualitative parameters of rivers.

Keywords: Gene expression programming (GEP), Supporting Vector Machines (SVM), water quality, Tajan river, sodium absorption ratio (SAR), Total dissolved solids (TDS)

* Corresponding author

Email: s_gholizadeh517@Shahroodut.ac.ir

Extended abstract

Introduction

Although there is a lot of water on the earth planet, the amount of water that can be used in drinking, industry and agriculture sectors; compared to the total volume of water available, it is very small. Quantitatively, the volume of water in the earth's hydrological cycle is almost constant, but due to the degradation of chemical quality, the volume of usable water is decreasing (Hesami Afshar et al., 2013). As a result, studying the quality of water resources is one of the most important priorities for environmental protection and one of the important factors in achieving sustainable development in most societies (Azermdel et al., 2014). In recent years, the use of artificial intelligence techniques, including artificial neural network, neurofuzzy, gene expression programming, and support vector machine, has increased to predict and estimate various parameters (Sattari et al., 2013). Many studies have been done in this field, for example, we can refer to the study (Musavi-Jahromi & Golabi, 2008) that in this study, the efficiency of artificial neural network for simulating the quality parameters of SAR, TDS and EC for some stations of the Karon River with more accuracy 90% were investigated.

Materials and methods

Case study area

The area studied in this research is the Tajen watershed, which has an area of about 4147.22 square kilometers and its average slope is 0.85%. This watershed is located in the geographical position of 53°7' E to 53°42' E and 36°17' N to 36°56' N. Tajen river passes through the city of Sari. This river enters the Caspian coast in Sari city and flows into the Caspian Sea in Farah Abad district (Fig. 1).

Research Methods

In the present study, the capability of two models, support vector machines (SVM) and gene expression programming model (GEP) were evaluated to estimate two water quality parameters namely TDS and SAR in Tajen River. To estimate TDS and SAR, easily measurable parameters were used as input parameters to these models. These parameters included water electrical conductivity, acidity, sodium, calcium, potassium, absorption ratio of sodium, magnesium, chlorine, sulfate, bicarbonate and river flow. Information on 12 monthly water quality variables for a period of 40 years (1977-2017) was collected from the regional water company of Mazandaran province, and in this regard, 80% of the data was assigned to the training stage and 20% of the data was assigned to the test stage. Also, in order to evaluate the effectiveness of the method used to predict TDS and SAR, three statistical indices including correlation coefficient (R^2), root mean square error (RMSE) and mean absolute value of errors (MAE) were used.

Results and discussion

Based on the regression analysis between independent and dependent variables using Stepwise Regression method, 5 different patterns of inputs were considered for TDS estimation of Tajen River. The desired and optimal model is the model that has the highest correlation coefficient and the lowest amount of error. The results showed that the GEP model with an explanation coefficient higher than 0.98 had a good ability to predict TDS in all the models used. But in comparing the models with each other, model 5, in which Ec, SO₄, SAR, HCO₃ and CL parameters were used as input parameters, having high coefficient, RMSE and MAE equal to 0.98, 17.2 and 54 6.0 in the training phase and 0.9927, 9.5, 6.83 in the test phase were better than other models. In order to estimate the absorption ratio of sodium (SAR) based on Stepwise Regression method, 4 input patterns were considered. According to the results, the value of R^2 for all models was more than 0.97 and also the maximum RMSE error was equal to 0.101 and 0.131 in the training and testing phase. The comparison of the obtained results showed that model 2 with input parameters Na and EC is the optimal model with the highest correlation coefficient equal to 0.987 in training and 0.989 in testing and the least error and also having low input parameters. The results indicate that the desired model has a high ability to estimate SAR by having a higher correlation coefficient, and in other words, it provides reliable results in

predicting sodium absorption ratio. In estimating the total dissolved solids (TDS) parameter, using model 1, having the lowest input parameter, has the highest correlation coefficient and the lowest error rate, so the same model was used in the implementation of the support vector machine model to compare the two models. Model 2 was also used in the estimation of sodium absorption ratio (SAR). The support vector machine model was implemented using the two mentioned patterns and the best result was obtained. The obtained results indicate that the SVM model with an explanation coefficient greater than 0.94 and 0.91 in the test phase has performed well for the TDS and SAR parameters, respectively. Also, the RMSE values for TDS and SAR parameters were obtained as 14.76 and 0.076, respectively. Although based on statistical criteria, the SVM model has acceptable accuracy, but compared to the GEP model, the obtained results indicate that the performance of the SVM model has decreased. In other words, the error value of RMSE and MAE has increased and the explanation coefficient R^2 has decreased.

Conclusion

In this research, for the TDS parameter, 5 models were used with the input variables of EC, SO₄, SAR, HCO₃, CL, and for the SAR parameter, 4 models were used that included the input variables of EC, SO₄, SAR, HCO₃, CL. Results showed that, in all the models used, the GEP model has the ability to predict the TDS parameter, so that as expected, with the increase in the number of input variables, because the number of degrees of freedom increases, the performance of the model in estimating the TDS parameter has increased. On the other hand, in order to reduce the costs, it is necessary to reduce the number of input variables, and in other words, the optimal model is a model that has a smaller number of input variables but relatively high model accuracy. The results of the implementation of the GEP model with different options of the input patterns showed that in the TDS estimation of the Tajan River, when using the model 1 in which only the EC parameter is used, the model with R^2 , RMSE and MAE equal to 0.981, 17.17 and 6.78 have relatively good accuracy. Also, model 2 with two input variables including Na and EC was selected as the superior model in estimating the sodium absorption ratio (SAR) of Tajan River. In this model, the values of R^2 , RMSE and MAE in the test phase were equal to 0.989, 0.069 and 0.043. Also, the implementation of the SVM model showed that this model has acceptable accuracy in estimating two parameters, TDS and SAR, but compared to the GEP model, the obtained results indicate that the performance of the SVM model has decreased.

Reference

- Hesami - Afshar, M.; Khalili, K.; Ali Golinia, T. & Zamanzad Gavidel, S. 2013. Prediction of TDS concentration of Zarinerood river based on the least number of input parameters with the help of genetic programming. The third international conference on environmental planning and management, University of Tehran (in persian).
- Azarmdel, H.; Mostafazadeh, R. & Ghasemi, A. 2010. Assessment of Water Quality Monitoring Stations in Gorgan-Roud River, Golestan Province. *Watershed Management Science & Engineering*. 4(10):57-61(in persian)
- Sattari, M.T.; Pal, M.; Apaydin, H. & Ozturk, F. 2013. M5 model tree application in daily river flow forecasting in Sohu Stream, Turkey. *Water Resources Management*. 40(3): 233-242
- Musavi-Jahromi, S. & Golabi, M. 2008. Application of artificial neural networks in the river water quality modeling: Karoon River, Iran. *Journal of Applied Sciences*. 8(12): 2324-2328.

برآورد پارامترهای کیفی رودخانه تجن مازندران با استفاده از مدل‌های ماشین‌های بردار پشتیبان و مدل برنامه‌ریزی بیان ژن

مهدی یوسفی^۱، صمد امام‌قلی زاده^{۲*}، هادی قربانی^۳، محبوبه ونک^۴

۱ دانشجوی کارشناسی ارشد مهندسی خاکشناسی، گروه آب و خاک دانشکده کشاورزی، دانشگاه صنعتی شاهرود، ایران

۲ استاد گروه آب و محیط زیست دانشکده مهندسی عمران، دانشگاه صنعتی شاهرود، ایران

۳ دانشیار گروه آب و خاک دانشکده کشاورزی، دانشگاه صنعتی شاهرود، ایران

۴ دانشجوی گروه آب و محیط زیست دانشکده مهندسی عمران، دانشگاه صنعتی شاهرود، ایران

تاریخ پذیرش مقاله: ۱۴۰۱/۱۰/۰۵

تاریخ وصول مقاله: ۱۴۰۱/۰۲/۱۱

چکیده

به دلیل اهمیت پیش‌بینی و پایش پارامترهای کیفی آب رودخانه‌ها، در پژوهش حاضر قابلیت دو مدل ماشین‌های بردار پشتیبان (SVM) و مدل برنامه‌ریزی بیان ژن (GEP) برای برآورد دو پارامتر کیفی آب یعنی TDS و SAR در رودخانه تجن مورد ارزیابی قرار گرفت. برای برآورد TDS و SAR از پارامترهای زودیافت یا به عبارت بهتر پارامترهایی که اندازه‌گیری آن‌ها آسان و کم هزینه بوده است به عنوان پارامترهای ورودی به مدل استفاده شد. این پارامترها شامل هدایت الکتریکی آب، اسیدیته، سدیم، کلسیم، پتاسیم، نسبت جذب سدیم، منیزیم، کلر، سولفات، بیکربنات و دی‌رومید بودند. جهت تعیین همبستگی بین متغیرهای مستقل و وابسته از نرم‌افزار SPSS استفاده شد. بر اساس آنالیز داده‌ها با استفاده از روش گام به گام (step by step)، سناریوهای مختلفی از ترکیب داده‌های ورودی برای پیش‌بینی TDS و SAR در نظر گرفته شد. مقایسه نتایج به دست آمده نشان داد استفاده از سه متغیر SO_4 ، EC و SAR در برآورد TDS در رودخانه تجن، دارای بالاترین ضریب همبستگی و کمترین میزان خطا بوده است. همچنین در برآورد پارامتر SAR استفاده از دو پارامتر Na و EC بهترین نتیجه را داده است. در مجموع نتایج به دست آمده نشان‌دهنده عملکرد بالای روش برنامه‌ریزی بیان ژن در مقایسه با مدل ماشین‌های بردار پشتیبان در برآورد پارامترهای کیفی رودخانه‌های تجن بوده است و بنابراین از مدل می‌توان جهت پیش‌بینی پارامترهای کیفی رودخانه‌ها مورد استفاده قرار داد.

کلید واژه‌ها: برنامه‌ریزی بیان ژن، ماشین‌های بردار پشتیبان، متغیر کیفی آب، رودخانه تجن، نسبت جذب سدیم، کل مواد محلول

سرآغاز

آب در کره زمین بسیار زیاد می‌باشد، اما حجم آبی که قابلیت استفاده در بخش‌های شرب، صنعت و کشاورزی را دارا باشد؛ در مقایسه با حجم کل آب بسیار ناچیز است. به لحاظ کمی، حجم آب در چرخه هیدرولوژی زمین تقریباً ثابت است ولی به دلیل تنزل کیفیت شیمیایی، حجم آب قابل استفاده در حال کم شدن می‌باشد (Hesami Afshar et al., 2013). در نتیجه مطالعه کیفیت منابع آب از مهم‌ترین اولویت‌های حفاظت محیط زیست و از عامل‌های مهم در دستیابی به توسعه پایدار در بیشتر جوامع است (Azermadel et al., 2014). در زمینه مدیریت بهتر برای حفظ کیفیت آب، مدل‌های بسیار زیادی گسترش یافته است. بیشتر این مدل‌ها نیازمند پارامترهای ورودی بسیاری هستند که یا دسترسی به آنها مشکل است و یا اینکه اندازه‌گیری آنها محتاج صرف هزینه و زمان زیادی می‌باشد (Dolat-Kordestani et al., 2017). در سال‌های اخیر استفاده از تکنیک‌های هوش مصنوعی شامل شبکه عصبی مصنوعی، نرو فازی، برنامه‌ریزی بیان ژن و ماشین پشتیبان‌بردار برای پیش‌بینی و برآورد پارامترهای مختلف زیاد شده است (Sattari et al., 2013). مطالعات زیادی در این زمینه انجام شده است که برای مثال می‌توان به مطالعات (Diamantopoulou et al., 2005) اشاره نمود که از مدل شبکه عصبی مصنوعی جهت پیش‌بینی پارامترهای کیفیت آب رودخانه اکیسوس (شمال یونان) استفاده کردند. نتایج آن‌ها نشان داد که شبکه عصبی مصنوعی برای پیش‌بینی پارامترهای کیفیت آب بسیار مناسب بوده و همچنین یک ابزار کارا برای برآورد مقادیر اندازه‌گیری نشده یا مفقوده می‌باشد (Musavi-Jahromi & Golabi, 2008). در پژوهشی کارایی شبکه عصبی مصنوعی برای شبیه‌سازی پارامترهای کیفی TDS، SAR و EC برای برخی ایستگاه‌های رودخانه کارون با دقت بیش از ۹۰ درصد بررسی کردند.

(Emamgholizadeh et al., 2014) با استفاده از دو مدل شبکه عصبی مصنوعی شامل پرسپترون چند لایه^(۱) (MLP) و تابع شعاعی^(۲) (RBF) و همچنین از سیستم استنتاج فازی-عصبی تطابقی^(۳) (ANFIS) برای پیش‌بینی اکسیژن محلول (DO)، میزان اکسیژن بیوشیمیایی (BOD) و میزان اکسیژن شیمیایی (COD) در رودخانه کارون استفاده کردند. آن‌ها از ۹ متغیر ورودی کلسیم، سدیم، منیزیم، نیتريت، نترات، فسفات،

EC، pH و کدورت جهت آموزش و صحت سنجی مدل استفاده کردند. نتایج به دست آمده نشان داد عملکرد مدل MLP بهتر از دو مدل RBF و ANFIS بوده است.

(Mojaradi et al., 2018) در پژوهشی به منظور برآورد شاخص NSFQI و تعیین ارتباط بین پارامترهای کیفی آب و شاخص کیفیت آب رودخانه تالار از روش‌های برنامه‌ریزی بیان ژن و شبکه‌های عصبی مصنوعی استفاده کردند و نتایج به دست آمده نشان داد که دقت شبکه‌های عصبی مصنوعی در تخمین شاخص NSFQI بیشتر از روش برنامه‌ریزی بیان ژن است.

(Bozorg-Haddad et al., 2017) به منظور مدل سازی پارامترهای کیفیت آب رودخانه سفید رود از روش‌های برنامه‌نویسی ژنتیک (GP) و رگرسیون وکتور حداقل مربعات (LSSVR) استفاده کردند و نتایج پژوهش نشان داد که روش رگرسیون حداقل مربعات از دقت بیشتری برخوردار است. (Karami & Golabi, 2012) در پژوهشی به منظور شبیه سازی و پیش‌بینی پارامترهای کیفی آب با استفاده از شبکه عصبی مصنوعی نشان دادند که در شبیه سازی TDS بیشترین تاثیر به ترتیب مربوط به SO_4 ، Ca، K، Na و Cl و کمترین ناشی از Q و CO_3 بوده است. در شبیه سازی EC بیشترین تاثیر به ترتیب مربوط به Ca، Cl و SO_4 کمترین ناشی از، HCO_3 ، K، Q، Mg و CO_3 بوده است. در شبیه سازی SAR بیشترین تاثیر به ترتیب مربوط به Ca، Na و Mg و کمترین ناشی از K و CO_3 بوده که با توجه به فرمول محاسبه SAR نتیجه به دست آمده قابل قبول می‌باشد. (Rezaei et al., 2019) در پژوهشی به منظور برآورد میزان مواد جامد محلول در آب رودخانه کشکان به ارزیابی مدل‌های هوشمند ماشین بردار پشتیبان، برنامه‌ریزی بیان ژن و شبکه عصبی بیزین پرداختند. نتایج نشان داد که مدل ماشین بردار پشتیبان بیشترین دقت را نسبت به مدل‌های دیگر دارد.

(Mehdipour et al., 2017) در پژوهشی برای پیش‌بینی غلظت اکسیژن محلول در آب دریاچه اومیر از برنامه‌ریزی بیان ژن استفاده کردند و با استفاده از داده‌های فسفر، نترات، فسفات، نیتروژن محلول، pH، دمای آب، کلروفیل محلول، هدایت الکتریکی، سرعت بارش، سرعت باد و دمای محیط به عنوان عوامل مستقل در پیش‌بینی اکسیژن محلول استفاده شدند. نتایج به دست آمده نشان داد مدل با ضریب تبیین (۰/۸۰۳۱) و

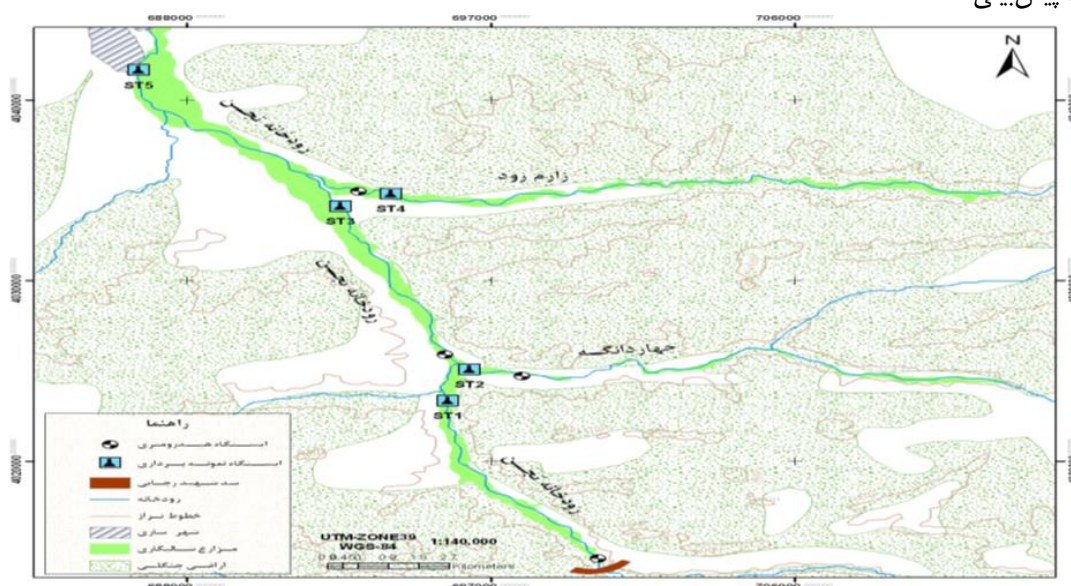
مواد و روش ها

منطقه مورد مطالعه

حوضه آبریز تجن حدود ۴۱۴۷/۲۲ کیلومتر مربع وسعت دارد و شیب متوسط آن ۰/۸۵ درصد است. این حوضه آبریز در موقعیت جغرافیایی ۵۳ درجه و ۷ دقیقه الی ۵۳ درجه و ۴۲ دقیقه طول شرقی و ۳۶ درجه و ۵۶ دقیقه و ۳۶ درجه و ۱۷ دقیقه عرض شمالی قرار دارد. رودخانه تجن از میان شهر ساری عبور می‌نماید. این رودخانه در محل شهرستان ساری وارد پهنه ساحلی خزر شده و در ناحیه فرج آباد به دریای خزر می‌ریزد. زیست بوم رودخانه تجن در نیم قرن اخیر به واسطه قرار گرفتن مرکز استان مازندران در میانه مسیر آن همواره مورد تعرض جدی قرار داشته است. پاک تراشی جنگل در بخش دودانگه (محمدآباد و فریم صحرا) و متعاقب آن فرسایش شدید خاک و حمل آن به بخش پایین دست اراضی جنگلی، احداث سد سلیمان تنگه، احداث و رونق نسبی جاده کیاسر- شه میرزاد، توسعه اراضی کشاورزی و چرای بی رویه دام در مناطق نامبرده، توسعه سریع و غیراصولی شهرستان ساری در سنوات اخیر، دفع غیراصولی زباله‌های شهری در حوضه آبریز رودخانه تجن و بلاخره احداث کارخانه چوب و کاغذ مازندران که سالانه چند صد هزارتن پساب تولید کاغذ را مستقیماً و بدون هرگونه فیلتراسیون به داخل رودخانه واریز می‌کند، گوشه‌ای از بحران آلودگی محیط‌زیستی رودخانه تجن در سال‌های اخیر می‌باشند (Louisa, 2014).

میانگین مربعات خطا (۰/۰۹۳۷) عملکرد خوبی در پیش‌بینی غلظت اکسیژن محلول داشته است. (Najafzadeh et al., 2019) با استفاده از برنامه‌ریزی بیان ژن (GEP)، رگرسیون چند متغیره تکاملی (EPR) و درخت تصمیم (MT) برای برآورد سه شاخص اکسیژن بیوشیمیایی (BOD)، اکسیژن محلول و میزان اکسیژن شیمیایی (COD) رودخانه کارون پرداختند. آن‌ها از ۹ متغیر ورودی کلسیم، سدیم، منیزیم، نیتريت، نترات، فسفات، pH، EC و کدورت جهت آموزش و صحت سنجی مدل استفاده کردند. بررسی عملکرد مدل‌ها نشان داد که مدل EPR دارای برتری نسبی در مقایسه با مدل‌های GEP و MT بوده است.

همچنین از مدل ماشین بردار پشتیبان در برآورد پارامترهای کیفی آب رودخانه‌ها، و آب زیرزمینی استفاده شده است. برای مثال (Esa-Zadeh et al., 2018) با استفاده از مدل ماشین بردار پشتیبان و شبکه عصبی مصنوعی به تخمین پارامترهای کیفی آب خوان دشت گیلان پرداختند. نتایج نشان داد که هر یک از دو مدل SVM و ANN توانایی بسیار زیادی در تخمین پارامترهای کیفی آب خوان‌ها دارند. همچنین عملکرد مدل SVM نسبت به مدل ANN، به ازای تعداد ورودی کمتر بهتر است و در تعداد ورودی بیشتر برعکس می‌باشد. (Kamyab- Talesh et al., 2019) از مدل ماشین بردار پشتیبان (SVM) برای پیش‌بینی شاخص کیفیت آب در رودخانه سفیدرود واقع در شمال کشور استفاده کردند. مدل SVM توانست با ضریب تبیین و خطای میانگین مربعات ریشه برابر با ۰/۸۷ و ۰/۶۱ شاخص کیفیت آب را پیش‌بینی کند.



شکل (۱): محدوده مورد مطالعه رودخانه تجن

داده مورد استفاده تحقیق

به منظور برآورد پارامترهای کیفی آب رودخانه تجن به کمک روش برنامه‌ریزی بیان ژن، اطلاعات ۱۲ متغیر کیفی ماهانه آب برای یک دوره ۴۰ ساله (۱۳۹۶-۱۳۵۶) از شرکت آب منطقه‌ای استان مازندران جمع‌آوری شد. نتایج آماری مربوط به میانگین متغیرهای مورد مطالعه در طول دوره آماری موردنظر در جدول (۲) ارائه شده است. به منظور برآورد پارامترها با استفاده از

برنامه‌ریزی بیان ژن ۸۰٪ داده‌ها را به مرحله آموزش و ۲۰٪ داده‌ها را نیز به مرحله آزمون اختصاص داده شد. به منظور بررسی کفایت داده‌های مورد استفاده در این تحقیق از آزمون ماکوس (Mackus) استفاده گردید. نتایج بررسی با استفاده از این آزمون نشان داد که تعداد داده‌ها در سطح آماری ۹۵ درصد مورد قابل قبول می‌باشند.

جدول (۱): پارامترهای آماری اندازه‌گیری‌های کیفی بر روی رودخانه تجن

TDS mg/lit	SAR	K	na	Mg	Ca	So4	Cl	Hco3	pH	Ec Micromohs/cm	Q m ³ /s	پارامتر
۴۴۲/۳۷	-/۱۸۷۰	۳/۰۵	۱/۴۴	۲/۱۵	۳/۲۲	۱/۴۹	۱/۲۸	۳/۹۶	۷/۷۹	۶۸۶/۱۲	۱۲/۳۴	میانگین
۱۲۱/۲۳۶	-/۶۴۷	-/۰۲۱۹	۱/۱۹	-/۶۷۳	-/۷۸۹	-/۷۳۰	۱/۰۶۱	-/۹۷۴	-/۲۹۹	۱۸۵/۴۴	۲۹/۰۲۵	انحراف معیار
۱۴۶۹۸/۲	-/۴۱۹	۰	۱/۴۱	-/۴۵۴	-/۶۲۰	-/۵۳۰	-/۱۲۷	-/۹۵۰	-/۰۹۰	۳۴۳۹۰/۶	۸۴۲/۴۷	واریانس
۳۰	-/۰۴۰	-/۰۱۰	-/۰۸۰	-/۵۰	۱	-/۱۰۰	-/۱۵۰	۱/۴۰	۶/۶۰	۲۴۰	-/۰۱	حداقل
۱۵۶۰	۷/۸۱	-/۳۶۰	۱۶/۲۰	۴/۷۰	۷/۴۰	۶/۰	۱۵/۵۰	۹/۸۰	۸/۹۰	۲۴۰۰	۴۸۹/۴۱	حداکثر

همبستگی خوبی با متغیر وابسته بود، به عنوان پارامتر ورودی برای مدلسازی مورد استفاده قرار گرفت. جدول (۴) ترکیب در نظر گرفته شده را نشان می‌دهد. به عبارتی ۵ و ۴ ترکیب به ترتیب جهت پیش‌بینی TDS و SAR مورد استفاده قرار گرفت.

جهت تعیین همبستگی بین متغیرهای مستقل و وابسته از روش گام به گام (Stepwise Regression) با استفاده از نرم افزار SPSS23 استفاده گردید. جدول (۳) نتایج به دست آمده را نشان می‌دهد. بر اساس نتایج این جدول و با حذف پارامترهایی که وابستگی نسبت به یکدیگر داشتند و یا این که پارامتری که دارای

جدول (۲): ضریب همبستگی متغیرهای مورد مطالعه رودخانه تجن

TDS	SAR	K	Na	Mg	Ca	SO ₄	Cl	HCO ₃	pH	EC	Q		
											۱	Q	
										۱	-/۰۲۶۴**	EC	
									۱	-/۰۱۲۴**	-/۰۰۶۴*	pH	
								۱	-/۰۲۲۳**	-/۰۶۹۴*	-/۰۱۷۵**	HCO ₃	
							۱	-/۰۳۸۱**	-/۰۰۴۱*	-/۰۸۲۶**	-/۰۱۵۵**	CL	
						۱	-/۰۱۰۱**	-/۰۱۴۰*	-/۰۰۷۲**	-/۰۳۵۵**	-/۰۲۰۹**	SO ₄	
					۱	-/۰۱۵۱**	-/۰۳۳۳**	-/۰۵۶۷**	-/۰۲۰۱**	-/۰۴۵۸**	-/۰۲۱۷**	CA	
				۱	-/۰۱۲**	-/۰۴۱۶**	-/۰۳۹۲**	-/۰۴۷۰**	-/۰۱۶۴**	-/۰۵۶۷**	-/۰۱۹۷**	MG	
			۱	-/۰۳۱۳**	-/۰۱۸۹**	-/۰۲۴۳**	-/۰۸۸۱**	-/۰۴۰۶**	-/۰۰۱۷**	-/۰۸۲۹**	-/۰۱۴۳**	NA	
		۱	-/۰۴۲۹**	-/۰۰۴۹**	-/۰۳۳۶**	-/۰۰۳۰**	-/۰۴۵۴**	-/۰۲۸۳**	-/۰۰۹۷**	-/۰۴۵۸**	-/۰۱۳۷**	-/۰۵۶۷**	K
	۱	-/۰۴۰۶**	-/۰۹۸۶**	-/۰۳۴۶**	-/۰۰۹۰**	-/۰۲۱۴**	-/۰۸۳۵**	-/۰۳۲۸**	-/۰۰۱۲**	-/۰۷۵۱**	-/۰۱۳۷**	-/۰۱۳۷**	SAR
۱	-/۰۷۴۹**	-/۰۴۴۸**	-/۰۸۲۵**	-/۰۵۶۷**	-/۰۵۳۷**	-/۰۳۶۳**	-/۰۸۱۵**	-/۰۶۸۷**	-/۰۱۳۱**	-/۰۹۹۱**	-/۰۲۵۹**	-/۰۵۳۷**	TDS

جدول (۳): الگوهای ورودی مورد استفاده جهت برآورد TDS

پارامتر(های) ورودی		شماره الگوهای ورودی
SAR	TDS	
Na	EC	۱
Na, EC	EC, SO ₄	۲
Na, EC, CL	EC, SO ₄ , SAR	۳
Na, EC, CL, Q	EC, SO ₄ , SAR, HCO ₃	۴
	EC, SO ₄ , SAR, HCO ₃ , CL	۵

ماشین‌های بردار پشتیبان

ماشین‌های بردار پشتیبان^(۴) (SVM) نیز یکی از روش‌های هوش مصنوعی و در واقع نوعی الگوریتم داده‌کاوی است. روش ماشین بردار پشتیبان بر پایه تئوری یادگیری آماری که از حداقل سازی ریسک ساختاری پیروی می‌کند استوار است (Cortes & Vapnik, 1995). در یک مدل رگرسیونی SVM لازم است وابستگی تابعی متغیر وابسته y به مجموعه‌ای از متغیرهای مستقل x تخمین زده شود. فرض بر این است که مانند دیگر مسایل رگرسیونی، رابطه بین متغیرهای وابسته و مستقل توسط یک تابع معین f به علاوه یک مقدار اضافی نویز^(۵) مشخص می‌شود (رابطه ۱).

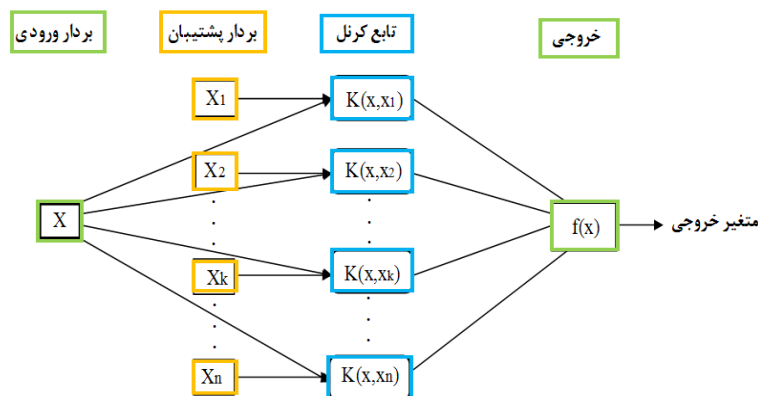
$$y = f(x) + \text{Noise} \quad (1)$$

بنابراین موضوع اصلی، پیدا کردن فرم تابع f است تا مدل SVM بتواند مواردی را که تاکنون تجربه نکرده است به صورت صحیح

پیش‌بینی کند. این تابع بوسیله آموزش مدل توسط مجموعه‌ای از داده‌ها به عنوان مجموعه آموزش با بهینه‌سازی تابع خطا به دست می‌آید. بر مبنای تعریف این تابع خطا، ۲ نمونه از مدل‌های شناخته شده SVM وجود دارد:

- SVM رگرسیونی نوع ۱ (که به ϵ -SVM یا ϵ -SVM معروف است).
- SVM رگرسیونی نوع ۲ (که به ν -SVM یا ν -SVM معروف است).

بررسی منابع نشان می‌دهد SVM رگرسیونی نوع (ν -SVM) دارای کاربرد بیشتری در مسایل رگرسیونی است و این مدل قادر است از طریق تغییر ابعاد مساله به وسیله توابع کرنل، مسایل غیرخطی را حل نماید (Azamathulla et al., 2018; Parsaie et al., 2019; Emamgholizadeh & Mohammadi, 2021). شکل (۲) ساختار ماشین بردار پشتیبان را نشان می‌دهد.

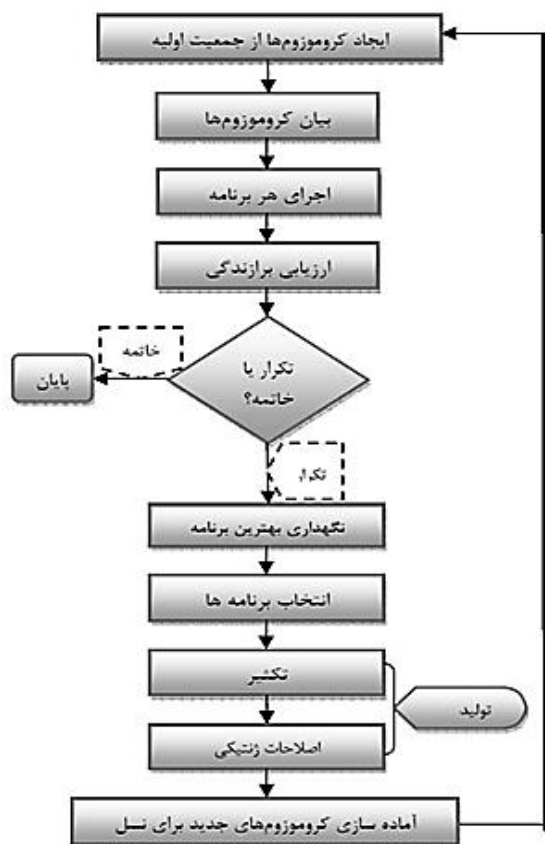


شکل (۲): ساختار ماشین بردار پشتیبان

تطبیقی موجود پیشی بگیرد (Ferreira, 2001). مقدار کمی که نمایش‌دهنده کیفیت یک کروموزوم به عنوان یک جواب از مسئله است و در مقابل هزینه قرار دارد و میزان شایستگی یک کروموزوم را در بین جمعیت کروموزومی نشان می‌دهد، به عنوان تابع برازش شناخته می‌شود. در واقع تابع برازش شاخصی برای انتخاب جفت‌های مناسب و جفت‌گیری آنها در جمعیت کروموزوم‌ها فراهم می‌آورد. فرایند GEP مانند سایر الگوریتم‌ها زمانی خاتمه می‌یابد که به ضابطه تعیین شده‌ای مثل میزان معین خطای میانگین کارایی اعضای جمعیت رسیده باشد و یا چرخه فرآیند به تعداد دلخواه ادامه یافته باشد. در این صورت

برنامه‌ریزی بیان ژن

مدل برنامه‌ریزی بیان ژن توسط فریرار در سال ۱۹۹۹ ابداع شد. در این مدل، اعضا رشته‌های خطی با طول ثابت هستند که سپس به شکل غیرخطی با اندازه و شکل و بیان درختی تعریف می‌شود. از جمله مهمترین مزایای GEP آن است که کروموزوم‌های خطی، فشرده، نسبتاً کوچک و آسان برای عملیات ژنتیکی تعریف شده‌اند. همچنین بیان درختی منحصراً کروموزوم‌های مربوطه که بر اساس سازگاری انتخاب شده‌اند را بیان می‌نماید، ایجاد این اعضای منحصر به فرد این قابلیت را برای الگوریتم فراهم می‌کند که با کارایی بالا، از تکنیک‌های



شکل (۳): الگوریتم برنامه‌ریزی بیان ژن (Ferreira, 2001)

بهترین راه حل یافت شده گزارش داده می‌شود و در غیر این صورت، بهترین راه حل از نسل حاضر نگه داشته می‌شود و بقیه راه حل‌ها به فرایند گزینشی واگذار می‌شوند که براساس آن بهترین افراد شانس بیشتری برای تولید فرزندان دارند. با تکرار روند ذکر شده و با پیشرفت نسل به جلو، کیفیت جمعیت بهبود یافته و به پاسخ بهینه نزدیک خواهد شد (Lopes & Weinret, 2004). در واقع تابع برازش برای ارزیابی عملکرد هر برنامه خروجی بر مبنای خطای نسبی انتخاب می‌شود و آن را به وسیله‌ی یک محدوده انتخابی و میزان دقت کاوش می‌کند. از لحاظ ریاضی، تابع برازش F_i به صورت رابطه (۱) بیان می‌شود که در آن R محدوده انتخابی، P_{ij} مقدار پیش‌بینی شده به وسیله برنامه انفرادی i برای مورد برازش j (از میان n برازش) و T_j مقدار هدف برای مورد برازش J است. عبارت داخل قدر مطلق متناظر با مقدار خطای نسبی است و دقت نامیده می‌شود. چنانچه $P_{ij} = T_j$ باشد، خطا صفر خواهد بود. بنابراین برای یک برازش کامل که در آن خطا صفر باشد، مقدار f_i از یک برنامه انفرادی i برابر با $f_i = f_{\max} = Nr$ خواهد بود (Ferreira, 2001).

$$f_i = \sum_{j=1}^n \left(R - \left| \frac{P_{(ij)} - T_j}{T_j} \cdot 100 \right| \right) \quad (2)$$

شکل (۳) الگوریتم برنامه‌ریزی بیان ژن را نشان می‌دهد.

در پژوهش حاضر از نرم افزار GeneXpro Tools4 جهت برآورد پارامترهای کیفی مورد نظر استفاده قرار گرفت. این مدل توسط محققین مختلفی برای برآورد دبی رسوب (Emamgholizadeh & Karimi 2019)، انتشارپذیری خاک (Emamgholizadeh et al., 2017)، پارامتر ظرفیت تبادل خاک (Emamgholizadeh et al., 2015) مورد استفاده قرار گرفت.

شاخص‌های آماری سنجش توانایی مدل GEP

به منظور ارزیابی و بررسی کارایی روش استفاده شده برای پیش‌بینی TDS و SAR در این تحقیق از سه معیارهای آماری شامل ضریب همبستگی (R^2)، مجذور میانگین مربعات خطا (RMSE) و میانگین قدر مطلق خطاها (MAE) استفاده شده است.

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

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

$$MAE = \frac{1}{N} \left| \sum_{i=1}^N (O_i - P_i) \right| \quad (5)$$

در این روابط، N تعداد داده‌ها، O مقدار مشاهده شده و P مقدار پیش‌بینی شده است. نماد بار نیز دلالت بر میانگین پارامتر مورد نظر دارد.

نتایج و بحث

همان طوری که در بخش مواد و روش‌ها توضیح داده شد بر

استفاده شده، با داشتن ضریب تبیین، RMSE و MAE برابر با ۰/۹۸، ۱۷/۲ و ۶/۵۴ در مرحله آموزش و همچنین ۰/۹۹۲۷، ۹/۵، ۶/۸۳ در مرحله آزمون بهتر از سایر الگوها بوده است. جهت مقایسه بهتر، نمودار پراکنش داده‌های اندازه‌گیری شده و پیش‌بینی شده با مدل GEP در شکل (۴) نشان داده شده است. اگرچه استفاده از الگوی ۵ موجب بهبود دقت پیش‌بینی مدل شده است، ولی نقطه ضعف استفاده از الگو در مقایسه با سایر الگوها به خصوص الگوی ۱ آن است که از پارامترهای بیشتری جهت پیش‌بینی استفاده می‌کند که این امر موجب افزایش هزینه خواهد شد. لذا استفاده از الگوی ۱ با تعداد پارامتر کمتر (فقط پارامتر EC) اگر چه ممکن است مقدار خطا را افزایش دهد، ولی موجب کاهش هزینه اندازه‌گیری جهت پیش‌بینی TDS خواهد شد. بنابراین این الگو با یک پارامتر ورودی را می‌توان جهت پیش‌بینی TDS پیشنهاد داد.

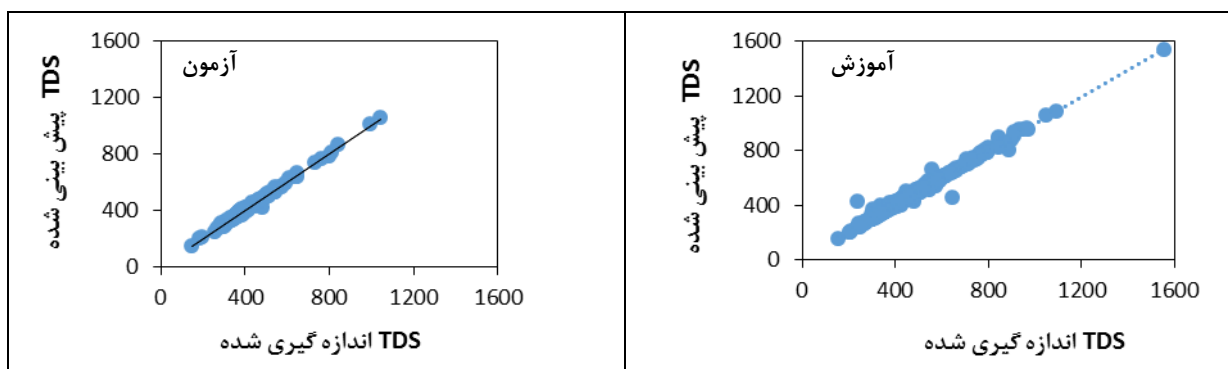
مبنای تحلیل رگرسیون بین متغیرهای مستقل و وابسته با استفاده از روش گام به گام (Stepwise Regression)، ۵ الگوی مختلف از ورودی‌ها برای برآورد TDS رودخانه تجن در نظر گرفته شد. در جدول (۶) مقادیر پارامترهای مورد استفاده برای اجرای مدل GEP در برآورد شاخص TDS و SAR رودخانه تجن آمده است. نتایج حاصل از اجرای روش برنامه ریزی بیان ژن شامل ضرایب همبستگی، جذر میانگین مربعات خطا و متوسط خطای میانگین در هر مرحله آموزش و آزمون در جدول (۷) ارائه گردیده است. مدل مطلوب و بهینه مدلی است که با دارا بودن بیشترین ضریب همبستگی، کمترین میزان خطا را به خود اختصاص دهد. نتایج جدول (۷) نشان می‌دهد مدل GEP با دارا بودن ضریب تبیین بالاتر از ۰/۹۸ توانایی خوبی در پیش‌بینی TDS در تمامی الگوهای مورد استفاده داشته است. اما در مقایسه الگوها با یکدیگر، الگوی ۵ که در آن از پارامترهای SAR، SO₄، Ec، HCO₃ و CL که به عنوان پارامتر ورودی

جدول (۴): مقادیر پارامترهای GEP در برآورد شاخص کیفی TDS و SAR رودخانه‌های تجن

مقدار	عملگرهای زنتیکی	مقدار	تنظیمات کلی
۰/۴۴	نرخ جهش	۳۰	تعداد کروموزوم‌ها
۰/۱	نرخ وارون‌سازی	۸	اندازه راس
۰/۱	نرخ ترانهش درج متوالی	۳	تعداد ژن‌ها در هر کروموزوم
۰/۱	نرخ ترانهش ریشه درج متوالی	۱۰۰۰	تعداد جمعیت تولیدی
۰/۳	نرخ ترکیب تک نقطه ای	جمع(+)	تابع پیوند
۰/۳	نرخ ترکیب دو نقطه ای	RMSE	معیار خطا
۰/۱	نرخ ترکیب ژن		
۰/۱	نرخ ترانهش ژن		

جدول (۵): نتایج مدل برنامه‌ریزی بیان ژن برای الگوهای ورودی برآورد TDS رودخانه تجن

الگو	مرحله اجرا	R ²	RMSE	MAE
۱	آموزش	۰/۹۹۲	۱۰/۰۶	۷/۳۲
	آزمون	۰/۹۱۸	۱۷/۱۷	۶/۷۸
۲	آموزش	۰/۹۹۲	۹/۷	۷/۰۶
	آزمون	۰/۹۸۱	۱۷/۳	۶/۹۹
۳	آموزش	۰/۹۹۱	۹/۸۲	۷/۱۴
	آزمون	۰/۹۸۱	۱۷/۱۶	۶/۷۹
۴	آموزش	۰/۹۹۲	۹/۵	۶/۶۷
	آزمون	۰/۹۸۱	۱۷/۲	۶/۵۴
۵	آموزش	۰/۹۹۲	۹/۵	۶/۸۳
	آزمون	۰/۹۸۱	۱۷/۰۶	۶/۱۶



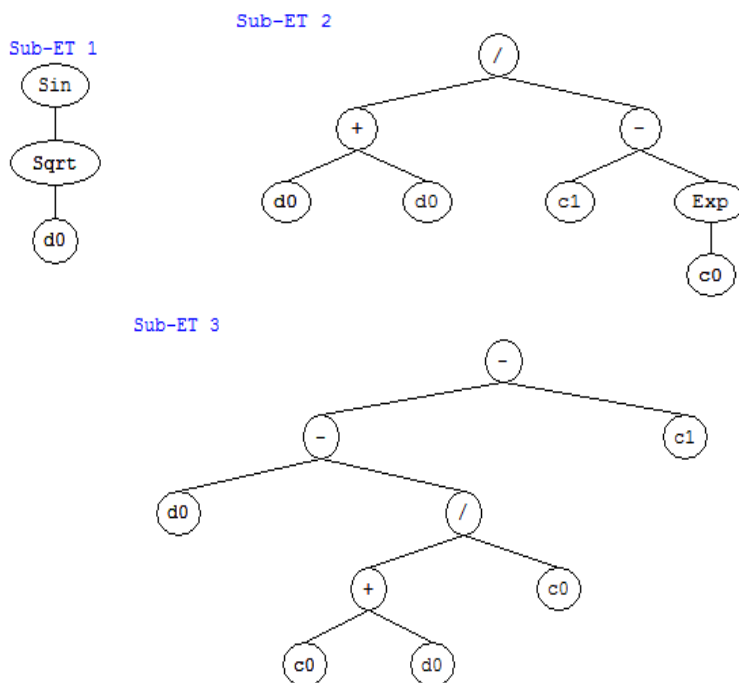
شکل (۴): رابطه بین TDS اندازه‌گیری شده و پیش‌بینی شده رودخانه تجن با استفاده از مدل برنامه‌ریزی بیان ژن با اجرای الگو ۵ در مراحل آموزش و آزمون

(۶)

$$TDS = \sin \sqrt{EC} + \frac{2EC}{-50.834} + EC - \left(\frac{3.128 + EC}{3.128} \right) + 3.476$$

رابطه ریاضی پیشنهادی که بر مبنای برنامه‌ریزی بیان ژن به دست آمد، نشان می‌دهد که TDS با EC رابطه خوبی دارد. این نتیجه در تطابق با نتایج سایر محققین مانند (Rezaei et al., 2019; Karami & Golabi, 2012) می‌باشد.

اگر از الگوی ۱ برای پیش‌بینی شده TDS استفاده شود، نمودار درختی برنامه‌ریزی بیان ژن به صورت شکل (۵) می‌باشد. با توجه به شکل (۵) نمودار درختی رودخانه تجن و با استفاده از جدول (۱۰) مقادیر ثابت برنامه‌ریزی بیان ژن رابطه ریاضی نهایی برآورد TDS رودخانه تجن را می‌توان پیشنهاد داد:



شکل (۵): نمودار درختی الگو ۱ برآورد TDS رودخانه تجن با استفاده از برنامه‌ریزی بیان ژن

جدول (۶): مقادیر ثابت در بیان درختی الگوی ۱، برآورد TDS رودخانه تجن

G3C1	G3C0	G2C1	G2C0	G1C1	G1C0
-۳/۴۷۷۶	۳/۱۲۸	۸/۵۳۰	۴/۰۸۳۷	۰/۴۶۱۱	-۸/۷۷۹

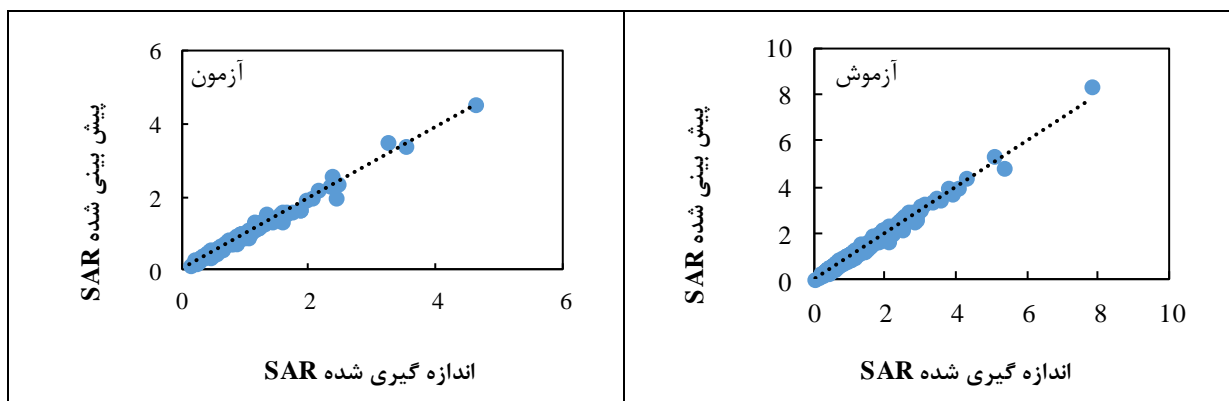
برآورد نسبت جذبی سدیم SAR

همان طوری که قبلا بیان شد برای برآورد نسبت جذبی سدیم SAR بر اساس روش گام به گام (Stepwise Regression)، الگوی ورودی جهت برآورد نسبت جذبی سدیم SAR در نظر گرفته شد که در جدول (۴) نشان داده شده است. نتایج حاصل از اجرای مدل برنامه‌ریزی بیان ژن برای الگوهای مذکور به تفکیک مراحل آموزش و آزمون در جدول (۱۱) ارائه گردیده است. بر اساس سه معیار آماری R^2 ، RMSE و MAE مدل مطلوب و بهینه مدلی است که با دارا بودن بیشترین ضریب همبستگی، کمترین میزان خطا را به خود اختصاص دهد. بر اساس نتایج این جدول، مقدار R^2 برای تمامی الگوها بیشتر از ۰/۹۷ بوده است و همچنین حداکثر خطای RMSE برابر با ۰/۱۰۱ و ۰/۱۳۱ در مرحله آموزش و آزمون بوده است. مقایسه

نتایج به دست آمده نشان داد که الگوی ۲ با پارامترهای ورودی Na و EC با دارا بودن بیشترین ضریب همبستگی برابر با ۰/۹۸۷ در آموزش و ۰/۹۸۹ در آزمون و کمترین خطا و همچنین با داشتن پارامتر ورودی کم مدل بهینه می‌باشد. نتایج حاکی از آن است که الگوی مورد نظر با دارا بودن ضریب همبستگی بیشتر، توانایی بالایی در برآورد SAR دارد و به عبارتی نتایج قابل اطمینانی را در پیش‌بینی نسبت جذبی سدیم ارائه می‌کند در شکل (۶) نمودارهای همبستگی پارامترهای اندازه‌گیری شده و پیش‌بینی شده در دو مرحله آموزش و آزمون و همچنین ساختار درختی ژن‌های حاصل از اجرای الگوی ۲ در شکل (۷) نشان داده شده است. همچنین در جدول (۱۴) مقادیر ثابت بیان درختی الگوی ۲ آمده است. بر اساس بیان درختی الگوی فوق و همچنین رابطه ریاضی برآورد SAR رودخانه تجن در ادامه آمده است:

جدول (۷): نتایج مدل برنامه‌ریزی بیان ژن برای الگوهای ورودی، برآورد SAR رودخانه تجن

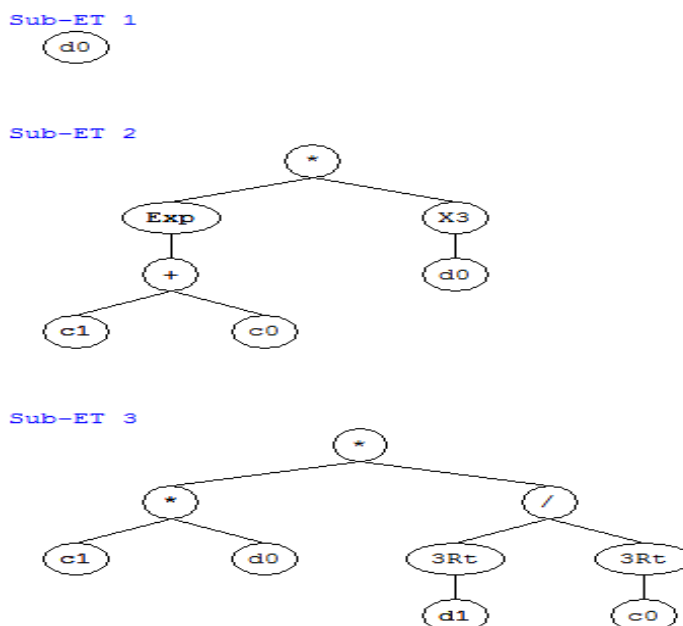
الگو	مرحله اجرا	R^2	RMSE	MAE
۱	آموزش	۰/۹۷۲	۰/۱۰۱	۰/۰۶۲
	آزمون	۰/۹۶۴	۰/۱۳۱	۰/۰۷۷
۲	آموزش	۰/۹۸۷	۰/۰۶۵	۰/۰۴۰
	آزمون	۰/۹۸۹	۰/۰۶۹	۰/۰۴۳
۳	آموزش	۰/۹۷۷	۰/۰۸۷	۰/۰۵۵
	آزمون	۰/۹۸۴	۰/۰۸۵	۰/۰۵۸
۴	آموزش	۰/۹۷۶	۰/۰۹۰	۰/۰۵۶
	آزمون	۰/۹۸۳	۰/۰۸۶	۰/۰۵۹



شکل (۶): رابطه بین SAR اندازه‌گیری شده و پیش‌بینی شده رودخانه تجن با استفاده از مدل برنامه‌ریزی بیان ژن با اجرای الگو ۲ در مراحل آموزش و آزمون

جدول (۸) مقادیر ثابت در بیان درختی الگوی ۲ برآورد SAR رودخانه تجن

G3C1	G3C0	G2C1	G2C0	G1C1	G1C0
-۰/۰۸۲	۶/۹۰۶	-۴/۹۴۴	-۳/۰۶۴	-۴/۷۴۲	-۳/۰۶۴



شکل (۷) نمودار درختی برآورد SAR رودخانه تجن با استفاده از برنامه‌ریزی بیان ژن

۱ با داشتن کمترین پارامتر ورودی، دارای بالاترین ضریب همبستگی و کمترین میزان خطا بوده است، لذا در اجرای مدل ماشین بردار پشتیبان نیز از همان الگو جهت مقایسه دو مدل استفاده گردید. همچنین در برآورد نسبت جذب سدیم (SAR) استفاده از الگوی ۲ استفاده شد. مدل ماشین بردار پشتیبان با استفاده از دو الگوی مذکور اجرا و بهترین نتیجه به دست آمده در جدول (۱۵) آمده است. نتایج به دست آمده بیانگر آن است، مدل SVM با داشتن ضریب تبیین بزرگتر از ۰/۹۴ و ۰/۹۱ در مرحله آزمون به ترتیب برای دو پارامتر TDS و SAR دارای عملکرد مناسبی بوده است. همچنین مقدار RMSE برای دو پارامتر TDS و SAR به ترتیب ۱۴/۷۶ و ۰/۰۷۶ به دست آمد. اگرچه بر اساس معیارهای آماری، مدل SVM دارای دقت قابل قبولی می‌باشد، اما در مقایسه با مدل GEP، نتایج به دست آمده بیانگر آن است عملکرد مدل SVM کاهش پیدا کرده است. به عبارتی مقدار خطا RMSE و MAE افزایش داشته و ضریب تبیین (R^2) کاهش داشته است.

(۷)

$$SAR = \left[NA + [0.000333(NA)^3] + \left[(-0.081665NA) \left(\frac{EC^{\frac{1}{3}}}{1.9043} \right) \right] \right]$$

رابطه فوق طبق الگو ۲ که کاملاً مطابق هدف ما مبنی بر کمترین پارامتر ورودی، بالاترین ضریب همبستگی و کمترین میزان خطا بود به دست آمد و مطابق با فرمول SAR ($SAR = \frac{NA}{\sqrt{\frac{Ca+Mg}{2}}}$) که بیشترین تاثیر را در این رابطه سدیم دارا می‌باشد و استفاده از EC در رابطه که یک پارامتر زود یافت می‌باشد توانست در کنار پارامتر NA موجب بهبود رابطه پیشنهادی در برآورد SAR شود.

برآورد کل مواد محلول (TDS) و نسبت جذب سدیم (SAR) رودخانه تجن

همان طوری که نتایج اجرای مدل GEP در بخش قبلی نشان داد، در برآورد پارامتر کل مواد محلول (TDS) استفاده از الگوی

جدول (۹) نتایج اجرای مدل ماشین بردار پشتیبان برای پیش‌بینی TDS و SAR رودخانه تجن

الگو	پارامتر	تعداد نرون بهینه	مرحله اجرا	R^2	RMSE	MAE
۱	TDS	۷	آموزش	۰/۹۵۶	۱۳/۲۳	۹/۸۷
		۷	آزمون	۰/۹۴۶	۱۴/۷۶	۱۰/۶۷
۲	SAR	۹	آموزش	۰/۹۳۲۷	۰/۰۸۴	۰/۰۵۶
		۹	آزمون	۰/۹۱۳	۰/۰۷۶	۰/۰۶۷

نتیجه گیری

با توجه به اهمیت دانستن پارامتر TDS و SAR در این تحقیق با استفاده از آمار رودخانه تجن به پیش‌بینی آن‌ها با استفاده از مدل GEP پرداخته شد. با توجه به اینکه ترکیب پارامترهای ورودی به مدل نقش تاثیرگذاری در عملکرد مدل GEP دارد. بنابراین، بر این مبنا الگوهای مختلفی از متغیرهای ورودی بر اساس روش آماری گام به گام در نظر گرفته شد. برای پارامتر TDS از ۵ الگو با متغیرهای ورودی EC , SO_4 , SAR , HCO_3 , CL که شامل پارامترهای زودیافت بودند، استفاده شد. همچنین برای پارامتر SAR از ۴ الگو که شامل متغیرهای ورودی EC , SO_4 , SAR , HCO_3 , CL بودند، استفاده شد.

نتایج به دست آمده از اجرای مدل نشان داد، در تمامی الگوهای مورد استفاده مدل GEP توانایی پیش‌بینی پارامتر TDS را دارد به طوری که همان‌طور که انتظار می‌رفت با افزایش تعداد متغیرهای ورودی، چون تعداد درجه آزادی افزایش می‌یابد عملکرد مدل در برآورد پارامتر TDS افزایش یافته است. از طرفی برای کاهش هزینه‌ها لازم است اندازه‌گیری تعداد متغیرهای ورودی کم گردد و به عبارت بهتر مدل بهینه، مدلی است که دارای تعداد متغیرهای ورودی کمتر ولی دقت مدل نسبتاً بالا باشد. نتایج حاصل از اجرای مدل GEP با گزینه‌های مختلف از الگوهای ورودی نشان داد در برآورد TDS رودخانه

تجن وقتی از الگوی ۱ که در آن فقط از پارامتر EC استفاده شده است، مدل با داشتن R^2 , RMSE و MAE برابر با ۰/۹۸۱، ۱۷/۱۷ و ۶/۷۸ دارای دقت نسبتاً خوبی می‌باشد. همچنین در برآورد نسبت جذبی سدیم رودخانه تجن (SAR) الگو ۲ با دو متغیرهای ورودی شامل Na و EC به عنوان مدل برتر انتخاب گردید. در این الگو مقادیر R^2 , RMSE و MAE در مرحله آزمون برابر با ۰/۹۸۹، ۰/۰۶۹ و ۰/۰۴۳ بوده است. همچنین اجرای مدل SVM نشان داد این مدل دارای دقت قابل قبولی می‌باشد در برآورد دو پارامتر TDS و SAR دارد، اما در مقایسه با مدل GEP، نتایج به دست آمده بیانگر آن است عملکرد مدل SVM کاهش پیدا کرده است.

در مجموع نتایج این تحقیق نشان دهنده عملکرد بالای روش برنامه‌ریزی بیان ژن در برآورد پارامترهای کیفی رودخانه‌های تجن می‌باشد و لذا از این مدل می‌توان برای برآورد و پیش‌بینی TDS و SAR مورد استفاده قرار داد.

یادداشت‌ها

1. Multi Perceptron Layer
2. Radial Bias Function
3. Adaptive Neuro-Fuzzy Inference System
4. Support vector machines
5. Noise

فهرست منابع

- Hesami- Afshar, M.; Khalili, K.; Ali Golinia, T. & Zamanzad Gavidel, S. 2013. Prediction of TDS concentration of Zarinerood river based on the least number of input parameters with the help of genetic programming. The third international conference on environmental planning and management, University of Tehran (in persian).
- Azarmdel, H.; Mostafazadeh, R. & Ghasemi, A. 2010. Assessment of Water Quality Monitoring Stations in Gorgan-Roud River, Golestan Province. *Watershed Management Science & Engineering*. 4(10):57-61 (in persian)
- Dolat-Kordestani, M.; Nohegar, A.; & Janizadeh, S. 2016. Evaluation of several artificial intelligence models for predicting the quality of underground water (case study: Dasht-Guru). *Scientific Research Journal of Desert Ecosystem* (in persian). 6(17): page 42-27.
- Diamantopoulou, M.J.; Papamichail, D.M. and Antonopoulos, V.Z. 2005. The use of a neural network technique for the prediction of water quality parameters. *Operational Research* 5: 115-125.
- Sattari, M.T.; Pal, M.; Apaydin, H. & Ozturk, F. 2013. M5 model tree application in daily river flow forecasting in Sohu Stream, Turkey. *Water Resources Management*. 40(3): 233-242.
- Musavi-Jahromi, S. & Golabi, M. 2008. Application of artificial neural networks in the river water quality modeling: Karoon River, Iran. *Journal of Applied Sciences*. 8(12): 2324-2328.
- Emamgholizadeh, S.; Kashi, H.; Marofpoor, I. & Zalaghi, E. 2014. Prediction of water quality parameters of Karoon River (Iran) by artificial intelligence-based models. *International Journal of Environmental Science Technology*. 11(3): 645-656.

- Mojaradi, B.; Alizadeh, F. & Samadi, M. 2017. Estimation of Talar river water quality index using artificial neural networks and gene expression programming. *Scientific-Research Journal of Watershed Science and Engineering of Iran* (in persian). 12 (41), 72-61
- 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.
- Karami, B. & Golabi, M. 2013. Simulation and prediction of water quality parameters using artificial neural network, fuzzy-neural method and statistical regression (case study: Karun river, Khuzestan province), the ninth International Seminar on River Engineering, Ahvaz, Shahid Chamran University, Ahvaz (in persian).
- Rezaei, A.; Shahi-nejad, B. & Yonesi H. 2010. Evaluation of intelligent models in estimating the amount of dissolved solids in Kashkan river water of Lorestan province. *Watershed engineering and management scientific-research journal* (in persian). 1(1): 165-147
- Mehdipour, V.; Memarianfard, M. & Homayounfar, F. 2017. Application of Gene Expression Programming to water dissolved oxygen concentration prediction. *International Journal of Human Capital in Urban Management*. 2(1): 1-10.
- Najafzadeh, M.; Ghaemi, A. & Emamgholizadeh, S. 2019. Prediction of water quality parameters using evolutionary computing-based formulations. *International Journal of Environmental Science Technology*. 16(10): 6377-6396.
- Esa-zadeh, M.; Be-azar, S.M.; Ashrafzadeh, A.; Khanjani, R. 2017. Estimation of quality parameters of Gilan plain aquifer using gamma test and support vector machine models and artificial neural network. *Environmental science and technology* (in Persian). 20(6), 1-21
- Kamyab-Talesh, F.; Mousavi, S.-F.; Khaledian, M.; Yousefi-Falakdehi, O. & Norouzi-Masir, M. 2019. Prediction of Water Quality Index by Support Vector Machine: a Case Study in the Sefidrud Basin, Northern Iran. *Water Resources Management*. 46(1): 112-116.
- Louisa, F. 2014. Investigation of the origin and concentration of heavy metal pollutants in the catchment area of Talar, Tajen, and Nekarud rivers." The 21st meeting of geological sciences of the Geological Organization of the country, Iran (in persian).
- Cortes, C. & Vapnik, V. 1995. Support-vector networks. *Machine learning*. 20(3): 273-297.
- Parsaie, A.; Haghiabi, A.H.; Emamgholizadeh, S. & Azamathulla, H.M. 2019. Prediction of discharge coefficient of combined weir-gate using ANN, ANFIS and SVM. *International Journal of Hydrology Science Technology* 9(4): 412-430.
- Azamathulla, H.M.; Parsaie, A.; Haghiabi, A.H. & Emamgholizadeh, S. 2018. Prediction of discharge coefficient of combined weir-gate using ANN, ANFIS and SVM. *International Journal of Hydrology Science and Technology*. 9: 412-430.
- Emamgholizadeh, S. & Mohammadi, B. 2021. New hybrid nature-based algorithm to integration support vector machine for prediction of soil cation exchange capacity. *Soft computing*. 25: 13451-13464.
- Ferreira, C. 2001. Gene expression programming: a new adaptive algorithm for solving problems. *Complex Syst*. 13 (2): 87-129.
- Lopes, H.S. & Weinret, W.R. 2004. EGYPSIS: An enhanced gene expression programming approach for symbolic regression problems, *Journal of Applied Mathematics and Computer Science*, 14(3): 375-384.
- Emamgholizadeh, S.; Bateni, S.M.; Shahsavani, D.; Ashrafi, T. & Ghorbani, H. 2015. Estimation of soil cation exchange capacity using Genetic Expression Programming (GEP) and Multivariate Adaptive Regression Splines (MARS). *Journal of Hydrology*, 529 (2015): 1590-1600.
- Emamgholizadeh, S.; Bahman, K.; Bateni, S.M.; Ghorbani, H.; Marofpoor, I. & Nielson, J.R. 2017. Estimation of soil dispersivity using soft computing approaches. *Neural Computing Applications* 28, 207-216.
- Emamgholizadeh, S. & Karimi, R. 2019. A comparison of artificial intelligence models for the estimation of daily suspended sediment load: a case study on the Telar and Kasilian rivers in Iran. *Water Supply* 19, 165-178.