



Evaluation of GPR-PSO and KNN-PSO data-mining models for prediction of suspended sediment concentration distribution

Mohsen Nasrabadi^{1✉} | Yaser Mehri² | Ali Abdalrazaq Sabbar³ | Mohammad Javad Nahvinia⁴

1. Corresponding Author, Department of Water Science and Engineering, Faculty of Agriculture and Environment, Arak University, Arak, Iran. E-mail: m-nasrabadi@araku.ac.ir
2. Department of Irrigation and Reclamation Engineering, College of Agriculture and Natural Resources, University of Tehran, Karaj, Iran. E-mail: yaser.mehri@ut.ac.ir
3. Department of Water Science and Engineering, Faculty of Agriculture and Environment, Arak University, Arak, Iran. E-mail: arazaag94@gmail.com
4. Department of Water Science and Engineering, Faculty of Agriculture and Environment, Arak University, Arak, Iran. E-mail: m-nahvinia@araku.ac.ir

Article Info

Article type: Research Article

Article history:

Received: Dec. 3, 2024

Revised: Jan. 10, 2024

Accepted: Jan. 26, 2024

Published online: May. 2025

Keywords:

Concentration Distribution,
Suspended Sediments,
Data Mining Models,
Gaussian Process Regression.

ABSTRACT

The vertical distribution of suspended sediment concentration (SSC) is one of the most important parameters in the hydraulics of sediment transport in rivers. This parameter plays an important role in calculating the total sediment discharge in channels and rivers. For this reason, accurate measurement of this parameter has always been one of the goals of researchers. One way to accurately predict this parameter is to use intelligent models. For this purpose, in this study, four data mining models, KNN, KNN-PSO, GPR, and GPR-PSO, have been used to predict the distribution of sediment concentration (C/Ca). All models were coded in the MATLAB software environment. According to the results, it was found that the optimization performed on the KNN and GPR models was effective and increased the performance of these models. By comparing the models, it was shown that the GPR-PSO model has more accuracy than other models. The accuracy of this model in the training phase is equal to $RMSE = 0.0297$, $R^2 = 0.9878$, and $KGE = 0.9776$, and in the testing phase equal to $RMSE = 0.0226$, $R^2 = 0.9907$, and $KGE = 0.9715$. After GPR-PSO, the KNN-PSO model was ranked with $RMSE = 0.0295$, $R^2 = 0.9870$, and $KGE = 0.9864$ in the training phase and $RMSE = 0.0374$, $R^2 = 0.9808$, and $KGE = 0.9569$ in the testing phase. After the aforementioned models, GPR and KNN were respectively ranked. Also, by analyzing the results, it was determined that the two parameters y/D and y/a are the most important parameters in determining the most accurate results.

Cite this article: Nasrabadi, M., Mehri, Y., Abdalrazaq Sabbar, A., & Nahvinia, M. J. (2025) Evaluation of GPR-PSO and KNN-PSO data-mining models for prediction of suspended sediment concentration distribution, *Iranian Journal of Soil and Water Research*, 56 (3), 701-714. <https://doi.org/10.22059/ijswr.2025.377078.669846>

© The Author(s).

Publisher: The University of Tehran Press.

DOI: <https://doi.org/10.22059/ijswr.2025.377078.669846>



EXTENDED ABSTRACT

Introduction

The vertical distribution of suspended sediment concentration (SSC) is one of the most fundamental parameters in the hydraulics of sediment transport in rivers. This parameter plays an important role in calculating the total sediment discharge in channels and rivers. For this reason, accurate measurement of this parameter has always been one of the goals of researchers. One way to accurately estimate this parameter is to use intelligent models.

The main aim of the present study is to present and evaluate highly accurate models for predicting suspended sediment concentration under laboratory conditions. Calculation of suspended sediment distribution is one of the important parameters in river engineering, and its accurate calculation leads to an acceptable estimate of suspended sediment discharge. Considering the necessity of accurate measurement of suspended sediment distribution and the difficulty of laboratory works, the use of intelligent models can be useful. To solve this problem, four intelligent models of KNN, KNN-PSO, GPR, GPR-PSO have been used in the present study.

Methodology

In this study, four intelligent models KNN, KNN-PSO, GPR, GPR-PSO were used to predict the distribution of suspended sediment concentration. For this purpose, all models were coded in the MATLAB software environment. Laboratory data of Vanoni (1946) and Einstein and Chien (1955) were used for modeling (Table 1). These data include 203-point data of sediment concentration distribution. Einstein and Chien's experiments were conducted in a flume with a width of 30.7 cm and a smooth bed with different slopes of 0.00185 and 0.0025. Sediment particles with diameters of 1.3, 0.94 and 0.274 mm were used. In addition, Vanoni's experiments were conducted in a flume with a rough bed with a width of 84.5 cm and a constant slope of 0.0025. The diameters of sediment particles used in this study were 0.1, 0.13, and 0.16 mm.

Results and Discussion

The data were first normalized and adjusted between 0 and 1. Then, 80% of the data were used for training and 20% of them were used for testing. The results showed that the optimization by PSO method has increased the accuracy of GPR and KNN models. It was found that the superior model is the GPR-PSO model, which has an accuracy of $RMSE = 0.0297$, $R^2 = 0.9878$ and $KGE = 0.9776$ in the training phase and $RMSE = 0.0226$, $R^2 = 0.9907$ and $KGE = 0.9715$ in the testing phase. After the aforementioned model, the KNN-PSO model was ranked. These results show that if the optimal values are selected for the GPR model, it will have more accuracy than KNN model. This is because the prediction level in KNN is discrete, which makes the search space for PSO limited, while in the GPR model the prediction level is continuous, which makes the PSO optimization search a larger space of parameters, which makes it more accurate in regression and interpolation problems. Also, the modeling method in KNN is linear, which makes the optimization performed less effective; while the GPR model has a greater ability to model nonlinear relationships such as sediment problems, which makes this model more flexible than other models. In addition, the GPR model has the ability to consider uncertainty in modeling, while this is not possible with KNN. This is because the GPR model is a probabilistic model that models the distribution of the output variable according to the input variable, while KNN only considers the K-nearest neighbors for the prediction.

Conclusions

According to the results obtained, it was found that PSO optimization had an important effect on the performance of the GPR and KNN models and increased the accuracy of the models. Among the studied models, the GPR-PSO model was recognized as the best model. The accuracy of this model in the test phase was equal to $RMSE = 0.0226$, $R^2 = 0.9907$ and $KGE = 0.9715$. Also, by analyzing the results of the studied models, it was determined that the parameters y/D and y/a in most models were recognized as important parameters in determining the highest accuracy.

The results indicate that PSO optimization impacted the performance of the GPR and KNN models, resulting in increased accuracy. The GPR-PSO model was identified as the most effective model, with test phase accuracy metrics of $RMSE = 0.0226$, $R^2 = 0.9907$, and $KGE = 0.9715$. Analysis of the results also revealed that the parameters y/D and y/a were consistently identified as crucial factors in achieving optimal accuracy across most models.

Author Contributions

"Conceptualization, Y.M. and M.N.; methodology, A.S.; software, Y.M.; validation, Y.M., M.N. and M.J.N.; formal analysis, Y.M.; investigation, M.N.; resources, A.S.; data curation, M.N.; writing—original draft preparation, Y.M.; writing—review and editing, M.N.; visualization, Y.M.; supervision, M.N.;

All authors have read and agreed to the published version of the manuscript."

Data Availability Statement

Not applicable

Acknowledgements

The authors would like to thank all participants of the present study.

Ethical considerations

The study was approved by the Ethics Committee of the Arak University. The authors avoided data fabrication, falsification, plagiarism, and misconduct.

Conflict of interest

The author declares no conflict of interest.

ارزیابی مدل‌های هوشمند GPR-PSO و KNN-PSO در برآورد توزیع غلظت رسوبات معلق

محسن نصرآبادی^۱ | یاسر مهری^۲ | علی عبدالرزاق صبار^۳ | محمدجواد محوی نیا^۴

۱. نویسنده مسئول، گروه علوم و مهندسی آب، دانشکده کشاورزی و محیط زیست، دانشگاه اراک، اراک، ایران. رایانامه: m-nasrabadi@araku.ac.ir

۲. گروه مهندسی آبیاری و آبادانی، دانشکده کشاورزی و منابع طبیعی، دانشگاه تهران، کرج، ایران. رایانامه: yaser.mehri@ut.ac.ir

۳. گروه علوم و مهندسی آب، دانشکده کشاورزی و محیط زیست، دانشگاه اراک، اراک، ایران. رایانامه: arazaag94@gmail.com

۴. گروه علوم و مهندسی آب، دانشکده کشاورزی و محیط زیست، دانشگاه اراک، اراک، ایران. رایانامه: m-nahvinia@araku.ac.ir

اطلاعات مقاله	چکیده
نوع مقاله: مقاله پژوهشی	توزیع عمودی غلظت رسوبات معلق یکی از اساسی‌ترین پارامترها در هیدرولیک انتقال رسوبات در رودخانه‌ها محسوب می‌شود. این پارامتر نقش مهمی در محاسبه دبی کل رسوبات در کانال‌ها و رودخانه‌ها دارد. به همین دلیل اندازه‌گیری دقیق این پارامتر همواره یکی از اهداف پژوهشگران بوده است. یکی از راه‌های برآورد دقیق این پارامتر، استفاده از مدل‌های هوشمند است. برای این منظور، در این تحقیق برای پیش‌بینی توزیع غلظت رسوبات (C/Ca)، چهار مدل داده‌کاوی KNN، KNN-PSO، GPR، GPR-PSO استفاده شده است. تمامی مدل‌ها در محیط نرم‌افزار MATLAB کدنویسی شدند. با توجه به نتایج مشخص شد که بهینه‌سازی انجام شده بر روی مدل KNN و GPR تاثیرگذار بوده و سبب افزایش عملکرد (دقت) این مدل‌ها شده است. با مقایسه بین مدل‌ها، نشان داده شد که مدل GPR-PSO دقت بیشتری نسبت به سایر مدل‌ها دارد. دقت این مدل در مرحله آموزش برابر با $RMSE = 0.0297$ ، $R^2 = 0.9915$ و $KGE = 0.9976$ و در مرحله آزمون برابر با $RMSE = 0.0226$ ، $R^2 = 0.9907$ و $KGE = 0.9878$ است. از لحاظ دقت، بعد از GPR-PSO، مدل KNN-PSO با $RMSE = 0.0295$ ، $R^2 = 0.9870$ و $KGE = 0.9864$ در مرحله آموزش و $RMSE = 0.0374$ ، $R^2 = 0.9808$ و $KGE = 0.9569$ در مرحله آزمون قرار گرفت. پس از مدل‌های یادشده، GPR و KNN قرار گرفتند. همچنین با تحلیل نتایج مشخص شد که دو پارامتر y/a و y/D مهم‌ترین پارامترها در تعیین نتایج دقیق‌تر هستند.
واژه‌های کلیدی:	توزیع غلظت، رسوبات معلق، مدل‌های داده‌کاوی، رگرسیون فرآیند گاوسی.

استناد: نصرآبادی؛ محسن، مهری؛ یاسر، عبدالرزاق صبار؛ علی، محوی نیا؛ محمدجواد، (۱۴۰۴). ارزیابی مدل‌های هوشمند GPR-PSO و KNN-PSO در برآورد توزیع

غلظت رسوبات معلق، مجله تحقیقات آب و خاک ایران، ۵۶ (۲)، ۷۱۴-۷۰۱. <https://doi.org/10.22059/ijswr.2025.377078.669846>



© نویسندگان.

ناشر: مؤسسه انتشارات دانشگاه تهران.

DOI: <https://doi.org/10.22059/ijswr.2025.377078.669846>

مقدمه

رسوبات رودخانه‌ای به‌طور معمول به دو بخش بار بستر و بار معلق تقسیم می‌شوند که در این تقسیم‌بندی حدود ۸۰ درصد از بار رسوبی در رودخانه‌ها را رسوبات معلق تشکیل می‌دهند (شفاعی بجستان، ۱۳۸۷؛ امید و نصرآبادی، ۱۳۹۱). بنابراین اندازه‌گیری و برآورد دبی بار رسوبی معلق اهمیت زیادی برای پژوهشگران و تصمیم‌گیران مسائل مرتبط با مهندسی و حفاظت رودخانه دارد. برای اندازه‌گیری دبی بار رسوبی معلق از رابطه کلی زیر استفاده می‌شود:

$$q_s = \int_a^D UC dy \quad \text{رابطه (۱)}$$

در این رابطه، q_s دبی رسوبات معلق، a ارتفاعی از بستر کانال که از آن ارتفاع به بعد، رسوبات به صورت معلق حرکت می‌کنند (شکل ۱)، D عمق جریان، U سرعت جریان در فاصله y در بالای بستر و C غلظت رسوبات در این نقطه است. با توجه به این رابطه، آگاهی از توزیع عمقی سرعت و غلظت رسوبات معلق در محاسبه دبی بار معلق جریان اهمیت زیادی دارد. راوس (۱۹۳۷) اولین معادله توزیع غلظت رسوبات معلق را با استفاده از معادله پخش ویلیام اشمیت (۱۹۲۵) برای جریان‌های دوبعدی یکنواخت و ماندگار در کانال‌های باز توسعه داد (Vanoni, 1946). این معادله به‌صورت زیر است:

$$\frac{C}{C_a} = \left(\frac{D-y}{y} \frac{a}{D-a} \right)^z \quad \text{رابطه (۲)}$$

در این معادله، D عمق جریان، C_a غلظت جریان در ارتفاع اختیاری $y=a$ (ارتفاع مرجع) و z عدد تعلیق یا عدد راوس است و به صورت زیر تعریف می‌شود:

$$z = \frac{w_s}{\beta k U_*} \quad \text{رابطه (۳)}$$

در این معادله، w_s سرعت ته‌نشینی رسوبات، U_* سرعت برشی، k ثابت ون کارمن (برابر با ۰/۴۱) $\beta = \varepsilon_s / \varepsilon_m$ ضریبی است که بر اساس نظر بیشتر پژوهشگران برابر با یک در نظر گرفته می‌شود که نشان‌دهنده برابری ضریب پخش رسوبات (ε_s) و ضریب پخش جریان (ε_m) خواهد بود (شفاعی بجستان، ۱۳۸۷؛ نصرآبادی و همکاران، ۱۳۹۲). ون راین (۱۹۹۳) محدوده این ضریب را بین ۱ تا ۳ بدست آورد.

تاکنون پژوهش‌های زیادی در ارتباط با اندازه‌گیری یا محاسبه توزیع عمودی غلظت رسوبات انجام شده است. بیشتر مطالعات انجام شده به‌صورت آزمایشگاهی بوده و روابطی در این زمینه ارائه شده است که این روابط جامع نبوده و هر یک دارای معایبی است. Prandtl (1932) و Rouse (1937) اولین تحقیقات مرتبط با توزیع بار معلق را انجام دادند. آن‌ها روابطی برای محاسبه توزیع بار معلق ارائه کردند که برای محدوده خاصی از جریان آشفته کاربرد داشت (Pal and Ghoshal., 2016). افزون بر این، یکی دیگر از ایرادهای این روابط محاسبه غلظت صفر در سطح جریان و مقدار زیاد غلظت در نزدیکی بستر است. از دیگر معایب روش‌های آزمایشگاهی به‌خصوص در حیطه مهندسی هیدرولیک و رسوب، هزینه و زمان زیاد مورد نیاز است (Nasrabadi et al., 2021; Mehri et al., 2021). بنابراین استفاده از روش‌های مدلسازی، به‌خصوص روش‌های هوشمند، می‌تواند راهکار مناسبی برای جایگزینی این موارد باشد. از مزایای این روش‌ها می‌توان به سرعت پردازش زیاد، هزینه کم و انعطاف‌پذیری زیاد اشاره کرد. انواع مختلفی از مدل‌های هوشمند وجود دارد که تاکنون بررسی کاملی در مورد آن‌ها برای پیش‌بینی پدیده‌های هیدرولیکی انجام نشده است. به برخی از این مطالعات در ادامه اشاره می‌شود. Zhu et al., (2007) دبی رسوبات رودخانه یانگ تسه در چین را با استفاده از مدل ANN پیش‌بینی کردند. آن‌ها نتیجه گرفتند که این مدل توانایی پیش‌بینی دبی رسوبات را دارد. Uike et al., (2009) برای پیش‌بینی بار رسوبی معلق از چهار مدل ANN، ANFIS، رگرسیون خطی و غیرخطی استفاده کردند. این پژوهشگران نشان دادند که ANN دقت بیشتری نسبت به سه روش دیگر دارد. Melesse et al., (2011) برای پیش‌بینی دبی بار رسوبی معلق از چهار مدل MLP، MLR، MNLR و ARIMA استفاده کردند. برای مدلسازی از داده‌های میدانی سه رودخانه در کشور ایالات متحده استفاده شد. آن‌ها نشان دادند که در بین ۴ مدل بررسی شده، مدل MLP دقت بیشتری نسبت به مدل‌های دیگر دارد. Hassanpour et al., (2019) برای پیش‌بینی بار رسوبی معلق از ۵ مدل FCM-SVR، SRC، ANN، ANFIS و SVR استفاده کردند. آن‌ها نشان دادند که مدل FCM-SVR نسبت به ۴ مدل دیگر دقت بیشتری دارد. Adnan et al., (2019) از سه مدل ANFIS-DE، ANFIS-FCM و MARS برای پیش‌بینی دبی بار رسوبی معلق استفاده کردند و به این نتیجه رسیدند که بهینه سازی DE بر روی ANFIS تاثیرگذار بوده و موجب افزایش دقت آن شده است. همچنین آن‌ها نتیجه گرفتند که ANFIS-DE برترین

مدل بین مدل‌های مورد استفاده است. (Khozani et al., (2020 در تحقیقی با استفاده از مدل‌های هوشمند GP، GEP و EGP حرکت رسوب را پیش‌بینی کردند. آن‌ها نتیجه گرفتند که EGP دقت بیشتری نسبت به سایر مدل‌ها دارد. (Kaveh et al., (2021 از سه مدل LSTM، FNN و ANFIS برای پیش‌بینی غلظت رسوبات معلق استفاده کردند. آن‌ها از داده‌های میدانی واقع در رودخانه Schuykill ایالات متحده استفاده کردند. آن‌ها نتیجه گرفتند که مدل LSTM دقت بیشتری نسبت به سایر مدل‌ها دارد. (Mehri et al., (2021 از ۴ مدل هوشمند ANFIS، GMDH، ANFIS-GA و ANFIS-PSO برای پیش‌بینی توزیع غلظت رسوبات معلق استفاده کردند. آن‌ها از داده‌های آزمایشگاهی برای مدلسازی استفاده کردند. این پژوهشگران نشان دادند که هر ۴ مدل مورد استفاده توانایی مدلسازی این پدیده را دارد. آن‌ها همچنین نتیجه گرفتند که بهینه‌سازی بر دقت روش‌های GA و PSO موثر بوده و سبب افزایش دقت مدل ANFIS شده است. آن‌ها همچنین نشان دادند که در بین مدل‌های مورد استفاده ANFIS-PSO دقت بیشتری دارد. Samantaray and Sahoo (2022) از دو روش بهینه‌سازی PSO و WOA برای بهینه کردن SVM استفاده کردند تا بتوانند غلظت بار رسوبی معلق را پیش‌بینی کند. آن‌ها نشان دادند که WOA تأثیر بیشتری بر افزایش عملکرد SVM دارد. (Heddam et al., (2024 از سه مدل RF، RT و M5P برای پیش‌بینی بار رسوبی معلق استفاده کردند. آن‌ها نشان دادند که M5P دقت بیشتری نسبت به دو مدل دیگر دارد.

هدف از این پژوهش، ارائه مدلی با دقت بالا برای پیش‌بینی غلظت بار رسوبی معلق در شرایط آزمایشگاهی است. محاسبه توزیع بار رسوبی معلق یکی از پارامترهای مهم در مهندسی رودخانه به حساب می‌آید که محاسبه دقیق آن به برآورد قابل قبول دبی رسوبات معلق منجر می‌شود. با توجه به لزوم اندازه‌گیری دقیق بار رسوبی معلق و دشوار بودن کارهای آزمایشگاهی، استفاده از مدل‌های هوشمند می‌تواند مفید باشد. برای حل این مشکل، در این تحقیق از ۴ مدل هوشمند KNN-PSO، KNN، GPR و GPR-PSO استفاده شده است.

مواد و روش‌ها

در این تحقیق از ۴ مدل هوشمند KNN-PSO، KNN، GPR و GPR-PSO برای پیش‌بینی توزیع غلظت رسوبات معلق استفاده شد. برای این منظور همه مدل‌ها در محیط نرم‌افزار MATLAB کدنویسی شد. برای مدلسازی از داده‌های آزمایشگاهی تحقیقات Einstein and Chein (1955) و Vanoni (1946) استفاده شد (جدول ۱). این داده‌ها شامل ۲۰۳ داده نقطه‌ای توزیع غلظت رسوبات است. آزمایش‌های Einstein and Chein در فلومی با عرض ۳۰/۷ سانتی‌متر و بستر صاف با شیب‌های مختلف ۰/۰۰۱۸۵ و ۰/۰۰۲۵ انجام شده است. از ذرات رسوبی با قطرهای ۱/۳، ۰/۹۴ و ۰/۲۷۴ میلی‌متر استفاده شده است. افزون بر این، آزمایش‌های Vanoni در فلومی با بستر زبر با عرض ۸۴/۵ سانتی‌متر و شیب ثابت ۰/۰۰۲۵ انجام شده است. قطر ذرات رسوب مورد استفاده در این مطالعه ۰/۱، ۰/۱۳ و ۰/۱۶ میلی‌متر بود.

پارامترهای ورودی و خروجی

پارامترهای ورودی و خروجی در برآورد توزیع غلظت رسوبات، را می‌توان در جدول ۲ مشاهده کرد.

جدول ۲. پارامترهای ورودی و خروجی	
Output	Input
	U^*/W_s
	y/a
C/C_a	d_{50}/D
	k_s/d_{50}
	y/D

که در این جدول، C غلظت رسوب، C_a غلظت رسوبات معلق در ارتفاع مرجع، $y = a$ ، U^* سرعت برشی، W_s سرعت ته‌نشینی ذرات رسوب، k_s ارتفاع زبری، d_{50} قطر معادل ذرات رسوب، a ارتفاع مرجع، D عمق جریان، و y فاصله عمودی از بستر کانال است.

مدل GPR

در مدل GPR تابع به صورت گوسی مدلسازی می‌شود که دارای مجموعه‌ای از متغیرهای تصادفی است که تعدادی از آن‌ها دارای توزیع گوسی مشترک است که به صورت زیر نشان داده می‌شود.

$$X = [x_1, x_2, \dots, x_n]^T \quad (\text{رابطه ۴})$$

که خصوصیات آن با استفاده از میانگین $\mu(x)$ و تابع کواریانس $k(x, x')$ نشان داده می شود. به طور کلی، مدل GPR را می توان به شکل زیر نشان داد:

$$f(x) \sim GPR(\mu(x), k(x, x')) \quad \text{رابطه (۵)}$$

که در آن، x و x' متغیرهای تصادفی هستند. مدل رگرسیون خطی در تابع گوسی به صورت زیر نشان داده می شود:

$$y = f(x) + \varepsilon, \quad f(x) = x^T \cdot w \quad \text{رابطه (۶)}$$

که در آن x بردار ورودی، y بردار خروجی، $f(x)$ نشان دهنده مقدار تابع هدف، w وزن بردار و ε نویز موجود است. ضرایب به صورت زیر تعریف می شوند:

$$w \sim N(0, \sum_P) \quad \text{رابطه (۷)}$$

$$\varepsilon \sim N(0, \sigma_n^2)$$

که در آن، $\sum_P$ کواریانس و σ_n^2 واریانس است. در مدل GPR تابع کواریانس شکل های مختلفی دارد که یکی از این حالت ها عبارت است از تابع کرنل نمایی مربعی که به صورت زیر نشان داده می شود:

$$k(x, x') = \theta \exp(-\lambda \|x - x'\|^2) \quad \text{رابطه (۸)}$$

که در آن، θ و λ پارامترهای تابع کرنل است. گفتنی است برای تشخیص همگرایی، از قیدهایی مانند تغییرات کوچک در تابع هزینه، تغییرات کوچک در پارامترها، حداکثر تعداد تکرار یا ترکیبی از اینها استفاده شده است. در GPR، پارامترهای کرنل مانند طول مقیاس، واریانس سیگنال و واریانس نویز، باید مثبت باشند. این محدودیت از ماهیت فیزیکی و ریاضی این پارامترها ناشی می شود.

جدول ۱. خلاصه داده های آزمایشگاهی Einstein and Chein (1955) و Vanoni (1946)

Run No.	C_a (g/lit)	d_{50} (mm)	U_* (cm/s)	D (mm)	نوع بستر
۱	۴۴/۴۴	۱/۳	۱۱/۴۸	۱۳/۷۸	صاف
۲	۱۶۴/۷۴	۱/۳	۱۲/۸۵	۱۱/۹۵	صاف
۳	۲۱۸/۵۲	۱/۳	۱۳/۲۷	۱۱/۶۴	صاف
۴	۳۷۳/۰۳	۱/۳	۱۴/۲۸	۱۱/۵۲	صاف
۵	۵۸۵/۸۴	۱/۳	۱۴/۵	۱۰/۹۷	صاف
۶	۲۱/۵۳	۰/۹۴	۱۱/۸۲	۱۴/۳	صاف
۷	۴۸/۴۲	۰/۹۴	۱۱/۷۳	۱۴/۲۳	صاف
۸	۵۶/۱۹	۰/۹۴	۱۱/۵۴	۱۳/۸۷	صاف
۹	۱۰۹/۴۹	۰/۹۴	۱۱/۸۶	۱۳/۵۳	صاف
۱۰	۱۸۳/۳۷	۰/۹۴	۱۲/۶	۱۳/۰۱	صاف
۱۱	۲۱	۰/۲۷۴	۱۰/۵۸	۱۳/۲۶	صاف
۱۲	۱۴۰	۰/۲۷۴	۱۰/۰۵	۱۳/۲	صاف
۱۳	۲۲۵	۰/۲۷۴	۱۰/۴۳	۱۳/۴۱	صاف
۱۴	۲۹۵	۰/۲۷۴	۱۲/۱۷	۱۲/۳۱	صاف
۱۵	۵۱۰	۰/۲۷۴	۱۲	۱۲/۴۴	صاف
۱۶	۵۲۰	۰/۲۷۴	۱۲/۴۵	۱۱/۸۹	صاف
۱۷	۴/۷۸	۰/۱	۴/۱۵	۱۴	Rough
۱۸	۱/۲۹	۰/۱	۲/۹۷	۷/۲	Rough
۱۹	۴/۲۷	۰/۱	۵/۸۸	۲۰	Rough
۲۰	۳/۴۲	۰/۱	۴/۱۵	۷/۲	Rough
۲۱	۶/۷۵	۰/۱	۴/۶۹	۹	Rough

Einstein and Chein (1955)

Vanoni (1946)

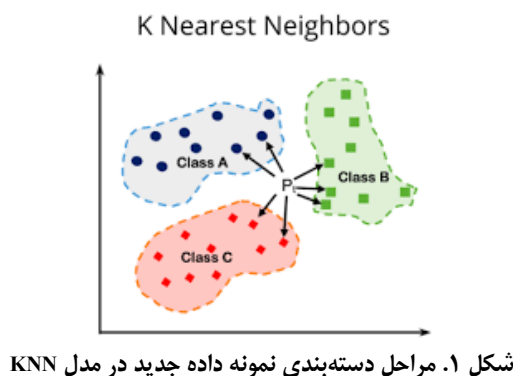
مدل KNN

KNN یک نوع الگوریتم یادگیری نظارت شده است که برای مسائل طبقه بندی و رگرسیون استفاده می شود. این مدل با یافتن k -نزدیک

ترین داده در مجموعه آموزشی به یک داده معین و پیش‌بینی خروجی براساس کلاس اکثریت k -نزدیک‌ترین همسایه‌ها کار می‌کند. مقدار k یک فرآپارامتر است که می‌تواند برای بهینه‌سازی عملکرد مدل تنظیم شود. در مدل KNN، فاصله بین داده‌های آزمون و داده‌های آموزشی با استفاده از معیارهای فاصله مختلف اندازه‌گیری می‌شود (Hassanzadeh and Abbaszadeh, 2023). اولین مرحله در الگوریتم KNN محاسبه فاصله بین نقطه داده جدید و تمام نقاط داده موجود در مجموعه آموزشی است. متداول‌ترین روش برای این کار استفاده از فاصله اقلیدسی است:

$$d(x, y) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \dots + (x_n - y_n)^2} \quad \text{رابطه ۹}$$

در این رابطه، $d(x, y)$ فاصله بین نقطه x و نقطه y ، n تعداد ابعاد (یا ویژگی‌ها) موجود در داده‌ها، x_i مقدار i -امین ویژگی برای نقطه x و y_i مقدار i -امین ویژگی برای نقطه y است. پس از محاسبه فاصله بین نقطه داده جدید و تمام نقاط داده آموزشی، الگوریتم k نزدیک‌ترین همسایه را بر اساس فاصله انتخاب می‌کند. این k نزدیک‌ترین همسایه نقاط داده آموزشی هستند که کمترین فاصله را با نقطه داده جدید دارند. با مشخص شدن k نزدیک‌ترین همسایه، پیش‌بینی برای نقطه داده جدید براساس برچسب‌های این همسایه‌ها انجام می‌شود. در شکل ۱ مراحل دسته‌بندی نمونه داده جدید در مدل KNN نشان داده شده است.



مدل بهینه‌سازی PSO

این روش با استفاده از مجموعه‌ای از ذرات تصادفی کار می‌کند. سپس وضعیت بهینه را با استفاده از به‌روزرسانی مکرر این ذرات جستجو می‌کند (Roushangar and Shahnazi, 2019). موقعیت هر ذره با استفاده از بردار آن تعیین می‌شود که به صورت زیر نشان داده می‌شود:

$$\vec{X}_i = [X_{i1} X_{i2} X_{i3} \dots X_{iN}] \quad \text{رابطه ۱۰}$$

که در آن شاخص یک ذره در ازدحام با i نشان داده می‌شود و N بعد فضای جستجو است. سرعت ذرات را به شکل زیر می‌توان نشان داد:

$$\vec{V}_i = [v_{i1} v_{i2} v_{i3} \dots v_{iN}] \quad \text{رابطه ۱۱}$$

هر ذره بهترین موقعیت را به حافظه می‌سپارد که به شکل زیر نشان داده می‌شود:

$$\vec{X}_{Pbest} = [X_{Pbest1} X_{Pbest2} X_{Pbest3} \dots X_{PbestN}] \quad \text{رابطه ۱۲}$$

بهترین موقعیت که توسط تمامی ذرات کشف شود به عنوان بهترین ذرات جهانی نامیده می‌شود:

$$\vec{X}_{Gbest} = [X_{Gbest1} X_{Gbest2} X_{Gbest3} \dots X_{GbestN}] \quad \text{رابطه ۱۳}$$

در هر تکرار از PSO، سرعت ذره بدست می‌آید:

$$v_i(t) = \omega v_i(t-1) + \rho_1 (X_{Pbest_i} - X_i(t)) + \rho_2 (X_{Gbest_i} - X_i(t)) \quad \text{رابطه ۱۴}$$

$$X_i(t) = X_i(t-1) + v_i(t) \quad \text{رابطه ۱۵}$$

که در آن، $\rho_1 = C_1 R_1$ و $\rho_2 = C_2 R_2$ و در آن C_1 و C_2 ضرایب ثابت هستند و R_1 و R_2 دو عدد تصادفی با توزیع یکنواخت بین ۰ و ۱ است.

نرمال سازی داده ها

برای نرمال سازی داده از رابطه ۱۵ استفاده شد. با استفاده از این رابطه داده ها بین ۰ و ۱ قرار گرفتند. سپس داده ها به دو قسمت آموزش (۸۰ درصد) و آزمون (۲۰ درصد) تقسیم بندی شدند. گفتنی است همه داده ها به صورت تصادفی انتخاب شدند:

$$X = \frac{X_0 - X_{\min}}{X_{\max} - X_{\min}} \quad \text{رابطه ۱۶}$$

که در آن، X داده نرمال شده، X_0 داده آزمایشگاهی، $X_{\max}$ حداکثر مقدار داده آزمایشگاهی و $X_{\min}$ حداقل مقدار داده آزمایشگاهی است.

شاخص های ارزیابی عملکرد

برای ارزیابی عملکرد مدل ها از جذر میانگین مربعات خط (RMSE)، جذر میانگین مربعات خطای نرمال شده (NRMSE)، ضریب همبستگی (R^2) و ضریب کلینگ گوپتا (KGE) استفاده شده است. روابط مورداستفاده برای محاسبه این شاخص ها به صورت زیر تعریف می شود:

$$R^2 = \left(\frac{\sum XY}{\sqrt{\sum X^2 \sum Y^2}} \right)^2 \quad \text{رابطه ۱۵}$$

$$RMSE = \left[\frac{\sum (X - Y)^2}{N} \right]^{0.5} \quad \text{رابطه ۱۶}$$

$$KGE = 1 - \sqrt{(R-1)^2 + (\beta-1)^2 + (\gamma-1)^2} \quad \text{رابطه ۱۷}$$

که در این روابط، X مقدار واقعی، Y مقدار پیش بینی شده، N تعداد نمونه، β نسبت میانگین مقدار پیش بینی شده به مقدار واقعی و γ انحراف معیار مقدار پیش بینی شده به مقدار واقعی است. با ارزیابی این شاخص ها برای هریک از مدل های مورد بررسی، عملکرد و کارایی آنها مورد مقایسه قرار خواهد گرفت. شاخص R^2 مقدار این شاخص بین صفر تا یک می باشد و اگر از ۶/۰ بیشتر باشد نشان می دهد متغیرهای مستقل تا حد زیادی توانسته اند تغییرات متغیر وابسته را تبیین کنند. دامنه مقادیر حاصل از RMSE از صفر تا مثبت بی نهایت متغیر است و همچنین واحد اندازه گیری یکسانی با متغیر وابسته یا هدف دارد. برای شاخص KGE از این دسته بندی استفاده شده است. بین ۰/۷ تا ۱ (عملکرد خیلی خوب)، بین ۰/۶ تا ۰/۷ (عملکرد خوب)، بین ۰/۵ تا ۰/۶ (عملکرد)، بین ۰/۴ تا ۰/۵ (عملکرد رضایت بخش)، و کوچک تر از ۰/۴ (عملکرد غیر رضایت بخش) (Mehri et al., 2019; Marashi et al, 2023; Abbaszadeh et al. 2023 2024; Hassanzadeh and Abbaszadeh., 2023; Hassanzadeh et al. 2024).

نتایج و بحث

در این تحقیق داده های مورد استفاده (۲۰۳ داده آزمایشگاهی) نرمال شده و بین ۰ و ۱ تنظیم شدند. سپس ۸۰ درصد داده ها برای آموزش و ۲۰ درصد داده ها برای آزمون استفاده شد. از ۴ مدل هوشمند KNN-PSO، GPR، GPR-PSO برای مدلسازی استفاده شد. همه مدل ها در محیط نرم افزار MATLAB کدنویسی شدند. جدول ۳ نتایج مدلسازی را نشان می دهد. همان طور که مشاهده می شود برای مدلسازی از ترکیب های مختلفی از I_1 تا I_6 استفاده شده است. در هر ترکیب از I_1 تا ترکیب I_5 یکی از پارامترهای موثر وجود نداشت. در ترکیب I_6 همه پارامترها وجود داشتند. در این جدول، مقدار RMSE بین ۰ و ۱ بوده و این شاخص هرچه قدر به صفر نزدیک باشد، دقت مدل بالاتر خواهد بود. مقادیر R^2 و KGE نیز همین گونه هستند، با این تفاوت که در صورت نزدیکی به ۱ دقت بالاتر خواهد بود. همان طور که مشاهده می شود بهینه سازی به روش PSO، دقت مدل های GPR و KNN را افزایش داده است. نتایج نشان داد، مدل برتر، مدل GPR-PSO است که دقتی برابر با ۰/۰۲۹۷، $RMSE = ۰/۰۲۹۷$ ، $R^2 = ۰/۹۸۷۸$ و $KGE = ۰/۹۷۷۶$ در مرحله آموزش و $RMSE = ۰/۰۲۲۶$ ، $R^2 = ۰/۹۹۰۷$ و $KGE = ۰/۹۷۱۵$ در مرحله آزمون دارد. پس از مدل یادشده، از لحاظ دقت مدل KNN-PSO قرار گرفت. این نتایج نشان می دهد در صورتی که مقادیر بهینه برای مدل GPR انتخاب شود دقتی بیشتر از KNN خواهد داشت. این موضوع به این دلیل است که سطح پیش بینی در KNN به صورت گسسته است که سبب می شود فضای جستجو برای PSO محدود باشد، در حالی که در مدل GPR سطح پیش بینی به صورت پیوسته است که موجب می شود بهینه سازی PSO فضای بیشتری از پارامترها را جستجو کند و این مورد سبب می شود که در مسائل رگرسیون و درونیابی دقیق تر باشد. همچنین نحوه مدلسازی در KNN به صورت خطی است که باعث می شود بهینه سازی انجام شده تاثیر کمتری داشته باشد؛ این در حالی است که مدل GPR توانایی بیشتری در مدلسازی روابط غیرخطی مانند مسائل رسوب دارد که موجب می شود این مدل انعطاف پذیری بیشتری نسبت به سایر مدل ها داشته باشد. افزون بر این موضوع، مدل GPR توانایی در نظر گرفتن عدم قطعیت در مدلسازی را دارد، در حالی که این مورد با KNN امکان پذیر نیست. این به این دلیل است که مدل GPR یک مدل احتمالی است که توزیع متغیر خروجی را با توجه به متغیر ورودی مدل می کند در حالی که KNN فقط K نزدیک ترین همسایه را

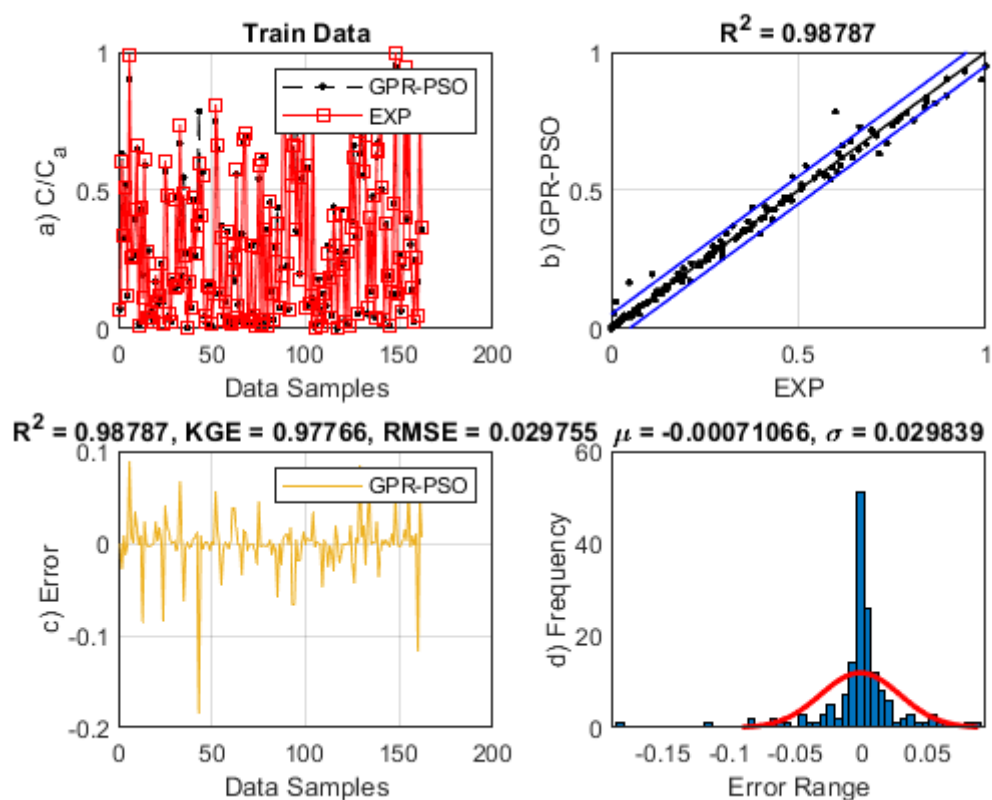


برای پیش‌بینی خود در نظر می‌گیرد.

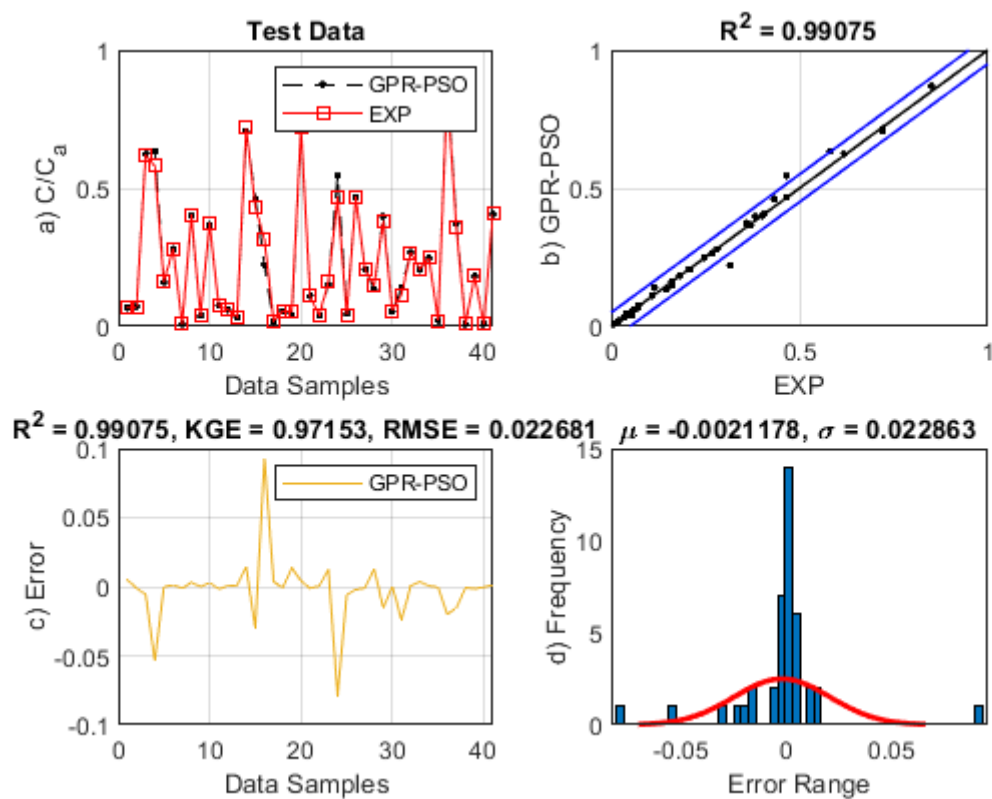
جدول ۳. نتایج مدل‌ها

مدل	نام گروه	پارامترها	آموزش RMSE	آموزش R^2	آموزش KGE	آزمون RMSE	آزمون R^2	آزمون KGE
GPR	I ₁	$y/a, d_{50}/D, k_s/d_{50}, y/D$	۰/۰۲۲۹	۰/۹۹۱۹	۰/۹۸۷۹	۰/۰۱۰۷	۰/۸۸۶۹	۰/۸۱۱۶
	I ₂	$U^*/W_s, d_{50}/D, k_s/d_{50}, y/D$	۰/۰۱۹۱	۰/۹۹۴۳	۰/۹۹۱۰	۰/۰۸۸۵	۰/۹۰۸۲	۰/۸۴۷۳
	I ₃	$U^*/W_s, y/a, k_s/d_{50}, y/D$	۰/۰۱۹۵	۰/۹۹۴۱	۰/۹۹۱۰	۰/۰۹۸۷	۰/۸۸۹۰	۰/۸۲۰۰
	I ₄	$U^*/W_s, y/a, d_{50}/D, y/D$	۰/۰۱۸۸	۰/۹۹۴۵	۰/۹۹۰۸	۰/۰۹۸۵	۰/۸۹۲۱	۰/۸۱۰۸
	I ₅	$U^*/W_s, y/a, d_{50}/D, k_s/d_{50}$	۰/۰۱۷۹	۰/۹۹۵۱	۰/۹۸۹۶	۰/۱۱۶۱	۰/۸۴۴۲	۰/۷۶۹۳
	I ₆	$U^*/W_s, y/a, d_{50}/D, k_s/d_{50}, y/D$	۰/۰۱۸۹	۰/۹۹۴۵	۰/۹۹۰۸	۰/۱۰۰۳	۰/۸۸۶۳	۰/۸۱۱۸
KNN	I ₁	$y/a, d_{50}/D, k_s/d_{50}, y/D$	.	۱	۱	۰/۰۸۹۷	۰/۸۹۵۳	۰/۹۳۸۶
	I ₂	$U^*/W_s, d_{50}/D, k_s/d_{50}, y/D$	.	۱	۱	۰/۰۷۷	۰/۸۳۹۹	۰/۸۷۴۹
	I ₃	$U^*/W_s, y/a, k_s/d_{50}, y/D$	.	۱	۱	۰/۰۹۵۳	۰/۸۶۳۷	۰/۹۰۶۹
	I ₄	$U^*/W_s, y/a, d_{50}/D, y/D$	.	۱	۱	۰/۰۹۸۵	۰/۸۷۸۵	۰/۸۹۳۵
	I ₅	$U^*/W_s, y/a, d_{50}/D, k_s/d_{50}$	.	۱	۱	۰/۱۰۹۰	۰/۸۷۰۸	۰/۸۵۷۸
	I ₆	$U^*/W_s, y/a, d_{50}/D, k_s/d_{50}, y/D$	.	۱	۱	۰/۰۸۸۵	۰/۸۷۸۸	۰/۹۱۴۷
KNN-PSO	I ₁	$y/a, d_{50}/D, k_s/d_{50}, y/D$	۰/۰۳۴۰	۰/۹۸۳۵	۰/۹۸۵۲	۰/۰۳۹۷	۰/۹۷۴۶	۰/۹۶۳۰
	I ₂	$U^*/W_s, d_{50}/D, k_s/d_{50}, y/D$	۰/۰۳۰۸	۰/۹۸۵۰	۰/۹۸۵۷	۰/۰۵۳۵	۰/۹۶۶۰	۰/۹۵۴۶
	I ₃	$U^*/W_s, y/a, k_s/d_{50}, y/D$	۰/۰۴۲۹	۰/۹۷۳۶	۰/۹۷۴۱	۰/۰۴۵۵	۰/۹۶۶۹	۰/۹۵۷۰
	I ₄	$U^*/W_s, y/a, d_{50}/D, y/D$	۰/۰۳۲۸	۰/۹۸۳۹	۰/۹۸۴۰	۰/۰۵۲۲	۰/۹۶۳۵	۰/۹۴۶۵
	I ₅	$U^*/W_s, y/a, d_{50}/D, k_s/d_{50}$	۰/۰۴۴۲	۰/۹۷۱۹	۰/۹۷۳۲	۰/۰۵۸۸	۰/۹۴۶۷	۰/۹۳۳۳
	I ₆	$U^*/W_s, y/a, d_{50}/D, k_s/d_{50}, y/D$	۰/۰۲۹۵	۰/۹۸۷۰	۰/۹۸۶۴	۰/۰۳۷۴	۰/۹۸۰۸	۰/۹۵۶۹
GPR-PSO	I ₁	$y/a, d_{50}/D, k_s/d_{50}, y/D$	۰/۰۳۴۶	۰/۹۸۳۰	۰/۹۸۱۸	۰/۰۲۸۹	۰/۹۸۷۱	۰/۹۷۶۷
	I ₂	$U^*/W_s, d_{50}/D, k_s/d_{50}, y/D$	۰/۰۳۹۳	۰/۹۷۴۸	۰/۹۸۳۶	۰/۰۴۵۵	۰/۹۷۸۱	۰/۹۸۱۸
	I ₃	$U^*/W_s, y/a, k_s/d_{50}, y/D$	۰/۰۳۲۸	۰/۹۸۳۵	۰/۹۸۰۸	۰/۰۳۲۸	۰/۹۸۶۷	۰/۹۶۶۸
	I ₄	$U^*/W_s, y/a, d_{50}/D, y/D$	۰/۰۳۰۰	۰/۹۸۵۲	۰/۹۸۵۷	۰/۰۳۸۰	۰/۹۸۴۸	۰/۹۶۴۰
	I ₅	$U^*/W_s, y/a, d_{50}/D, k_s/d_{50}$	۰/۰۳۷۴	۰/۹۷۹۹	۰/۹۶۶۳	۰/۰۵۴۷	۰/۹۵۶۵	۰/۹۴۴۹
	I ₆	$U^*/W_s, y/a, d_{50}/D, k_s/d_{50}, y/D$	۰/۰۲۹۷	۰/۹۸۷۸	۰/۹۷۷۶	۰/۰۲۲۶	۰/۹۹۰۷	۰/۹۷۱۵

با ارزیابی جدول ۳ مشاهده می‌شود در استفاده از ترکیب‌های مختلف از I₁ تا I₆ برترین دقت در مدل‌های بهینه‌شده، زمانی نشان داده شد که همه پارامترها (ترکیب I₆) مورد استفاده قرار گرفتند. در پژوهشی دیگر، مهری و همکاران (۲۰۲۱) نشان دادند زمانی که برای مدلسازی از همه پارامترها استفاده می‌شود دقت مدل افزایش پیدا می‌کند. در ترکیب I₅ زمانی که از پارامتر y/D استفاده نشده بود، دقت مدل‌ها، پایین‌ترین مقدار را از خود نشان داد. این موضوع اهمیت پارامتر y/D را نشان می‌دهد. به همین ترتیب دومین پارامتر مهم، پارامتر y/a است. شکل‌های ۲ و ۳ عملکرد مدل GPR-PSO را در مرحله آموزش و آزمون با زیرشکل‌های a، b و c و d نشان می‌دهد. شکل a روند پیش‌بینی با مدل GPR-PSO را برای هر مورد از داده‌ها نشان می‌دهد. همان‌طور که مشاهده می‌شود مدل GPR-PSO عملکرد مناسبی در مدلسازی در دو مرحله آموزش و آزمون داشته است. در شکل b عملکرد مدل در پیش‌بینی نشان داده شده است. خط ۴۵ درجه دقت ۱۰۰ درصدی را نشان می‌دهد و دو خط آبی نشان‌دهنده دقت ۹۵ درصدی هستند. شکل c مقدار خطا را در هر مورد از داده‌ها نشان می‌دهد علت وجود نوسان، وجود مقادیر خطای متفاوت در پیش‌بینی هریک از داده‌ها است و شکل d نمودار توزیع گوسی خطا را نشان می‌دهد. همان‌طور که مشاهده می‌شود مقدار انحراف معیار و واریانس آن نزدیک به صفر است. همچنین مشاهده می‌شود داده‌های خطا از توزیع نرمال یا توزیع گوسی تبعیت می‌کنند و بیشتر داده‌ها در اطراف محدوده خطای صفر قرار دارند.



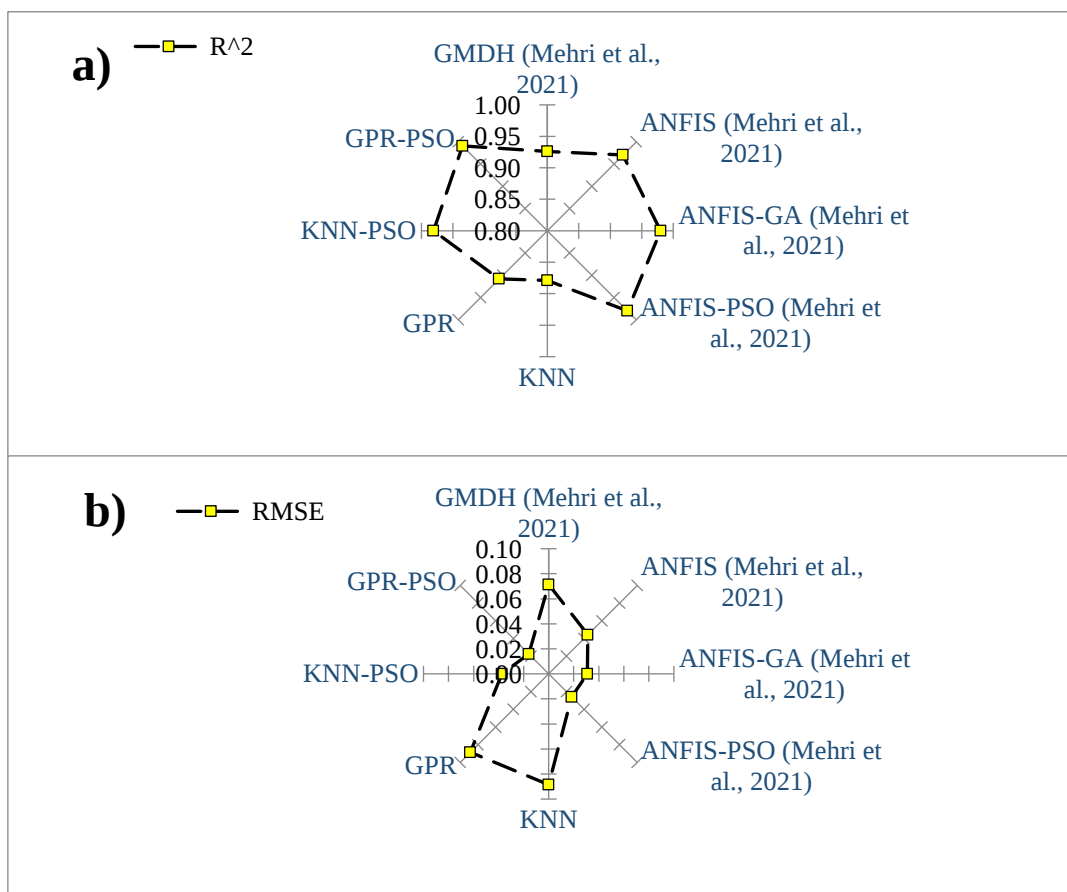
شکل ۲. عملکرد مدل GPR-PSO در مرحله آموزش



شکل ۳. عملکرد مدل GPR-PSO در مرحله آزمون

در شکل ۴ مقایسه‌ای بین مدل‌های ارائه‌شده در پژوهش حاضر با مطالعه Mehri et al., (2021) انجام شده است. همان‌طور که در

این شکل مشاهده می‌شود دقت مدل توسعه داده شده در این تحقیق یعنی مدل GPR-PSO دقیق‌تر از مدل برتر ارائه شده در پژوهش یادشده است. دقت مدل برتر ANFIS-PSO در تحقیق یادشده در مرحله آزمون برابر با $RMSE = 0.026$ و $R^2 = 0.98$ بود؛ درحالی که مدل GPR-PSO دقتی برابر با $RMSE = 0.026$ و $R^2 = 0.997$ نشان داده است. این تحقیق نشان می‌دهد در صورتی که بهینه سازی مناسبی بر روی مدل GPR انجام شود دقت این مدل می‌تواند از مدل‌های مبتنی بر سیستم‌های استنتاج فازی بیشتر شود. علت این موضوع این است که مدل GPR یک روش ناپارامتریک است به این معنا که می‌تواند روابط پیچیده بین متغیرها را بدون هیچ گونه فرضی در مورد توزیع اساسی داده‌ها مدل کند. این انعطاف‌پذیری به GPR اجازه می‌دهد تا الگوی داده‌ها را راحت‌تر مدل‌سازی کند که یافتن راه‌حل‌های بهینه را برای PSO آسان‌تر می‌کند. در مقابل ANFIS یک روش با انعطاف‌پذیری کمتر و پارامتریک‌تر است به این معنا که مفروضات خاصی در مورد توزیع داده‌ها مانند شکل توابع عضویت در سیستم منطق فازی ایجاد می‌کند. این مفروضات می‌تواند کار را برای PSO سخت‌تر کند تا پارامترهای مدل ANFIS را بهینه کند. همچنین همان‌طور که گفته شد، GPR تمایل دارد سطوح پیوسته‌تری داشته باشد که موجب می‌شود عملکرد PSO برای جستجو و یافتن راه‌حل‌های بهینه مناسب‌تر شود. این مورد به این دلیل است که PSO با به روزرسانی موقعیت هر ذره بر اساس بهترین موقعیت خود و بهترین موقعیت ازدحام کار می‌کند. اگر سطح تابع هدف صاف و پیوسته باشد، PSO می‌تواند به راحتی فضای جستجو را کاوش کرده و راه‌حل‌های بهینه را بیابد. در مقابل ANFIS تمایل به ناپیوستگی‌ها و لبه‌های تیز بیشتری دارد که می‌تواند مسیریابی و یافتن راه‌حل‌های بهینه را برای PSO دشوارتر کند.



شکل ۴. مقایسه نتایج مدل‌های تحقیق حاضر با تحقیقات پیشین

نتیجه‌گیری

توزیع غلظت رسوبات معلق یکی از پارامترهای مهم در هیدرولیک انتقال رسوب و مهندسی محیط‌زیست است. بنابراین اندازه‌گیری و پیش بینی دقیق آن بسیار لازم و ضروری است. برای این منظور، در این تحقیق از چهار مدل KNN، KNN-PSO، GPR، GPR-PSO برای پیش‌بینی مقدار غلظت رسوبات معلق استفاده شد. تمامی مدل‌ها در محیط نرم‌افزار MATLAB کدنویسی شدند. سپس از ۲۰۳ داده آزمایشگاهی نقطه‌ای برای مدل‌سازی استفاده شد. در ابتدا، داده‌ها نرمال شده و بین صفر و ۱ قرار داده شدند. سپس داده‌ها رندوم شده و ۸۰

درصد آن برای آموزش و ۲۰ درصد آن برای آزمون استفاده شد. با توجه به نتایج مشخص شد که بهینه سازی PSO بر روی نتایج مدل GPR و KNN تاثیرگذار بوده و سبب افزایش دقت مدل ها شده است. در بین مدل ها، مدل GPR-PSO به عنوان برترین مدل شناخته شد. دقت این مدل در مرحله آزمون برابر با $RMSE = 0.0226$ ، $R^2 = 0.9907$ و $KGE = 0.9715$ است. همچنین با تحلیل نتایج مدل ها مشخص شد که پارامترهای y/a و y/D در اکثر مدل ها به عنوان پارامتر مهم در تعیین بیشترین دقت شناخته شد.

"هیچ گونه تعارض منافع بین نویسندگان وجود ندارد"

فهرست منابع

- امید، محمدحسین. و نصرآبادی، محسن. (۱۳۹۱). مهندسی رسوب. انتشارات دانشگاه تهران. چاپ اول. ۷۹۰ صفحه.
- نصرآبادی، محسن، ریاحی، صالح، صمدی بروجنی، حسین. (۱۳۹۳). ارزیابی معادلات توزیع غلظت رسوبات معلق در کانال های باز، *پژوهش آب/یران*، ۱۸، ۱۸۵-۱۷۵.
- شفاعی بجستان، محمود. (۱۳۸۷). تئوری و کاربرد هیدرولیک انتقال رسوب. انتشارات دانشگاه شهید چمران اهواز. ۵۵۰ صفحه.

REFERENCES

- Abbaszadeh, H., Daneshfaraz, R., Sume, V., & Abraham, J. 2024. Experimental investigation and application of soft computing models for predicting flow energy loss in arc-shaped constrictions. *AQUA—Water Infrastructure, Ecosystems and Society*, 73(3), 637-661.
- Abbaszadeh, H., Norouzi, R., Sume, V., Kuriqi, A., Daneshfaraz, R., & Abraham, J. 2023. Sill role effect on the flow characteristics experimental and regression model analytical. *Fluids*, 8(8), p. 235.
- Adnan, R.M., Liang, Z., El-Shafie, A., Zounemat-Kermani, M. and Kisi, O., 2019. Prediction of suspended sediment load using data-driven models. *Water*, 11(10), p.2060.
- Chien, N., 1955. Effects of Heavy Sediment Concentration Near the Bed on Velocity and Sediment Distribution. Missouri River Division, Corps of Engineers, US Army.
- Cigizoglu, H.K., 2004. Estimation and forecasting of daily suspended sediment data by multi-layer perceptrons. *Advances in Water Resources*, 27(2), pp.185-195.
- Dong, X., Yu, Z., Cao, W., Shi, Y. and Ma, Q., 2020. A survey on ensemble learning. *Frontiers of Computer Science*, 14, pp.241-258.
- Gupta, L.K., Pandey, M., Raj, P.A. and Shukla, A.K., 2023. Fine sediment intrusion and its consequences for river ecosystems: a review. *Journal of Hazardous, Toxic, and Radioactive Waste*, 27(1), p.04022036.
- Hassanpour, F., Sharifazari, S., Ahmadaali, K., Mohammadi, S. and Sheikhalipour, Z., 2019. Development of the FCM-SVR hybrid model for estimating the suspended sediment load. *KSCE Journal of Civil Engineering*, 23, pp.2514-2523.
- Hassanzadeh, Y. & Abbaszadeh, H. 2023. Investigating Discharge Coefficient of Slide Gate-Sill Combination Using Expert Soft Computing Models. *Journal of Hydraulic Structures*, 9(1), pp.63-80.
- Hassanzadeh, Y., 2007. Evaluation of sediment load in a natural river. *Water International*, 32(1), pp.145-154.
- Hassanzadeh, Y., Abbaszadeh, H. 2023. Investigating Discharge Coefficient of Slide Gate-Sill Combination Using Expert Soft Computing Models', *Journal of Hydraulic Structures*, 9(1), pp. 63-80. doi: 10.22055/jhs.2023.43683.1251
- Hassanzadeh, Y., Abbaszadeh, H., Abedi, A., & Abraham, J. 2024. Numerical simulation of the effect of downstream material on scouring-sediment profile of combined spillway-gate. *AQUA—Water Infrastructure, Ecosystems and Society*, jws2024360.
- Heddarn, S., Naghibi, A., Khosravi, K. and Singh, S.K., 2024. Suspended sediment load prediction and tree-based algorithms. In *Remote Sensing of Soil and Land Surface Processes* (pp. 257-269). Elsevier.
- Kaveh, K., Kaveh, H., Bui, M.D. and Rutschmann, P., 2021. Long short-term memory for predicting daily suspended sediment concentration. *Engineering with Computers*, 37, pp.2013-2027.
- Khozani, Z.S., Safari, M.J.S., Mehr, A.D. and Mohtar, W.H.M.W., 2020. An ensemble genetic programming approach to develop incipient sediment motion models in rectangular channels. *Journal of Hydrology*, 584, p.124753.
- Marashi, A., Kouchakzadeh, S. & Yonesi, H.A. 2023. Rotary gate discharge determination for inclusive data from free to submerged flow conditions using ENN, ENN-GA, and SVM-SA. *Journal of Hydroinformatics*. 25(4), 1312-1328.



- Mehri, Y., Nasrabadi, M. and Omid, M.H., 2021. Prediction of suspended sediment distributions using data mining algorithms. *Ain Shams Engineering Journal*, 12(4), pp.3439-3450.
- Mehri, Y., Soltani, J. & Khashehchi, M. 2019. Predicting the coefficient of discharge for piano key side weirs using GMDH and DGMDH techniques. *Flow Measurement and Instrumentation*, 65, pp. 1-6.
- Melesse, A.M., Ahmad, S., McClain, M.E., Wang, X. and Lim, Y.H., 2011. Suspended sediment load prediction of river systems: An artificial neural network approach. *Agricultural Water Management*, 98(5), pp.855-866.
- Nasrabadi, M., Mehri, Y., Ghassemi, A. and Omid, M.H., 2021. Predicting submerged hydraulic jump characteristics using machine learning methods. *Water Supply*, 21(8), pp.4180-4194.
- Nasrabadi, M., Riahi, S., Samadi Borujeni, H. 2014. Evaluation of the Distribution Equations of Suspended Sediment Concentration in Open Channels, *Iranian Water Research*, 8(1), 175-185 (in Persian).
- Omid, M. H. and Nasrabadi, M. 2012. *Sedimentation Engineering*. Tehran University Press. First Edition. 790 pages (in Persian).
- Pal, D. and Ghoshal, K., 2016. Vertical distribution of fluid velocity and suspended sediment in open channel turbulent flow. *Fluid Dynamics Research*, 48(3), p.035501.
- Prandtl, L., 1932. Zur turbulenten Strömung in Rohren und längs Platten. In *Ergebnisse der aerodynamischen Versuchsanstalt zu Göttingen Lfg. 4* (pp. 18-29). De Gruyter.
- Ribeiro, M.H.D.M. and dos Santos Coelho, L., 2020. Ensemble approach based on bagging, boosting and stacking for short-term prediction in agribusiness time series. *Applied soft computing*, 86, p.105837.
- Rouse, H., 1937. Modern conceptions of the mechanics of fluid turbulence. *Transactions of the American Society of Civil Engineers*, 102(1), pp.463-505.
- Samantaray, S. and Sahoo, A., 2022. Prediction of suspended sediment concentration using hybrid SVM-WOA approaches. *Geocarto International*, 37(19), pp.5609-5635.
- Shafai-Bejestan, M. 2008. *Theory and Application of Sediment Transport Hydraulics*. Shahid Chamran University Press, Ahvaz. 550 pages (in Persian).
- Taki, M., Rohani, A., Soheili-Fard, F. and Abdeslahi, A., 2018. Assessment of energy consumption and modeling of output energy for wheat production by neural network (MLP and RBF) and Gaussian process regression (GPR) models. *Journal of cleaner production*, 172, pp.3028-3041.
- Trojovský, P. and Dehghani, M., 2022. Pelican optimization algorithm: A novel nature-inspired algorithm for engineering applications. *Sensors*, 22(3), p.855.
- Ulke, A., Tayfur, G. and Ozkul, S., 2009. Predicting suspended sediment loads and missing data for Gediz River, Turkey. *Journal of Hydrologic Engineering*, 14(9), pp.954-965.
- Vanoni, V.A., 1946. Transportation of suspended sediment by water. *Transactions of the American Society of Civil Engineers*, 111(1), pp.67-102.
- Zenko, B., Todorovski, L. and Dzeroski, S., 2001, November. A comparison of stacking with meta decision trees to bagging, boosting, and stacking with other methods. In *Proceedings 2001 IEEE international conference on data mining* (pp. 669-670). IEEE.
- Zhu, Y.M., Lu, X.X. and Zhou, Y., 2007. Suspended sediment flux modeling with artificial neural network: An example of the Longchuanjiang River in the Upper Yangtze Catchment, China. *Geomorphology*, 84(1-2), pp.111-125.
- Zounemat-Kermani, M., Batelaan, O., Fadaee, M. and Hinkelmann, R., 2021. Ensemble machine learning paradigms in hydrology: A review. *Journal of Hydrology*, 598, p.126266.