Shahid Chamran University of Ahvaz

Irrigation Sciences and Engineering (JISE)
Vol. 48, No. 3, Fall 2025, p. 1-20

Research Paper

Journal homepage: https://jise.scu.ac.ir

EXTENDED ABSTRACT

Estimation of sodium absorption ratio (SAR) and electrical conductivity (EC) using hybrid models of support vector regression and random forest-case study of Qatur river, west of Iran

M. Sharafi[1], J. Behmanesh[2*], V. Rezaverdinejad[3] and S. Samadianfard[4]

1- Ph.D. Student, Department of Water Engineering, Faculty of Agriculture, Urmia University, Urmia, Iran.
2- Corresponding Author, Professor, Department of Water Engineering, Faculty of Agriculture, Urmia University, Urmia, Iran. (j.behmanesh@urmia.ac.ir)*
3- Professor, Department of Water Engineering, Faculty of Agriculture, Urmia University, Urmia, Iran.
4- Associate Professor, Department of Water Engineering, Faculty of Agriculture, University of Tabriz, Tabriz, Iran.

ARTICLE INFO

Artlcle history:
Received: 29 July 2024
Revised: 18 November 2024
Accepted: 20 November 2024
Available online: 05 November 2025

Keywords:
Irrigation water quality, Sodium absorption ratio, Machine learning, Data mining, Wilcox diagram, Hydrology.

TO CITE THIS ARTICLE:

Sharafi, M., Behmanesh, J., Rezaverdinejad, V., Samadianfard, S. 2025. 'Estimation of sodium absorption ratio (SAR) and electrical conductivity (EC) using hybrid models of support vector regression and random forest-case study of Qatur river, west of Iran', *Irrigation Sciences and Engineering*, 48(3), pp. 1-20. doi: 10.22055/jise.2024.47590.2131

Abstract

In the field of irrigation, water quality has an effective role in determining the compatibility of water with soil and different crops. As a valuable tool for water quality assessment, the Wilcox diagram enables stakeholders to promote sustainable agricultural practices by reducing potential risks related to water salinity and sodium content. However, measuring sodium absorption ratio (SAR) and electrical conductivity (EC) parameters is associated with various challenges such as cost and time, calibration error, and maintenance problems. So, this research was carried out to predict SAR and EC parameters using easily accessible parameters including sulfate and chloride in Qatur river located in the northwest of Iran from 1993 to 2019. The models used included random forest (RF) and support vector regression (SVR) which have been optimized by genetic algorithm (GA), fruit fly algorithm and Golden ball optimization (GBO). The analysis of the results of this research showed that the RF-GA model had the least error in the Qatur river station with root mean square error of 0.809 meq.lit-1 and 0.146 dS.m-¬1 for SAR and EC parameters, respectively. Also, The SVR-GBO model with the root mean square error of 0.809 meq.lit-1 and 0.146 dS.m-1 and SVR-FOA with the root mean square error of 0.998 meq.lit-1 and 0.169 dS.m-1 for SAR and EC parameters, respectively had suitable performance. Finally, the evaluation of the utilized parameters indicated that chloride plays an important role in predicting SAR and EC parameters, and if sulfate is not measured, it can be used as a promising alternative.

Introduction

The process of measuring both EC and SAR parameters is a laborious task and necessitates considerable monetary investment, especially when dealing with a substantial quantity of samples (Waqas et al., 2024). Consequently, the utilization of alternative methodologies has gained paramount importance. The present research aims to accurately predict the EC and SAR parameters in the Qatur River located in the Aras River catchment, Iran, using the support vector regression (SVR) and random forest (RF) models optimized with fruit fly (FOA), genetic (GA), and golden ball (GBO) algorithms. Finally, the models will be evaluated using different evaluation criteria.

Methodology

Study area

In recent years, the quality of the Aras River has potentially decreased due to factors such as intense agricultural activities and the need to monitor water quality has become more important than in the past. Therefore, to monitor the water quality, the condition of Qatur River, in the right branch of Aras River, Iran, has been investigated.

Random forest (RF)

The random forest algorithm is a potent ensemble learning algorithm widely used for diverse tasks such as regression, classification, and unsupervised learning. It was presented by Breiman and is known for its superior performance (Breiman 2001).

Support vector regression (SVR)

Support vector regression (SVR) is an algorithm belonging to the domain of machine learning, specifically designed for the purpose of regression tasks. SVR was introduced by Vapnik and Chervonenkis (1974).

Genetic algorithm (GA)

Genetic algorithms (GAs) are a type of non-deterministic models that are inspired by Darwinian theory (Hüe 1997). Unlike other optimization methods, GA focuses on evaluating the objective function (McKinney and Lin 1994).

Fruit fly optimization algorithm (FOA)

The Fruit Fly Optimization Algorithm (FOA) is a nature-inspired optimization algorithm based on the behavior of fruit flies, first proposed by Pan (2012). It is a metaheuristic algorithm that aims to find the best solution to a given problem (Wang et al., 2022).

Golden Ball Optimization (GBO)

The Golden Ball Optimization algorithm is an innovative meta-heuristic method inspired by soccer game tactics and strategies. In this algorithm, players are considered as search agents in the problem space that pursue a common goal through cooperation and interaction (Osaba et al., 2014).

Also, in this research, three different scenarios have been utilized to evaluate the role of input parameters on SAR and EC parameters. Table (1) also indicates the scenarios utilized to evaluate the role of input parameters.

Table 1- The input combinations of the studied models.

Scenario	Input		Output
1	SO_4		SAR-EC
2	Cl		SAR-EC
3	SO_4	Cl	SAR-EC

Results and Discussion

Tables (2) presents the results of the models utilized in the test period based on R, RMSE, NRMSE, NS, and WI criteria for Qatur River station. The evaluation of the studied models of the SAR parameter prediction indicated that the RF-GA-3 model with criteria R = 0.983, RMSE = 0.809 (meq/lit), NRMSE = 0.024, NS = 0.965, and WI = 0.991 has had the best performance compared to other models. For the EC parameter, the results also indicated that the RF-GA-3 model with criteria R = 0.995, RMSE = 0.146 (dS/m), NRMSE = 0.012, NS = 0.991, and WI = 0.998 has the highest accuracy. Also, the results are in agreement with the results of Hassan et al., (2021) who predicted water quality parameters utilizing classification algorithms and indicated that the MLR model with an accuracy of 99.83% is known as a valuable model in EC parameter prediction.

Conclusions

Wilcox diagram divides water into different classes using SAR and EC parameters. This investigation has been carried out to predict SAR and EC parameters utilizing easily accessible

parameters including sulfate and chloride in Qatur River station. The utilized models include RF, RF-GA, SVR, and SVR-FOA, which were evaluated utilizing different evaluation criteria. The evaluation of the outcomes of this examination indicated that the RF-GA-3 model with R = 0.983, and 0.995, respectively, for SAR and EC parameters had the highest accuracy in the Qatur River station. One of the limitations of the present investigation is the lack of access to data from other stations to evaluate the accuracy of the models used. Therefore, it is suggested to utilize data from different stations to check the ability of models for future studies.

Table 2- The evaluation results of the models used in Qatur River station.

Parameter	**SAR**					**EC**				
Scenario	R	RMSE (meq/lit)	NRMSE	NS	WI	R	RMSE (dS/m)	NRMSE	NS	WI
RF-1	0.862	2.348	0.069	0.708	0.925	0.876	0.813	0.064	0.708	0.931
RF-2	0.969	1.078	0.032	0.938	0.984	0.989	0.219	0.017	0.979	0.995
RF-3	**0.977**	**0.923**	**0.027**	**0.955**	**0.989**	**0.992**	**0.215**	**0.017**	**0.900**	**0.995**
RF-GA-1	0.896	1.927	0.057	0.803	0.943	0.918	0.598	0.047	0.842	0.957
RF-GA-2	0.977	0.939	0.028	0.953	0.988	0.990	0.213	0.017	0.980	0.995
RF-GA-3	**0.983**	**0.809**	**0.024**	**0.965**	**0.991**	**0.995**	**0.146**	**0.012**	**0.991**	**0.998**
RF-GBO-1	0.887	2.011	0.059	0.786	0.938	0.902	0.661	0.052	0.807	0.948
RF-GBO-2	0.974	0.981	0.029	0.949	0.987	0.989	0.225	0.018	0.978	0.994
RF-GBO-3	**0.980**	**0.877**	**0.026**	**0.959**	**0.990**	**0.995**	**0.154**	**0.012**	**0.990**	**0.997**
SVR-1	0.869	2.194	0.065	0.745	0.915	0.869	0.764	0.060	0.743	0.914
SVR-2	0.966	1.128	0.033	0.933	0.983	0.987	0.241	0.019	0.974	0.993
SVR-3	**0.970**	**1.068**	**0.032**	**0.940**	**0.984**	**0.991**	**0.229**	**0.018**	**0.977**	**0.995**
SVR-FOA-1	0.913	1.840	0.054	0.821	0.942	0.908	0.645	0.051	0.817	0.942
SVR-FOA-2	**0.974**	**0.998**	**0.028**	**0.947**	**0.987**	0.996	0.234	0.019	0.976	0.993
SVR-FOA-3	0.973	1.050	0.031	0.942	0.984	**0.994**	**0.169**	**0.013**	**0.987**	**0.999**
SVR-GBO-1	0.874	2.166	0.064	0.752	0.916	0.988	0.235	0.019	0.976	0.994
SVR-GBO-2	0.968	1.097	0.032	0.936	0.983	0.993	0.182	0.014	0.985	0.996
SVR-GBO-3	**0.972**	**1.042**	**0.031**	**0.943**	**0.984**	**0.994**	**0.176**	**0.014**	**0.986**	**0.997**

Acknowledgments

The authors of the article hereby express their gratitude to the regional water company of West Azerbaijan province for receiving the data and also to the research and technology vice-chancellor of Urmia University for the financial support of this study with grant number 19621.

Conflict of Interest

The authors declared no potential conflicts of interest concerning the research, authorship, and publication of this article. Confirmation.

Funding

The financial support of this research has been provided by Grant number 1961 and we hereby thank the Honorable Vice Chancellor of Research from Urmia University for their support.

Data Availability Statements

The datasets generated and/or analyzed during the current study are available from the corresponding author on reasonable request.

Authors' contribution

First Author: Milad Sharafi, Collecting data and writing the first draft of the article.
Second Author: Javad Behmanesh, Reading and revising the initial draft.
Third Author: Vahid Rezaverdinejad, Reading and revising the initial draft.
Fourth Author: Saeed Samadian Fard, Reading and revising the initial draft.

References

1- Breiman, L., 2001. Random forests. *Machine learning*, 45(4), pp.5-32. https://doi.org/10.1023/A:1010933404324.

2- Hassan, M., Hassan, M., Akter, L., Rahman, M., Zaman, S., Khan, H., Jahan, N., Smrity, R.N., Farhana, J. and Raihan, M., 2021. Efficient prediction of water quality index (WQI) using machine learning algorithms. *Human-Centric Intelligent Systems*, 1(3), pp.86-97. https://doi.org/10.2991/hcis.k.211203.001.

3- Hüe, X., 1997. Genetic algorithms for optimization. Citeseer.

4- McKinney, D.C. and Lin, M.D., 1994. Genetic algorithm solution of groundwater management models. *Water Resources Research*, 30(6), pp.1897-1906. https://doi.org/10.1029/94WR00554.

5- Osaba, E., Diaz, F. and Onieva, E., 2014. Golden ball: a novel meta-heuristic to solve combinatorial optimization problems based on soccer concepts. *Applied intelligence*, 41(1), pp.145-166. https://doi.org/10.1007/s10489-013-0512-y.

6- Pan, W.-T., 2012. A new fruit fly optimization algorithm: taking the financial distress model as an example. *Knowledge-Based Systems*, 26(1), pp.69-74. https://doi.org/10.1016/j.knosys.2011.07.001.

7- Vapnik, V. and Chervonenkis, A.Y., 1974. The method of ordered risk minimization, I. *Avtomatika i Telemekhanika*, 8(1), pp.21-30.

8- Wang, R.-Y., Hu, P., Hu, C.-C. and Pan, J.-S., 2022. A novel Fruit Fly Optimization Algorithm with quasi-affine transformation evolutionary for numerical optimization and application. *International Journal of Distributed Sensor Networks*, 18(2), pp.155-168. https://doi.org/10.1177/15501477211073037.

9- Waqas, M., Humphries, U.W., Wangwongchai, A., Dechpichai, P., Zarin, R. and Hlaing, P.T., 2024. Incorporating novel input variable selection method for in the different water basins of Thailand. *Alexandria Engineering Journal*, 86(1), pp.557-576. https://doi.org/10.1016/j.aej.2023.11.046.

© 2025 The Author(s). Published by Shahid Chamran University of Ahvaz. This is an open-access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

علوم و مهندسی آبیاری جلد ۴۸، شماره‌ی۳، پاییز ۱۴۰۴، مقاله پژوهشی، ص. ۲۰-۱

تخمین نسبت جذب سدیم (SAR) و هدایت الکتریکی (EC) با استفاده از مدل‌های هیبریدی رگرسیون بردار پشتیبان و جنگل تصادفی-مطالعه موردی رودخانه قطور، غرب ایران

میلاد شرفی[۱]، جواد بهمنش[۲*]، وحید رضاویردی‌نژاد[۳] و سعید صمدیان‌فرد[۴]

۱- دانشجوی دکتری، گروه مهندسی آب، دانشکده کشاورزی، دانشگاه ارومیه، ارومیه، ایران.

۲*- نویسنده مسئول، استاد، گروه مهندسی آب، دانشکده کشاورزی، دانشگاه ارومیه، ارومیه، ایران j.behmanesh@urmia.ac.ir

۳- استاد، گروه مهندسی آب، دانشکده کشاورزی، دانشگاه ارومیه، ارومیه، ایران.

۴- دانشیار، گروه مهندسی آب، دانشکده کشاورزی، دانشگاه تبریز، تبریز، ایران.

دریافت: ۱۴۰۳/۰۵/۰۸ بازنگری: ۱۴۰۳/۰۸/۲۸ پذیرش: ۱۴۰۳/۰۸/۳۰ انتشار: ۱۴۰۴/۰۸/۱۴

چکیده

در زمینه آبیاری، کیفیت آب نقش مؤثری در تعیین سازگاری آب با خاک و محصولات مختلف دارد. نمودار Wilcox به‌عنوان یک ابزار ارزشمند برای ارزیابی کیفیت آب، ذینفعان را قادر می‌سازد تا با کاهش خطرات بالقوه مربوط به شوری آب و محتوای سدیم، شیوه‌های کشاورزی پایدار را ترویج کنند. با این حال، اندازه‌گیری پارامترهای نسبت جذب سدیم (SAR) و هدایت الکتریکی (EC) با چالش‌های مختلفی مانند هزینه و زمان، خطای کالیبراسیون و مشکلات نگهداری همراه است. بر این اساس، این تحقیق به‌منظور پیش‌بینی پارامترهای SAR و EC با استفاده از پارامترهای دردسترس شامل سولفات و کلرید در رودخانه قطور واقع در شمال غرب ایران از سال ۱۳۷۲ تا ۱۳۹۸ انجام شد. مدل‌های مورد استفاده شامل جنگل تصادفی (RF)، رگرسیون بردار پشتیبان (SVR) بودند که توسط الگوریتم‌های ژنتیک (GA)، مگس میوه (FOA) و بهترین توپ طلایی (GBO) بهینه‌سازی شدند. تجزیه و تحلیل یافته‌های حاصل از این تحقیق نشان داد که مدل RF-GA کم‌ترین خطا را در ایستگاه رودخانه قطور با داشتن جذر میانگین مربعات خطای ۰/۸۰۹ meq/lit و ۰/۱٤٦ dS/m به‌ترتیب برای پارامترهای SAR و EC داشته است. همچنین مدل SVR-GBO با داشتن جذر میانگین مربعات خطای ۰/۸۷۷ meq/lit و ۰/۱٥٤ dS/m و مدل SVR-FOA نیز با داشتن جذر میانگین مربعات خطای ۰/۹۹۸ meq/lit و ۰/۱٦۹ dS/m به‌ترتیب برای پارامترهای SAR و EC عملکرد مناسبی را داشتند. درنهایت ارزیابی پارامترهای مورد استفاده نشان داد که کلرید نقش مهمی در پیش‌بینی پارامترهای SAR و EC دارد و در صورت عدم اندازه‌گیری سولفات، می‌توان از آن به‌عنوان یک جایگزین امیدوارکننده استفاده کرد.

کلید واژها: نسبت جذب سدیم، یادگیری ماشینی، داده‌کاوی، هیدرولوژی.

مقدمه

رودخانه‌ها منبع اصلی تأمین آب سطحی در جهان هستند و مقدار آب مورد نیاز بخش‌های مختلف را تأمین می‌کنند (Giri 2021). رودخانه‌ها به‌عنوان یکی از بخش‌های حیاتی چرخه هیدرولوژیکی، اکوسیستم‌ها را حفظ می‌کنند و نشان می‌دهند که با تقویت تنوع زیستی، نقش به‌سزایی در اکوسیستم‌ها دارند (Palmer and Ruhi 2019). رودخانه‌ها علاوه‌بر نقشی که در اکوسیستم‌ها دارند، به‌عنوان منبع اصلی تامین آب شیرین در جوامع انسانی نیز ضروری هستند (Wang and He 2022). کمیت و کیفیت آب رودخانه‌ها در سال‌های اخیر به‌دلایلی مانند رشد جمعیت، بی‌توجهی به نگرانی‌های زیست‌محیطی و مصرف بیش از حد منابع آب سطحی به‌طور قابل توجهی کاهش یافته است (Mousazadeh et al., 2021). این کاهش کیفیت در رودخانه‌ها، چالش‌های مهمی را برای بخش کشاورزی، که بزرگترین مصرف‌کننده آب شیرین در سطح جهان است، ایجاد کرده است و تقریباً ۸۵ درصد از مصرف را به خود اختصاص داده است (Qadir et al., 2007). از جمله مسائل اصلی مرتبط با استفاده از آب با کیفیت پایین در کشاورزی، افزایش سطح شوری و سدیم در خاک است (Machado and Serralheiro 2017; Pincay and Cantos 2024). استفاده از آب شور به تجمع نمک در خاک کمک می‌کند و در نتیجه مانع از رشد محصولات حساس می‌شود (Alkharabsheh et al., 2021). علاوه بر این، آبیاری با آب حاوی سدیم بالا بر ساختار خاک‌های رسی تأثیر منفی می‌گذارد و در نتیجه مانع رشد گیاه می‌شود (Rahimi et al., 2021). بنابراین، با توجه به نقش مهم نمک و سدیم در آب آبیاری، خاک‌ها بر اساس میزان هدایت الکتریکی (Electrical conductivity) و نسبت جذب سدیم (Sodium adsorption ration) به طبقات مختلفی طبقه‌بندی می‌شوند (Amezketa and de Lersundi 2008). اندازه‌گیری هدایت الکتریکی در آب معمولاً با استفاده از

 DOI: 10.22055/jise.2024.47590.2131

یک EC متر انجام می‌شود (Cavins et al., 2000). درحالی‌که تعیین نسبت جذب سدیم پس از ارزیابی غلظت کلسیم، منیزیم و سدیم انجام می‌شود (Sposito and Mattigod 1977). فرآیند اندازه‌گیری هر دو پارامترEC و SAR یک کار پرزحمت بوده و نیاز به سرمایه‌گذاری پولی قابل‌توجهی دارد، به‌ویژه هنگامی‌که با مقدار قابل‌توجهی از نمونه‌ها سروکار داریم. لذا، استفاده از روش‌های جایگزین اهمیت زیادی را پیدا می‌کند. در سال‌های اخیر، مدل‌های هوشمند به‌دلیل توانایی آن‌ها در به حداقل رساندن خطاها و بهینه‌سازی زمان پیش‌بینی، نسبت به رویکردهای تجربی اهمیت قابل‌توجهی پیدا کرده‌اند (Waqas et al., 2024; Woźniak et al., 2024).

Nouraki et al. (2021) پارامترهای کیفیت آب شامل SAR، سختی کل (Total hardness) و کل جامدات محلول (Total dissolved solids) را با استفاده از مدل‌های جنگل تصادفی (Random forest)، M5P، رگرسیون بردار پشتیبان (Support vector regression)و رگرسیون خطی چندگانه (Multi-layer regression)در رودخانه کارون پیش‌بینی کردند. یافته‌های این تحقیق نشان داد که مدل‌هایRF، SVR و MLR به‌ترتیب با داشتن ضریب تعیین ۰/۹۸، ۰/۹۹ و ۰/۹۹ عملکرد مناسبی را در تخمین پارامترهای TDS، SAR و TH داشته‌اند. در تحقیقی Sepahvand et al. (2021) عملکرد مدل‌های درختی شامل M5P، RF، M5P کیسه‌ای (bagged M5P) و روش گروهی برای پردازش داده‌ها (group method for data handling) را برای پیش‌بینی SAR در سه حوزه آبخیز آلستار، بیرانشهر و خرم‌آباد واقع در ایران ارزیابی کردند. این تحقیق نشان داد که مدل BM5P با داشتن جذر میانگین مربعات خطا ۰/۰۰۹ از دقت بیشتری نسبت به سایر مدل‌های مورد استفاده برخوردار بوده است. در یک بررسی، El Bilali و Taleb (2020) پارامترهای کیفیت آب آبیاری را با استفاده از مدل‌های MLR، RF، SVR، تقویت تطبیقی (adaptive boosting)، نزدیک‌ترین همسایگی (k- k nearest neighbor)، درخت تصمیم (decision tree)، شبکه عصبی مصنوعی (Artificial neural network) و گرادیان نزولی تصادفی (stochastic gradient descent) پیش‌بینی کردند. همچنین داده‌های مورد نیاز از پایش نه ایستگاه در چهار رودخانه واقع در حوضه بورگرگ مراکش به‌دست آمد. نتایج این مطالعه نشان داد که مدل‌های مورد استفاده (به جز KNN وSVR) پارامترهای کیفیت آب رودخانه را با ضریب همبستگی ۰/۹ با موفقیت پیش‌بینی کردند. در مطالعه‌ای Mokhtar et al. (2022) شاخص‌های کیفیت آب آبیاری بحرالبقر مصر را بر اساس مدل‌های یادگیری ماشین و رگرسیون چندگانه پیش‌بینی کردند. ارزیابی نتایج این تحقیق نشان داد که بر اساس شاخص پراکندگی (scatter index)، تمامی مدل‌های مورد استفاده با مقادیر خطای کمتر از ۰/۱ دقت بالایی برای همه پارامترها به جز کربنات سدیم باقی‌مانده نشان دادند. در تحقیقی دیگر Haghiabi et al. (2018) به پیش‌بینی کیفیت آب با استفاده از روش‌های یادگیری ماشین در رودخانه تیره واقع در جنوب غرب ایران پرداختند. ارزیابی دقت مدل‌های اعمال‌شده نشان داد که SVM دقیق‌ترین مدل بوده و کم‌ترین خطا را نسبت به سایر مدل‌ها داشته است. Qishlaqi et al. (2017) به ارزیابی هیدروشیمیایی کیفیت آب رودخانه تیره رود واقع در حوضه آبریز کرخه پرداختند. ارزیابی نتایج این مطالعه نشان داد که (به‌جز پارامتر سولفات) منیزیم، کلسیم و باقی آنیون‌ها و کاتیون‌ها در محدوده استاندارد سازمان جهانی بهداشت بودند.

در سال‌های اخیر، افزایش قابل‌توجهی در استفاده از مدل‌های هیبریدی در مقایسه با مدل‌های مستقل وجود داشته است (Rad and Kasaeian 2024). این تغییر را می‌توان به مزایای متعددی از جمله بهبود عملکرد، افزایش استحکام، انعطاف‌پذیری، سازگاری و هم‌افزایی نسبت داد که آن‌ها را به انتخابی ارجح نسبت به مدل‌های مستقل در زمینه‌ها و کاربردهای مختلف تبدیل کرده است (Manikandan and Pant 2021). توانایی در مهار نقاط قوت رویکردها و روش‌های مختلف، آن‌ها را به ابزاری قدرتمند برای مقابله با مشکلات پیچیده و دست‌یابی به نتایج دقیق‌تر و قابل اطمینان‌تر تبدیل می‌کند. در تحقیقی Ahmadianfar et al. (2020) مدل‌سازی EC رودخانه سفیدرود در ایران را با استفاده از رگرسیون خطی وزنی موجک هیبریدی در طی ۲۰ سال بررسی کرد. ارزیابی نتایج این تحقیق نشان داد که مدل پیشنهادی با ریشه میانگین مربعات خطا ۸۹/۷۸ به‌عنوان بهترین مدل در حوزه مورد بررسی تعیین شد. در مطالعه‌ای Kisi et al. (2019) پارامترهای کیفیت آب از جمله EC، SAR و TH را با هیبریداسیون مدل سیستم استنتاج عصبی–فازی تطبیقی (Adaptive neural-fuzzy inference system) با الگوریتم‌های تکاملی مدل‌سازی کرد. نتایج تحقیق نشان داد که الگوریتم ژنتیک پیوسته (Continuous genetic algorithm) بهترین عملکرد را با افزایش ضریب تعیین مدل ANFIS در حدود ۰/۱۴، ۰/۱۳ و ۰/۱۵ برای پارامترهای TH، EC و SAR داشت. در مطالعه‌ای دیگر Jamei et al. (2023) هدایت الکتریکی و بی‌کربنات آب رودخانه کارون را با استفاده از یک مدل ترکیبی مبتنی‌بر شبکه عصبی بازگشتی عنصری (elemental recurrent neural network) پیش‌بینی کرد. این تحقیق نشان داد که پیش‌بینی هدایت الکتریکی و شاخص‌های یون بی‌کربنات با استفاده از مدل‌های ترکیبی می‌تواند به‌طور موثر مورد استفاده قرار گیرد.

نظارت دقیق و مداوم بر کیفیت آب در منابع آب سطحی به‌دلیل تأثیر بالقوه بر سلامت انسان و حفظ اکوسیستم حیاتی است. رودخانه‌ها به‌عنوان منابع ارزشمند آب شیرین با افزایش جمعیت و گسترش کشاورزی در بیش‌تر موارد به‌عنوان پساب کارخانه‌ها و زهکش‌های کشاورزی عمل کرده و به‌طور فزاینده‌ای

در معرض خطر کاهش کیفیت قرار می‌گیرند. همچنین اندازه‌گیری پارامترهای کیفی مستلزم صرف زمان و هزینه برای انجام آزمایش‌ها و آوردن نمونه‌ها به آزمایشگاه در ظروف مخصوص بوده و در برخی موارد فاصله زیاد تا آزمایشگاه نیز باعث تغییر در کیفیت نمونه‌ها می شود. بنابراین استفاده از روش‌هایی که پارامترهای کیفیت آب را در زمان واقعی در رودخانه‌ها با دقت بالایی ارائه می‌کنند، به‌عنوان امری مؤثر تلقی می‌شوند. از سوی دیگر، نمودار ویلکاکس (Wilcox) با طبقه‌بندی کیفیت آب آبیاری با استفاده از پارامترهای SAR و EC، به‌عنوان راهنمای عملی برای کشاورزان برای مدیریت کارآمد و پایدار منابع آب در شیوه‌های کشاورزی عمل می‌کند. در نمودار Wilcox پارامتر SAR در قسمت عمودی نمودار و پارامتر EC در قسمت افقی قرار داشته و همچنین کل نمودار به ۱۶ بخش از C1-S1 (کیفیت بسیار خوب) تا C4-S4 (کیفیت بسیار بد) تقسیم شده است. بنابراین، پیش‌بینی این پارامترها با استفاده از عناصر در دسترس، باعث صرفه‌جویی در زمان و هزینه می‌شود. با توجه به پیچیدگی‌های زمین‌شناسی و استفاده گسترده از منابع آبی رودخانه قطور در فعالیت‌های کشاورزی، بررسی کیفیت آب آن از اهمیت بالایی برخوردار است. تاکنون تحقیقات محدودی به بررسی دقیق رابطه میان پارامترهای کیفی آب مانند نسبت جذب سدیم (SAR)و هدایت الکتریکی (EC) در این منطقه پرداخته‌اند. از این‌رو، نیاز به مدل‌های پیش‌بینی دقیق برای مدیریت بهینه منابع آب و جلوگیری از کاهش کیفیت آب بیش از پیش مورد نیاز می‌باشد. لذا، این پژوهش با استفاده از مدل‌های یادگیری ماشین شامل جنگل تصادفی (RF) و رگرسیون بردار پشتیبان (SVR) بهینه‌سازی شده با الگوریتم‌های ژنتیک (GA)، مگس میوه (FOA) و بهترین توپ طلایی (GBO) تلاش می‌کند دقت پیش‌بینی این پارامترها را ارتقا داده و به بهبود مدیریت منابع آبی منطقه کمک کند. در نهایت، مدل‌ها با استفاده از معیارهای ارزیابی مختلفی مورد ارزیابی قرار خواهند گرفت.

مواد و روش‌ها

منطقه مطالعه

در سال‌های اخیر کیفیت رودخانه ارس به‌دلیل عواملی مانند فعالیت‌های شدید کشاورزی و ضایعات کارخانه‌جات آبمیوه‌سازی به‌طور بالقوه کاهش یافته و کیفیت آب دو رودخانه قطور و آق در شاخه سمت راست رود ارس را تحت تأثیر قرار داده است (Nasehi et al., 2013). در تحقیق حاضر، تنها رودخانه قطور مورد بررسی قرار گرفته است. قطور رودخانه‌ای در شرق ترکیه و شمال غربی ایران است که از شهر سرای در شرق ترکیه سرچشمه می‌گیرد و سرانجام در نزدیکی شهر جلفا ایران به رود ارس می‌پیوندد. رودخانه قطور در منطقه‌ای با زمین‌شناسی پیچیده واقع شده است که عمدتاً شامل سازندهای رسوبی و آذرین می‌باشد. این منطقه بخشی از زون زمین‌ساختی زاگرس به شمار می‌رود و در برخی نواحی تحت تأثیر فعالیت‌های تکتونیکی قرار گرفته است. ساختارهای زمین‌شناسی منطقه شامل لایه‌های آهکی، شیل و ماسه‌سنگ است که به‌دلیل فرسایش، تنوع مورفولوژیکی قابل توجهی ایجاد کرده‌اند. کاربری اراضی اطراف رودخانه قطور به‌طور عمده شامل فعالیت‌های کشاورزی و باغداری است، به‌ویژه در نواحی پایین‌دست که منابع آبی رودخانه به آبیاری زمین‌های زراعی اختصاص می‌یابد. در بخش‌های کوهستانی و بالادست، مراتع و جنگل‌های پراکنده وجود دارد که دارای توپوگرافی شیب‌دار و خاک‌های نسبتاً ناپایدار می‌باشند. رودخانه آق نیز از ارتفاعات شهرستان چالدران در مرز ایران و ترکیه سرچشمه می‌گیرد و پس از عبور از دره قره، وارد دشت چاپیاره شده و در روستای مراکان به رودخانه قطور می‌پیوندد (Aghapour and Mahvi 2005).

شکل (۱) موقعیت جغرافیایی ایستگاه مورد بررسی را نشان می‌دهد. علاوه‌بر این، جدول (۱) آمار مجموعه داده‌های آموزش و آزمایش در رودخانه قطور را برای هر چهار پارامتر نشان می‌دهد. داده‌های مورد استفاده در این تحقیق از سازمان آب منطقه‌ای استان آذربایجان غربی و مربوط به سال‌های ۱۳۹۸-۱۳۷۲ بوده است. همچنین ۶۷ درصد داده‌ها برای آموزش مدل‌ها و ۳۳ درصد به‌طور تصادفی برای آزمایش استفاده شده است.

جنگل تصادفی (RF)

الگوریتم جنگل تصادفی یک الگوریتم یادگیری گروهی قدرتمند است که به‌طور گسترده برای کارهای متنوعی مانند رگرسیون، طبقه‌بندی و یادگیری بدون نظارت استفاده می‌شود. جنگل تصادفی توسط بریمن ارائه شد و به‌دلیل عملکرد برترش شناخته شده است (Breiman 2001). این الگوریتم با ساخت مجموعه‌ای از درخت‌ها انجام می‌شود، به‌طوری‌که درخت تکی بر روی زیرمجموعه‌های مختلفی از داده‌های آموزشی منحصر به فرد آموزش داده می‌شود (Biau and Scornet 2016). این روش به‌عنوان "کیسه‌بندی" (bagging) شناخته می‌شود و اجازه می‌دهد تا تنوع کنترل شده را در بین درختان معرفی کند (Altman and Krzywinski 2017). یکی از ویژگی‌های کلیدی جنگل تصادفی استفاده از نمونه‌های خارج از کیسه (out-of-bag) است (Janitza and Hornung 2018). این نمونه‌ها نقاط داده‌ای هستند که در آموزش یک درخت خاص وجود ندارند و برای ارزیابی عملکرد تعمیم مدل استفاده می‌شوند. علاوه‌بر این، تکنیک جنگل تصادفی معیاری با اهمیت متغیر ارائه می‌دهد که به‌صورت عددی تأثیر هر مشخصه ورودی را بر فاز پیش‌بینی ارزیابی می‌کند.

DOI: 10.22055/jise.2024.47590.2131

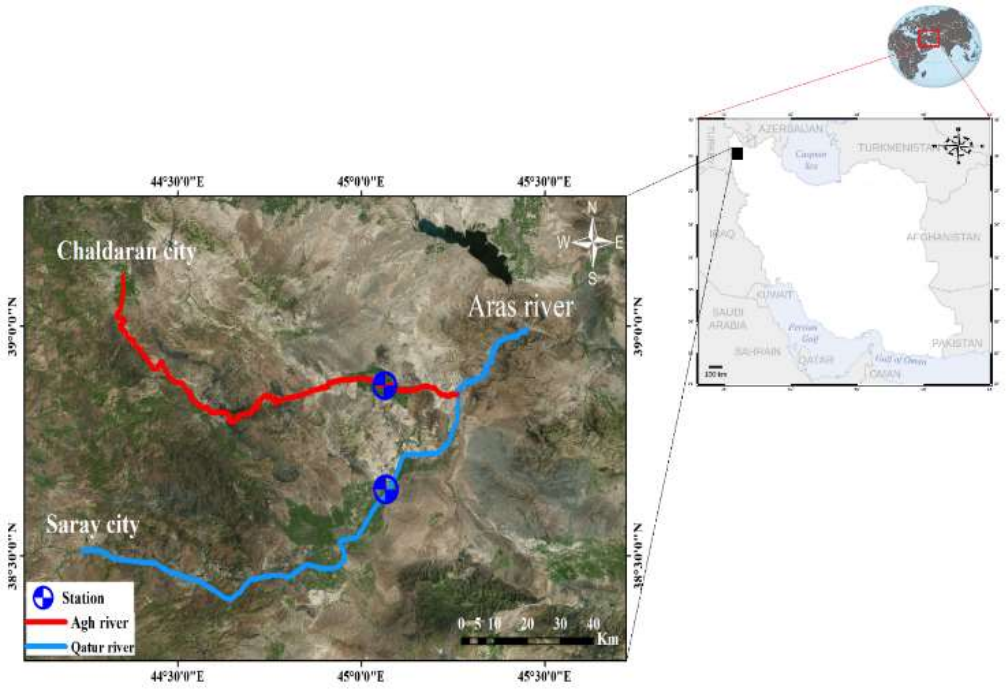


Fig. 1- Geographical location of the Aras basin.
شکل ۱- موقعیت جغرافیایی حوضه ارس

جدول ۱- آمار مجموعه داده‌های آموزش و آزمایش در رودخانه قطور
Table 1- Statistics of training and testing datasets in Qatur river

Statistical parameters	Training set				Testing set			
	SO_4	Cl	SAR	EC	SO_4	Cl	SAR	EC
Unit	mg/lit	mg/lit	-	dS/m	mg/lit	mg/lit	-	dS/m
Mean	2.12	4.82	2.53	1148.19	2.15	4.7	2.53	1132.7
Minimum	0.1	0.05	0.06	200	0.22	0.1	0.06	240
Maximum	10.28	170	40.76	20150	26.63	114	33.86	12870
Standard deviation	2.91	14.16	4.5	1723.12	3.13	12.22	4.35	1505.6

رگرسیون بردار پشتیبانی (SVR)

رگرسیون بردار پشتیبان (SVR) یک الگوریتم متعلق به حوزه یادگیری ماشین است که به‌طور خاص برای اهداف رگرسیونی طراحی شده است. این الگوریتم از الگوریتم ماشین بردار پشتیبان (Support vector machine) مشتق شده است که عمدتاً برای مسائل طبقه‌بندی استفاده می‌شود. SVR توسط Vapnik و Chervonenkis (1974) معرفی شد. رگرسیون بردار پشتیبان (SVR) به‌طور خاص برای رسیدگی به مشکلات رگرسیون طراحی شده است، که در آن هدف اولیه پیش‌بینی یک متغیر پیوسته است (Hamdi et al., 2018). از یک تابع به نام تابع ضرر غیرحساس (e-insensitive loss) برای محاسبه خطاها در داده‌های آموزشی استفاده می‌کند (Karal 2017). این تابع اجازه می‌دهد تا یک حد تحمل یا حاشیه خطا در پیش‌بینی‌ها مشخص شود.

الگوریتم ژنتیک (GA)

الگوریتم‌های ژنتیک (Genetic Algorithm's) نوعی از مدل‌های غیرقطعی هستند که از نظریه داروین الهام گرفته شده‌اند (Hüe 1997). برخلاف سایر روش‌های بهینه‌سازی، GAها نیازی به محاسبه مشتقات جزئی ندارند یا با مسائل اعداد صحیح مختلط سروکار دارند (McKinney and Lin 1994). در عوض، GAها بر ارزیابی تابع هدف تمرکز می‌کنند. یکی از مزایای GA این است که آن‌ها راه‌حل‌ها را از چندین نقطه شروع، جستجو می‌کنند نه فقط از یک نقطه به‌خصوص. این به کاهش خطر گیر افتادن در یک بهینه محلی کمک می‌کند (Albadr et

al., 2020). GAها از قوانین انتقال احتمالی برای هدایت فرآیند جستجو استفاده می‌کنند و از کاوش متنوع در فضای راه‌حل اطمینان می‌دهند (Costa-Carrapiço et al., 2020). پیشرفت سودمندترین راه‌حل‌ها در الگوریتم‌های ژنتیک با به‌دست آوردن نمونه‌های نماینده از کل طیف راه‌حل‌های احتمالی انجام می‌شود. نمونه‌ها متعاقباً از طریق بهره‌برداری از عملگرهای ژنتیکی، مانند تقاطع و جهش، برای تولید راه‌حل‌های جدید با هم ترکیب می‌شوند. این سیر عمل تا زمانی که معیارهای فسخ مقرر برآورده نشود، ادامه می‌یاب (Koohestani 2020). پارامترهای GA حاصل شامل انتخاب والدین، نرخ جهش و نرخ تقاطع است (Hassanat et al., 2019). نرخ تقاطع پتانسیل اجرای عملگر تقاطع را برای راه‌حل‌های والد انتخابی روشن می‌کند، درحالی‌که نرخ جهش احتمال اتخاذ یک عنصر ژنتیکی انتخاب شده به‌طور تصادفی را در یک راه‌حل فرزند تازه تشکیل شده مشخص می‌کند (Schmitt 2001).

الگوریتم بهینه سازی مگس میوه (FOA)

الگوریتم بهینه‌سازی مگس میوه (FOA) یک الگوریتم بهینه‌سازی الهام گرفته از طبیعت است که بر اساس رفتار مگس‌های میوه است که برای اولین بار توسط Pan (2012) ارائه شد، عمل می‌کند. این یک الگوریتم فراابتکاری است که هدف آن یافتن بهترین راه‌حل برای یک مشکل معین با تقلید از رفتار جستجوگر مگس میوه است (Wang et al., 2022). در طبیعت، مگس‌های میوه با حرکت تصادفی در محیط خود رفتار منحصر به‌فردی در جستجوی منابع غذایی از خود نشان می‌دهند. آن‌ها بر اساس رایحه خود جذب منابع غذایی می‌شوند و تمایل دارند در اطراف امیدوارکننده‌ترین آن‌ها جمع شوند (Shao et al., 2020).

الگوریتم بهینه‌سازی توپ طلایی (GBO)

الگوریتم بهینه‌سازی توپ طلایی (Golden Ball Optimization) یک روش نوآورانه فراابتکاری است که از تاکتیک‌ها و استراتژی‌های بازی فوتبال الهام گرفته است. در این الگوریتم، بازیکنان به‌عنوان عوامل جستجو در فضای مسئله تلقی می‌شوند که با همکاری و تعامل، هدف مشترک (یافتن بهترین راه‌حل) را دنبال می‌کنند (Osaba et al., 2014). توپ طلا نمادی از بهینه‌ترین راه‌حل ممکن در این فضاست که بازیکنان (عوامل) با تبادل اطلاعات و موقعیت‌یابی‌های بهینه سعی در نزدیک شدن به آن دارند. این روش از دو فرآیند اکتشاف (گستردگی جستجو در فضای راه‌حل‌ها) و بهره‌برداری (متمرکز شدن بر نواحی امیدبخش) بهره می‌برد تا تعادلی میان یافتن راه‌حل‌های جدید و اصلاح آن‌ها ایجاد کند. ویژگی برجسته بهینه‌سازی توپ طلا در توانایی آن برای فرار از بهینه‌های محلی و دسترسی به نواحی بهینه‌تر فضای جستجو نهفته است (Worawattawechai et al., 2022). GBO با الهام از تصمیم‌گیری‌های گروهی و تاکتیک‌های فوتبال، مدلی قدرتمند برای حل مسائل بهینه‌سازی است که پیچیدگی و تعاملات بین عوامل را به‌خوبی در نظر می‌گیرد.

توسعه مدل‌ها

در این تحقیق از الگوریتم ژنتیک برای افزایش دقت مدل جنگل تصادفی مستقل استفاده شد. برای پیاده‌سازی مدل هیبریدی RF-GA ابتدا داده‌ها به دو قسمت آموزشی و آزمایشی تقسیم شده و وارد مدل می‌شوند. سپس جامعه اولیه برای داده‌ها در نظر گرفته شده و پس از انتخاب پارامترهای بهینه‌سازی، مدل جنگل تصادفی داده‌ها را طبقه‌بندی کرده و تحلیل برازش از داده‌ها گرفته می‌شود. این تحلیل تا آخرین نسل که بهترین دقت به‌دست می‌آید، ادامه دارد. در نهایت، با استفاده از معیارهای ارزیابی عملکرد، مدل ترکیبی محاسبه شده و نتایج مدل‌سازی نمایش داده می‌شود. مدل RF-GA در نرم‌افزار RapidMiner پیاده‌سازی شده است و مراحل اجرای الگوریتم ژنتیک در شکل (۲) نشان داده شده است.

FOA یک الگوریتم بهینه‌سازی مبتنی‌بر رفتار مگس میوه بوده که برای پیاده‌سازی این مدل داده‌ها به دو قسمت آموزشی و آزمون تقسیم شده و وارد مدل می‌شوند. برای اجرای مدل، پارامترهای بهینه‌سازی شامل اندازه جمعیت و حداکثر تولید توسط کاربر برای مدل تعریف می‌شود. سپس فاصله و جهت جستجوی غذا به‌صورت تصادفی برای هر مگس میوه تعریف می‌شود. در مرحله بعد پارامتر S_i توسط مدل محاسبه شده و برازش مناسب به‌دست می‌آید. سپس بهترین غلظت بو در نقاط X و Y یافت می‌شود. در مرحله بعد، اگر حداکثر تکرار در نظر گرفته شده توسط کاربر بیشتر از نسل باشد، مراحل قبلی تکرار می‌شود، در غیر این‌صورت با استفاده از ارزیابی عملکرد محاسبه می‌شود. مدل هیبریدی SVR-FOA در نرم افزار MATLAB پیاده‌سازی شده و فرآیند اجرای الگوریتم مگس میوه در شکل (۳) نشان داده شده است.

 DOI: 10.22055/jise.2024.47590.2131

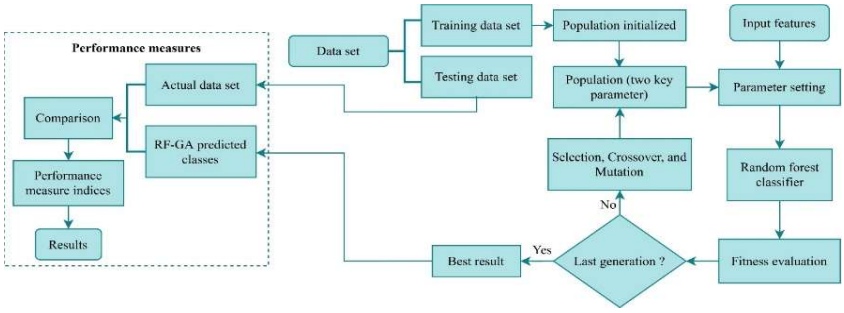


Fig. 2- General process of GA model
شکل ۲- فرآیند کلی مدل GA

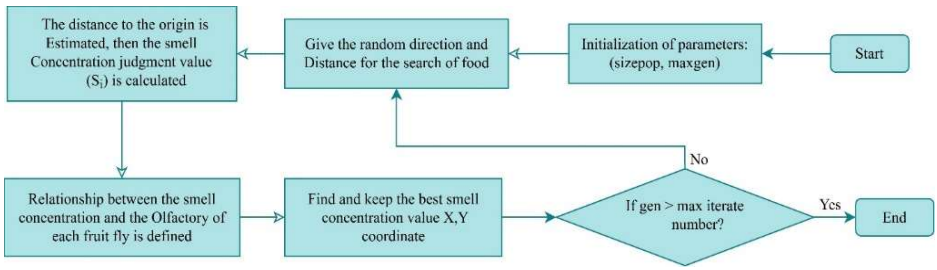


Fig. 3- General process of FOA model
شکل ۳- فرآیند کلی مدل FOA

مدل‌های هیبریدی RF-GBO و SVR-GBO در نرم‌افزار Python پیاده‌سازی شده و فلوچارت مربوط به الگوریتم بهینه‌سازی بهترین توپ طلایی در شکل (۴) آورده شده است. برای اجرای این الگوریتم در ابتدا جمعیت بازیکنان (population) به‌طور تصادفی تولید می‌شوند. اندازه جمعیت نیز در این مرحله مشخص می‌شود. سپس جمعیت بازیکنان به چند تیم تقسیم می‌شوند تا در هر فصل(iteration) با یکدیگر رقابت کنند. بازیکنان با استفاده از داده‌ها آموزش داده می‌شوند (training dataset) تا بازی شروع شود و در هر مرحله امتیازاتی به هر تیم تعلق می‌گیرد تا حداکثر امتیاز حاصل شود. در نهایت اگر تعداد تکرارها بیش‌تر از حداکثر تکرار (max iteration) نباشد مدل به تقسیم‌بندی بازیکنان ارجاع داده می‌شود تا بقیه مراحل از سر گرفته شوند.

پارامترهای مورد نیاز برای دست‌یابی به حداکثر کارایی در مطالعه حاضر برای اجرای مدل‌های SVR و SVR-FOA در جدول (۲) آورده شده است. در هر دو مدل SVR و SVR-FOA، پارامترهای ε، γ، و C توابع محوری هستند. مدل ε، همچنین به‌عنوان اپسیلون شناخته می‌شود، نشان‌دهنده حاشیه تحمل در اطراف خط رگرسیون است. γ، که به‌عنوان گاما شناخته می‌شود، پارامتری است که تأثیر یک نمونه تمرین فردی را مشخص می‌کند. از سوی دیگر، C مربوط به پارامتر منظم‌سازی است که بر تعادل بین دستیابی به حداقل خطای آموزشی و پیچیدگی مدل مرزی حاکم است. همچنین در مدل هیبریدی SVR-FOA، نرخ یادگیری (Learning Rate) و برد پرواز (Flight Range) دو متغیر اساسی هستند که در فرآیند بهینه‌سازی اهمیت محوری دارند. LR که به‌عنوان نرخ یادگیری نیز شناخته می‌شود، میزان هر مرحله یا سرعتی را که الگوریتم در آن دانش را از تکرارهای قبلی کسب می‌کند، تعیین می‌کند. برعکس، FR نشان‌دهنده حداکثر بازه‌ای است که مگس میوه می‌تواند در طول هر تکرار بپوشاند. پارامترهای maxgen و sizepop در هر سه مدل ترکیبی نشان‌دهنده حداکثر تولید و اندازه جمعیت است که پس از آزمون و خطا به‌دست می‌آید. Exploration factor میزان تعادل بین جستجوی فضای جدید و بهینه‌سازی فضای قبلی را تعیین می‌کند و Learning coefficient نشان‌دهنده نرخ یادگیری در هر تکرار می‌باشد. Inertia weight میزان اینرسی (ماندگاری حرکت) را در طول تکرارها کنترل می‌کند و Convergence tolerance محدوده همگرایی الگوریتم را مشخص می‌کند.

همچنین در این تحقیق از سه سناریو مختلف برای ارزیابی نقش پارامترهای ورودی بر پارامترهای SAR و EC استفاده شده که در جدول (۳) آورده شده است. ترتیب سناریوهای در نظر گرفته شده بر اساس ضریب همبستگی پیرسون بوده به‌طوری‌که سناریوی اول کم‌ترین همبستگی و سناریوی آخر بیش‌ترین همبستگی را با پارامترهای SAR و EC داشته‌اند. دلیل انتخاب ضریب همبستگی پیرسون به سادگی و قابلیت تفسیر آن بازمی‌گردد که امکان ارزیابی سریع و واضح روابط خطی بین متغیرها را فراهم می‌کند.

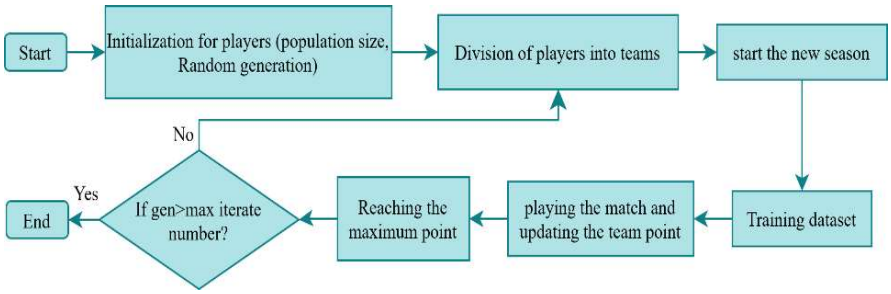


Fig. 4- General process of GBO algorithm
شکل ٤- فرآیند کلی الگوریتم GBO

جدول ۲- مقادیر هایپرپارامترهای مورد استفاده برای مدل‌های مورد مطالعه
Table 2- The values of hyperparameters used for the studied models

parameter	SVR	FOA	GA	GBO
LR	-	0.1	-	-
FR	-	0.5	-	-
maxgen	-	100	200	200
sizepop	-	30	50	50
kernel	RBF	-	-	-
C	1	-	-	-
ε	0.001	-	-	-
γ	0.01	-	-	-
Exploration factor	-	-	-	0.7
Learning coefficient	-	-	-	1.5
Inertia weight	-	-	-	0.7
Convergence tolerance	-	-	-	10^{-5}

جدول ۳- سناریوهای مورد استفاده برای مدل‌های مورد مطالعه
Table 3- Scenarios used for the studied models

Scenario	Input		Output
1	SO_4		SAR-EC
2	Cl		SAR-EC
3	SO_4	Cl	SAR-EC

معیارهای ارزیابی

در مطالعه حاضر برای پیش‌بینی پارامترهای SAR و EC، معیارهای ارزیابی ضریب همبستگی (Correlation coefficient)، جذر میانگین مربعات خطا (Root mean squared error)، جذر میانگین مربعات خطای نرمال شده (Normalized Root mean squared error)، ضریب نش‌ساتکلیف (Nash-Sutcliffe coefficient) و شاخص توافق ویلموت (Willmott's index of agreement) استفاده شد که در معادله‌های (۱) تا (۵) آورده شده است.

$$R = \frac{\sum_{i=1}^{N}(O_i-\overline{O_i})(P_i-\overline{P_i})}{\sqrt{\sum_{i=1}^{N}(O_i-\overline{O_i})^2\sum_{i=1}^{N}(P_i-\overline{P_i})^2}} \quad (۱)$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(P_i-O_i)^2} \quad (۲)$$

$$NRMSE = \frac{RMSE}{Max(O_i)-Min(O_i)} \quad (۳)$$

$$NS = 1-\left[\frac{\sum_{i=1}^{N}(O_i-P_i)^2}{\sum_{i=1}^{N}(O_i-\overline{O_i})^2}\right] \quad (۴)$$

$$WI = 1-\left[\frac{\sum_{i=1}^{N}(O_i-P_i)^2}{\sum_{i=1}^{N}(|P_i-\overline{O_i}|+|O_i-\overline{O_i}|)^2}\right] \quad (۵)$$

در روابط بالا N تعداد مشاهدات، P_i و O_i به‌ترتیب مقادیر پیش‌بینی شده و مشاهداتی را نشان می‌دهند. همچنین $\overline{P_i}$ و $\overline{O_i}$ به‌ترتیب میانگین مقادیر پیش‌بینی شده و مشاهداتی را نشان می‌دهند.

 DOI: 10.22055/jise.2024.47590.2131

نتایج و بحث

در تحقیق حاضر، سه سناریو با استفاده از متغیرهای پیش‌بینی کننده شامل Cl و SO_4 به‌عنوان ورودی مدل‌ها در نظر گرفته شد (جدول ۳). عملکرد مدل‌های RF، RF-GA، SVR، و SVR-FOA در این مجموعه داده تأیید شد. نتایج مدل‌های مورد استفاده در دوره آزمایشی را بر اساس معیارهای R، RMSE، NRMSE، NS و WI برای ایستگاه رودخانه قطور در جدول (۴) ارائه شده است. ارزیابی سناریوهای مختلف RF برای پارامتر SAR نشان می‌دهد که سناریوی سوم با R= ۰/۹۷۷، RMSE = ۰/۹۲۳ (meq/lit)، NRMSE = ۰/۰۲۷، NS = ۰/۹۵۵ و WI = ۰/۹۸۹ دقت بالاتری نسبت به سایر سناریوهای RF داشته است. برای مدل RF-GA، مقایسه خطاهای پیش‌بینی برای هر سه سناریو نشان می‌دهد که RF-GA-3 کم‌ترین خطا (RMSE =۰/۸۰۹ meq/lit) را نسبت به سایر سناریوهای این مدل داشته است. در میان سناریوهای مختلف SVR، سناریوی اول دارای بیش‌ترین خطا (RMSE =۲/۱۹۴ meq/lit، و NRMSE =۰/۰۶۵) و سناریوی سوم دارای کم‌ترین خطا (RMSE =۱/۰۶۸ meq/lit، و NRMSE =۰/۰۳۲) بوده است. برای مدل ترکیبی SVR-FOA، دو سناریوی آخر تقریباً ضریب همبستگی یکسانی دارند، با این حال، سناریوی دوم دارای مقدار خطای نسبتاً پایین‌تری (RMSE =۰/۹۹۸ meq/lit) و از این رو دارای دقت بالاتری است.

ارزیابی اولین سناریوی مدل RF با مدل RF-GA برای پیش‌بینی هدایت الکتریکی نشان می‌دهد که در مدل ترکیبی، معیارهای خطای RMSE و NRMSE به‌ترتیب ۳۰/۵ و ۳۰/۶ درصد کاهش یافته‌اند. در سناریوی دوم مدل‌های RF منفرد و ترکیبی، مقدار معیارهای R، NS، WI و NRMSE بدون تغییر باقی‌ماند، درحالی‌که مقدار RMSE در مدل ترکیبی اندکی کاهش یافت. همچنین برای مدل‌های SVR منفرد و هیبرید شده با مگس میوه، ارزیابی سناریوی اول دو مدل نشان می‌دهد که مقدار شاخص‌های خطای RMSE و NRMSE به‌ترتیب ۱۶/۹ و ۱۶/۲ درصد کاهش یافته است. برای سناریوی دوم، شاخص‌های R و RMSE در مدل ترکیبی به‌ترتیب یک و سه درصد بهبود یافته‌اند و سایر شاخص‌ها تقریباً بدون تغییر باقی مانده‌اند. در سناریوی سوم، مقایسه دو مدل نشان داد که در مدل ترکیبی، شاخص‌های خطای RMSE و NRMSE به‌ترتیب ۳۰/۲ و ۳۲/۳ درصد کاهش و سایر شاخص‌ها اندکی افزایش یافتند.

مقایسه نتایج مدل‌های مختلف SVR در پیش‌بینی پارامتر SAR نشان می‌دهد که SVR-FOA-2 توانسته دقت مدل SVR-1 را حدود ۷۴/۹۴ درصد افزایش دهد (با کاهش RMSE). همچنین RF-GA-3 نیز توانست دقت مدل RF-1 را حدود ۹۷/۵ درصد افزایش دهد. همچنین برای پیش‌بینی پارامتر EC، مدل SVR-FOA-3 توانست دقت مدل SVR-1 را حدود ۱۲۷/۵۵ درصد افزایش دهد و RF-GA-3 نیز توانست دقت مدل RF-1 را حدود ۱۳۹/۱ درصد افزایش دهد.

در نهایت، ارزیابی مدل‌های مورد مطالعه از پیش‌بینی پارامتر SAR نشان داد که مدل RF-GA-3 با داشتن معیارهای R =۰/۹۸۳، RMSE = ۰/۸۰۹ (meq/lit)، NRMSE ۰/۰۲۴ =، NS = ۰/۹۶۵ و WI = ۰/۹۹۱ بهترین عملکرد را نسبت به مدل‌های دیگر داشته است. برای پارامتر EC نیز نتایج نشان داد که مدل RF-GA-3 با معیارهای R =۰/۹۹۵، (meq/lit) RMSE = ۰/۱۴۶، NRMSE = ۰/۰۱۲، NS = ۰/۹۹۱ و WI = ۰/۹۹۸ بالاترین دقت را داشته است. نتایج این تحقیق با نتایج Rahnama et al. (2020) که کاربرد مدل ANFIS را با سیستم فازی خوشه‌بندی c-means برای پیش‌بینی SAR در نقاط مختلف ایران بررسی کردند و نشان دادند که مدل پیشنهادی مقدار SAR را با ضریب تعیین ۰/۴۸ پیش‌بینی می‌کند، همخوانی دارد. همچنین این نتایج در تطابق با نتایج Hassan et al. (2021) می‌باشد، که پارامترهای کیفیت آب را با استفاده از الگوریتم‌های طبقه‌بندی پیش‌بینی کردند و نشان دادند که مدل MLR با دقت ۹۹/۸۳ درصد به‌عنوان یک مدل ارزشمند در پیش‌بینی پارامتر EC شناخته می‌شود. همچنین از محدودیت‌های تحقیق حاضر می‌توان به عدم دسترسی به داده‌های سایر ایستگاه‌ها برای ارزیابی دقت مدل‌های مورد استفاده اشاره نمود. بنابراین پیشنهاد می‌شود از داده‌های ایستگاه‌های مختلف برای بررسی توانایی مدل‌ها برای مطالعه‌های آتی استفاده شود.

ارزیابی و مقایسه نتایج مدل‌های منفرد و هیبریدی نشان می‌دهد که مدل‌های هیبریدی به‌شکل مؤثری سبب کاهش خطای پیش‌بینی مدل‌های منفرد شده‌اند. به‌طوری‌که ارزیابی سناریوهای مختلف مدل RF-GBO نشان می‌دهد سناریو سوم در پیش‌بینی نسبت جذب سدیم بهترین عملکرد را در میان تمام سناریوها داشته است. ضریب همبستگی بالای ۰/۹۸ نشان‌دهنده همخوانی بسیار قوی بین مقادیر پیش‌بینی شده و مشاهده شده SARاست، که این امر به‌دقت بالای مدل اشاره دارد. همچنین، مقدار RMSE پایین (۰/۸۷۷) وNRMSE برابر با ۰/۰۲۶ حاکی از این است که این مدل از خطاهای بسیار کمتری نسبت به دیگر مدل‌ها برخوردار بوده است. مقدار ضریب کارایی نش–ساتکلیف (NS) برابر با ۰/۹۵۹ و شاخص ویلموت (WI) معادل ۰/۹۹ نیز نشان می‌دهد که مدل RF-GBO-3 در بازتولید داده‌های مشاهده‌شده، کارایی بسیار بالایی داشته است. همچنین، مدل‌های RF-GBO-2 و SVR-GBO-3 با ضریب همبستگی‌های نزدیک به هم (۰/۹۷۴ و ۰/۹۷۲)، عملکرد قابل قبولی داشتند، اما دقت آن‌ها اندکی کمتر از RF-GBO-3 بوده است.

جدول ٤- نتایج ارزیابی مدل‌های مورد استفاده در ایستگاه رودخانه قطور

Table 4- The evaluation results of the models used in Qatur river station

Parameter	SAR					EC				
Scenario	R	RMSE (meq/lit)	NRMSE	NS	WI	R	RMSE (dS/m)	NRMSE	NS	WI
RF-1	0.862	2.348	0.069	0.708	0.925	0.876	0.813	0.064	0.708	0.931
RF-2	0.969	1.078	0.032	0.938	0.984	0.989	0.219	0.017	0.979	0.995
RF-3	**0.977**	**0.923**	**0.027**	**0.955**	**0.989**	**0.992**	**0.215**	**0.017**	**0.900**	**0.995**
RF-GA-1	0.896	1.927	0.057	0.803	0.943	0.918	0.598	0.047	0.842	0.957
RF-GA-2	0.977	0.939	0.028	0.953	0.988	0.990	0.213	0.017	0.980	0.995
RF-GA-3	**0.983**	**0.809**	**0.024**	**0.965**	**0.991**	**0.995**	**0.146**	**0.012**	**0.991**	**0.998**
RF-GBO-1	0.887	2.011	0.059	0.786	0.938	0.902	0.661	0.052	0.807	0.948
RF-GBO-2	0.974	0.981	0.029	0.949	0.987	0.989	0.225	0.018	0.978	0.994
RF-GBO-3	**0.980**	**0.877**	**0.026**	**0.959**	**0.990**	**0.995**	**0.154**	**0.012**	**0.990**	**0.997**
SVR-1	0.869	2.194	0.065	0.745	0.915	0.869	0.764	0.060	0.743	0.914
SVR-2	0.966	1.128	0.033	0.933	0.983	0.987	0.241	0.019	0.974	0.993
SVR-3	**0.970**	**1.068**	**0.032**	**0.940**	**0.984**	**0.991**	**0.229**	**0.018**	**0.977**	**0.995**
SVR-FOA-1	0.913	1.840	0.054	0.821	0.942	0.908	0.645	0.051	0.817	0.942
SVR-FOA-2	**0.974**	**0.998**	**0.028**	**0.947**	**0.987**	0.996	0.234	0.019	0.976	0.993
SVR-FOA-3	0.973	1.050	0.031	0.942	0.984	**0.994**	**0.169**	**0.013**	**0.987**	**0.999**
SVR-GBO-1	0.874	2.166	0.064	0.752	0.916	0.988	0.235	0.019	0.976	0.994
SVR-GBO-2	0.968	1.097	0.032	0.936	0.983	0.993	0.182	0.014	0.985	0.996
SVR-GBO-3	**0.972**	**1.042**	**0.031**	**0.943**	**0.984**	**0.994**	**0.176**	**0.014**	**0.986**	**0.997**

در پیش‌بینی هدایت الکتریکی نیز، مدل RF-GBO-3 مجدداً عملکرد بهتری را نسبت به سایر سناریوهای الگوریتم بهینه‌سازی GBO داشته است. این مدل با داشتن ضریب همبستگی ۰/۹۹۵، RMSE برابر با ۰/۱۵۴ و NRMSE معادل ۰/۰۱۲، دقت مناسبی را در پیش‌بینیEC ارائه می‌دهد. همچنین، مقادیر NS و WI به‌ترتیب برابر با ۰/۹۹ و ۰/۹۹۷ نشان می‌دهد که مدل به‌خوبی توانسته است داده‌های مشاهده‌شده را پیش‌بینی کرده و بهینه‌ترین پیش‌بینی‌ها را فراهم آورد. مدل-SVR GBO-3نیز با ضریب همبستگی ۰/۹۹۴ و RMSE برابر ۰/۱۷۶ عملکردی نزدیک به RF-GBO-3 داشته است.

مقایسه عملکرد مدل‌های هیبریدی و منفرد در تحقیقات پیشین نیز نشان‌دهنده برتری قابل ملاحظه مدل‌های هیبریدی می‌باشد. برای مثال Wu و Wang (2022) در تحقیقی از دو مدل جدید یادگیری ماشین شامل جنگل و تصادفی و جنگل تصادفی ترکیب شده با تجزیه حالت تجربی مجموعه کامل با نویز تطبیقی (CEEMDAN) برای پیش‌بینی کیفیت آب در حوضه رودخانه توالاتین استفاده کردند. مقایسه نتایج این مطالعه نشان داد که مدل CEEMDAN-RF با داشتن کم‌ترین جذر میانگین درصد خطا (۳/۹ درصد) بهترین عملکرد را داشته و سبب بهبود پایداری مدل جنگل تصادفی منفرد گردید. Liu et al. (2013) در مطالعه‌ای با استفاده از یک رویکرد ترکیبی، شامل رگرسیون بردار پشتیبان ترکیب شده با الگوریتم ژنتیک با ارزش واقعی (real-value genetic algorithm) به پیش‌بینی کیفیت آب کارخانه‌های آبزی YiXing ، در چین پرداختند. نتایج این بررسی نشان داد که RGA-SVR از مدل‌های SVR و انتشار پس‌پاپ (back-propagation)بر اساس ریشه میانگین مربعات خطا بهتر عمل کرد. Su et al. (2022) در تحقیقی با استفاده از مدل رگرسیون بردار پشتیبان هیبرید شده با الگوریتم جستجوی گنجشک بهبودیافته (improved sparrow search algorithm) به پیش‌بینی کیفیت آب پرداختند. نتایج این مطالعه نشان داد که مدل هیبریدی با داشتن ضریب تعیین ۰/۹۸ دقت بالاتری را نسبت به مدل SVR (با داشتن ضریب تعیین ۰/۹۳) داشته است.

با توجه به این که الگوریتم‌های بهینه‌سازی مانند GBO در تحقیقات معدودی مورد استفاده قرار گرفته‌اند. لذا، مطالعه‌های آینده می‌توانند از این الگوریتم در ترکیب با سایر مدل‌های یادگیری ماشین برای مسائل مشابه استفاده کنند. این پیشنهاد می‌تواند به افزایش درک از توانایی این الگوریتم‌ها در مسائل مختلف کمک کند.

شکل (۵) نمودارهای پراکنش برای بهترین سناریو در هر مدل برای پیش بینی SAR در ایستگاه رودخانه قطور را نشان می‌دهد. خط قرمز ۱:۱ بوده و هر چه نقاط به این خط نزدیک‌تر باشند دقت آن مدل بیش‌تر خواهد شد. مقایسه ضریب تعیین در هر چهار مدل نشان می‌دهد که مدل RF-GA-3 با ضریب تعیین ۰/۹۶۵۵ بهتر از سایر مدل‌ها عمل کرده است. پس از مدل -RF GA، مدل RF-GBO-3 نیز دارای مقدار همبستگی بالاتری نسبت به مدل‌های دیگر با ضریب تعیین ۰/۹۵۹۵ می‌باشد.

 DOI: 10.22055/jise.2024.47590.2131

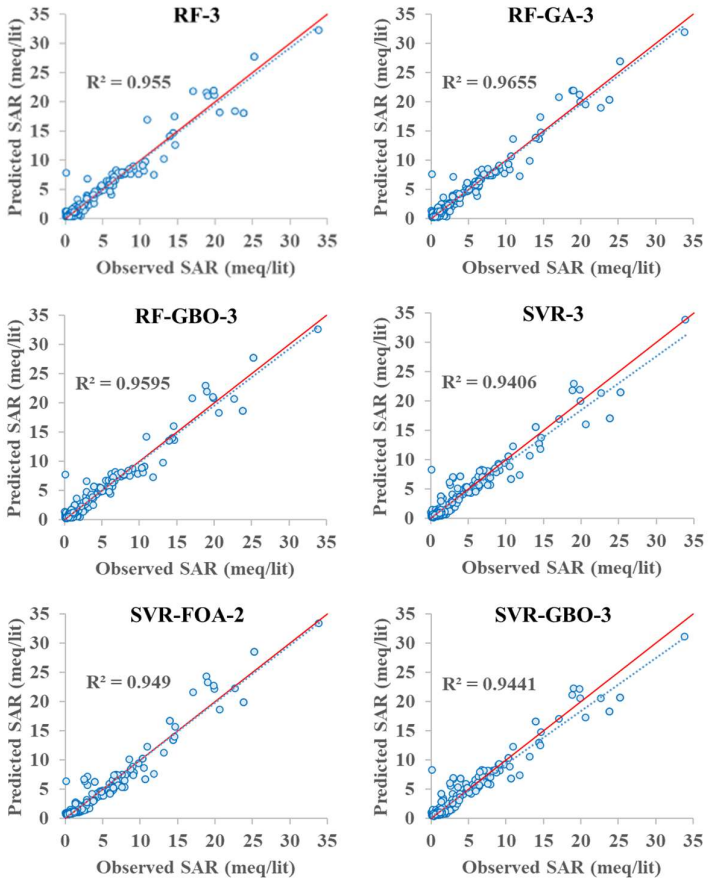


Fig. 5- Scatter plots of the best scenario for predicting the SAR parameter at Qatur River station

شکل ۵- نمودارهای پراکنش بهترین سناریو برای پیش‌بینی پارامتر SAR در ایستگاه رودخانه قطور

شکل (۶) نمودارهای پراکنش را برای بهترین سناریو در هر مدل برای پیش‌بینی EC در ایستگاه رودخانه قطور نشان می‌دهد. خط آبی ۴۵ درجه بوده و هر چه نقاط به این خط نزدیک‌تر باشند، مقدار پیش‌بینی شده به مقدار مشاهداتی نزدیک‌تر خواهد بود. مقایسه ضریب تعیین در هر چهار مدل نشان می‌دهد که مدل RF-GA-3 با ضریب تعیین ۰/۹۹۰۸ بهترین عملکرد را داشته است. پس از مدل RF-GA، مدل SVR-GBO-3 نیز با ضریب تعیین ۰/۹۹۰۳ از دقت بالاتری نسبت به مدل‌های دیگر برخوردار است.

شکل (۷) نمودارهای جعبه‌ای را برای تمام سناریوهای هر مدل در رودخانه قطور نشان می‌دهد. در هر نمودار، جعبه بنفش مقدار مشاهداتی و سایر جعبه‌ها مقدار پیش‌بینی شده می‌باشند. مقایسه جعبه‌های مدل‌های مورد استفاده برای پیش‌بینی SAR با جعبه مشاهداتی نشان می‌دهد که RF-GA-3 و RF-GBO-3 به‌ترتیب نزدیک‌ترین میانگین و میانه را به مقدار مشاهده شده داشته‌اند. همچنین در مدل‌های RF-3 و SVR-FOA-3 اگرچه میانگین و میانه به مقدار مشاهده شده نزدیک بود، اما مقایسه دامنه آن‌ها نشان می‌دهد که هر دو قسمت پایین و بالایی دارای کم‌برآوردی بوده است. برای پارامتر EC، مقایسه جعبه‌ها نشان می‌دهد که سناریوی سوم مدل‌های RF-GA و سپس RF-GBO نزدیک‌ترین میانگین و میانه را نسبت به مقدار مشاهده شده داشته‌اند. در نهایت، مدل SVR-2 کم‌ترین مقدار تخمینی را در بین تمام سناریوها دارا بوده است.

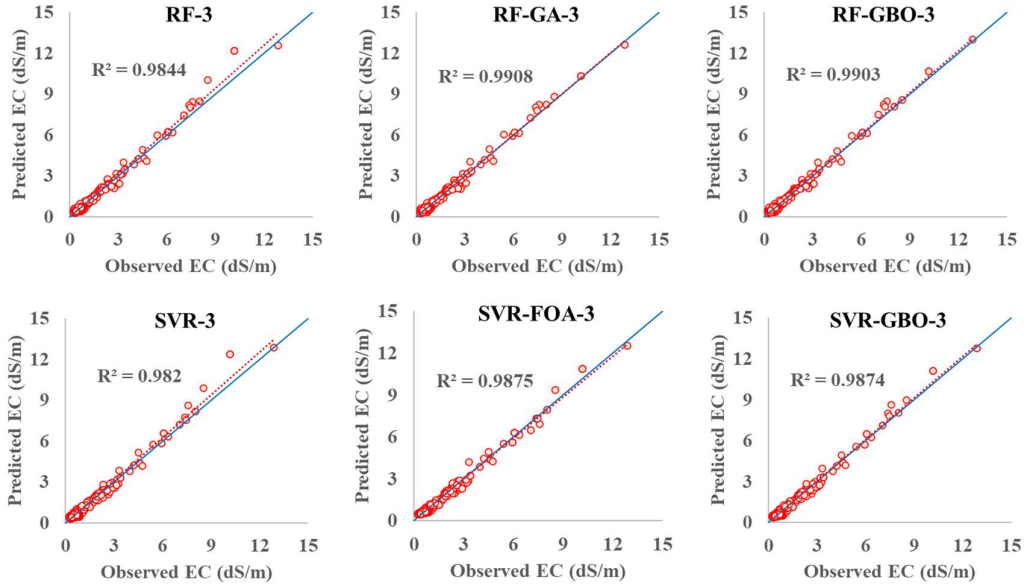


Fig. 6- Scatter plots of the best scenario for predicting the EC parameter at Qatur River station.

شکل ٦- نمودارهای پراکنش بهترین سناریو برای پیش‌بینی پارامتر EC در ایستگاه رودخانه قطور

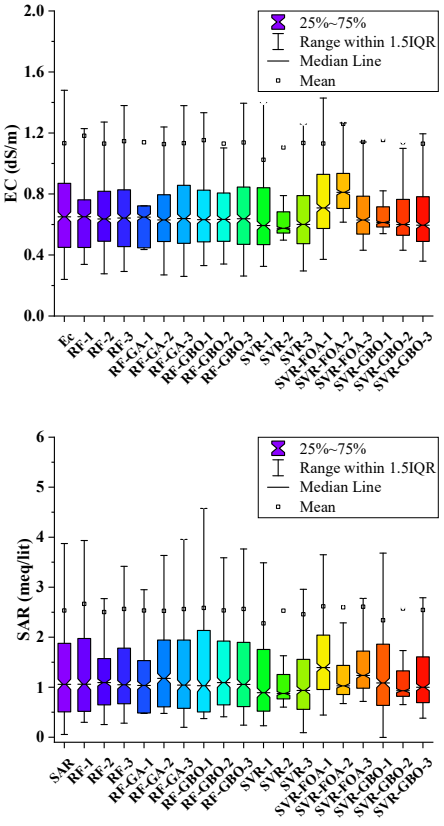


Fig. 7- Boxplots for all scenarios of each model in Qatur River.

شکل ۷- نمودارهای جعبه‌ای برای تمام سناریوهای هر مدل در رودخانه قطور

نتیجه‌گیری

نمودار Wilcox ابزار ارزشمندی است که در ارزیابی مناسب بودن آب برای اهداف آبیاری مورد استفاده قرار می گیرد. نمودار Wilcox آب را با استفاده از پارامترهای SAR و EC به کلاس‌های مختلف تقسیم می‌کند. اما اندازه‌گیری این پارامترها علاوه بر گران بودن، با چالش‌های مختلفی روبه‌رو هستند. بنابراین، این تحقیق به‌منظور پیش‌بینی پارامترهای SAR و EC با استفاده از پارامترهای سهل الوصول شامل سولفات و کلرید در ایستگاه رودخانه قطور انجام شده است. مدل‌های مورد استفاده شامل RF، RF-GA، RF-GBO, SVR, SVR-FOA و SVR-GBO هستند که با استفاده از معیارهای ارزیابی مختلف مورد ارزیابی قرار گرفتند. ارزیابی نتایج این بررسی نشان داد که مدل RF-GA با داشتن ضریب همبستگی ۰/۹۸۳ و ۰/۹۹۵ به‌ترتیب برای پارامترهای SAR و EC بیش‌ترین دقت را در ایستگاه رودخانه قطور داشته است همچنین مدل RF-GBO با داشتن ضریب همبستگی ۰/۹۸ و ۰/۹۹۵ و مدل SVR-FOA نیز با داشتن ضریب همبستگی ۰/۹۷۴ و ۰/۹۹۴ به‌ترتیب برای پارامترهای SAR و EC دقت مناسبی را از خود نشان دادند. ارزیابی نتایج مدل‌ها نشان داد که مدل‌های هیبریدی RF-GA، RF-GBOو SVR-FOA می‌توانند پارامترهای SAR و EC را به‌عنوان یک ابزار قابل اعتماد پیش‌بینی کنند. مدل‌های ترکیبی مورد استفاده پتانسیل قابل‌توجهی برای کمک به کشاورزان و مدیران سیستم‌های آبیاری در تصمیم‌گیری کارآمدتر دارند زیرا پیش‌بینی دقیق پارامترهای SAR و EC منجر به درک صحیح کیفیت آب مصرفی و عملکرد صحیح سیستم‌ها می‌شود.

تشکر و قدردانی

به این وسیله نویسندگان مقاله مراتب تقدیر و تشکر خود را از شرکت آب منطقه‌ای استان آذربایجان غربی در دریافت آمار و نیز از معاونت پژوهش و فناوری دانشگاه ارومیه به خاطر حمایت مالی این مطالعه ابراز می‌نمایند.

تضاد منافع نویسندگان

نویسندگان این مقاله اعلام می‌دارند که هیچ گونه تضاد منافعی در خصوص نگارش و انتشار مطالب و نتایج این پژوهش ندارند.

منابع مالی

حمایت های مالی این تحقیق از محل اعتبار پژوهشی با شماره پژوهانه ۱۹۶۲۱صورت گرفته است. بدینوسیله از حمایت معاونت محترم پژوهشی دانشگاه ارومیه تقدیر می‌شود.

دسترسی به داده‌ها

مجموعه داده‌های تولید شده و یا تحلیل شده در طول مطالعه حاضر، در صورت درخواست معقول، از نویسنده مسئول در دسترس هستند.

مشارکت نویسندگان

نویسنده اول: میلاد شرفی، جمع آوری داده‌ها و نوشتن درفت اولیه مقاله.

نویسنده دوم: جواد بهمنش، راهنمایی نوشتن مقاله، خواندن و اصلاح درفت اولیه.

نویسنده سوم: وحید رضاوردی‌نژاد، راهنمایی نوشتن مقاله، خواندن و اصلاح درفت اولیه.

نویسنده چهارم: سعید صمدیان فرد، راهنمایی نوشتن مقاله، خواندن و اصلاح درفت اولیه.

References

1- Aghapour, A.A. and Mahvi, A.H., 2005. Investigating the chemical quality of Qatur River water and determining the cause of its changes along the way. In *the 8th National Environmental Health Conference*, *University of Tehran, Tehran, Iran*.

2- Ahmadianfar, I ,.Jamei, M. and Chu, X., 2020. A novel hybrid wavelet-locally weighted linear regression (W-LWLR) model for electrical conductivity (EC) prediction in surface water. *Journal of Contaminant Hydrology*, 232(1), pp.103-121. https://doi.org/10.1016/j.jconhyd.2020.103641.

3- Albadr, M.A., Tiun, S., Ayob, M. and Al-Dhief, F., 2020. Genetic algorithm based on natural selection theory for optimization problems. *Symmetry*, 12(11), pp.1758-1771. https://doi.org/10.3390/sym12111758.

4- Alkharabsheh, H.M., Seleiman, M.F., Hewedy, O.A., Battaglia, M.L., Jalal, R.S., Alhammad, B.A., Schillaci, C., Ali, N. and Al-Doss, A., 2021. Field crop responses and management strategies to mitigate soil salinity in modern agriculture: A review. *Agronomy*, 11(11), pp.2299-2317. https://doi.org/10.3390/agronomy11112299.

5- Altman, N. and Krzywinski, M., 2017. Ensemble methods: bagging and random forests. *Nature Methods*, 14(10), pp.933-935 .

6- Amezketa, E. and de Lersundi, J.D.V., 2008. Soil classification and salinity mapping for determining restoration potential of cropped riparian areas. *Land Degradation & Development*, 19(2), pp.153-164. https://doi.org/10.1002/ldr.820.

7- Biau, G. and Scornet, E., 2016. A random forest guided tour. *Test*, 25(2), pp.197-227. https://doi.org/10.1007/s11749-016-0481-7.

8- Breiman, L., 2001. Random forests. *Machine learning*, 45(4), pp.5-32. https://doi.org/10.1023/A:1010933404324.

9- Cavins, T.J., Gibson, J.L., Whipker, B.E. and Fonteno, W.C., 2000. pH and EC meters-tool for substrate analysis. *North Carolina State University*, 1(1), pp.1-14.

10- Costa-Carrapiço, I., Raslan, R. and González, J.N., 2020. A systematic review of genetic algorithm-based multi-objective optimisation for building retrofitting strategies towards energy efficiency. *Energy and Buildings*, 210(12), pp.109-130. https://doi.org/10.1016/j.enbuild.2019.109690.

11- El Bilali, A. and Taleb, A., 2020. Prediction of irrigation water quality parameters using machine learning models in a semi-arid environment. *Journal of the Saudi Society of Agricultural Sciences*, 19(7), pp.439-451. https://doi.org/10.1016/j.jssas.2020.08.001.

12- Giri, S., 2021. Water quality prospective in Twenty First Century: Status of water quality in major river basins, contemporary strategies and impediments: A review. *Environmental Pollution*, 271(4), pp.116-132. https://doi.org/10.1016/j.envpol.2020.116332.

13- Haghiabi, A.H., Nasrolahi, A.H. and Parsaie, A., 2018. Water quality prediction using machine learning methods. *Water Quality Research Journal*, 53(1), pp.3-13. https://doi.org/10.2166/wqrj.2018.025.

14- Hamdi, T., Ali, J.B., Di Costanzo, V., Fnaiech, F., Moreau, E. and Ginoux, J.-M., 2018. Accurate prediction of continuous blood glucose based on support vector regression and differential evolution algorithm. *Biocybernetics and Biomedical Engineering* ,38(2) ,pp.362-372. https://doi.org/10.1016/j.bbe.2018.02.005.

15- Hassan, M., Hassan, M., Akter, L., Rahman, M., Zaman, S., Khan, H., Jahan, N., Smrity, R.N., Farhana, J. and Raihan, M., 2021. Efficient prediction of water quality index (WQI) using machine learning algorithms. *Human-Centric Intelligent Systems*, 1(3), pp.86-97. https://doi.org/10.2991/hcis.k.211203.001.

16- Hassanat, A., Almohammadi, K., Alkafaween, E.a., Abunawas, E., Hammouri, A. and Prasath, V.S., 2019. Choosing mutation and crossover ratios for genetic algorithms—a review with a new dynamic approach. *Information*, 10(12), pp.390-402. https://doi.org/10.3390/info10120390.

17- Hüe, X., 1997. Genetic algorithms for optimization. Rapport technique, Citeseer.

18- Jamei, M., Ali, M., Karimi, B., Karbasi, M., Farooque, A.A. and Yaseen, Z.M., 2023. Surface water electrical conductivity and bicarbonate ion determination using a smart hybridization of optimal Boruta package with Elman recurrent neural network. *Process Safety and Environmental Protection*, 174(5), pp.115-134. https://doi.org/10.1016/j.psep.2023.03.062.

19- Janitza, S. and Hornung, R., 2018. On the overestimation of random forest's out-of-bag error. *PloS one*, 13(8), pp.201-217. https://doi.org/10.1371/journal.pone.0201904.s001.

20- Karal, O., 2017. Maximum likelihood optimal and robust support vector regression with lncosh loss function. *Neural networks*, 94(3), pp.1-12. https://doi.org/10.1016/j.neunet.2017.06.008.

21- Kisi, O., Azad, A., Kashi, H., Saeedian, A., Hashemi, S.A.A. and Ghorbani, S., 2019. Modeling groundwater quality parameters using hybrid neuro-fuzzy methods. *Water Resources Management*, 33(1), pp.847-861. https://doi.org/10.1007/s11269-018-2147-6.

22- Koohestani, B., 2020. A crossover operator for improving the efficiency of permutation-based genetic algorithms. *Expert Systems with Applications*, 151(4), pp.113-130. https://doi.org/10.1016/j.eswa.2020.113381.

23- Liu, S., Tai, H., Ding, Q., Li, D., Xu, L. and Wei, Y., 2013. A hybrid approach of support vector regression with genetic algorithm optimization for aquaculture water quality prediction. *Mathematical and Computer Modelling*, 58(3-4), pp.458-465. https://doi.org/10.1016/j.mcm.2011.11.021.

24- Machado, R.M.A. and Serralheiro, R.P., 2017. Soil salinity: effect on vegetable crop growth. Management practices to prevent and mitigate soil salinization. *Horticulturae*, 3(2), pp.30-42. https://doi.org/10.3390/horticulturae3020030.

25- Manikandan, M. and Pant, R.S., 2021. Research and advancements in hybrid airships—A review. *Progress in Aerospace Sciences*, 127(17), pp.100-118. https://doi.org/10.1016/j.paerosci.2021.100741.

26- McKinney, D.C. and Lin, M.D., 1994. Genetic algorithm solution of groundwater management models. *Water Resources Research*, 30(6), pp.1897-1906. https://doi.org/10.1029/94WR00554.

27- Mokhtar, A., Elbeltagi, A., Gyasi-Agyei, Y., Al-Ansari, N. and Abdel-Fattah, M.K., 2022. Prediction of irrigation water quality indices based on machine learning and regression models. *Applied Water Science*, 12(4), pp.76-91. https://doi.org/10.1007/s13201-022-01590-x.

28- Mousazadeh, M ,.Paital, B., Naghdali, Z., Mortezania, Z., Hashemi, M., Karamati Niaragh, E., Aghababaei, M., Ghorbankhani, M., Lichtfouse, E. and Sillanpää, M., 2021. Positive environmental effects of the coronavirus 2020 episode: a review. *Environment, Development and Sustainability*, 23(2), pp.12738-12760. https://doi.org/10.1007/s10668-021-01240-3.

29- Nasehi, F., Hassani, A., Monavvari, M., Karbassi, A. and Khorasani, N., 2013. Evaluating the metallic pollution of riverine water and sediments: a case study of Aras River .*Environmental Monitoring and Assessment*, 185(4), pp.197-203. https://doi.org/10.1007/s10661-012-2543-y.

30- Nouraki, A., Alavi, M., Golabi, M. and Albaji, M., 2021. Prediction of water quality parameters using machine learning models: A case study of the Karun River, Iran. *Environmental Science and Pollution Research*, 28(40), pp.57060-57072. https://doi.org/10.1007/s11356-021-14560-8.

31- Osaba, E., Diaz, F. and Onieva, E., 2014. Golden ball: a novel meta-heuristic to solve combinatorial optimization problems based on soccer concepts. *Applied Intelligence*, 41(1), pp.145-166. https://doi.org/10.1007/s10489-013-0512-y.

32- Palmer, M. and Ruhi, A., 2019. Linkages between flow regime, biota, and ecosystem processes: Implications for river restoration. *Science*, 365(64), pp.2087-2099. https://doi.org/10.1126/science.aaw2087.

33- Pan, W.-T., 2012. A new fruit fly optimization algorithm: taking the financial distress model as an example. *Knowledge-Based Systems*, 26(1), pp.69-74. https://doi.org/10.1016/j.knosys.2011.07.001.

34- Pincay, J.M.C. and Cantos, M.F.P., 2024. Analysis of Soil Salinization as an Environmental Issue in Latin America. *Journal of Ecological Engineering*, 25(1), pp.146-152. http://dx.doi.org/10.12911/22998993/174378.

35- Qadir, M., Wichelns, D., Raschid-Sally, L., Minhas, P.S., Drechsel, P., Bahri, A., McCornick, P.G., Abaidoo, R.C., Attia, F. and El-Guindy, S., 2007. Agricultural use of marginal-quality *Water: Opportunities and Challenges. IWMI*, 12(1), pp.1-34 .

36- Qishlaqi, A., Kordian, S. and Parsaie, A., 2017. Hydrochemical evaluation of river water quality—a case study. *Applied Water Science*, 7, pp.2337-2342. https://doi.org/10.1007/s13201-016-0409-0.

37- Rad, M.A.V. and Kasaeian, A., 2024. Evaluation of stand-alone hybrid renewable energy system with excess electricity minimizer predictive dispatch strategy. *Energy Conversion and Management*, 299(18), pp.117-131. https://doi.org/10.1016/j.enconman.2023.117898.

38- Rahimi, L., Amanipoor, H. and Battaleb-Looie, S., 2021. Effect of salinity of irrigation water on soil properties (abadan plain, SW Iran). *Geocarto International*, 36(16), pp.1884-1903. https://doi.org/10.1080/10106049.2019.1678678.

39- Rahnama, E., Bazrafshan, O. and Asadollahfardi, G., 2020. Application of data-driven methods to predict the sodium adsorption rate (SAR) in different climates in Iran. *Arabian Journal of Geosciences*, 13(1), pp.1-19. https://doi.org/10.1007/s12517-020-06146-4.

40- Schmitt, L.M., 2001. Theory of genetic algorithms. *Theoretical Computer Science*, 259(1), pp.1-61. https://doi.org/10.1016/S0304-3975(00)00406-0.

41- Sepahvand, A., Singh, B., Sihag, P., Nazari Samani, A., Ahmadi, H. and Fiz Nia, S., 2021. Assessment of the various soft computing techniques to predict sodium absorption ratio (SAR). *ISH Journal of Hydraulic Engineering*, 27(1), pp.124-13. https://doi.org/10.1080/09715010.2019.1595185.

42- Shao, Z., Pi, D. and Shao, W., 2020. Hybrid enhanced discrete fruit fly optimization algorithm for scheduling blocking flow-shop in distributed environment. *Expert Systems with Applications*, 145(17), pp.113-147. https://doi.org/10.1016/j.eswa.2019.113147.

43- Sposito, G. and Mattigod, S.V., 1977. On the chemical foundation of the sodium adsorption ratio. *Soil Science Society of America Journal*, 41(2), pp.323-329. https://doi.org/10.2136/sssaj1977.03615995004100020030x.

44- Su, X., He, X., Zhang, G., Chen, Y. and Li, K., 2022. Research on SVR water quality prediction model based on improved sparrow search algorithm. *Computational Intelligence and Neuroscience*, 2022(1), pp.7327072. https://doi.org/10.1155/2022/7327072.

45- Vapnik, V. and Chervonenkis, A.Y., 1974. The method of ordered risk minimization, I. *Avtomatika i Telemekhanika*, 8(1), pp.21-30.

46- Wang, H. and He, G., 2022. Rivers: Linking nature, life, and civilization. *River*, 1(1), pp.25-36. https://doi.org/10.1002/rvr2.7.

47- Wang, R.-Y., Hu, P., Hu, C.-C. and Pan, J.-S., 2022. A novel Fruit Fly Optimization Algorithm with quasi-affine transformation evolutionary for numerical optimization and application. *International Journal of Distributed Sensor Networks*, 18(2), pp.155-168. https://doi.org/10.1177/15501477211073037.

48- Waqas, M., Humphries, U.W., Wangwongchai, A., Dechpichai, P., Zarin, R. and Hlaing, P.T., 2024. Incorporating novel input variable selection method for in the different water basins of Thailand. *Alexandria Engineering Journal*, 86(1), pp.557-576. https://doi.org/10.1016/j.aej.2023.11.046.

49- Worawattawechai, T., Intiyot, B., Jeenanunta, C. and Ferrell Jr, W.G., 2022. A learning enhanced golden ball algorithm for the vehicle routing problem with backhauls and time windows. *Computers & Industrial Engineering*, 168, pp.108044. https://doi.org/10.1016/j.cie.2022.108044.

50- Woźniak, M., Szczotka, J., Sikora, A. and Zielonka, A., 2024. Fuzzy logic type-2 intelligent moisture control system. *Expert Systems with Applications*, 238(18), pp.121-137. https://doi.org/10.1016/j.eswa.2023.121581.

51- Wu, J. and Wang, Z., 2022. A hybrid model for water quality prediction based on an artificial neural network, wavelet transform, and long short-term memory. *Water*, 14(4), pp.610. https://doi.org/10.3390/w14040610.