Advanced Technologies in Water Efficiency

homepage:https://atwe.razi.ac.ir

Online ISSN:2783-4964

Nitrate concentration modeling in water resources using support machine regression and metaheuristic optimization algorithm Harris Hawks

Shahab Naderi[1], and **Saeid Shabanlou**[2]✉

1. Department of Civil Engineering, Taft Branch, Islamic Azad University, Taft, Iran. E-mail: sh.naderi66@gmail.com
2. Corresponding author, Department of Water Engineering, Kermanshah Branch, Islamic Azad University, Kermanshah, Iran. E-mail: saeid.shabanlou@gmail.com

Article Info

Article type:
Research Article

Article history:
Received 29 August 2024
Received in revised form 25 November 2024
Accepted 20 January 2025
Available online 22 June 2025

Keywords:
nitrate,
artificial neural network,
groundwater,
pollution,
machine learning.

ABSTRACT

Objective: The objective of this study is to develop a machine learning model for simulating the nitrate concentration. Simulating and predicting nitrate concentration has always been one of the most important issues in the field of water resources management.

Method: In this research, after collecting the data, the nitrate concentration data are first clustered using the JNB, then, an SVR model is used for each cluster. The SFFS algorithm is used to select the input variables for the model simultaneously with the training process of this model, then, based on the results of these three models, the average value of the error indices for the training stage (RMSE = 0.2387, MAE = 0.2236, R^2=0.9874) and test (RMSE = 0.2474, MAE = 0.2350, R^2=0.9841) are calculated. In this case, the trial and error procedure is used for this work. In the next step, the HHO algorithm is used to determine the optimal value of the parameters of the kernel functions. In this case, the values of R^2, MAE and RMSE for the training phase are 0.9961, 0.1169, and 0.1502, respectively, and their values for the test phase are 0.9845, 0.1308, and 0.9978, respectively.

Results: Based on the results of this study, firstly, the use of HHO to predict nitrate concentration can significantly increase the accuracy of the SVR model, secondly, the use of different machine learning models together can play an effective role in increasing the accuracy of regression models such as SVR. The results of this study show that the use of data clustering before developing machine learning models can improve the accuracy of nitrate concentration prediction. The HHO-SVR hybrid model has performed better in different clusters with proper selection of kernel function and has provided optimal results. Also, this study emphasizes that the different statistical characteristics of each cluster have a significant effect on the performance of the models. Therefore, to more accurately predict nitrate concentration in groundwater, it is recommended to first cluster the data and then develop a specific model for each cluster.

Conclusions: The results of this study show that the use of data clustering before developing machine learning models can improve the accuracy of nitrate concentration prediction. The HHO-SVR hybrid model has performed better in different clusters with proper selection of kernel function and has provided optimal results. Also, this study emphasizes that the different statistical characteristics of each cluster have a significant effect on the performance of the models. Therefore, to more accurately predict nitrate concentration in groundwater, it is recommended to first cluster the data and then develop a specific model for each cluster.

Cite this article: Naderi, Sh., & Shabanlou, S. (2025). Nitrate concentration modeling in water resources using support machine regression and metaheuristic optimization algorithm Harris Hawks. *Advanced Technologies in Water Efficiency, 5* (2), 1-21. https://doi.org/10.22126/atwe.2024.11300.1143

Publisher: Razi University.
https://doi.org/10.22126/atwe.2024.11300.1143

Introduction

Nitrate is one of the contaminants which can pollute groundwater in different ways. So, we always try to use accurate methods to simulate and predict its concentration. The objective of this study is to develop a machine learning model for simulating the nitrate concentration.

Method

In this research, the SVM model is used for simulating the nitrate concentration and Harris Hawks optimization (HHO) algorithm to determine the parameters of the kernel function of the SVR model. The linear kernel, RBF, sigmoid and polynomial kernel functions are used to develop the SVR model. The data used in this study include 1453 water samples (from 2000 to 2020) for each of which the physical and chemical parameters are measured. After performing the necessary pre-processing including identifying outlier data and correcting incomplete information to avoid biasing the model, all data are normalized between zero and one. First, all the data are divided into three clusters using the JNB algorithm, then, a HHO-SVR mode is developed for each cluster. Based on the results of this study, the HHO-SVR model has the best performance for the first, second and third clusters with the PK, RBF, and SK kernel functions, respectively. Based on the results of this study, due to the differences in the results of the HHO-SVR model in different clusters, it is better to cluster the nitrate data before developing the HH-SVR model and then develop a model for each cluster because the samples belonging to each cluster have different statistical characteristics.

Results

Simulating and predicting nitrate concentration has always been one of the most important issues in the field of water resources management. In this research, after collecting the data, the nitrate concentration data are first clustered using the JNB, then, an SVR model is used for each cluster. The SFFS algorithm is used to select the input variables for the model simultaneously with the training process of this model, then, based on the results of these three models, the average value of the error indices for the training stage (RMSE = 0.2387, MAE = 0.2236, R^2=0.9874) and test (RMSE = 0.2474, MAE = 0.2350, R^2=0.9841) are calculated. In this case, the trial and error procedure is used for this work. In the next step, the HHO algorithm is used to determine the optimal value of the parameters of the kernel functions. In this case, the values of R^2, MAE and RMSE for the training phase are 0.9961, 0.1169, and 0.1502, respectively, and their values for the test phase are 0.9845, 0.1308, 0.9978, respectively. Based on the results of this study, firstly, the use of HHO to predict nitrate concentration can significantly increase the accuracy of the SVR model, secondly, the use of different machine learning models together can play an effective role in increasing the accuracy of regression models such as SVR.

Conclusions

The results of this study show that the use of data clustering before developing machine learning models can improve the accuracy of nitrate concentration prediction. The HHO-SVR hybrid model has performed better in different clusters with proper selection of kernel function and has provided optimal results. Also, this study emphasizes that the different statistical characteristics of each cluster have a significant effect on the performance of the models.

Therefore, to more accurately predict nitrate concentration in groundwater, it is recommended to first cluster the data and then develop a specific model for each cluster.

Author Contributions

All authors contributed equally to the conceptualization of the article and writing of the original and subsequent drafts.

Data Availability Statement

Data Availability Statement

Ethical Considerations

The authors avoided data fabrication, falsification, plagiarism, and misconduct.

Funding

Not applicable.

Conflict of Interest

The authors declare no conflict of interest.

فناوری های پیشرفته در بهره وری آب
وبگاه نشریه: https://atwe.razi.ac.ir
شاپا الکترونیکی: ۲۷۸۳-۴۹۶۴

مدل‌سازی غلظت نیترات در منابع آب با بهره‌گیری از رگرسیون ماشین پشتیبان و الگوریتم بهینه‌سازی متا هیوریستیک شاهین هریس

شهاب نادری [۱]، و سعید شعبانلو [۲]✉

۱. گروه مهندسی عمران، واحد تفت، دانشگاه آزاد اسلامی، تفت، ایران. رایانامه: sh.naderi66@gmail.com
۲. نویسنده مسئول، گروه مهندسی آب، واحد کرمانشاه، دانشگاه آزاد اسلامی، کرمانشاه، ایران. رایانامه: saeid.shabanlou@gmail.com

اطلاعات مقاله

نوع مقاله:
مقاله پژوهشی

تاریخ دریافت: ۱۴۰۳/۰۶/۰۸
تاریخ بازنگری: ۱۴۰۳/۰۹/۰۵
تاریخ پذیرش: ۱۴۰۳/۱۱/۰۱
تاریخ انتشار: ۱۴۰۴/۰۴/۰۱

کلیدواژه‌ها:
نیترات،
شبکه عصبی مصنوعی،
آب‌های زیرزمینی،
آلودگی،
یادگیری ماشین.

چکیده

هدف: هدف این مطالعه توسعه یک مدل یادگیری ماشین برای شبیه‌سازی غلظت نیترات است. شبیه‌سازی و پیش‌بینی غلظت نیترات همیشه از مهم‌ترین مسائل در حوزه مدیریت منابع آب بوده است.

روش پژوهش: در این تحقیق بعد از جمع‌آوری داده‌ها ابتدا داده‌های مربوط به غلظت نیترات با استفاده از JNB خوشه‌بندی شدند، سپس برای هر خوشه یک مدل SVR توسعه داده شد، همزمان با فرآیند آموزش این مدل از الگوریتم SFFS برای انتخاب متغیرهای ورودی به مدل استفاده شد، سپس بر اساس نتایج حاصل از این سه مدل مقدار متوسط شاخص‌های خطا برای مرحله آموزش و تست محاسبه شدند، در این حالت از روش سعی و خطا برای این کار استفاده شد. در گام بعد از الگوریتم HHO برای تعیین مقدار بهینه پارامترهای توابع کرنل استفاده شد.

یافته‌ها: بر اساس نتایج حاصل از این سه مدل مقدار متوسط شاخص‌های خطا برای مرحله آموزش (RMSE = 0.2387 , MAE = 0.2236 , R^2=0.9874) و تست (RMSE = 0.2474 , MAE = 0.2350 , R^2=0.9841) نتایج حاصل از الگوریتم HHO شامل مقدار شاخص‌های R^2 و RMSE, MAE برای مرحله آموزش به ترتیب برابر ۰/۹۹۶۱، ۰/۱۱۶۹ و ۰/۱۵۰۲ است و مقدار آن‌ها برای مرحله تست به ترتیب برابر ۰.۹۸۴۵،۰.۱۳۰۸، ۰.۹۹۷۸ است. بر اساس نتایج حاصل از این مطالعه اولا استفاده از HHO برای پیش‌بینی غلظت نیتزات میتواند باعث افزایش چشمگیر دقت مدل SVR می‌شود، دوما استفاده به جا از مدل‌های یادگیری ماشین مختلف در کنار هم میتوانند نقش موثری در افزایش دقت مدل‌های رگرسیونی مانند SVR داشته باشد.

نتیجه‌گیری: این تحقیق نشان داد که با انتخاب مناسب متغیرها، حتی در صورت استفاده از یک مدل نسبتاً ساده مانندSVR ، می‌توان به نتایج بسیار دقیقی دست‌یافت. این نشان می‌دهد که کیفیت داده‌ها و انتخاب ورودی‌های مناسب به‌اندازه پیچیدگی خود مدل اهمیت دارد .در نهایت، این تحقیق تأیید می‌کند که ترکیب الگوریتم‌های بهینه‌سازی مانند HHO با مدل‌های یادگیری ماشین و به‌کارگیری روش‌های انتخاب ویژگی، می‌تواند به‌عنوان یک راهکار مؤثر برای بهبود دقت پیش‌بینی‌ها در مسائل مرتبط با منابع آب، به‌ویژه در پیش‌بینی آلاینده‌های مهم، مورداستفاده قرار گیرد. این راهبردها می‌توانند در مدیریت بهتر منابع آب و کاهش آلودگی‌های ناشی از نیترات و سایر آلاینده‌ها به کار گرفته شوند.

استناد: نادری، شهاب؛ و شعبانلو، سعید. (۱۴۰۴). مدل‌سازی غلظت نیترات در منابع آب با بهره گیری از رگرسیون ماشین پشتیبان و الگوریتم بهینه‌سازی متاهیوریستیک شاهین هریس. *فناوری های پیشرفته در بهره وری آب* ، ۵ (۲)، ۲۱-۱.
https://doi.org/10.22126/atwe.2024.11300.1143

مقدمه

در بسیاری از نقاط دنیا، آب‌های زیرزمینی به‌عنوان منبع اصلی آب آشامیدنی در نواحی شهری و روستایی مورداستفاده قرار می‌گیرند. بااین‌حال، در سال‌های اخیر گسترش فعالیت‌های صنعتی و کشاورزی منجر به افزایش غلظت آلاینده‌های سمی مانند آنیون‌های معدنی، یون‌های فلزی، مواد شیمیایی خطرناک در آب‌های زیرزمینی شده‌اند (عادلجو و همکاران [1]، ۲۰۲۱). در میان آلاینده‌های مختلف آنیون‌های معدنی از اهمیت بالایی برخوردارند؛ زیرا حتی غلظت‌های بسیار پایین، در حدppb این آنیون‌ها سمی و برای بدن انسان و سایر حیوانات مضر است (وو و همکاران [2] ، ۲۰۱۰)، از طرفی حضور مقادیر بسیار کم آن‌ها در آب آشامیدنی معمولاً تغییرات ظاهری قابل تشخیصی در آن ایجاد نمی‌کند، بنابراین گاهی شناسایی آن‌ها بر اساس تغییرات ویژگی‌های فیزیکی آب بسیار دشوار است (مزرعه و همکاران [3] ، ۲۰۲۳). غلظت نیترات در آب‌های سطحی و زیرزمینی باتوجه‌به شرایط جغرافیایی (نوع ساختارهای زمین‌شناسی هر منطقه)، مدیریت ضایعات انسانی و حیوانی، مقادیر استفاده از کودهای شیمیایی ازته و تخلیه ترکیبات ازته توسط صنایع متفاوت خواهد بود (ژانگ و همکاران [4] ، ۲۰۲۱). غلظت مجاز نیترات در آب آشامیدنی در چند سال اخیر به‌عنوان یک موضوع سلامت عمومی موردتوجه است (مزرعه و همکاران، ۲۰۲۳). باتوجه‌به مطالعات به‌عمل‌آمده توسط سازمان بهداشت جهانی در مورد نیترات، این سازمان حداکثر مجاز نیترات را (۰۹ mg L-1برحسب نیترات) اعلام نموده است (سازمان بهداشت جهانی[5]، ۲۰۲۲).

ادبیات موضوع و پیشینه پژوهش

به دلیل پیشرفت‌های چشمگیر علم هوش مصنوعی در سال‌های اخیر، امروزه بسیاری از مطالعات برای پیش‌بینی غلظت نیترات مبتنی بر روش‌های یادگیری ماشین هستند (وو و همکاران، ۲۰۱۰). دررابطه‌با پژوهش حاضر، مطالعاتی صورت‌گرفته است که در ادامه تعدادی از این مطالعات که در خارج از کشور مورد بررسی قرار گرفته است، ارائه شده است.

ریاضی و همکاران [6] (۲۰۱۹) برای تهیه نقشه آسیب‌پذیری آبخوان یک MABLR را توسعه دادند، آن‌ها برای توسعه این مدل از ۱۵ فاکتور مختلف که با استفاده از الگوریتم FSLR انتخاب شدند، استفاده کردند. همچنین از الگوریتم FLS برای بهینه سازی آن استفاده کرند. بعد از توسعه این مدل تعمیم پذیری آن را با بررسی عملکرد آن را در مقیاس بزرگ بررسی کردند. آن‌ها برای بررسی دقت این مدل علاوه بر محاسبه شاخص‌های خطا مانند RMSE از مقایسه عملکرد آن با مدل‌های شبکه عصبی پرشپترون چند لایه (MLP)، رگرسیون لجستیک (LR) و مدل ماشین بردار پشتیبان (SVM) نیز استفاده کردند. بر اساس نتایج حاصل از این مطالعه، این مدل روشی مناسب برای تهیه GAPM و کاهش بایاس و واریانس مدل است.

لاهجوج و همکاران [7] (۲۰۲۰) در یک مطالعه موردی برای بررسی اسب پذیری آب‌های زیرزمینی دشت ساسیس در شمال مراکش و تهیه نقشه آسیب‌پذیری آبخوان از RF استفاده کردند. بر اساس نتایج حاصل از این مطالعه در صورت داشتن اطلاعات کافی به‌خوبی می‌توان از این مدل برای تهیه نقشه آسیب‌پذیری آبخوان‌های آلوده به نیترات استفاده کرد.

الزاین و همکاران [8] (۲۰۲۱) کارایی مدل ANFIS برای شبیه‌سازی غلظت نیترات و تهیه نقشه آسی پذیری آبخوان را بررسی کرندند، سپس کارایی الگوریتم‌های PSO , DE , GA برای بهبود دقت آن را بررسی کردند. در این مطالعه از جذر میانگین مربعات خطا (RMSE)، ضریب همبستگی (R^2) و خطای میانگین مربعات (MSE) برای ارزیابی نتایج استفاده شد. بر اساس نتایج حاصل مدل ANFIS-PSO نسبت به همه مدل‌ها عملکرد بهتری دارد.

1. Adeloju et al
2. Wu et al
3. Mazraeh et al
4. Zhang et al
5. World Healt Organization et al
6. Rizeei et al
7. Lahjouj et al
8. Elzain et al

ال امری و همکاران [۱] (۲۰۲۲) برای آنالیز غلظت نیترات در آبخوان‌های کم‌عمق، شناسایی مناطق آسیب‌پذیر و پیش‌بینی غلظت نیترات از مدل‌های ANN and ARIMA استفاده کردند. در واقع آن‌ها در این مطالعه از مدل ANN برای شبیه‌سازی غلظت نیترات و از مدل ARIMA برای پیش‌بینی غلظت آن در آینده استفاده کردند. بر اساس نتایج حاصل از این مطالعه به‌خوبی می‌توان از این مدل‌ها برای مدیریت آبخوان استفاده کرد.

از دیگر تحقیقات در زمینه موضوع این تحقیق می‌توان به تحقیقات عزیزی و همکاران [۲] (۲۰۲۳)؛ امیری و همکاران [۳] (۲۰۲۱)؛ شعبانلو[۴] (۲۰۱۸)؛ اسماعیلی و همکاران [۵] (۲۰۲۱)؛ فلاحی و همکاران [۶] (۲۰۲۳) و پناهی و همکاران [۷] (۲۰۲۲) اشاره کرد.

روش پژوهش

این پژوهش در محدوده مطالعاتی دشت بهار در استان همدان انجام شد. دشت بهار بین طول‌های جغرافیایی ۴۸° ۲۶' تا ۴۸° ۳۶' شرقی و عرض‌های جغرافیایی ۳۴° ۴۸' تا ۳۴° ۵۸' شمالی واقع شده است. این دشت در بخش مرکزی استان همدان قرار دارد و شامل مساحتی در حدود ۱۵۰۰ کیلومتر مربع است. بخش عمده‌ای از این محدوده شامل زمین‌های زراعی و کشاورزی بوده و سایر مناطق آن شامل ارتفاعات و اراضی طبیعی است. منابع آبی این دشت از جمله سفره‌های آب زیرزمینی و رودخانه‌های محلی برای تأمین آب کشاورزی و شرب منطقه حیاتی هستند. بالاترین نقطه ارتفاعی در محدوده دشت بهار همدان در حدود ۲۴۰۰ متر از سطح دریا واقع شده و پایین‌ترین نقطه آن در حدود ۱۶۰۰ متر از سطح دریا قرار دارد. از مراکز جمعیتی مهم این محدوده می‌توان به شهر بهار و روستاهای اطراف آن اشاره کرد. در این منطقه، تعداد ۸۵۰ حلقه چاه برای برداشت سالانه حدود ۷۵ میلیون متر مکعب آب شناسایی شده است. همچنین ۱۲۰ دهنه چشمه با تخلیه سالانه ۱۰ میلیون متر مکعب و ۸ رشته قنات با تخلیه ۳ میلیون متر مکعب نیز شناسایی شده‌اند. حجم کل برداشت منابع آب زیرزمینی در این محدوده مطالعاتی حدود ۸۸ میلیون متر مکعب در سال است که بخش عمده آن مربوط به دشت بهار بوده و مابقی به مناطق کوهپایه‌ای و ارتفاعات مجاور اختصاص دارد. در سطح آبخوان آبرفتی دشت بهار تعداد ۸۰۰ حلقه چاه، ۱۰ دهنه چشمه و ۵ رشته قنات وجود دارد که به‌صورت مجموع سالانه ۸۵ میلیون متر مکعب آب برداشت و تخلیه می‌کنند.

۱. رگرسیون با ماشین بردار پشتیبان

یکی از بهترین و دقیق‌ترین روش‌های یادگیری با نظارت است که اولین‌بار در سال ۱۹۶۸ توسط وپنیک معرفی شد. در این روش هدف اصلی جداسازی داده‌ها با استفاده از یک ابر صفحه است. موفعیت این ابرصفحه بر اساس فاصله بین داده‌های دو کلاس تعیین می‌شود. دو صفحه دیگر به موازات این ابر صفحه و وجود دارند، موقعیت این صفحه ها بر اساس موقعیت بردارای پشتیبان تعیین می‌شود، در نهایت هدف نهایی SVM حداکثر کردن فاصله این دو صفحه باتوجه‌به موقعیت بردارهای پشتیبان است (نوبل[۸]، ۲۰۰۶). اگر داه های دو کلاس به صورت خطی تفکیک پذیر باشند می‌توان از مدل HM-SVMs[۹] داده‌های دو کلاس را از یکدیگر تفکیک کرد؛ اما با استفاده از این مدل نمیتوان داده‌های غیرخطی را تفکیک کرد؛ لذا برای تفکیک داده‌های غیرخطی باید از مدل SM-SVMs [۱۰] استفاده کرد (واپنیک و چروونکیس[۱۱]، ۱۹۷۱).

1. El Amri et al
2. Azizi et al
3. Amiri et al
4. Shabanlou
5. Esmaeili et al
6. Fallahi et al
7. Panahi et al
8. Noble
9. Hard-margin support vector machines
10. Soft-margin support vector machines (SM-SVMs)
11. Vapnik and Chervonenkis

اگر $S=\{(x_1,y_1),(x_2,y_2),(x_3,y_3),\dots (x_l,y_l)\}$ مجموعه ای از داده‌های آموزشی باشد به طوری که $x_i \in \mathbb{R}^n$ ورودی و $y_i = \mathbb{R}^n$ خروجی است و $i=1,2,3,\dots l$ است و $\mathbb{R}^n$ ابعاد بردار ورودی است هدف مدل SVR یافتن تابع $f(x)$ است، به طوری که برای همه نمونه های آموزشی حداکثر به میزان ε از مقادیر مشاهده شده (y_O) اختلاف داشته باشد. دوما تا حد ممکن فلت[1] باشد یا به عبارتی مقدار w بسیار ناچیز باشد(معادله (۱))، (هرست[2] و همکاران، ۱۹۹۸).

$$f(x)= w^Tx+b \; x,w \in \mathbb{R}^n, b \in \mathbb{R} \qquad (۱)$$

که در آن x ورودی، w وزن و b بایاس است، باتوجه‌به این که یک راه برای فلت شدن تابع معادله (۱) حداقل کردن نرم w است. بنابراین مسئله SVR را می‌توان به‌صورت یک مسئله بهینه‌سازی محدب نوشت (معادله (۲)).

$$\text{Minimize } \frac{1}{2}\|w\|^2$$
$$\text{Subject to } \begin{cases} y_i-(w^Tx)-b\le \varepsilon \\ y_i-(w^Tx)-b\le \varepsilon \end{cases} \qquad (۲)$$

در حالت عادی فرض می‌شود تابع f برای همه داده‌های واقعی وجود دارد و می‌تواند همه داده‌های آموزشی را حداکثر با اختلاف ε تقریب بزند، بنابراین این مسئله بهینه‌سازی محدب قابل‌حل است. اما گاهی ممکن است معادله (۲) برای همه داده‌های آموزشی برقرار نباشد برای حل چنین مسائلی می‌توان از متغیر کمبود[3] استفاده کرد، در این حالت معادله (۲)به‌صورت زیر بازنویسی می‌شود.

$$\text{Minimize } \frac{1}{2}\|w\|^2+C\sum_{i=1}^{l}(\xi_i+\xi_j^*)$$
$$\text{Subject to } \begin{cases} y_i-(w^Tx)-b\le \varepsilon+\xi_i \\ y_i-(w^Tx)-b\le \varepsilon+\xi_j^* \\ \xi_i,\xi_j^* \ge 0 \end{cases} \qquad (۳)$$

ثابت C یک توازنی بین میزان فلت بودن تابع f و حداکثر مقدار انحراف بیشتر از ε را برقرار می‌کند. این مسئله را می‌توان با استفاده از معادله (۴) توصیف کرد.

$$|\xi|_\varepsilon := \begin{cases} 0 \text{ if } |\xi| \le \varepsilon \\ |\xi|- \varepsilon \text{ Otherwise} \end{cases} \qquad (۴)$$

با تبدیل تابع هدف به یک تابع لاگرانژی و سپس تعمیم مسئله خطی به یک مسئله غیرخطی با استفاده از یک تابع نگاشت یک تابع جدید دارای نقطه زینی[4] حاصل می‌شود، در نهایت با استفاده از معادله (۵) مقدار این تابع را می‌توان محاسبه کرد. علت استفاده از تابع نگاشت این است که از معادله (۱) تنها می‌توان برای شبیه‌سازی مسائل تفکیک پذیرخطی استفاده کرد و با استفاده از آن نمیتوان مسائل تفکیک پذیر غیز خطی را شبیه‌سازی کرد، لذا برای شبیه‌سازی اینگونه مسائل باید داده‌ها را از بعد فضای ورودی به فضای ویژگی با ابعاد بالاتر (فضای هیلبرت با ابعاد محدود یا نامحدود) نگاشت داد.

$$f(x)= \sum_{i=1}^{l}(a_i- a_i^*)k(x_i,x)+b \qquad (۵)$$

1. Flat
2. Hearst
3. Slack variables
4. Saddle point

که در آن a_i and a_i^* ضرایب لاگرانژی هستند که هنگام حل مسئله بهینه سازی حاصل می‌شوند. $k(x_i,x)$ تابع کرنل است که که مقدار آن برابر ضرب داخلی دو بردار x_i and x است. این تابع یکی از مهم‌ترین بخش های مدل SVR است،زیرا تعیین نوع آن و مقدار پارامترهای آن نقش مهمی در میزان دقت مدل SVR دارد.تاکنون توابع کرنل مختلفی برای توسعه مدل SVR معرفی شده‌اند (هرست و همکاران [1] ، ۱۹۹۸). در این مطالعه از توابع خطی، RBF، سیگموئید و چندجمله‌ای برای توسعه مدل SVR استفاده شد که در جدول (۱) قابل مشاهده است. برای تنظیم پارامترهای هر یک از این توابع از الگوریتم بهینه‌سازی شاهین هریس(HHO) استفاده شد،سپس با استفاده از شاخص‌های خطا کارایی مدل SVR با هر یکی از این توابع ارزیابی شد.

جدول ۱. توابع کرنل استفاده شده برای توسعه مدل SVR

تابع کرنل	معادله ریاضی	پارامتر قابل‌تنظیم
Linear kernel(LK)	$k(x,y) = x^T.y + C$	C
Radial Basis Function(RBF)	$k(x,y) = exp(-\alpha \|x - y\|^2 + C)$	α, C
Sigmoid Kernel(SK)	$k(x,y) = tanh(\beta(x^T.y) + C)$	β, C
Polynomial Kernel(PK)	$k(x,y) = (\gamma(.x^T.y) + C)^d$	γ, C, d
Multi Quadric(MQ)	$\sqrt{\|x-y\|^2 + C^2}$	C

۲. الگوریتم بهینه‌سازی شاهین هریس [2]

الگوریتم بهینه‌سازی شاهین هریس (HHO) یک الگوریتم فراابتکاری الهام‌گرفته از رفتار شکار شاهین‌های هریس است که برای حل مسائل بهینه‌سازی عددی پیوسته و گسسته به کار می‌رود. این الگوریتم بر اساس رفتار گروهی و شکار هماهنگ شاهین‌های هریس طراحی شده است که در حیات‌وحش به‌صورت دسته‌جمعی و هوشمندانه طعمه خود را به دام می‌اندازند (حیدری و همکاران [3]، ۲۰۱۹) در فرآیند بهینه‌سازی، شاهین‌های هریس از دو فاز اصلی استفاده می‌کنند: شناسایی اولیه طعمه و مرحله حمله نهایی. این دو فاز به مدل‌سازی جستجوی محلی و جهانی در فضای بهینه‌سازی کمک می‌کند، به طوری که با جستجوی اولیه در فضای کلی و سپس تمرکز بر نقاط پتانسیل بالا، همگامی موثری بین اکتشاف و استخراج برقرار می‌شود. در الگوریتمHHO ، رفتار شکار به صورت ریاضی به‌عنوان یک فرآیند تصادفی مبتنی بر انرژی مدل‌سازی می‌شود. در مراحل اولیه، شاهین‌ها به دنبال شناسایی مکان بهینه طعمه هستند و این فاز معادل جستجوی گسترده در فضای راه‌حل‌های ممکن است. سپس، در مراحل بعدی، بسته به فاصله و میزان انرژی طعمه، حملات نهایی به سمت هدف صورت می‌گیرد که معادل جستجوی محلی و دقیق‌تر است. فرمولاسیون ریاضی این حمله‌ها به گونه‌ای طراحی شده است که تعادلی بین فرآیندهای اکتشاف (جستجوی گسترده) و استخراج (جستجوی دقیق) برقرار می‌شود، به طوری که از گیر افتادن در بهینه‌های محلی جلوگیری شده و همگرایی بهینه بهبود یابد. HHO به دلیل سادگی، سرعت همگرایی بالا و توانایی حل مسائل پیچیده با ابعاد بالا، به طور گسترده در حوزه‌های مختلف علمی و صنعتی مورداستفاده قرار گرفته است. این الگوریتم می‌تواند به طور موثر با بهینه‌سازی مسائل چندهدفه و چندمحدوده‌ای سازگار شود و عملکرد قابل توجهی در مقایسه با سایر الگوریتم‌های بهینه‌سازی همچون GA، PSOو الگوریتم رقابت استعماری (CA) نشان داده است.

در این الگوریتم انتقال بین مرحله اکتشاف و بهره‌برداری در هر تکرار بر اساس پارامتر انرژی فرار [4] انجام می‌شود (رابطه ۱). اگر $E \geq 1$ الگوریتم وارد فاز اکتشاف می‌شود؛ ولی اگر $|E| \leq 1$ الگوریتم وارد فاز بهره‌برداری می‌شود ، معادله (۶)، (حیدری و همکاران، ۲۰۱۹) :

$$E=2E_0\left(1-\frac{t}{T}\right) \qquad (۶)$$

1. Hearst et al
2. Harris Hawks Optimization algorithm
3. Heidari et al
4. Escape Energy

که در آن E انرژی آزادسازی است، E_0 عدد تصادفی بین 1 and -1 ، T حداکثر تعداد تکرار است، t شماره تکرار فعلی است.
در مرحله اکتشاف شاهین‌ها بر اساس دو استراتژی موقعیت طعمه‌ها را شناسایی می‌کنند، در استراتژی اول موقعیت هر طعمه به‌صورت تصادفی تعیین می‌شود ($p \geq 0.5$). اما در استراتژی دوم موقعیت هر طعمه بر اساس موقعیت سایر طعمه‌ها تعیین می‌شود (معادله ۷) ($p < 0.5$) (حیدری و همکاران، ۲۰۱۹).

$$X_i^{t+1} = \begin{cases} X_{rand}^t - r_1 |X_{rand}^t - 2r_2 X_i^t| & p \geq 0.5 \\ (X_{rabbit}^t - X_m^t) - r_3(lb + r_4(ub - lb)) & p < 0.5 \end{cases} \quad (۷)$$

که در آن X_i^t، X_i^{t+1} به ترتیب موقعیت شاهین در تکرار t و t+1 است. X_{rabbit}^t موقعیت طعمه در تکرار tام است. r_1, r_2, r_3, r_4، اعداد تصادفی در بازه ۰ و ۱ هستند. ub و lb حد پایین و بالای متغیرها هستند. X_{rand}^t و X_m^t به ترتیب میانگین موقعیت جمعیت شاهین‌ها و موقعیت تصادفی شاهین‌ها در تکرار Tام است. p احتمال وقوع هر یک از استراتژی های شماره یک و دو است.
استراتژی شکار شاهین‌ها در مرحله اکتشاف بر اساس پارامترهای انرژی فرار و r مشخص می‌شود، r احتمال موفقیت یا عدم موفقیت طعمه هنگام فرار است.
اگر E≤۰.۵ الگوریتم ورد مرحله محاصره نرم می‌شود. اما اگر E ≥۰.۵ الگوریتم وارد مرحله محاصره سخت می‌شود.
در ابتدای حمله که انرژی خرگوش زیاد است شاهین‌ها به‌آهستگی به طعمه نزدیک می‌شوند (محاصره نرم)، باگذشت زمان و خسته شدن خرگوش حلقه محاصره شاهین‌ها تنگ‌تر می‌شود (محاصره سخت). بر اساس پارامترهای E and r رفتار شاهین‌ها در مرحله بهره‌برداری را می‌توان به چهار دسته تقسیم کرد (آلابول و همکاران [1] ، ۲۰۲۱).

الف) محاصره نرم

در این مرحله باوجوداینکه خرگوش انرژی کافی برای فرار دارد؛ اما در نهایت نمی‌تواند از دست شاهین‌ها فرار کند. ($|E| \geq 0.5, r \geq 0.5$).در این حالت موقعیت هر شاهین با استفاده از معادله (۸) آپدیت می‌شود.

$$X_i^{t+1} = \Delta X_i^t - E \left| J X_{rabbit} - X_i^t \right| \quad (۸)$$

$$\Delta X_i^t = X_{rabbit} - X_i^t \quad (۹)$$

$$J = 2(1 - r_5) \quad (۱۰)$$

که در آن ΔX_i^t اختلاف بین بردار موقعیت طعمه و موقعیت کنونی در تکرار tام است. r_5 یک عدد تصادفی در بازه ۰ و ۱ است. J اندازه تصادفی گام خرگوش هنگام فرار است که در هر تکرار مقدار آن را با استفاده از معادله (۱۰) محاسبه می‌شود.

ب) محاصره سخت

در این مرحله شکار به‌اندازه کافی خسته شده است و شاهین‌ها آن را به طور سخت محاصره می‌کنند ($|E| < 0.5, r \geq 0.5$).در این حالت موقعیت هر خرگوش را می‌توان با استفاده از معادله (۱۱) محاسبه کرد.

$$X_i^{t+1} = X_{rabbit}^t - E \left| \Delta X_i^t \right| \quad (۱۱)$$

ج) محاصره نرم با شیرجه‌های سریع پیش‌رونده [2]

در این حالت خرگوش انرژی کافی برای فرار دارد و محاصره نرم هنوز پا برجاست ($|E| \geq 0.5, r < 0.5$) ، در این حالت موقعیت هر خرگوش را با استفاده از معادله (۱۲) می‌توان محاسبه کرد.

1. Alabool et al
2. Soft besiege with progressive rapid dives

$$X_i^{t+1}=\begin{cases}Y \text{ if } f(Y)<f(X^t)\\ Z \text{ if } f(Z)<f(X^t)\end{cases} \quad (۱۲)$$

که در آن مقدار Y,Z با استفاده از معادله (۱۳) و (۱۴) محاسبه می‌شوند.

$$Y=X_{rabbit}^t-E\left|JX_{rabbit}^t-X_i^t\right| \quad (۱۳)$$

$$Z=Y+S\times levy_f(D) \quad (۱۴)$$

که در آن D اندازه بعد مسئله است. S یک بردار تصادفی با اندازه $1\times D$، $levy_f$ نیز یک تابع است که خروجی آن با استفاده از معادله (۱۵) قابل محاسبه است.

$$levy(x)=0.01\times\frac{u\times\sigma}{|v|^{\frac{1}{\alpha}}},\ \sigma=\left(\frac{\Gamma(1+\alpha)\times\sin\left(\frac{\pi\alpha}{2}\right)}{\Gamma\left(\frac{1+\alpha}{2}\right)\times2^{\left(\frac{\alpha-1}{2}\right)}}\right)^{\frac{1}{\alpha}} \quad (۱۵)$$

که در آن v یک عدد تصادفی بین ۰ و یک ۱ است. α یک عدد ثابتی است که مقدار آن ۱.۵ در نظرگرفته می‌شود.

د) محاصره سخت با شیرجه‌های سریع پیش‌رونده[1]

در این مرحله طعمه انرژی کافی برای فرار ندارد و شاهین‌ها تلاش می‌کنند با محاصره سخت طعمه را شکار کنند ($|E|<0.5, r<0.5$). در این حالت، موقعیت خرگوش در هر تکرار را با استفاده از معادله (۱۶) می‌توان محاسبه کرد.

$$X_i^{t+1}=\begin{cases}Y \text{ if } f(Y)<f(X^t)\\ Z \text{ if } f(Z)<f(X^t)\end{cases} \quad (۱۶)$$

که در آن مقدار Z را با استفاده از معادله (۱۰) می‌توان محاسبه کرد. مقدار Y نیز با استفاده ازمعادله (۱۷) قابل محاسبه است.

$$Y=X_{rabbit}^t-E\left|JX_{rabbit}^t-X_m^t\right| \quad (۱۷)$$

۳. مدل HHO-SVR

ابتدا داده‌های مربوط به غلظت نیترات با استفاده از الگوریتم JNB[2] خوشه‌بندی شدند (C_1, C_2, C_3) (جدول ۱)، با استفاده از این الگوریتم متوان مجموعه ای از داده‌های عددی را به گونه ای خوشه‌بندی کرد که SD[3] هر خوشه با سایر خوشه ها حداقل شود، در این حالت تفکیک داده‌ها بر اساس تابع مطلوبیت انجام شد (بوچیر و همکاران[4]، ۲۰۲۲). در ادامه توسعه برای هر خوشه یک مدل SVR توسعه داده شد و عملکرد آن با توابع کرنل مختلف ارزیابی شد (SVM^{KF})، . در این حالت پارامترهای هر تابع کرنل با استفاده از روش سعی و خطا(TE)[5] تنظیم شدند($C_1\text{-}SVM_{TE}^{KF}, C_2\text{-}SVM_{TE}^{KF}, C_3\text{-}SVM_{TE}^{KF}, C_4\text{-}SVM_{TE}^{KF}$).یکی از چالش ها هنگام آموزش هر نوع شبکه عصبی روش انتخاب بهترین متغیرهای ورودی به مدل است،اگر تعداد متغیرها بیش از اندازه کم باشد، نمیتوان انتظار نتایج قابل قبولی را از مدل داشت،اگر هم تعداد متغیرهای مدل بیش از حد زیاد باشد به دلیل پیچیده تر شدن محاسبت ممکن است مدل همگرا نشود (پریتو و همکاران[6]، ۲۰۱۶). در این مطالعه بهترین متغیرهای ورودی به هر مدل با استفاده از تکنیک SFFS[7] انتخاب شدند($C_1\text{-}S\text{-}SVM_{TE}^{KF}, C_2\text{-}S\text{-}SVM_{TE}^{KF}, C_3\text{-}S\text{-}SVM_{TE}^{KF}$)، بر این اساس هر مدل با استفاده از

1. Hard besiege with progressive rapid dives
2. Jenks Natural Breaks
3. Squared deviation
4. Bouchair et al
5. Trial and error
6. Prieto et al
7. Sequential forward floating selection

ترکیب مختلفی از متغیرهای ورودی توسعه داده می‌شود،سپس با مقایسه شاخص‌های خطا مدل‌های مختلف بهترین متغیرها برای شبیه‌سازی غلظت نیترات انتخاب می‌شوند. توابع کرنل از مهم‌ترین بخش های موثر در فرآیند یادگیری مدل SVR هستند،هر یک از این توابع پارامترهایی دارند که تعیین مقدار دقیق آن‌ها تاثیر قابل توجهی در کارایی مدل SVR دارد،به‌عنوان مثال C و ε دو پارامتر مشترک بین بسیاری از توابع کرنل هستند (دنگ و همکاران [1] ، ۲۰۱۹).

C یک پارامتر قابل‌تنظیم است که دامنه تغییرات آن بین صفر و مثبت بی‌نهایت است، وقتی به این پارامتر مقادیر بزرگی اختصاص داده شود SVR اجازه وقوع هیچ‌گونه خطای را در داده‌های آموزش نمی‌دهد، در نتیجه قدت تعمیم مدل کاهش می‌یابد؛ از طرفی هر چقدر مقدار C به صفر نزدیک‌تر باشد سخت‌گیری مدل نسبت به پذیرش خطا کمتر می‌شود. به همین دلیل تعیین دقیق مقدار C از اهمیت ویژه‌ای برخوردار است (نوبل، ۲۰۰۶). دامنه تغییرات پارامتر ε نیز مانند C بین صفر و مثبت بی‌نهایت است. این پارامتر نقش مهمی در تعیین موقعیت بردارهای پشتیبان دارد، هر چه مقدار آن بیشتر باشد تعداد بردارهای پشتیبان کمتر و پیچیدگی مدل نیز کمتر خواهد شد. اما هر چه مقدار این پارامتر کمتر باشد، تعداد بردارهای پشتیان بیشتر و در نتیجه فرآیند آموزش مدل نیز پیچیده تر می‌شود (لیانگ و ژانگ[2]، ۲۰۲۱). در صورت استفاده از تابع کرنل چندجمله ای یکی از پارامترهایی که مقدار آن باید مشخص شود d است، با استفاده از این پارامتر درجه چند جمله ای مشخص می‌شود. هر چه مقدار آن بیشتر باشد سری زمانی را با دقت بیشتری می‌توان شبیه کرد؛ اما مدل SVR پیچیده تر و در نتیجه سرعت همگرا شدن آن به دلیل افزایش محاسبات کمتر و زمان آموزش مدل نیز طولانی تر می‌شود. اما هرچه درجه چندحمله ای کمتر باشد سری زمانی را با در نظر گرفتن جزئیات کم تری می‌توان شبیه‌سازی کرد؛ اما سرعت همگرا شدن مدل بیشتر و مدت زمان فرآیند آموزش کمتر می‌شود (دبنات و تاکاهاشی[3]، ۲۰۰۴). در این مطالعه یک مدل هیبریدی جدیدی با در نظر گرفتن همه محدودیت های بالا توسعه داده شد. در این مدل از الگوریتم HHO برای تعیین دقیق تر مقدار پارامترهای هر تابع کرنل در مدل SVR استفاده شد ($C_1 - S -$ $SVM_{HHO}^{KF}, C_2 - S - SVM_{HHO}^{KF}, C_3 - S - SVM_{HHO}^{KF}$). برای توسعه همه مدل‌ها از ۸۰ درصد داده‌ها برای آموزش و از سایر داده‌های باقیمانده برای تست استفاده شد.در نهایت برای همه مدل‌ها آنالیز حساسیت انجام شد و حساسیت آن‌ها نسبت به پارامترهای مختلف ارزیابی شد.

جدول ۲. خروجی الگوریتم JNB

Cluster name	Lover bund	Upper bund	Number of data
C_1	7.43	13.98	530
C_2	14	32.1	485
C_3	33.58	70.5	439

۴. شاخص‌های مورداستفاده برای صحت سنجی مدل

بعد از توسعه مدل‌های مختلف از شاخص‌های زیر برای ارزیابی عملکرد مدل‌ها استفاده شد.

$$RMSE= \left[N^{-1}\sum_{i=1}^{N}(x_{obs}- x_{sim})^2\right]^{.5} ,(0,+\infty] \quad (۱۸)$$

$$MAE= \frac{1}{n}\sum_{i=1}^{n}| x_{obs}- x_{sim}|,(0,+\infty] \quad (۱۹)$$

$$R^2= \frac{\sum_{i=1}^{N}(x_{sim}- \bar{x}_{obs})^2}{\sum_{i=1}^{N}(x_{obs}- \bar{x}_{obs})^2}, (-\infty,1] \quad (۲۰)$$

1. Deng et al
2. Liang and Zhang
3. Debnath and Takahashi

که در آن x_{sim} و x_{obs} به ترتیب مقدار مشاهده شده و شبیه‌سازی شده است. $\bar{x}_{obs}$ مقدار متوسط داده‌های مشاهده ای است. Nتعداد داده‌های مشاهده شده است.

یافته‌های پژوهش

۱. خوشه اول

بر اساس نتایج حاصل از الگوریتم JNB، ۳۷ درصد از کل داده‌ها به خوشه اول تعلق دارند و متوسط غلظت نیترات در این خوشه ۱۱.۴۸۷ است، در جدول (۳) مشخصات اماری نمونه های متعلق به این خوشه نشان داده شده‌اند. ابتدا مدل SVR_{TE} برای این خوشه توسعه داده شد و عملکرد آن با توابع کرنل مختلف ارزیابی شد. بر اساس نتایج حاصل، برای این خوشه این مدل با تابع کرنل SK بهترین عملکرد را دارد (جدول (۳)). همچنین بر اساس نتایج حاصل از اجرای متد SFFS هم‌زمان با فرآیند اموزش EC, pH, TDS, HCO3, SO4 بهترین متغیرها برای شبیه‌سازی غلظت نیترات در این خوشه هستند، در جدول (۴) پارامترهای مدل $C_1\text{-S-SVR}_{TE}^{SK}$ نشان داده شده‌اند. همچنین اشکال (۴) و (۵) سری زمانی داده‌های مشاهده ای و سری زمانی حاصل از این مدل را در مرحله آموزش و تست نشان می‌دهد. در مرحله بعد مدل SVR_{HHO} برای این خوشه توسعه داده شد و مانند مدل SVR_{TE} عملکرد آن با توابع کرنل مختلف ارزیابی شد. با این تفاوت که در این حالت بهینه ترین مقدار پارامترهای توابع کرنل مختلف مدل SVR توسط الگوریتم HHO محاسبه شدند و سپس عملکرد مدل‌های SVR_{HHO} مختلف بایکدیگر مقایسه شدند، براساس نتایج حاصل، مدل SVR_{HHO} نیز با تابع کرنل SK بهترین عملکرد را برای شبیه‌سازی غلظت نیترات در مرحله آموزش و تست داده‌های این خوشه دارد ($C_1\text{-S-SVR}_{HHO}^{SK}$)، در جدول (۶) نتایج حاصل از توسعه این مدل نشان داده شده‌اند. در جدول (۷) پارامترهای این مدل و در شکل های (۶) و (۷) سری زمانی داده‌های مشاهده ای و سری زمانی حاصل از این مدل در مرحله آموزش و تست نشان داده شده‌اند.

جدول ۳. ویژگی‌های آماری متغیر وابسته و متغیرهای مستقل در خوشه ۱

Variable	Unit	Mean	StDev	CoefVar	Minimum	Maximum
Na	mg/L	26.65	13.91	46.46	1.81	51.50
Ca	mg/L	56.48	19.71	33.15	21.63	91.42
Mn	mg/L	0.13	0.03	0.61	0.12	0.13
K	mg/L	5.33	3.02	47.65	0.66	10.91
Fe	mg/L	0.209	0.00	0.28	0.20	0.21
Al	mg/L	0.11	0.00	0.40	0.11	0.11
F	mg/L	9.19	4.00	40.52	2.60	15.80
Cl	mg/L	19.58	10.93	49.73	0.88	38.34
SO4	mg/L	66.39	36.67	52.57	0.95	132.06
CO3	mg/L	1.49	0.78	51.48	0.50	2.46
HCO3	mg/L	230.36	69.71	27.53	107.88	352.80
PO4	mg/L	2.43	0.73	27.75	1.51	3.35
NO2	mg/L	1.59	0.81	51.07	0.58	2.62
Br	mg/L	0.01	0.00	17.34	0.042	0.07
CO2	mg/L	11.15	5.46	48.99	2.15	20.18
WT[1]	C°	21.27	4.87	22.91	12.80	29.59
AT[2]	C°	25.00	6.65	26.61	13.5	36.39
pH	-	6.59	0.04	0.68	6.50	6.67
Alkalinity	mg/L	256.50	76.61	29.87	120.06	394.31
TH[3]	mg/L	339.99	148.41	42.34	84.18	596.75
TDS	mg/L	291.65	65.34	18.55	166.81	416.62
EC	µmohs/cm	636.68	139.25	19.90	395.26	878.08
NO3	mg/L	11.48	1.43	12.51	7.63	13.97
Turbidity	NTU	0.80	0.20	23.98	0.43	1.16
Mg	mg/L	26.59	11.19	38.29	7.16	45.92

جدول ٤ . نتایج شاخص‌های ارزیابی مدل C_1-S-SVR_{TE}^{SK} در مرحله آموزش و تست

Training						Testing					
RMSE	MAE	WI	EVS	KGE	R^2	RMSE	MAE	WI	EVS	KGE	R^2
0.2423	0.2389	0.9954	0.9720	0.9781	0.9814	0.2534	0.2407	0.9924	0.9656	0.9744	0.9823

جدول ۵. مقدار بهینه پارامترهای مدل C_1-S-SVR_{TE}^{SK} در مرحله آموزش و تست

Model	Parameter	Value
SFFS	Best input Combination	EC, pH ,TDS, HCO3, SO4
SVR	Kernel Function	SK
	C	0.00894
	β	0.05913

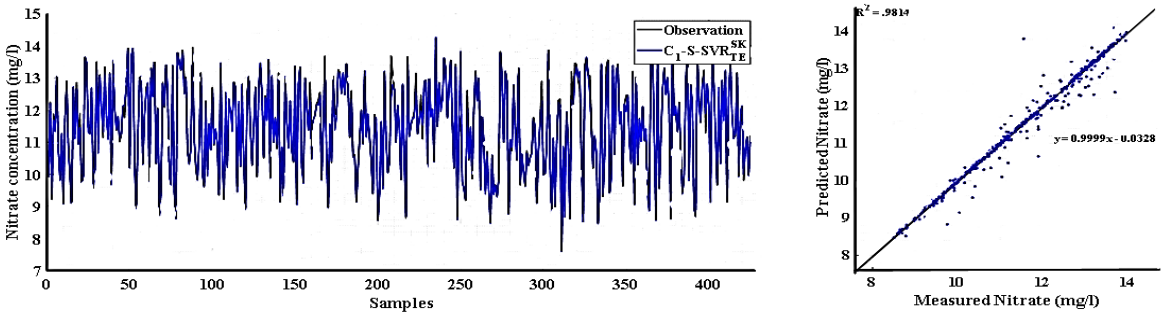
جدول ٦. نتایج شاخص‌های ارزیابی مدل $C_1 - S - SVR_{HHO}^{SK}$ در مرحله آموزش و تست

Training						Testing					
RMSE	MAE	WI	EVS	KGE	R^2	RMSE	MAE	WI	EVS	KGE	R^2
0.1489	0.1187	0.9977	0.9926	0.9956	0.9969	0.1304	0.0747	0.9984	0.9967	0.9974	0.9983

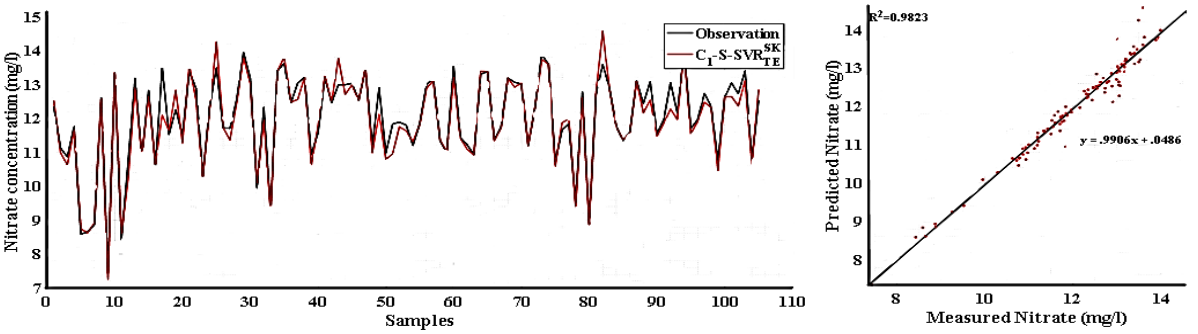
1. Water temperature
2. Air temperature
3. Total hardness

جدول ۷. مقدار بهینه پارامترهای مدل C_1-S-SVR^{SK}_{HHO} در مرحله آموزش و تست

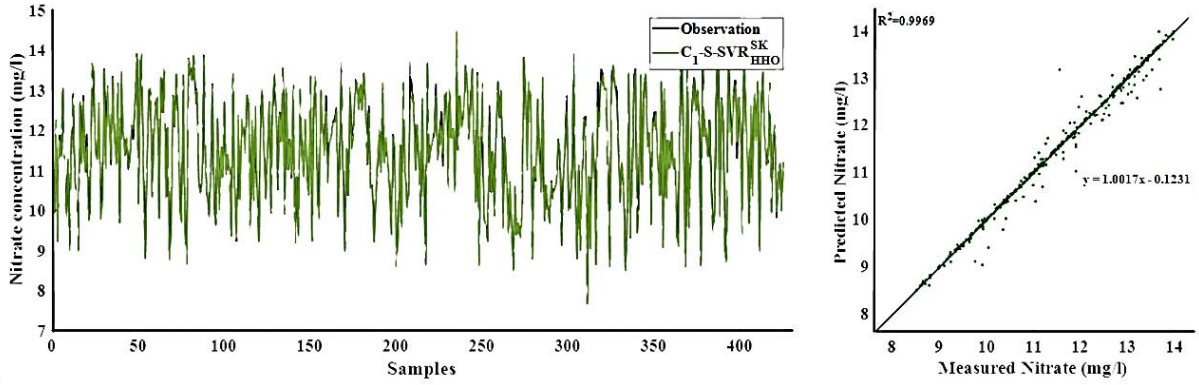
Model	Parameter	Value
SFFS	Best input combination	EC, pH , TDS, HCO3, SO4
HHO	Number of search agents	100
	Maximum number of iterations	1000
SVR	Kernel Function	SK
	C	0.02691
	β	0.5061



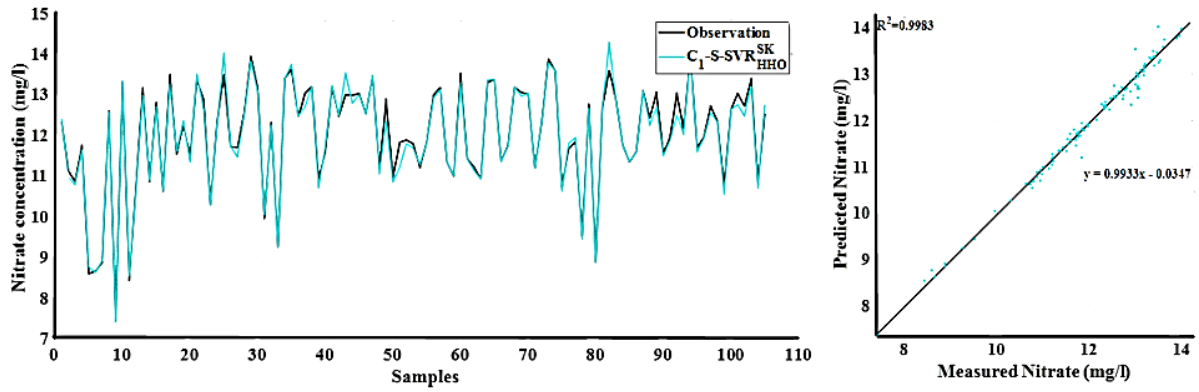
شکل ٤. عملکرد مدل C_1-S-SVR^{SK}_{TE} در مرحله آموزش برای پیش‌بینی غلظت نیترات



شکل ٥. عملکرد مدل C_1-S-SVR^{SK}_{TE} در مرحله تست برای پیش‌بینی غلظت نیترات



شکل ٦. عملکرد مدل C_1-S-SVR^{SK}_{HHO} در مرحله آموزش برای پیش‌بینی غلظت نیترات



شکل ۷. عملکرد مدل C_1-S-SVR^{SK}_{HHO} در مرحله تست برای پیش‌بینی غلظت نیترات

۲. خوشه دوم

روند شبیه‌سازی داده‌های خوشه دوم (C_2) شبیه روند شبیه‌سازی داده‌های م خوشه اول است، ۳۴ درصد داده‌ها متعلق به این خوشه هستند و متوسط غلظت نیترات در این خوشه ۲۲/۲۸۳ است که در جدول (۸) نشان داده شده است. برای این خوشه نیز ابتدا عملکر مدل SVR_{RBF} با توابع کرنل مختلف ارزیابی شد، بر اساس نتایج حاصل، این مدل با تابع کرنل SK بهترین عملکرد را در مرحله آموزش و تست داده‌های این خوشه را دارد(C_2-S-SVR^{RBF}_{TE}) که در جدول (۹) نتایج ارائه شده است. همچنین باتوجه‌به اجرای متد SFFS هم‌زمان با فرآیند آموزش EC, pH , TDS, HCO3, Cl بهترین متغیرها برای شبیه‌سازی غلظت نیترات در این خوشه هستند. در جدول (۱۰) پارامترهای مدل C_2-S-SVR^{RBF}_{TE} نشان داده شده‌اند، همچنین در شکل های (۸) و (۹) سری زمانی داده‌های مشاهده ای وسری زمانی حاصل از این مدل را نشان می‌دهد. در مرحله بعد مدل SVR_{HHO} برای این خوشه توسعه داده شد. بر اساس نتایج حاصل از بهینه سازی پازامترهای توبع کرنل مختلف مدل HHO SVR مدل SVR_{HHO} با تابع کرنل RBF بهترین عملکرد را برای شبیه‌سازی غلظت نیترات در مرحله آموزش و تست این خوشه دارد (C_2-S-SVR^{RBF}_{HHO})، نتایج حاصل از توسعه این مدل در جدول (۱۱) نشان داده شده‌اند. در جدول (۱۲) پارامترهای مختلف این مدل و در شکل های (۱۰) و (۱۱) سری زمانی حاصل داده‌های مشاهده ای وسری زمانی حاصل از این مدل را نشان داده شده است. بر اساس نتایج حاصل از آنالیز حساسیت این مدل همانند مدل C_1-S-SVR^{SK}_{HHO} نسبت به EC بیشترین حساسیت را دارد، اما نسبت بهTDS کم ترین حساسیت را دارد.

جدول ۸. ویژگی‌های آماری متغیر وابسته و متغیرهای مستقل در خوشه ۲

Variable	Unit	Mean	StDev	CoefVar	Minimum	Maximum
Na	mg/L	29.99	16.28	54.31	2.44	57.78
Ca	mg/L	59.55	21.62	36.30	22.60	95.98
Mn	mg/L	0.10	0.00	0.58	0.099	0.101
K	mg/L	6.36	3.59	56.44	0.80	12.86
Fe	mg/L	0.21	0.00	0.29	0.20	0.21
Al	mg/L	0.12	0.00	0.44	0.12	0.13
F	mg/L	9.90	4.31	43.53	2.80	16.98
Cl	mg/L	22.00	12.38	56.28	0.99	43.00
SO4	mg/L	70.16	40.20	57.30	1.00	139.00
CO3	mg/L	1.5	0.81	53.15	0.52	2.55
HCO3	mg/L	253.55	70.21	27.69	120.77	389.35
PO4	mg/L	2.68	0.81	30.53	1.66	3.69
NO2	mg/L	1.36	0.69	51.13	0.49	2.22
Br	mg/L	0.06	0.01	19.46	0.04	0.08
CO2	mg/L	11.19	5.49	49.12	2.15	20.19
WT	C°	21.24	4.89	23.04	12.85	29.71
AT	C°	25.00	6.632	26.53	13.60	36.41
pH	-	7.33	0.05	0.79	7.23	7.42
Alkalinity	mg/L	257.04	76.08	29.60	120.32	394.48
TH	mg/L	350.15	130.62	37.30	60.76	637.84
TDS	mg/L	352.77	93.19	26.42	202.40	503.24
EC	µmohs/cm	699.74	355.61	50.82	261.8	1138.90
NO3	mg/L	22.28	4.48	20.11	13.85	32
Turbidity	NTU	0.85	0.22	26.21	0.47	1.28
Mg	mg/L	29.23	12.39	42.40	8.23	50.60

جدول ۹. نتایج شاخص‌های ارزیابی مدل C_2-S-SVR$_{TE}^{RBF}$ در مرحله آموزش و تست

Training						Testing					
RMSE	MAE	WI	EVS	KGE	R^2	RMSE	MAE	WI	EVS	KGE	R^2
0.2547	0.2476	0.9904	0.9633	0.9679	0.9816	0.2608	0.2488	0.9884	0.9625	0.9657	0.9801

جدول ۱۰. مقدار بهینه پارامترهای مدل C_2-S-SVR$_{TE}^{RBF}$ در مرحله آموزش و تست

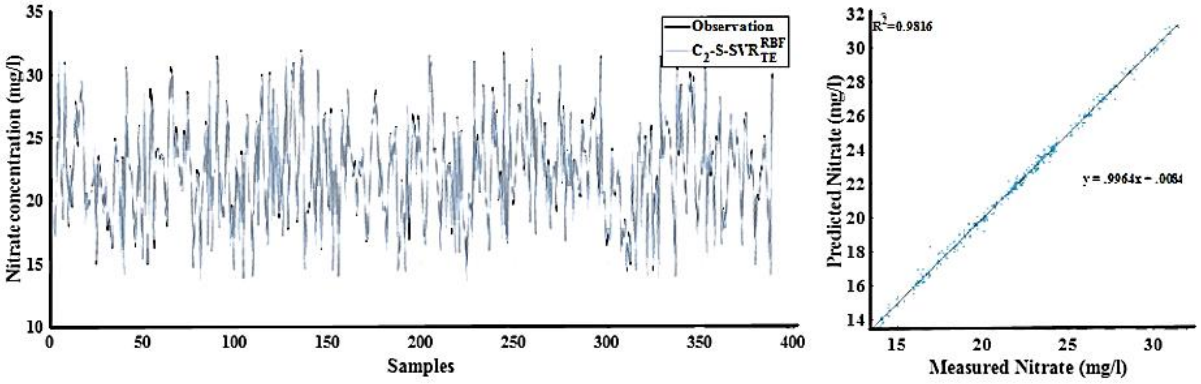
Model	Parameter	Value
SFFS	Best input Combination	EC, pH , TDS, HCO3, Cl
	Kernel Function	RBF
SVR	α	0.04196
	C	3

جدول ۱۱. نتایج شاخص‌های ارزیابی مدل C_2-S-SVR$_{HHO}^{RBF}$ در مرحله آموزش و تست

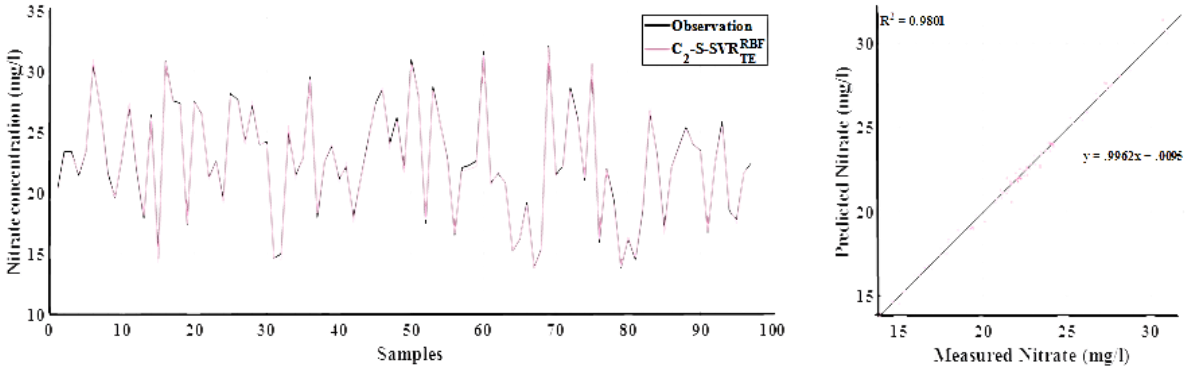
Training						Testing					
RMSE	MAE	WI	EVS	KGE	R^2	RMSE	MAE	WI	EVS	KGE	R^2
0.1698	0.2476	0.1303	0.9951	0.9924	0.9942	0.1585	0.1299	0.9959	0.9917	0.9940	0.9955

جدول ۱۲. مقدار بهینه پارامترهای مدل C_2-S-SVR$_{HHO}^{RBF}$ در مرحله آموزش و تست

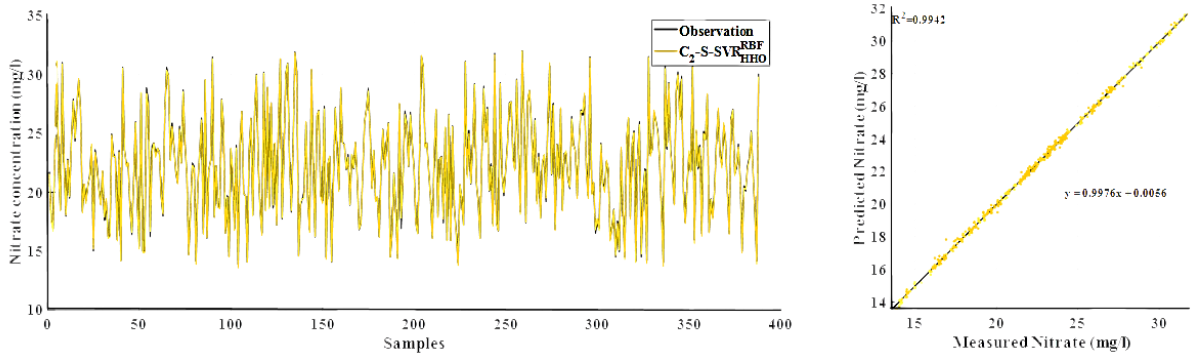
Model	Parameter	Value
SFFS	Best input combination	EC, pH, TDS, HCO3, Cl
HHO	Number of search agents	100
	Maximum number of iterations	1000
SVR	Kernel Function	RBF
	α	0.07004
	C	4



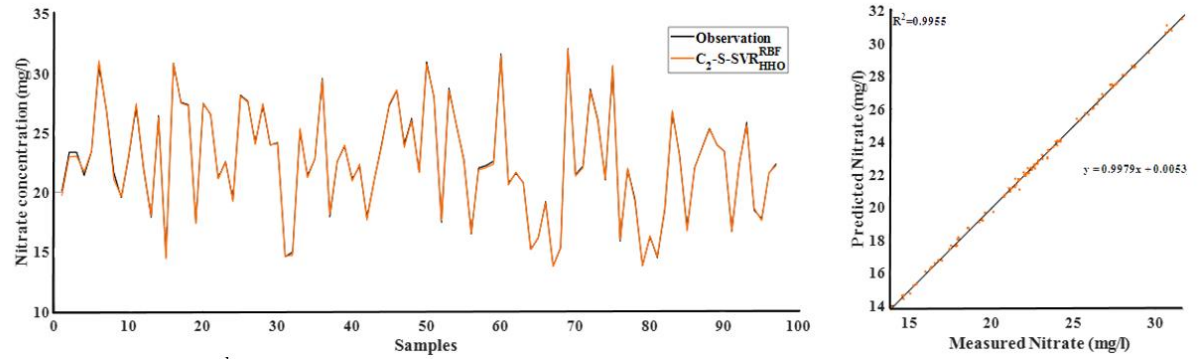
شکل ۸. عملکرد مدل C_2-S-SVR$^{RBF}_{TE}$ در مرحله آموزش برای پیش‌بینی غلظت نیترات



شکل ۹. عملکرد مدل C_2-S-SVR$^{RBF}_{TE}$ در مرحله تست برای پیش‌بینی غلظت نیترات



شکل ۱۰. عملکرد مدل C_2-S-SVR$^{RBF}_{HHO}$ در مرحله آموزش برای پیش‌بینی غلظت نیترات



شکل ۱۱. عملکرد مدل C_2-S-SVR$^{RBF}_{HHO}$ در مرحله تست برای پیش‌بینی غلظت نیترات

۳. خوشه سوم

۳۰ درصد داده‌ها به خوشه سوم تعلق دارند و متوسط غلظت نیترات در این خوشه ۴۴/۱۱۹ است که در جدول (۱۲) نشان داده شده است. بر اساس نتایج حاصل از توسعه مدل $SVRte$ و مطابق جدول (۱۳)، این مدل برای این خوشه با تابع کرنل PK بهترین عملکرد را در مرحله آموزش و تست را دارد (C_3-S-SVR_{TE}^{PK}). همچنین بر اساس نتایج حاصل از اجرای متد SFFS، HCO3, Ca, Na, EC, pH ,TDS, بهترین متغیرها برای شبیه‌سازی غلظت نیترات در این خوشه هستند. در جدول (۱۴) پارامترهای مختلف این مدل نشان داده شده‌اند (C_3-S-SVR_{TE}^{PK}). همچنین در شکل های (۱۲) و (۱۳) سری زمانی داده‌های مشاهده ای و شبیه‌سازی شده توسط این مدل نشان داده شده‌اند. در مرحله اخر بخش SVR_{HHO} برای این خوشه توسعه داده شد، بر اساس نتایج حاصل مدل SVR_{HHO} با تابع کرنل PK برای این خوشه (C_3-S-SVR_{HHO}^{PK}) بهترین عملکرد را در مرحله آموزش و تست دارد که در جدول (۱۵) نشان داده شده است. در جدول (۱۶) پارامترهای مختلف این مدل نشان داده شده‌اند. شکل های (۱۴) و (۱۵) نیز سری زمانی داده‌های مشاهده ای برای خوشه سوم و سری زمانی حاصل از مدل (C_3-S-SVR_{HHO}^{PK}) را نشان می‌دهد.

جدول ۱۲. ویژگی‌های آماری متغیر وابسته و متغیرهای مستقل در خوشه ۳

Variable	Unit	Mean	StDev	CoefVar	Minimum	Maximum
Na	mg/L	27.90	14.58	48.62	1.93	53.56
Ca	mg/L	57.09	19.70	33.14	21.85	92.63
Mn	mg/L	0.09	0.00	3.47	0.02	0.10
K	mg/L	6.13	3.45	54.07	0.78	12.42
Fe	mg/L	0.20	0.00	0.27	0.20	0.211
Al	mg/L	0.10	0.00	0.41	0.10	0.11
F	mg/L	8.89	3.88	39.37	2.52	15.28
Cl	mg/L	20.46	11.52	52.40	0.92	39.99
SO4	mg/L	62.30	35.63	50.90	0.89	123.71
CO3	mg/L	1.40	0.73	48.32	0.47	2.32
HCO3	mg/L	238.44	75.88	29.92	111.39	364.25
PO4	mg/L	2.53	0.76	28.99	1.57	3.51
NO2	mg/L	2.07	1.06	51.10	0.75	3.40
Br	mg/L	0.05	0.00	17.34	0.03	0.07
CO2	mg/L	11.17	5.45	48.85	2.15	20.19
WT	C°	21.29	4.94	23.22	12.82	29.70
AT	C°	25.01	6.64	26.58	13.56	36.43
pH	-	7.17	0.04	0.76	7.08	7.27
Alkalinity	mg/L	256.49	82.64	32.22	119.87	391.85
TH	mg/L	315.58	155.99	44.48	55.74	575.80
TDS	mg/L	334.61	77.94	22.13	191.46	478.42
EC	µmohs/cm	698.20	564.10	80.80	21.18	1377.29
NO3	mg/L	44.11	8.53	19.34	33.58	70.50
Turbidity	NTU	0.80	0.21	26.00	0.44	1.21
Mg	mg/L	27.77	11.76	40.24	7.44	48.04

جدول ۱۳. نتایج شاخص‌های ارزیابی مدل C_3-S-SVR_{TE}^{PK} در مرحله آموزش و تست

Training						Testing					
RMSE	MAE	WI	EVS	KGE	R^2	RMSE	MAE	WI	EVS	KGE	R^2
0.2191	0.1843	0.1843	0.9993	0.9970	0.9993	0.2282	0.2156	0.9965	0.9984	0.9963	0.9901

جدول ١٤. مقدار بهینه پارامترهای مدل C_3-S-SVR^{PK}_{TE} در مرحله آموزش و تست

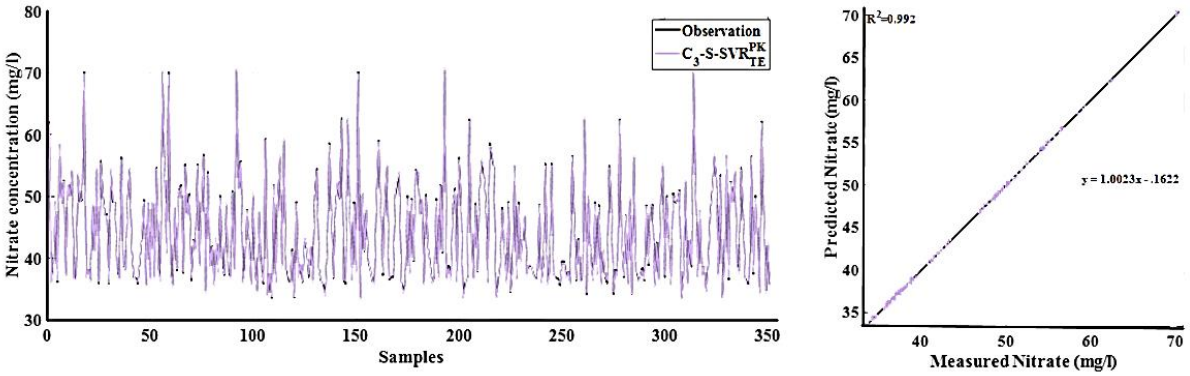
Model	Parameter	Value
SFFS	Best input Combination	EC, pH ,TDS,HCO3, Ca,Na
SVR	Kernel Function	SK
	C	0.0727
	Γ	0.3066
	D	4

جدول ۱۵. نتایج شاخص‌های ارزیابی مدل C_3-S-SVR^{PK}_{HHO} در مرحله آموزش و تست

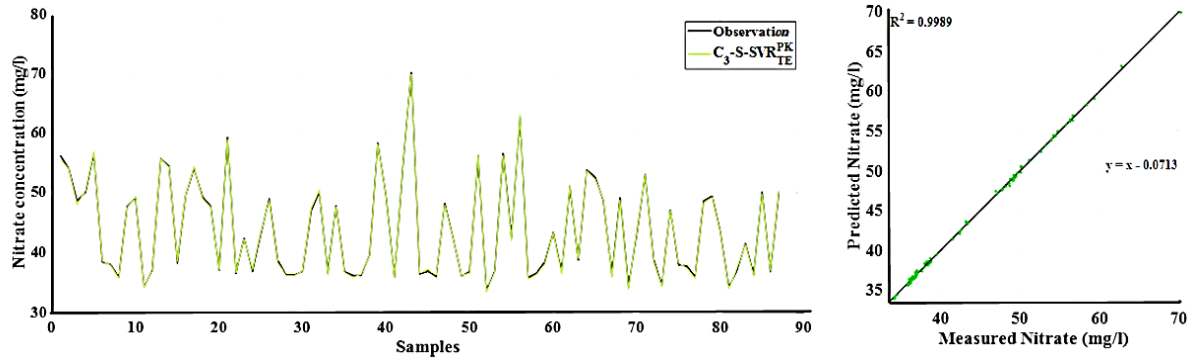
Training						Testing					
RMSE	MAE	WI	EVS	KGE	R^2	RMSE	MAE	WI	EVS	KGE	R^2
0.1319	0.1018	0.9981	0.9958	0.9968	0.9973	0.1217	0.0490	0.9999	0.9997	0.9985	0.9997

جدول ١٦. مقدار بهینه پارامترهای مدل C_3-S-SVR^{PK}_{HHO} در مرحله آموزش و تست

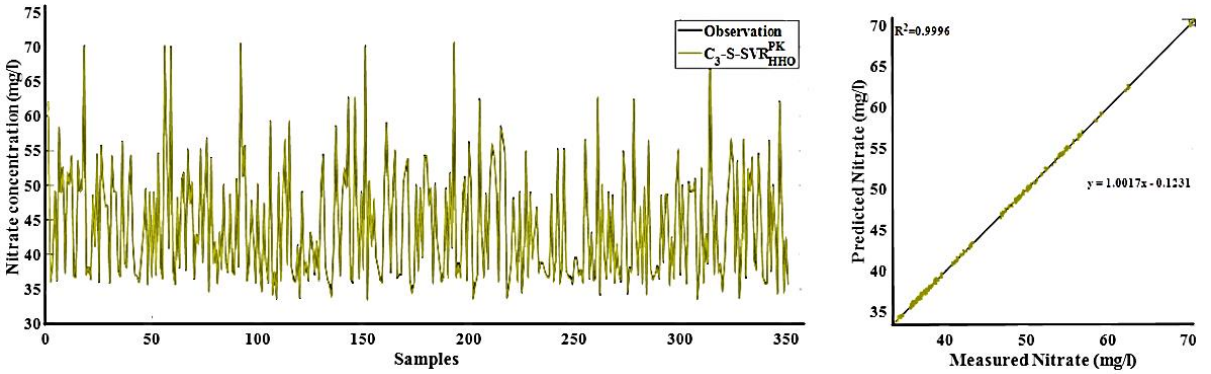
Model	Parameter	Value
SFFS	Best input combination	EC, pH ,TDS,HCO3, Ca,Na
HHO	Number of search agents	100
	Maximum number of iterations	1000
SVR	Kernel Function	PK
	C	0.01196
	Γ	0.2750
	D	4



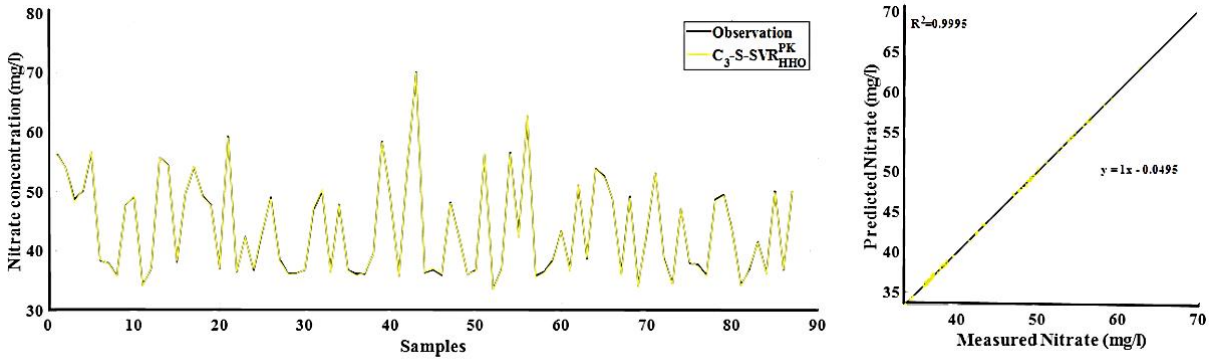
شکل ۱۲. عملکرد مدل C_3-S-SVR^{PK}_{TE} در مرحله آموزش برای پیش‌بینی غلظت نیترات



شکل ۱۳. عملکرد مدل C_3-S-SVR^{PK}_{TE} در مرحله تست برای پیش‌بینی غلظت نیترات



شکل ۱۴. عملکرد مدل C_3-S-SVR$^{PK}_{HHO}$ در مرحله آموزش برای پیش‌بینی غلظت نیترات



شکل ۱۵. عملکرد مدل C_3-S-SVR$^{PK}_{HHO}$ در مرحله تست برای پیش‌بینی غلظت نیترات

بحث

شبیه‌سازی و پیش‌بینی غلظت نیترات همیشه از مهم‌ترین مسائل در حوزه مدیریت منابع آب بوده است، در این تحقیق بعد از جمع‌آوری داده‌ها ابتدا داده‌های مربوط به غلظت نیترات با استفاده از JNB خوشه‌بندی شدند، سپس برای هر خوشه یک مدل SVR توسعه داده شد، هم‌زمان با فرآیند آموزش این مدل از الگوریتم SFFS برای انتخاب متغیرهای وردی به مدل استفاده شد، سپس بر اساس نتایج حاصل از این سه مدل مقدار متوسط شاخص‌های خطا برای مرحله آموزش (RMSE=0.2387, MAE=0.2236, R^2=0.9874) و تست (RMSE=0.2474, MAE= 0.2350, R^2=0.9841) محاسبه شدند، در این حالت از روش سعی و خطا برای این کار استفاده شد. در گام بعد از الگوریتم HHO برای تعیین مقدار بهینه پارامترهای توابع کرنل استفاده شد، در این حالت مقدار شاخص‌های R^2 و RMSE, MAE برای مرحله آموزش به ترتیب برابر۰.۰۹۹۶۱،۰.۱۱۶۹،۰.۱۵۰۲ است و مقداران ها برای مرحله تست به ترتیب برابر ۰.۹۸۴۵،۰.۱۳۰۸، ۰.۹۹۷۸ است .بر اساس نتایج حاصل از این مطالعه اولا استفاده از HHO برای پیش‌بینی غلظت نیتزات میتواند باعث افزایش چشمگیر دقت مدل SVR می‌شود،دوما استفاده به جا از مدل‌های یادگیری ماشین مختلف در کنار هم میتوانند نقش موثری در افزایش دقت مدل‌های رگرسیونی مانند SVR داشته باشد.

نتیجه‌گیری

باتوجه‌به نتایج به‌دست‌آمده، استفاده از الگوریتم HHO به طور قابل‌توجهی دقت مدل SVR را در پیش‌بینی غلظت نیترات بهبود بخشیده است. این بهبود نه‌تنها در شاخص‌های مرحله آموزش بلکه در مرحله تست نیز به‌وضوح مشاهده می‌شود. به‌کارگیری الگوریتم‌های بهینه‌سازی در تعیین پارامترهای مدل‌های یادگیری ماشین همچون SVR نشان‌دهنده اهمیت بالای انتخاب بهینه پارامترها در افزایش کارایی مدل‌ها است. این امر به‌ویژه در حوزه منابع آب و پیش‌بینی آلاینده‌های مهمی مانند نیترات می‌تواند راهگشا باشد. علاوه بر این، خوشه‌بندی داده‌ها پیش از مدل‌سازی نقش کلیدی در بهبود عملکرد مدل ایفا کرده است. داده‌های مختلف با ویژگی‌های آماری متفاوت، اگر به‌درستی خوشه‌بندی و به‌صورت مستقل مدل‌سازی شوند، می‌توانند منجر به نتایج دقیق‌تر

و قابل‌اعتمادتری شوند. این تحقیق نشان داد که خوشه‌بندی داده‌ها با استفاده از روش JNB و به‌کارگیری مدل‌های جداگانه برای هر خوشه می‌تواند به بهبود قابل‌توجه پیش‌بینی‌ها منجر شود و استفاده از این رویکرد در تحقیقات آینده توصیه می‌شود.

همچنین استفاده از الگوریتم SFFS نیز برای انتخاب متغیرهای ورودی مدل به‌عنوان یکی دیگر از دستاوردهای این تحقیق برجسته شد. انتخاب بهینه و دقیق متغیرهای ورودی می‌تواند تأثیر زیادی بر عملکرد مدل داشته باشد و این تحقیق نشان داد که با انتخاب مناسب متغیرها، حتی در صورت استفاده از یک مدل نسبتاً ساده مانندSVR ، می‌توان به نتایج بسیار دقیقی دست‌یافت. این نشان می‌دهد که کیفیت داده‌ها و انتخاب ورودی‌های مناسب به‌اندازه پیچیدگی خود مدل اهمیت دارد .در نهایت، این تحقیق تأیید می‌کند که ترکیب الگوریتم‌های بهینه‌سازی مانند HHO با مدل‌های یادگیری ماشین و به‌کارگیری روش‌های انتخاب ویژگی، می‌تواند به‌عنوان یک راهکار مؤثر برای بهبود دقت پیش‌بینی‌ها در مسائل مرتبط با منابع آب، به‌ویژه در پیش‌بینی آلاینده‌های مهم، مورداستفاده قرار گیرد. این راهبردها می‌توانند در مدیریت بهتر منابع آب و کاهش آلودگی‌های ناشی از نیترات و سایر آلاینده‌ها به کار گرفته شوند.

ملاحظات اخلاقی

پیروی از اصول اخلاق پژوهش

نویسندگان اصول اخلاقی را در انجام و انتشار این پژوهش علمی رعایت نموده‌اند و این موضوع مورد تأیید همه آنهاست.

تعارض منافع

بنا بر اظهار نویسندگان این مقاله تعارض منافع ندارد.

References

Adeloju, S.B., Khan, S., & Patti, A.F. (2021). Arsenic Contamination of Groundwater and Its Implications for Drinking Water Quality and Human Health in Under-Developed Countries and Remote Communities—A Review. *Appl. Sci.* 11, 1926. https://doi.org/10.3390/app11041926

Alabool, H.M., Alarabiat, D., Abualigah, L., Heidari, A.A. (2021). Harris hawks optimization: a comprehensive review of recent variants and applications. *Neural Comput & Applic*, 33, 8939–8980. https://doi.org/10.1007/s00521-021-05720-5

Amiri, S., Rajabi, A., Shabanlou, S., Yosefvand, F., & Izadbakhsh, M.A. (2023). Prediction of groundwater level variations using deep learning methods and GMS numerical model. *Earth Sci Inform*, 16, 3227–3241. https://doi.org/10.1007/s12145-023-01052-1

Azizi, E., Yosefvand, F., Yaghoubi, B., Izadbakhsh, M.A., & Shabanlou, S. (2023) Modelling and prediction of groundwater level using wavelet transform and machine learning methods: A case study for the Sahneh Plain, Iran. *Irrigation and Drainage*, 72(3), 747–762. https://doi.org/10.1002/ird.2794

Bouchair, A., Yagoubi, B., & Makhlouf, S.A. (2022). A Cluster-Oriented Policy for Virtual Network Embedding in SDN-Enabled Distributed Cloud. *International Journal of Computing and Digital Systems*, 11(1), 365-353. https://dx.doi.org/10.12785/ijcds/120129

Chai, T., & Draxler, R.R. (2014) Root Mean Square Error (RMSE) or Mean Absolute Error (MAE)?—Arguments against Avoiding RMSE in the Literature. *Geoscientific Model Development*, 7, 1247-1250. https://dx.doi.org/10.5194/gmd-7-1247-2014

Chen, J., Xin, B., Peng, Zh., Dou, L., & Zhang, J. (2009). Optimal contraction theorem for exploration–exploitation tradeoff in search and optimization. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, 39(3), 680-691. https://dx.doi.org/10.1109/TSMCA.2009.2012436

Cortes, C., & Vapnik, V. (1995). Support-Vector Networks. *Machine Learning*, 20, 273-297. https://dx.doi.org/10.1007/BF00994018

Debnath, R., & Takahashi, H. (2004). Kernel selection for the support vector machine. *IEICE transactions on information and systems*, 87(12), 2903-2904. https://www.researchgate.net/publication/220237100_Kernel_selection_for_the_support_vector_machine

Deng, W., Yao, R., Zhao, H.,Yang, X., &Li, G. (2019). A novel intelligent diagnosis method using optimal LS-SVM with improved PSO algorithm. *Soft Comput*, 23, 2445–2462. https://doi.org/10.1007/s00500-017-2940-9

Di Bucchianico, A. (2008). Coefficient of determination (R^2). *Encyclopedia of statistics in quality and reliability*, 1, Wiley Publicatins. https://doi.org/10.1002/9780470061572.eqr173

El Amri, A., M'nassri, S., Nasri, N., Nsir, H., & Majdoub, R. (2022). Nitrate concentration analysis and prediction in a shallow aquifer in central-eastern Tunisia using artificial neural network and time series modelling. *Environmental Science and Pollution Research*, 29(28), 43300-43318. https://doi.org/10.1007/s11356-021-18174-y

Elzain, H. E., Chung, S.Y., Park, K.H., Senapathi, V., Sekar, S., Sabarathinam, Ch., Hassan, M. (2021). ANFIS-MOA models for the assessment of groundwater contamination vulnerability in a nitrate contaminated area. *Journal of Environmental Management*, 286, 112162. https://doi.org/10.1016/j.jenvman.2021.112162

Esmaeili, F., Shabanlou, S., & Saadat, M. (2021). A wavelet-outlier robust extreme learning machine for rainfall forecasting in Ardabil City, Iran. *Earth Sci Inform*, 14, 2087–2100. https://doi.org/10.1007/s12145-021-00681-8

Fallahi, M.M., Shabanlou, S., Rajabi, A., Yosefvand, F., & IzadBakhsh, M.A. (2023). Effects of climate change on groundwater level variations affected by uncertainty (case study: Razan aquifer). *Appl Water Sci*, 13(143). https://doi.org/10.1007/s13201-023-01949-8

Fried, J.J. (1975) Groundwater pollution. Elsevier Scientific Publishing Company, Amsterdam. https://www.scirp.org/reference/referencespapers?referenceid=102612

Hearst, M. A., Dumais, S.T., Osuna, E., Platt, J., & Scholkopf, B. (1998). Support vector machines. *IEEE Intelligent Systems and their applications*, 13(4), 18-28. https://Doi:10.1109/5254.708428

Heidari, A. A., Mirjalili, S.A., Faris, H., Aljarah, I., Mafarja, M., & Chen, H. (2019). Harris hawks optimization: Algorithm and applications. *Future generation computer systems*, 97, 849-872. https://doi.org/10.1016/j.future.2019.02.028

Lahjouj, A., Hmaidi, A.E. Bouhafa, K., & Boufala. M. (2020). Mapping specific groundwater vulnerability to nitrate using random forest: case of Sais basin, Morocco. *Modeling Earth Systems and Environment*, 6(3), 1451-1466. https://link.springer.com/article/10.1007/s40808-020-00761-6

Liang, Z., & Zhang, L. (2021). Support vector machines with the ε-insensitive pinball loss function for uncertain data classification. *Neurocomputing*, 457, 117-127. https://doi.org/10.1016/j.neucom.2021.06.044

Mazraeh, A., Bagherifar, M., Shabanlou, S., & Ekhlasmand, R. (2023). A hybrid machine learning model for modeling nitrate concentration in water sources. *Water, Air, & Soil Pollution*, 234(11), 721. https://doi.org/10.1007/s11270-023-06745-3

Mazraeh, A., Bagherifar, M., Shabanlou, S., & Ekhlasmand, R. (2024). A novel committee-based framework for modeling groundwater level fluctuations: A combination of mathematical and machine learning models using the weighted multi-model ensemble mean algorithm. *Groundwater for Sustainable Development*, 24, 101062. https://doi.org/10.1016/j.gsd.2023.101062

Noble, W. (2006). What is a support vector machine? *Nat Biotechnol*, 24, 1565–1567. https://doi.org/10.1038/nbt1206-1565

Panahi, J., Mastouri, R., & Shabanlou, S. (2022). Insights into enhanced machine learning techniques for surface water quantity and quality prediction based on data pre-processing algorithms. *Journal of Hydroinformatics*, 24(4), 875–897. https://doi.org/10.2166/hydro.2022.022

Prieto, A., Prieto, B., Ortigosa, E.M., Ros, E., Pelayo, F., Ortega, J., & Rojas, I. (2016). Neural networks: An overview of early research, current frameworks and new challenges. *Neurocomputing*, 214, 242-268. https://doi.org/10.1016/j.neucom.2016.06.014

Rizeei, H. M., Pradhan, B., Saharkhiz, A., & Lee, S. (2019). Groundwater aquifer potential modeling using an ensemble multi-adoptive boosting logistic regression technique. *Journal of Hydrology*, 579, 124172. https://dx.doi.org/10.1016/j.jhydrol.2019.124172

Shabanlou, S. (2018). Improvement of extreme learning machine using self-adaptive evolutionary algorithm for estimating discharge capacity of sharp-crested weirs located on the end of circular channels. *Flow Measurement and Instrumentation*, 59, 63-71. https://doi.org/10.1016/j.flowmeasinst.2017.11.003

Vapnik, V., & Chervonenkis, A. (1971). On the uniform convergence of relative frequencies of events to their probabilities. *Theory of Probability & Its Applications*, 16(2), 264-280. https://doi.org/10.1137/1116025

World Health Organization. (2022). Guidelines for drinking-water quality: incorporating the first and second addenda, *WHO publications*, Geneva, Switzerland. https://www.who.int/publications/i/item/9789240045064

Wu, Q., Zhang, T., Sun, H., & Kannan, K. (2010). Perchlorate in tap water, groundwater, surface waters, and bottled water from China and its association with other inorganic anions and with disinfection byproducts. *Archives of environmental contamination and toxicology* , 58(3), 543-550. http://dx.doi.org/10.1007/s00244-010-9485-6

Zhang, Q., Qian, H., Xu, P., & Li, W. (2021). Effect of hydrogeological conditions on groundwater nitrate pollution and human health risk assessment of nitrate in Jiaokou Irrigation District. *Journal of Cleaner Production*, 298, 126783. http://dx.doi.org/10.1016/j.jclepro.2021.126783