

مقاله علمی-پژوهشی

تخمین بار رسوب معلق با استفاده از مدل‌های مبتنی بر یادگیری ماشین (مطالعه موردی: ایستگاه ده ملا رودخانه زهره)

امیرحسین رضائی فریز^۱، یوسف رضائی^{۲*}، محمد ناظری تهرودی^۳، دیپاک جاجاریا^۴

تاریخ دریافت: ۱۴۰۴/۰۴/۰۲ تاریخ پذیرش: ۱۴۰۴/۰۵/۱۳

چکیده

پیش‌بینی دقیق بار رسوب معلق، به‌عنوان یکی از مهم‌ترین پارامترهای مدیریتی در حوزه‌های آبخیز، نقش کلیدی در طراحی سازه‌های هیدرولیکی، کنترل سیلاب، حفاظت منابع آبی و کاهش خسارات ناشی از رسوب‌گذاری ایفا می‌کند. با توجه به پیچیدگی‌های رفتاری این پدیده و اثرگذاری عوامل متعدد بر آن، استفاده از مدل‌های دقیق و قابل اعتماد ضرورت دارد. در این پژوهش، عملکرد سه مدل یادگیری ماشین شامل جنگل تصادفی (RF)، رگرسیون فرآیند گاوسی (GPR) و مدل K-Star در پیش‌بینی بار رسوب معلق رودخانه زهره در ایستگاه هیدرومتری ده‌ملا مورد بررسی قرار گرفته است. برای ارزیابی دقت مدل‌ها از شاخص‌های آماری RMSE، R²، NSE و KGE در دو فاز آموزش و آزمایش استفاده شد. همچنین، نمودارهای پراکنندگی و دیاگرام تیلور به‌منظور تحلیل گرافیکی عملکرد مدل‌ها و سنجش میزان قطعیت پیش‌بینی‌ها به‌کار گرفته شدند. نتایج حاصل نشان داد که مدل جنگل تصادفی با ثبت مقادیر R² = ۰/۹۷، NSE = ۰/۹۶، KGE = ۰/۸۸ و RMSE = ۲۶۹۸۲ تن در روز در فاز آزمایش، بهترین عملکرد را ارائه کرده و نزدیک‌ترین موقعیت را به نقطه مرجع در دیاگرام تیلور داشت. مدل GPR نیز دقت نسبتاً بالایی به‌ویژه در پیش‌بینی‌های با نوسانات ملایم از خود نشان داد و در جایگاه دوم قرار گرفت. در مقابل، مدل K-Star به‌ویژه در تخمین مقادیر زیاد بار رسوب معلق، عملکرد ضعیف‌تری داشت و نسبت به نوسان‌های آماری حساس‌تر عمل کرد. یافته‌های این تحقیق نشان داد که مدل‌های پیشرفته یادگیری ماشین، به‌ویژه مدل‌های انعطاف‌پذیر و مقاوم مانند RF، می‌توانند به‌عنوان ابزارهایی مؤثر در پیش‌بینی رفتار پیچیده رسوبات معلق در مطالعات مهندسی رودخانه‌ها مورد استفاده قرار گیرند و در تصمیم‌گیری‌های مدیریتی نقش مهمی ایفا کنند.

واژه‌های کلیدی: جنگل تصادفی، رگرسیون فرآیند گاوسی، شبیه‌سازی، مدل‌های یادگیری ماشین، K-Star.

مقدمه

زیستگاه‌های آبی را ایجاد کنند (Khalilivavdareh et al, 2022). به ویژه در مناطقی که دبی آب نوسانات زیادی دارد، پیش‌بینی دقیق بار رسوب معلق می‌تواند به مدیریت مؤثرتر منابع آبی کمک کند و اثرات منفی آن را کاهش دهد. بار رسوب معلق در رودخانه‌ها می‌تواند تأثیرات منفی زیادی بر عملکرد سازه‌های آبی، مدیریت منابع آب و کیفیت زیست‌محیطی داشته باشد (Kisi and Shiri, 2012). به‌طور خاص، رسوبات معلق می‌توانند موجب کاهش ظرفیت ذخیره‌سازی سدها، تخریب زیستگاه‌های آبی، افت کارایی سیستم‌های آبیاری و افزایش هزینه‌های تصفیه آب شوند. در نتیجه، پیش‌بینی دقیق بار رسوب معلق برای مدیریت منابع آب و کاهش اثرات منفی این پدیده ضروری است. پیش‌بینی بار رسوب معلق در رودخانه‌ها به‌ویژه در شرایط تغییرات اقلیمی و سیلاب‌های شدید، از اهمیت ویژه‌ای برخوردار است.

علمی و همکاران (۱۳۸۸) برای برآورد بار رسوب ایستگاه آخولا

رسوبات معلق در رودخانه‌ها تأثیرات زیادی بر محیط‌های آبی و زیرساخت‌های مرتبط با منابع آب دارند. این رسوبات عمدتاً شامل ذرات ریز سیلت، رس و ماسه هستند که در جریان آب به صورت معلق باقی می‌مانند. این ذرات می‌توانند مشکلاتی همچون کاهش کیفیت آب، آسیب به سازه‌های آبی و سیستم‌های آبیاری و تهدید

۱- دانشجوی کارشناسی ارشد، گروه علوم و مهندسی آب، دانشگاه بیرجند، بیرجند،

ایران

۲- دانشیار، گروه علوم و مهندسی آب، دانشگاه بیرجند، بیرجند، ایران

۳- استادیار، گروه علوم و مهندسی آب، دانشگاه لرستان، خرم‌آباد، ایران

۴- استاد، گروه مهندسی حفاظت آب و خاک، دانشگاه مرکزی کشاورزی،

گنگتوک، هند.

*- نویسنده مسئول: (Email: y.ramezani@birjand.ac.ir)

رودخانه‌ای در کالیفرنیا (رودخانه ایل^۷) استفاده کردند. مدل‌های فازی مختلفی برای اعتبارسنجی ساخته شده و نتایج نشان‌دهنده برتری مدل فازی تکاملی در پیش‌بینی بار رسوب نسبت به سایر مدل‌ها بود، که به‌ویژه برای مهندسی رودخانه کاربرد است (Kisi and Yaseen, 2019). ایلماز و همکاران، برای برآورد بار رسوب معلق از مدل شبکه عصبی مصنوعی با الگوریتم‌های کلونی زنبور عسل (ANN-ABC^۸) و بهینه‌سازی مبتنی بر آموزش و یادگیری^۹ (TLBO) استفاده کردند. نتایج نشان داد که مدل‌های ANN-ABC و TLBO دقت بالاتری نسبت به مدل سنتی ANN بهینه شده با الگوریتم پس انتشار^{۱۰} (BP) دارند (Yilmaz et al, 2019). احترام و همکاران، از مدل‌های سیستم استنتاج عصبی-فازی تطبیقی و شبکه عصبی پیشخور چندلایه^{۱۱} به‌طور گسترده برای شبیه‌سازی بار رسوب معلق استفاده کردند و دقت این مدل‌ها را با استفاده از الگوریتم‌های خفاش^{۱۲} (BA) و علف‌هرز^{۱۳} (WA) بهبود دادند. این مدل‌ها با استفاده از داده‌های دبی رودخانه، بارش و بار رسوب معلق ماهانه برای حوضه اترک در ایران اعتبارسنجی شدند. نتایج نشان داد که مدل ANFIS-BA با دقت بالا و عملکرد قابل اعتماد، بهترین پیش‌بینی را در مقایسه با سایر مدل‌ها ارائه می‌دهد (Ehteram et al, 2019). رضایی و ودیعی، در ایستگاه اندازه‌گیری سیرا در رودخانه کرج، به مقایسه روش‌های مختلف داده‌محور شامل مدل‌های استنتاج عصبی-فازی تطبیقی، شبکه عصبی مصنوعی و ماشین بردار پشتیبان پرداخته و نشان دادند که مدل سیستم استنتاج عصبی-فازی تطبیقی در پیش‌بینی بار رسوب معلق بر سایر مدل‌ها برتری دارند (Rezaei and Vadiati, 2020). ابدا و همکاران، از رویکردهای هوش مصنوعی مانند شبکه عصبی مصنوعی با بهینه‌سازی ازدحام ذرات^{۱۴}، سیستم استنتاج عصبی-فازی تطبیقی با بهینه‌سازی ازدحام ذرات، جنگل تصادفی^{۱۵} و شبکه عصبی با حافظه طولانی کوتاه‌مدت^{۱۶} برای شبیه‌سازی بار رسوب معلق در دوره‌های سیلابی استفاده کردند. مقایسه دقت این مدل‌ها نشان داد که مدل شبکه عصبی مصنوعی با بهینه‌سازی ازدحام ذرات با کمترین خطای RMSE در دو حوضه مورد مطالعه، بهترین عملکرد را داشته است (Abda et al, 2021). محمدی و همکاران، یک رویکرد ترکیبی جدید برای برآورد بار رسوب

واقع بر رودخانه آجی‌چای در استان آذربایجان شرقی از شبکه عصبی مصنوعی استفاده کردند. برای بررسی تأثیر رسوب و دبی ایستگاه‌های بالادست، علاوه بر داده‌های ایستگاه آخولا از داده‌های ایستگاه‌های ونیار و مرکید نیز برای آموزش شبکه عصبی استفاده کردند. نتایج نشان داد تخمین بار رسوب در چند ایستگاه با استفاده از شبکه‌های عصبی، کارایی بهتری نسبت به استفاده از منحنی سنجه رسوب دارد. اولیائی و همکاران، در تحقیق خود دقت سه روش محاسبات نرم شامل شبکه عصبی مصنوعی^۱، سیستم استنتاج عصبی-فازی تطبیقی (انفیس)^۲ و شبکه عصبی تطبیقی وزن‌دار^۳ را برای پیش‌بینی بار رسوب معلق در مقایسه با روش سنتی منحنی سنجه رسوب بررسی کردند. نتایج نشان داد که مدل شبکه عصبی تطبیقی وزن‌دار بهترین دقت را در پیش‌بینی بار رسوب معلق داشته و نسبت به سایر مدل‌ها به‌ویژه در پیش‌بینی بار رسوب تجمعی و مقادیر اوج بار رسوب معلق عملکرد بهتری داشته است. مدل شبکه عصبی تطبیقی وزن‌دار توانست به‌طور قابل توجهی دقیق‌تر از سایر مدل‌ها داده‌های اندازه‌گیری شده را شبیه‌سازی کند (Olyaie et al, 2015). نورانی و عندلیب، کارایی مدل موجک-ماشین بردار پشتیبان را برای پیش‌بینی بار رسوب معلق روزانه و ماهانه رودخانه می‌سی‌سی‌پی بررسی کردند. نتایج نشان داد که مدل ماشین بردار پشتیبان با وزن‌های کاهش یافته^۴ نسبت به مدل‌های ماشین بردار پشتیبان با رگرسیون خطی^۵ و شبکه عصبی مصنوعی در پیش‌بینی روزانه بار رسوب معلق دقت بیشتری دارد و افزودن موجک به این مدل‌ها دقت پیش‌بینی‌ها را افزایش می‌دهد. همچنین، مدل ماشین بردار پشتیبان با وزن‌های کاهش یافته در پیش‌بینی‌های چند مرحله‌ای بار رسوب معلق عملکرد بهتری نسبت به مدل شبکه عصبی تطبیقی وزن‌دار داشت (Nourani and Andalib, 2015). دنیادیده و رستمی راوری (۱۳۹۵) در تخمین بار رسوب معلق رودخانه از سه روش راس، لین-کالینسی و انشتین استفاده کردند که با توجه به نتایج به‌دست آمده می‌توان گفت دقت معادلات مورد مطالعه در رودخانه دالکی قابل قبول می‌باشد و معادله راس از دقت بیشتری نسبت به سایر معادلات برخوردار می‌باشد. فضل‌الهی (۱۳۹۵) روش معمولی منحنی سنجه و روش‌های مختلف شبکه عصبی مصنوعی را جهت برآورد بار رسوب معلق مورد مقایسه قرار داد. نتایج نشان داد که شبکه عصبی مصنوعی کارایی بالاتری در برآورد بار رسوب معلق دارد. کیسی و یاسین، از یک روش هوش ترکیبی جدید مبتنی بر فازی تکاملی^۶ برای پیش‌بینی بار رسوب در

7- Eel River

8- Artificial Neural Network – Artificial Bee Colony

9- Teaching-Learning-Based Optimization

10- Back-Propagation

11- Multilayer Feedforward Neural Network

12- Bat Algorithm

13- Weed Algorithm

14- Particle Swarm Optimization

15- Random Forest

16- Long Short-Term Memory

1- Artificial Neural Network

2- Adaptive Neuro-Fuzzy Inference System

3- Weight Agnostic Neural Network

4- Weighted Least Squares Support Vector Machine

5- Support Vector Machine with Linear Regression

6- Evolutionary Fuzzy Model

مورچه‌ها^۸، مورچه گیر^۹، سنجاقک^{۱۰} و ازدحام سالپ^{۱۱} تخمین زدند. بهترین مدل، رگرسیون بردار پشتیبان بهینه‌شده با الگوریتم ازدحام سالپ با ۲۵ تکرار بود که مقادیر NSE ، R^2 و $RMSE$ به ترتیب برابر با ۱ و ۱ و ۱۰/۲ تن در روز را ارائه کرد (Alizadeh Gharaei et al, 2024). سامانتاری و همکاران، مدل ترکیبی رگرسیون بردار پشتیبان بهینه‌شده با الگوریتم بهینه‌سازی جمعی سالپ^{۱۲} را برای پیش‌بینی بار رسوب معلق در رودخانه براهمنی^{۱۳} در هند معرفی کردند. نتایج نشان داد که این مدل با ورودی‌های رسوب با تأخیر ۳ ماه و دبی در زمان‌های مختلف، دقت بالاتری نسبت به مدل‌های دیگر دارد و با $RMSE$ و MAE برابر با ۱۵/۵ و ۱۵/۴ تن در روز بهترین عملکرد را ارائه داد. این مدل علاوه بر دقت بالا، زمان محاسباتی را نیز کاهش می‌دهد و برای کاربردهای مهندسی مناسب است (Samantaray et al, 2024). بزاک و همکاران، مدل یادگیری ماشین XGBoost را برای پیش‌بینی بار رسوب معلق در حوضه رودخانه سورا^{۱۴} در اسلوونی به کار بردند. نتایج نشان داد که این مدل نسبت به روش منحنی سنجه رسوب عملکرد بهتری دارد. پیش‌بینی‌ها نشان داد که بار رسوب معلق در آینده ۲۰-۱۵ درصد افزایش خواهد داشت و تعداد روزهای با غلظت رسوب بیشتر از ۲۵ میلی‌گرم در لیتر افزایش می‌یابد که بر کیفیت آب تأثیر خواهد گذاشت (Bezak et al, 2025).

مطالعات مختلف نشان داده‌اند که مدل‌های یادگیری ماشین، مانند شبکه عصبی مصنوعی، ماشین بردار پشتیبان و رگرسیون فرآیند گاوسی، دقت بالاتری در پیش‌بینی بار رسوب معلق نسبت به روش‌های سنتی دارند. همچنین، مدل‌های ترکیبی و بهینه‌شده مانند رگرسیون بردار پشتیبان بهینه‌شده با الگوریتم‌های بهینه‌سازی نیز توانسته‌اند عملکرد بهتری در پیش‌بینی بار رسوب معلق ارائه دهند. با وجود پیشرفت‌های چشمگیر در استفاده از مدل‌های مختلف برای پیش‌بینی بار رسوب معلق، هنوز هم نیاز به مدل‌هایی که بتوانند پیچیدگی‌های بیشتر و رفتار غیرخطی این پدیده را شبیه‌سازی کنند، احساس می‌شود.

رودخانه زهره و ایستگاه ده‌ملا به دلیل تغییرات مکرر دبی و ویژگی‌های هیدرولوژیکی خاص، برای مطالعه بار رسوب معلق اهمیت ویژه‌ای دارند. داده‌های دقیق ثبت‌شده در ایستگاه ده‌ملا شامل اطلاعات دبی و بار رسوب معلق، بستری مناسب برای توسعه

معلق ارائه دادند که در آن شبکه عصبی چندلایه^۱ با بهینه‌سازی ازدحام ذرات و الگوریتم تکامل تفاضلی^۲ را تلفیق کرده و آن را با نام MLP-PSODE معرفی کردند. این مدل را برای رودخانه مهاباد در شمال غرب ایران پیاده‌سازی کردند و عملکرد آن با تکنیک‌های شبکه عصبی چندلایه، ترکیب شبکه عصبی چندلایه با الگوریتم بهینه‌سازی ازدحام ذرات^۳ (PSO)، تابع پایه شعاعی^۴ (RBF) و ماشین بردار پشتیبان مقایسه کردند که نتایج نشان داد مدل MLP-PSODE با دقت بالاتری نسبت به سایر مدل‌ها، بار رسوب را با ورودی دبی جریان (Q) برآورد می‌کند و عملکرد بهتری در برآورد مقادیر حدی دارد (Mohammadi et al, 2021). رضایی و همکاران، کارایی چهار مدل هوش مصنوعی شبکه عصبی مصنوعی، سیستم استنتاج عصبی- فازی تطبیقی، روش گروهی برای پردازش داده‌ها^۵ و ماشین بردار پشتیبان حداقل مربعات^۶ را برای پیش‌بینی بار رسوب معلق ماهانه در ایستگاه رودک رودخانه جاجرود بررسی کردند. نتایج نشان داد که استفاده از داده‌های دبی ماهانه دوره قبل دقت پیش‌بینی بار رسوب را افزایش می‌دهد. مدل ماشین‌های بردار پشتیبان حداقل مربعات بهترین عملکرد را در پیش‌بینی بار رسوب معلق نسبت به سایر مدل‌ها نشان داد (Rezaei et al, 2021). ثابقیان (۱۴۰۰) در راستای ارائه مدل بهینه برای برآورد بار رسوب معلق رودخانه، از روش‌های پیش‌پردازش سری زمانی به همراه روش‌های هوشمند مبتنی بر کرنل ماشین بردار پشتیبان و رگرسیون فرآیند گاوسی^۷ جهت گاوسی^۷ جهت کاهش مقدار خطا استفاده کرده که نتایج حاصل، دقت بالای روش‌های تلفیقی به کار رفته در تحقیق را در تخمین بار رسوب معلق به خوبی نشان داد. کرمی و همکاران، با استفاده از داده‌های اندازه‌گیری شده و الگوریتم‌های هوشمند، بار رسوب در حوضه آبریز سد ماکو را پیش‌بینی کردند. داده‌های جریان رودخانه و غلظت رسوب بالاترین همبستگی را با بار رسوب نشان دادند. نتایج نشان داد که مدل سیستم استنتاج عصبی- فازی تطبیقی ترکیب شده با الگوریتم بهینه‌سازی کلونی مورچگان بهترین دقت را داشته و عملکرد مدل‌های هوشمند دیگر نیز بهبود قابل توجهی را در پیش‌بینی بار رسوب نشان داد (Karami et al, 2022). عزیزاده قرائی و همکاران، بار رسوب معلق در رودخانه زهره را با استفاده از رگرسیون بردار پشتیبان و چهار الگوریتم بهینه‌سازی هوش ازدحامی شامل کلونی

- 8- Ant Colony Optimization
- 9- Ant Lion Optimizer
- 10- Dragonfly Algorithm
- 11- Salp Swarm Algorithm
- 12- Support Vector Machine Optimized by Salp Swarm Algorithm
- 13- Brahmani River
- 14- Sora River

- 1- Multi-Layer Perceptron
- 2- Differential Evolution
- 3- Particle Swarm Optimization
- 4- Radial Basis Function
- 5- Group Method of Data Handling
- 6- Least Squares Support Vector Machine
- 7- Gaussian Process Regression

فرایند گاوسی جهت پیش‌بینی بار رسوب معلق استفاده شده است. مدل‌ها در محیط نرم‌افزار Weka اجرا شدند. از ۷۰ درصد داده‌ها برای آموزش^۲ مدل‌ها و ۳۰ درصد داده‌ها برای آزمایش^۳ مدل‌ها استفاده گردید. برای آموزش بهتر مدل‌ها نیاز است از بخش بیشتری از داده‌ها برای آن استفاده شود. همچنین، از مشکلاتی نظیر بیش‌برازش در مدل‌سازی جلوگیری می‌شود (علی‌زاده قرائی و همکاران، (۲۰۲۴)).

Weka، به‌عنوان یک ابزار جامع یادگیری ماشین برای تحلیل داده‌ها انتخاب شده است. این نرم‌افزار با ارائه کتابخانه گسترده‌ای از الگوریتم‌های یادگیری ماشین و قابلیت پردازش داده‌های حجیم، برای مطالعه حاضر مناسب است. از جمله امکاناتی که در این تحقیق استفاده شده عبارتند از:

- پیش‌پردازش داده‌ها: با استفاده از ابزارهای داخلی برای پاک‌سازی و نرمال‌سازی داده‌ها.
- پیاده‌سازی مدل‌ها: اجرای مدل‌های هوش مصنوعی با تنظیمات مختلف.
- تحلیل و ارزیابی مدل‌ها: استفاده از معیارهای ارزیابی مختلف برای انتخاب بهترین مدل.

جنگل تصادفی

جنگل تصادفی، یک الگوریتم یادگیری ماشین مبتنی بر جنگل تصمیم‌گیری است که برای مسائل طبقه‌بندی و رگرسیون استفاده می‌شود (بیات و همکاران، ۱۴۰۱). این الگوریتم مطابق شکل (۲) با ایجاد چندین درخت تصمیم‌گیری و ترکیب پیش‌بینی‌های آن‌ها از طریق روش‌هایی مانند رأی‌گیری اکثریت (برای طبقه‌بندی) یا میانگین‌گیری (برای رگرسیون) عمل می‌کند. تصادفی بودن در انتخاب نمونه‌های داده و ویژگی‌ها برای ساخت هر درخت باعث کاهش بیش‌برازش^۴ می‌شود و مدل را بسیار مقاوم و دقیق می‌سازد.

K-Star

K-Star یکی از الگوریتم‌های طبقه‌بندی مبتنی بر نمونه^۵ است که از شباهت آماری بین نمونه‌ها استفاده می‌کند. این الگوریتم به جای استفاده از فاصله‌های متریک معمول، از یک معیار پیچیدگی انتقالی^۶ بهره می‌برد. K-Star احتمال انتقال نمونه هدف به نمونه‌های نمونه‌های آموزشی را محاسبه می‌کند و نمونه‌ای که بیشترین شباهت آماری را داشته باشد، به‌عنوان طبقه‌بندی انتخاب می‌شود. مزیت-K-Star این است که از تغییرات غیرخطی و پیچیده در داده‌ها بهتر از روش‌های متریک معمول پشتیبانی می‌کند و نسبت به داده‌های

مدل‌های پیش‌بینی فراهم می‌کنند. در تحقیق حاضر، مدل‌های جنگل تصادفی، رگرسیون فرآیند گاوسی و K-Star که کمتر در شبیه‌سازی بار رسوب معلق مورد استفاده قرار گرفته‌اند، مورد بررسی و ارزیابی قرار گرفته‌اند. این مدل‌ها برخلاف روش‌های مرسوم، دارای ماهیت متفاوتی بوده و کاربرد دست‌می‌باشند، به طوری که امکان پیش‌بینی‌های دقیق را با حداقل پیچیدگی برای کاربر فراهم می‌آورند. ویژگی اصلی این مدل‌ها، توانایی آن‌ها در مدل‌سازی روابط پیچیده و غیرخطی و پردازش داده‌های نوسانی است که در مدل‌های سنتی به خوبی پوشش داده نمی‌شوند. در این تحقیق، از مدل‌های جنگل تصادفی، رگرسیون فرآیند گاوسی و K-Star برای پیش‌بینی بار رسوب معلق استفاده شده است.

مواد و روش‌ها

منطقه مورد مطالعه

رودخانه زهره یکی از طولی‌ترین رودخانه‌های ایران است که طول تقریبی آن ۴۹۰ کیلومتر می‌باشد و نقش حیاتی در تأمین منابع آبی برای مصارف شرب، کشاورزی و صنعتی در منطقه ایفا می‌کند (شکل ۱). مساحت حوضه آبریز این رودخانه ۱۶۰۳۳ کیلومتر مربع است که ۱۰۷۸۹ کیلومتر مربع آن به مناطق کوهستانی و ۵۲۴۴ کیلومتر مربع باقی‌مانده به اراضی دشتی اختصاص دارد (Pronoos Sedighi et al, 2023). همچنین با توجه به شکل (۱) ایستگاه ده‌ملا در بخش هندیجان استان خوزستان و در حوالی روستای ده‌ملا قرار دارد. این منطقه به دو بخش شمالی و جنوبی تقسیم شده و توسط رودخانه زهره تغذیه می‌شود. ایستگاه ده‌ملا یکی از نقاط کلیدی برای اندازه‌گیری داده‌های هیدرولوژیکی و رسوبی رودخانه زهره است و به‌طور ویژه برای ارزیابی بار رسوب معلق استفاده می‌شود. این ایستگاه در مختصات جغرافیایی ۳۰/۳۰ درجه شمالی و ۴۰/۴۹ درجه شرقی واقع شده است.

در این مطالعه ۵۷۸ داده از ایستگاه هیدرومتری ده‌ملا که به صورت رخداد^۱ (با توجه به دبی جریان، بار رسوب معلق متناظر آن اندازه‌گیری شده است) از سال ۱۳۶۲ تا ۱۳۹۷ ثبت شده است، استفاده گردید. داده‌های استخراج شده از ایستگاه هیدرومتری ده‌ملا شامل دبی بر حسب متر مکعب بر ثانیه و بار رسوب معلق بر حسب تن در روز است. خصوصیات آماری داده‌های مورد استفاده شده به شرح جدول ۱ و موقعیت مکانی ایستگاه مورد مطالعه نیز به صورت شکل ۱ قابل مشاهده می‌باشد.

مدل‌های مورد استفاده

در این مطالعه از سه مدل جنگل تصادفی، K-Star و رگرسیون

1- Event

2- Train

3- Test

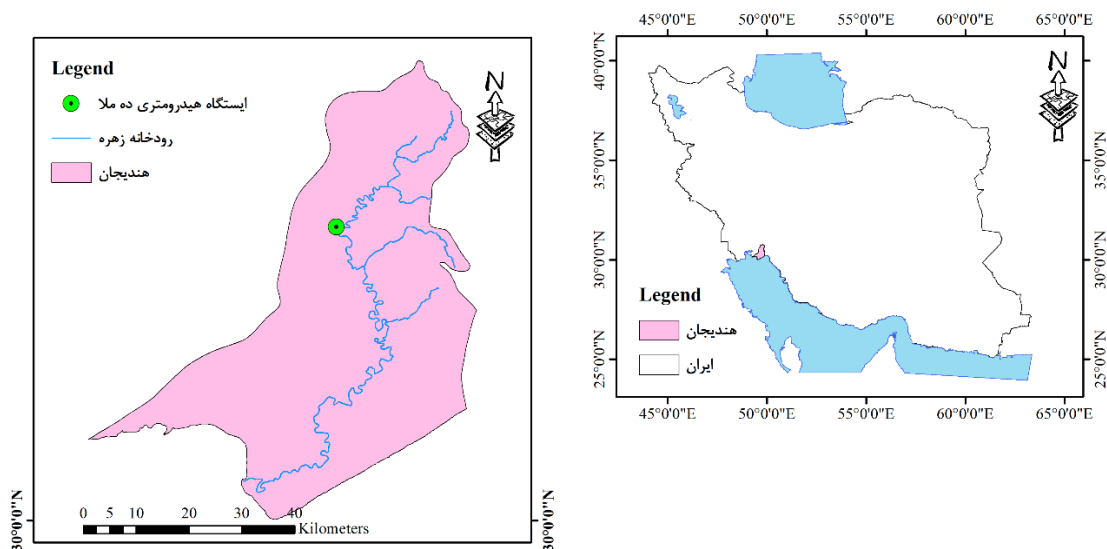
4- Overfitting

5- Instance-Based Learning

6- Transformational Complexity

رابطه p^* را می‌توان به شرح زیر بیان کرد (Ekmekecioğlu et al, 2022):

نوسانی مقاوم‌تر است (Sadeghian Agkandy et al, 2024). اگر a و b به عنوان نمونه‌های مورد بررسی در نظر گرفته شود، p^* را به عنوان احتمال هر مسیر از a و b می‌توان توصیف کرد. پس

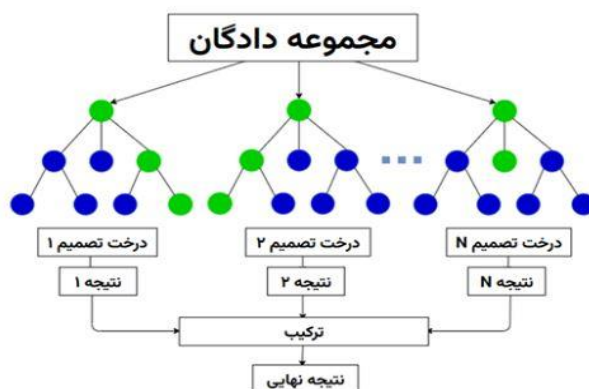


شکل ۱- نقشه موقعیت ایستگاه هیدرومتری ده‌ملا رودخانه زهره

جدول ۱- خصوصیات آماری داده‌های استفاده شده در تحقیق

متغیر	بیشینه	کمینه	میانگین	انحراف از معیار
دبی (m^3/s)	۱۱۳۹	۲/۱	۹۵/۵۶	۱۲۹/۹
بار رسوب معلق (ton/day)	۱۲۷۸۰۱۳	۸/۲	۲۶۳۸۵	۱۰۹۴۵۶

جنگل تصادفی



شکل ۲- الگوریتم جنگل تصادفی

به شکل زیر تعریف می‌شود:

$$m(x) = E(f(x)) \quad (۷)$$

$$k(x, x') = E(f(x)) - m(x)(f(x') - m(x')) \quad (۸)$$

$k(x, x')$ تابع کواریانس بوده که در نقاط x و x' محاسبه می‌شود.

فرایند گاوسی $f(x)$ می‌تواند به صورت زیر بیان گردد:

$$f(x) \sim GP(m(x), k(x, x')) \quad (۹)$$

آزمون من-کندال اصلاح شده

یکی از روش‌های پرکاربرد برای تحلیل روند در داده‌های سری زمانی، آزمون من-کندال^۱ است که برای نخستین بار توسط من و کندال معرفی شد (Man, 1945 and Kendal, 1948). این آزمون ناپارامتری بوده و برای تشخیص روندهای یکنواخت صعودی یا نزولی در داده‌ها، بدون نیاز به فرض توزیع نرمال، به کار می‌رود. به منظور حذف اثر خودهمبستگی پیاپی در آزمون و بهبود دقت آن، حامد و رائو، روشی برای اصلاح واریانس آماره آزمون من-کندال ارائه دادند (Hamed and Rao, 1998) که در آن با استفاده از ضرایب همبستگی رتبه‌ای پیاپی و با محاسبه اندازه نمونه مؤثر^۲ (ESS)، واریانس آماره آزمون به صورت تجربی اصلاح می‌شود تا اثر خودهمبستگی کاهش یابد. در این مطالعه نیز از آزمون من-کندال اصلاح شده با حذف اثرات خودهمبستگی داخلی داده‌ها جهت بررسی روند استفاده شد (Hamed, 2009 and Khalili et al, 2016).

معیارهای ارزیابی و کارایی مدل‌ها

عملکرد مدل‌های مورد بررسی با استفاده از آماره‌های ریشه میانگین مربعات خطا (RMSE)، نش-ساتکلیف (NSE)، ضریب تعیین (R^2) و کلینگ گوپتا (KGE) مورد بررسی قرار می‌گیرند. معادلات این آماره‌ها به شرح ذیل است:

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (y_i - \bar{y}_i)^2}{n}} \quad (۱۰)$$

$$NSE = 1 - \frac{\sum_{i=1}^n (y_i - \bar{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y}_i)^2} \quad (۱۱)$$

$$R^2 = \frac{\sum_{i=1}^n (y_i - \bar{y}_i)(y_i - \bar{y}_i)}{\sqrt{\sum_{i=1}^n (y_i - \bar{y}_i)^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y}_i)^2}} \quad (۱۲)$$

1 Mann-Kendall
2 Effective Sample Size

$$p^*(a, b) = \sum_{t \in pT(a)=b}^N p(t) \quad (۱)$$

t نشان دهنده مقدار آن است و p یک تابع احتمال است. با در نظر گرفتن شرایط زیر برای p^* :

$$\sum_b^N p^*(ab) = 1 \quad (۲)$$

$$0 \leq p^*(a, b) \leq 1 \quad (۳)$$

با توجه به روابط فوق، تابع K-Star به شرح زیر بیان می‌شود:

$$k^*(ab) \leq 1 - \log p^*(ab) \quad (۴)$$

ویژگی‌های رابطه فوق:

$$k^*(a, b) \geq 0 \quad (۵)$$

$$k^*(ab) + k^*(ca) \geq k^*(cb) \quad (۶)$$

مدل K-Star یک مدل غیرپارامتری است که برای پیش‌بینی داده‌های پیچیده و غیرخطی به کار می‌رود. این مدل از روش‌های مبتنی بر فاصله برای پیش‌بینی استفاده می‌کند و به جای استفاده از معادلات پارامتریک، پیش‌بینی‌ها را بر اساس شباهت داده‌ها انجام می‌دهد. به عبارت دیگر، مدل K-Star با شبیه‌سازی روابط پیچیده و غیرخطی بین داده‌ها، به پیش‌بینی دقیق‌تری دست می‌یابد. این مدل برای داده‌های پراکنده و نوسان‌دار بسیار مؤثر است و قادر به مدل‌سازی رابطه‌های پیچیده بین ورودی‌ها و خروجی‌ها می‌باشد. یکی از ویژگی‌های برجسته این مدل این است که می‌تواند بدون نیاز به فرضیات پارامتری، پیش‌بینی‌های دقیق‌تری انجام دهد (Kohavi and John, 1997).

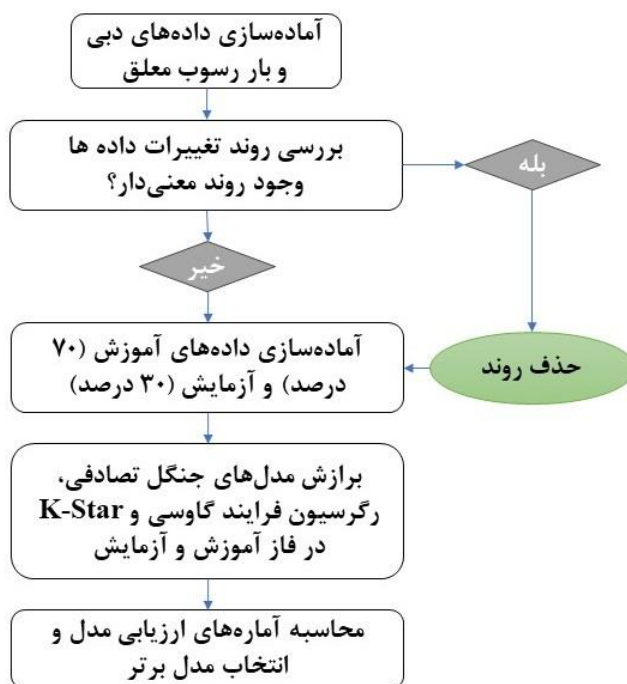
رگرسیون فرایند گاوسی

مدل رگرسیون فرایند گاوسی یک روش یادگیری ماشین است که بر پایه فرایندهای گاوسی برای مدل‌سازی توزیع احتمالی داده‌ها عمل می‌کند. این مدل به‌ویژه برای مسائل رگرسیونی به کار می‌رود و یک رویکرد غیرپارامتریک است که به جای تخمین پارامترهای خاص، توزیع احتمال توابعی را یاد می‌گیرد که می‌تواند داده‌های مشاهده‌شده را توضیح دهند. در مدل رگرسیون فرایند گاوسی، فرض بر این است که توزیع خروجی داده‌ها در هر نقطه ورودی، از یک فرایند گاوسی تبعیت می‌کند. فرایند گاوسی با یک میانگین $m(x)$ و یک تابع کواریانس $k(x, x')$ تعریف می‌شود. پیش‌بینی‌ها با استفاده از این توابع و با بهینه‌سازی تابعی به دست می‌آید که همواره عدم قطعیت پیش‌بینی را نیز نشان می‌دهد (روشنگر و اخگر، ۱۳۹۸). فرایند گاوسی $f(x)$ توسط توابع میانگین و کواریانس

انحراف معیار داده‌های شبیه‌سازی شده بر انحراف معیار داده‌های اندازه‌گیری شده و γ نسبت میانگین داده‌های شبیه‌سازی شده و اندازه‌گیری شده می‌باشد. به طور خلاصه در شکل ۳ تمام مراحل انجام این تحقیق در قالب یک فلوجارت ارائه شده است.

$$KGE = 1 - \sqrt{(R-1)^2 + (\beta-1)^2 + (\gamma-1)^2} \quad (14)$$

که در این معادلات، n تعداد داده‌ها، y_i مقدار اندازه‌گیری شده، \hat{y}_i مقدار شبیه‌سازی شده، \bar{y} میانگین مقادیر اندازه‌گیری شده، R همبستگی بین داده‌های شبیه‌سازی شده و اندازه‌گیری شده، β تقسیم



شکل ۳- فلوجارت مراحل انجام تحقیق

مدل‌ها، ارزیابی عملکرد آن‌ها با استفاده از معیارهای مختلف صورت گرفت. برای این منظور، مقادیر $RMSE$ ، NSE ، R^2 و KGE محاسبه شدند تا دقت پیش‌بینی مدل‌ها در فاز آموزش و آزمایش به صورت مجزا ارزیابی شود.

در این تحقیق، عملکرد سه مدل جنگل تصادفی، رگرسیون فرایند گاوسی و K -Star برای پیش‌بینی بار رسوب معلق با استفاده از چهار شاخص آماری مهم شامل $RMSE$ ، R^2 ، NSE و KGE در دو فاز آموزش و آزمایش ارزیابی شده است. جداول ۲ و ۳ مقادیر این شاخص‌ها را به تفکیک هر مدل و در هر دو فاز به صورت کامل نشان می‌دهند. جدول ۲ مربوط به مقادیر شاخص‌های عملکرد مدل‌ها در مرحله آموزش بوده و جدول ۳ نتایج مرحله آزمایش را نمایش می‌دهد. این جداول نشان می‌دهند که مدل‌ها در فاز آزمایش عملکرد متفاوتی داشته‌اند و مدل جنگل تصادفی با بیشترین دقت و کمترین خطا، برتری قابل توجهی نسبت به سایر مدل‌ها نشان داده است.

نتایج و بحث

در این پژوهش، به منظور بررسی روند در داده‌های بار رسوب معلق، از آزمون من-کندال اصلاح شده استفاده شد. اجرای آزمون با استفاده از کدنویسی در محیط MATLAB انجام گرفت و مقادیر آماره‌های مربوطه محاسبه شد. مقدار آماره S برابر با -۵۰۸۳ و واریانس اصلاح شده برابر با ۶۷۴۹۰۳۲ به دست آمد. بر این اساس، آماره نهایی Z معادل $-۰/۶۱۹$ محاسبه شد که نتایج، روند تغییرات کاهشی مقادیر بار رسوب معلق را نشان می‌دهد.

با توجه به سطح اطمینان ۹۵ درصد، مقدار Z در محدوده بحرانی $(\pm 1/۹۶)$ قرار ندارد. بنابراین، می‌توان نتیجه گرفت روند معناداری در داده‌های بار رسوب معلق مشاهده نمی‌شود.

همانگونه که قبلاً بیان شد، از مجموع ۵۷۸ رخدادهای ثبت شده، ۷۰ درصد داده‌ها برای آموزش و ۳۰ درصد باقی‌مانده برای آزمایش در نظر گرفته شدند. مدل‌های مورد استفاده در این تحقیق شامل جنگل تصادفی، رگرسیون فرایند گاوسی و K -Star بودند. پس از اجرای

جدول ۲ - مقایسه آماره‌های ارزیابی مدل‌ها در فاز آموزش

مدل	R ²	NSE	RMSE (ton/day)	KGE
جنگل تصادفی	۰/۹۴	۰/۹۳	۲۳۵۳۱	۰/۸۹
رگرسیون فرایند گاوسی	۰/۷۴	۰/۷۴	۴۶۲۷۱	۰/۸
K-Star	۰/۷۶	۰/۶۹	۵۱۳۰۴	۰/۴۸

جدول ۳ - مقایسه آماره‌های ارزیابی مدل‌ها در فاز آزمایش

مدل	R ²	NSE	RMSE (ton/day)	KGE
جنگل تصادفی	۰/۹۷	۰/۹۶	۲۶۹۸۲	۰/۸۸
رگرسیون فرایند گاوسی	۰/۹۳	۰/۹۳	۳۷۲۶۴	۰/۹۵
K-Star	۰/۹۲	۰/۹۲	۳۹۶۷۲	۰/۷۹

نتایج ارزیابی دقت و عملکرد مدل رگرسیون فرایند گاوسی در شبیه‌سازی بار رسوب معلق

مدل رگرسیون فرایند گاوسی (GPR) در این تحقیق عملکرد متوسطی در پیش‌بینی بار رسوب معلق رودخانه زهره از خود نشان داد. تحلیل نمودارهای پراکندگی در فاز آموزش و آزمایش (شکل‌های ۶ و ۷) بیانگر آن است که این مدل توانسته است روند کلی داده‌ها را نسبتاً خوب بازسازی کند، اما در بازه‌های دارای نوسانات شدید و مقادیر اوج، دقت پیش‌بینی کاهش یافته است. دیاگرام تیلور (شکل ۱۰) نیز نشان می‌دهد که انحراف معیار پیش‌بینی‌های این مدل کمتر از داده‌های اندازه‌گیری شده است، که این موضوع نمایانگر محافظه‌کاری مدل در پیش‌بینی مقادیر اوج می‌باشد. این ویژگی باعث شده است که مدل GPR در برخی نقاط نتواند نوسانات شدید بار رسوب را به خوبی پیش‌بینی کند.

از لحاظ کمی، در فاز آموزش، مقادیر شاخص‌های عملکرد مدل رگرسیون فرایند گاوسی شامل RMSE برابر با ۴۶۲۷۱ تن در روز، R² برابر با ۰/۷۴، NSE برابر ۰/۷۴ و KGE برابر ۰/۸ گزارش شده است. در فاز آزمایش، این مقادیر به ترتیب به RMSE معادل ۳۷۲۶۴ تن در روز، R² برابر ۰/۹۳، NSE برابر ۰/۹۳ و KGE برابر ۰/۹۵ بهبود یافته‌اند که نشان‌دهنده بهبود قابل توجه عملکرد مدل در مرحله آزمایش است. این نتایج نشان‌دهنده آن است که مدل GPR به خصوص در شرایط داده‌های نسبتاً یکنواخت، قادر به ارائه پیش‌بینی‌های دقیق‌تری است، ولی حساسیت به نوسانات شدید داده‌ها و محافظه‌کاری آن در پیش‌بینی مقادیر اوج، محدودیت‌هایی را ایجاد می‌کند. مقایسه نتایج مدل GPR با پژوهش‌های پیشین مانند مطالعه صادقیان آگکندی و همکاران (۲۰۲۴) که عملکرد رضایت‌بخشی برای این مدل تنها در شرایط داده‌های یکنواخت و کم‌نوسان گزارش شده بود، همخوانی دارد. به طور کلی، مدل رگرسیون فرایند گاوسی می‌تواند به عنوان گزینه‌ای مناسب در شرایطی که داده‌ها دارای

نتایج ارزیابی دقت و عملکرد مدل جنگل تصادفی در شبیه‌سازی بار رسوب معلق

مدل جنگل تصادفی در این پژوهش توانست بار رسوب معلق رودخانه زهره را با دقت خوبی شبیه‌سازی کند. نمودارهای پراکندگی در فاز آموزش و فاز آزمایش (شکل‌های ۴ و ۵) نشان می‌دهند که پیش‌بینی‌های مدل به طور منظم در اطراف خط ۴۵ درجه پراکنده شده‌اند که نشان‌دهنده انطباق قوی بین مقادیر پیش‌بینی شده و داده‌های اندازه‌گیری شده است. همچنین، دیاگرام تیلور (شکل ۱۰) موقعیت بهینه این مدل را نزدیک نقطه مرجع نشان می‌دهد و همبستگی بالا و انحراف معیار مشابهی با داده‌های اندازه‌گیری شده دارد. این موارد نشان می‌دهد که جنگل تصادفی نه تنها روند کلی بار رسوب معلق را به خوبی بازسازی کرده، بلکه توانایی مناسبی در پیش‌بینی مقادیر اوج و نوسانات داده‌ها دارد.

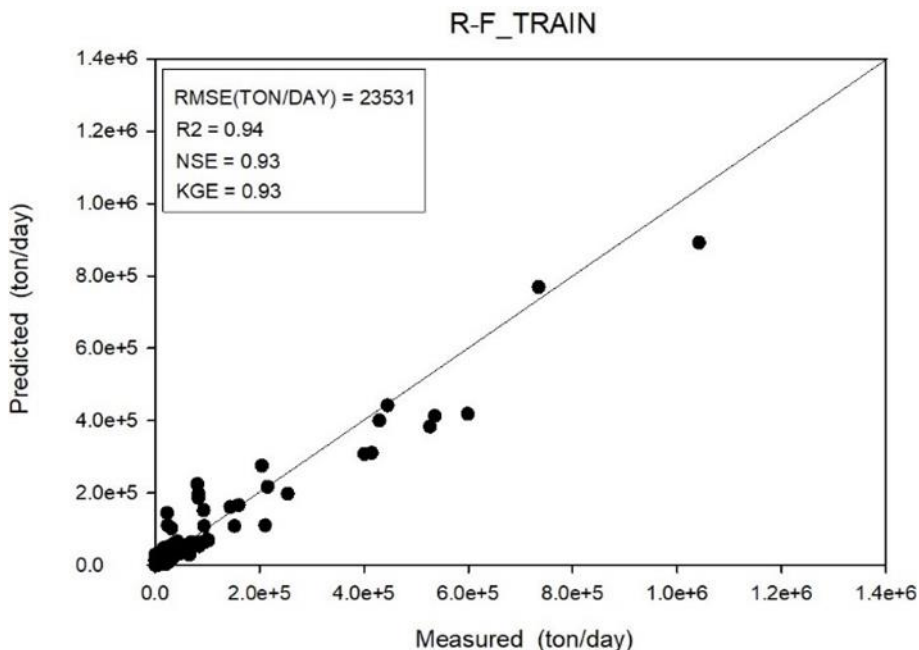
از نظر کمی، مدل جنگل تصادفی در فاز آموزش با RMSE برابر ۲۳۵۳۱ تن در روز، R² برابر با ۰/۹۴، NSE برابر ۰/۹۳ و KGE برابر ۰/۸۹ عملکرد بسیار خوبی نشان داد. این عملکرد در فاز آزمایش نیز با RMSE معادل ۲۶۹۸۲ تن در روز، R² برابر ۰/۹۷، NSE برابر ۰/۹۶ و KGE برابر ۰/۸۸ بهبود یافت. این شاخص‌ها نشان‌دهنده دقت بالای مدل در پیش‌بینی و قابلیت تعمیم مناسب آن هستند. مقایسه نتایج این تحقیق با پژوهش‌های علیزاده قرائی و همکاران (۲۰۲۴) (با داده‌های مشابه این مطالعه) که در آن مدل SVR بهینه‌شده با الگوریتم ازدحام سالپ مورد استفاده قرار گرفته بود و RMSE برابر با ۱۰/۲ تن در روز و R² برابر ۱ داشت نشان می‌دهد که مدل جنگل تصادفی توانسته دقت بسیار خوبی ارائه دهد. این نتایج نشان می‌دهد که جنگل تصادفی به دلیل ساختار منعطف خود و توانایی مقابله با نوسانات و داده‌های غیرخطی، گزینه‌ای مناسب برای پیش‌بینی بار رسوب معلق در حوضه‌های رودخانه‌ای است و می‌تواند در مطالعات مهندسی منابع آب به عنوان ابزاری کارآمد مورد استفاده قرار گیرد.

داده‌های اندازه‌گیری شده است که بیانگر هم‌راستایی کمتر مدل با داده‌های اندازه‌گیری شده است. در فاز آزمایش، عملکرد مدل بهبود یافته (شکل ۹) ولی همچنان پراکندگی قابل توجهی وجود دارد. همچنین، دیاگرام تیلور (شکل ۱۰) نشان می‌دهد که مدل K-Star نسبت به نقطه ایده‌آل فاصله بیشتری دارد و در پیش‌بینی مقادیر اوج بار رسوب معلق ضعیف‌تر عمل کرده است.

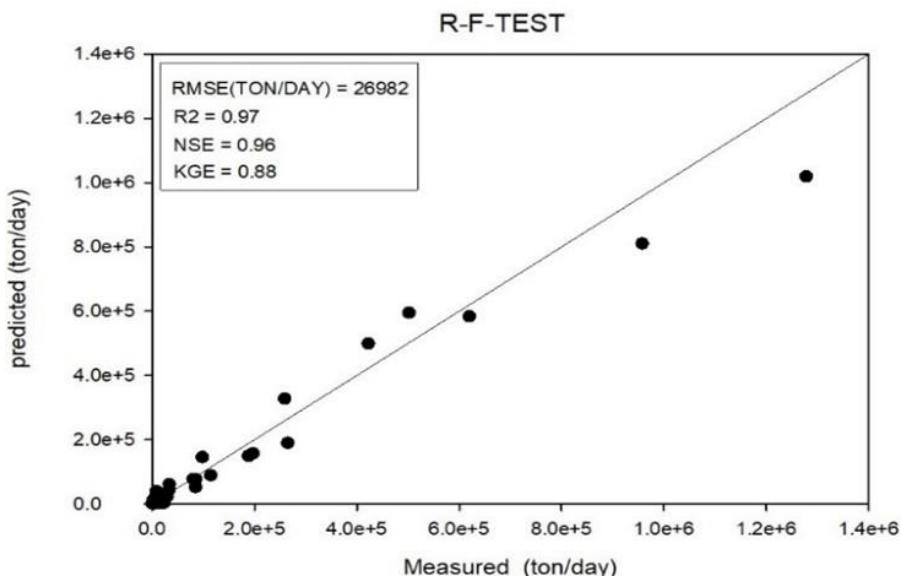
تغییرات ملایم و نسبتاً یکنواخت هستند مدنظر قرار گیرد، اما در مواجهه با داده‌های پرنوسان و مقادیر اوج بار رسوب معلق محدودیت‌هایی دارد.

نتایج ارزیابی دقت و عملکرد مدل K-Star در شبیه‌سازی بار رسوب معلق

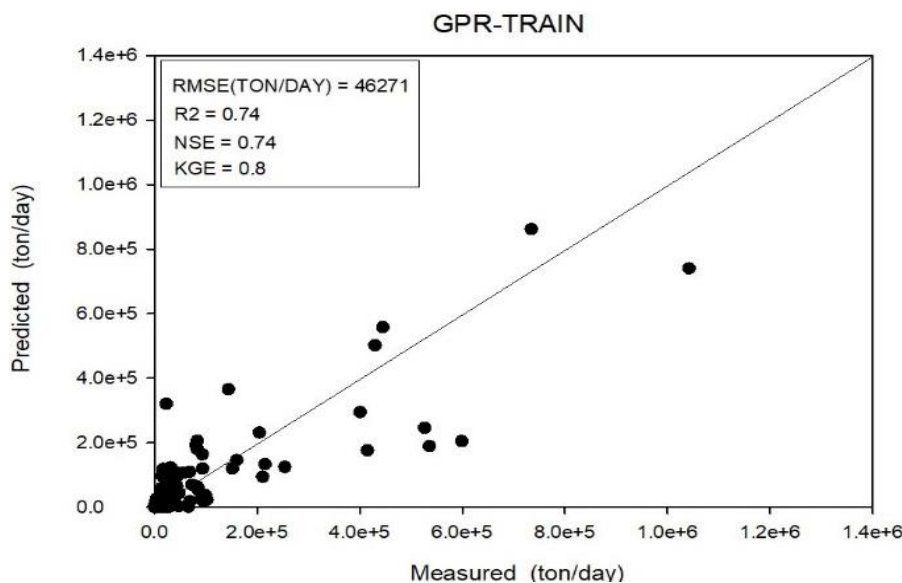
نمودارهای پراکندگی مدل K-Star در فاز آموزش (شکل ۸) نشان‌دهنده پراکندگی گسترده‌تر نقاط پیش‌بینی شده نسبت به



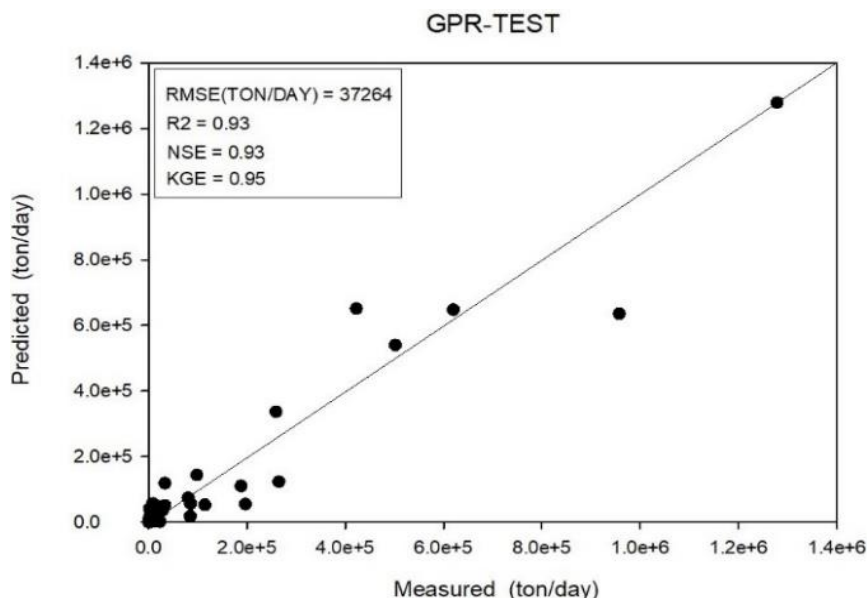
شکل ۴- پراکندگی مدل جنگل تصادفی در فاز آموزش



شکل ۵- پراکندگی مدل جنگل تصادفی در فاز آزمایش



شکل ۶- نمودار پراکندگی مدل GPR در فاز آموزش

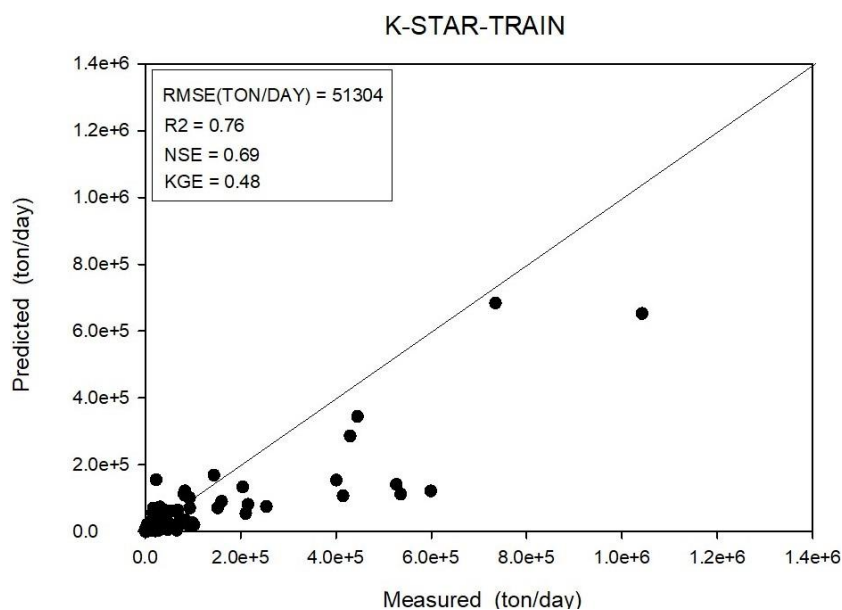


شکل ۷- نمودار پراکندگی مدل GPR در فاز آزمایش

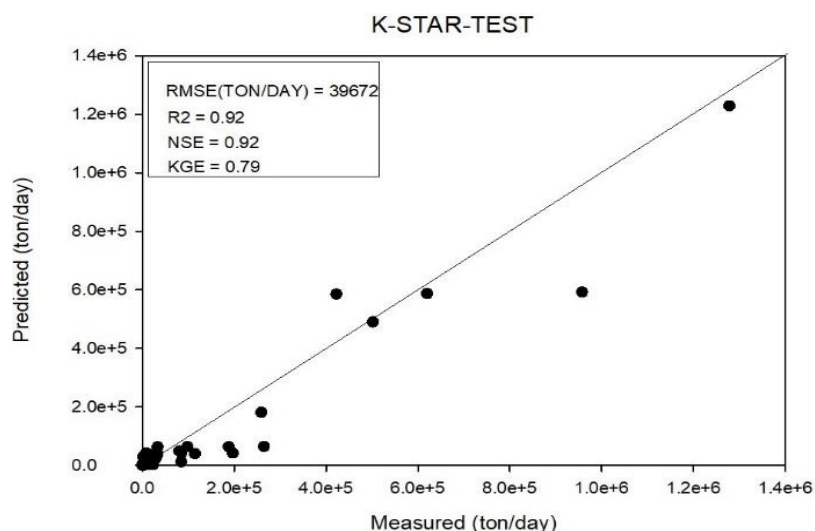
پیش‌بینی شده است. همچنین، وابستگی مدل به معیارهای انتقالی و پیچیدگی محاسباتی، احتمالاً منجر به عملکرد ضعیف‌تر آن در پیش‌بینی بار رسوب معلق شده است. نتایج مدل K-Star در این پژوهش مشابه یافته‌های صالح و همکاران است که در آن مدل K-Star ضعیف‌ترین عملکرد را در میان مدل‌های مقایسه شده داشت. همچنین، این مدل در مطالعاتی که داده‌ها دارای نوسانات شدید بودند، عملکرد قابل قبولی نداشته است (Salih et al, 2020) این موضوع نشان‌دهنده محدودیت‌های الگوریتم K-Star در مواجهه با داده‌های پیچیده و پراکنده است.

از نظر کمی، مدل K-Star در فاز آموزش با RMSE برابر ۵۱۳۰۴ تن در روز، R^2 برابر با ۰/۷۶، NSE برابر ۰/۶۹ و KGE برابر ۰/۴۸ عملکرد نسبتاً ضعیفی داشت. در فاز آزمایش، دقت مدل بهبود یافت و با RMSE معادل ۳۹۶۷۲ تن در روز، R^2 برابر ۰/۹۲، NSE برابر ۰/۹۲ و KGE برابر ۰/۷۹ همراه بود. با این حال، این شاخص‌ها هنوز از دو مدل دیگر پایین‌تر است و بیانگر حساسیت مدل K-Star به نوسانات شدید داده‌ها می‌باشد.

الگوریتم K-Star مبتنی بر شباهت آماری نمونه‌ها است که برای داده‌های همگن و کم‌نوسان مناسب‌تر است. حساسیت بالای این مدل به تغییرات ناگهانی و نوسان موجود در داده‌ها باعث کاهش دقت



شکل ۸- نمودار پراکندگی مدل K-Star در فاز آموزش

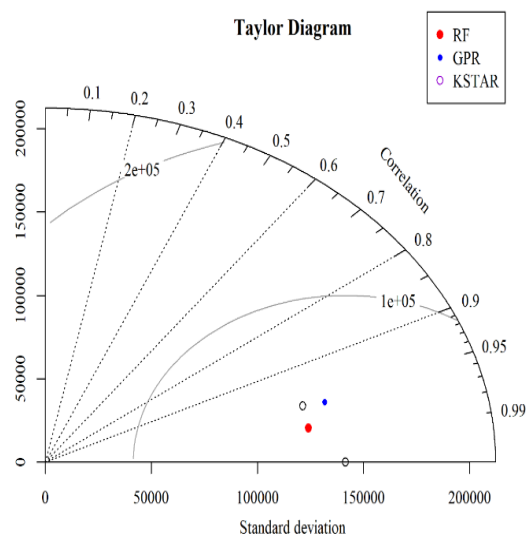


شکل ۹- نمودار پراکندگی مدل K-Star در فاز آزمایش

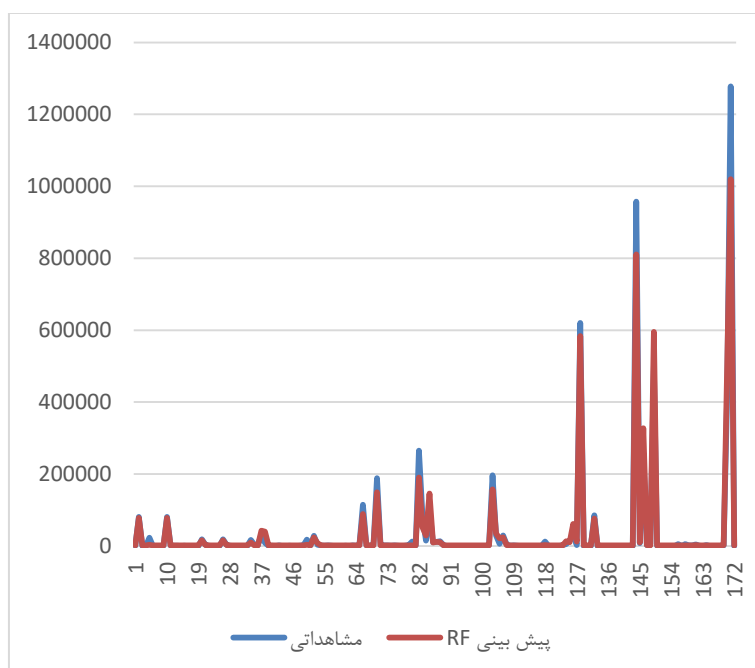
دقیق‌تر است. بنابراین، با وجود عملکرد خوب هر سه مدل در ثبت روند داده، RF و GPR از نظر دقت و پایداری پیش‌بینی نسبت به K-Star برتری دارند. این تحلیل تأکید می‌کند که انتخاب مدل باید بر اساس تعادل بین همبستگی و میزان خطا صورت گیرد و در این مطالعه RF و GPR گزینه‌های مناسب‌تری به شمار می‌روند. همچنین، در رابطه با بررسی مقادیر حدی، نتایج پیش‌بینی بار رسوب معلق هر مدل در فاز آزمایش با مقادیر مشاهداتی مورد بررسی قرار گرفت (شکل‌های ۱۱ تا ۱۳). نتایج نشان داد مدل جنگل تصادفی و رگرسیون فرایند گاوسی در پیش‌بینی مقادیر حدی بهتر از مدل K-Star عمل کرده‌اند.

در دیاگرام تیلور ارائه شده (شکل ۱۰)، سه مدل RF، GPR و K-Star همبستگی بالای ۹۰ درصد هستند که نشان‌دهنده توانایی آن‌ها در دنبال کردن دقیق روند داده‌های اصلی می‌باشد. با توجه به خطوط شعاعی که مقادیر جذر میانگین مربعات تفاضل‌ها^۱ (RMSE) را نمایش می‌دهند، هر سه مدل زیر منحنی 5×10^5 قرار دارند و این موضوع نشان‌دهنده خطای نسبتاً کم در پیش‌بینی‌های آن‌ها است. با این حال، مدل‌های RF و GPR در مقایسه با K-Star انحراف معیار کمتری دارند که نشان‌دهنده پیش‌بینی‌های پایدارتر و

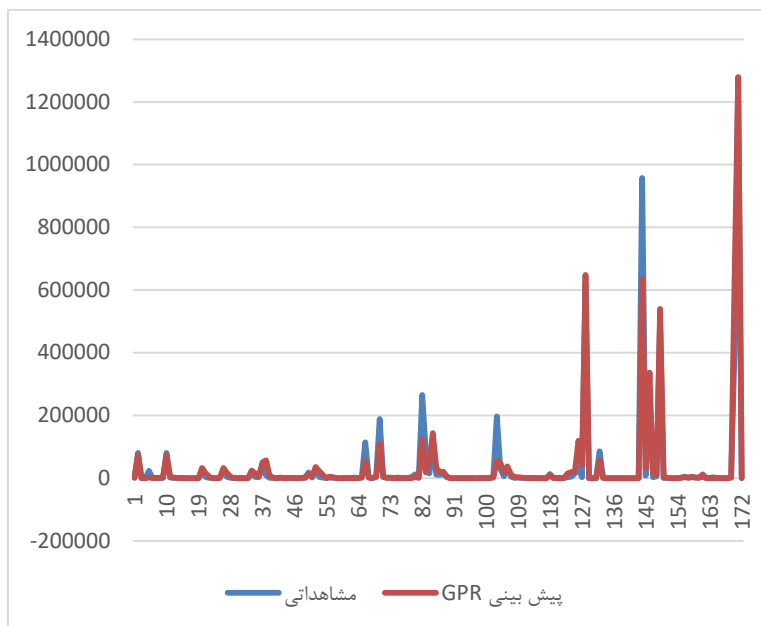
¹ Root Mean Square Difference



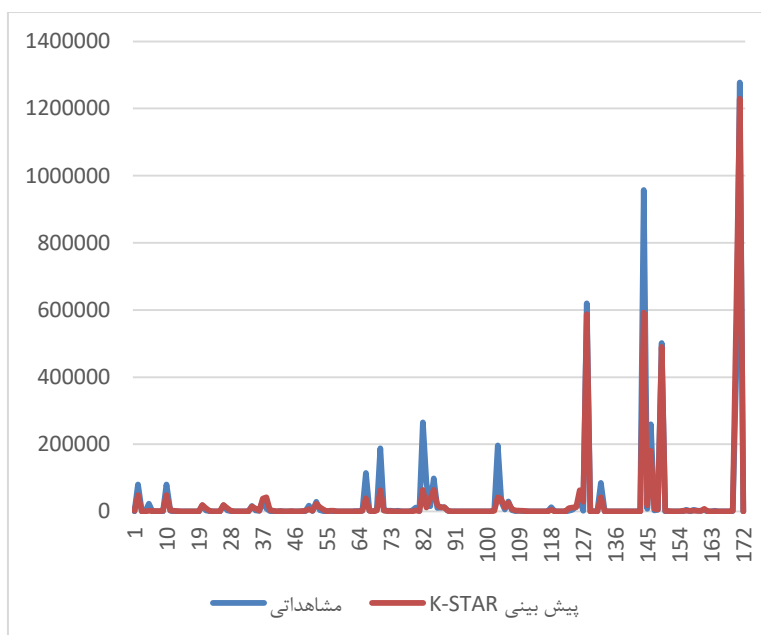
شکل ۱۰- نتایج دیاگرام تیلور در بررسی عملکرد مدل‌های مورد مطالعه در پیش‌بینی مقادیر مشاهداتی بار رسوب معلق



شکل ۱۱- مقایسه مقادیر مشاهداتی و پیش‌بینی مدل جنگل تصادفی در فاز آزمایش



شکل ۱۲- مقایسه مقادیر مشاهداتی و پیش بینی مدل رگرسیون فرآیند گاوسی در فاز آزمایش



شکل ۱۳- مقایسه مقادیر مشاهداتی و پیش بینی مدل K-Star در فاز آزمایش

نتیجه گیری

تصادفی در هر دو فاز آموزش و آزمایش عملکرد دقیقی داشت و شاخص‌هایی نظیر RMSE کم و R^2 نزدیک به یک را ثبت کرد. نمودار تیلور نیز برتری این مدل را از نظر همبستگی و تطابق آماری تأیید کرد. بر اساس مقایسه شاخص RMSE در فاز آزمایش، مدل جنگل تصادفی نسبت به مدل رگرسیون فرآیند گاوسی حدود ۳۷/۵ درصد و نسبت به مدل K-Star بیش از ۳۲ درصد بهبود در دقت پیش‌بینی داشته است. این نشان می‌دهد که جنگل تصادفی به دلیل

در این پژوهش، سه مدل یادگیری ماشین شامل جنگل تصادفی، رگرسیون فرآیند گاوسی و الگوریتم K-Star برای پیش‌بینی بار رسوب معلق در ایستگاه ده‌ملا رودخانه زهره مورد بررسی قرار گرفتند. هدف اصلی، مقایسه توانایی این مدل‌ها در بازسازی رفتار پیچیده بار رسوب معلق بر اساس دبی جریان بود. نتایج نشان داد که مدل جنگل

- Bezak, N., Lebar, K., Bai, Y. and Rusjan, S. 2025. Using Machine Learning to Predict Suspended Sediment Transport under Climate Change. *Water Resources Management*. 1-16.
- Ehteram, M., Ghotbi, S., Kisi, O., Najah Ahmed, A., Hayder, G., Ming Fai, C. ... and EL-Shafie, A. 2019. Investigation on the potential to integrate different artificial intelligence models with metaheuristic algorithms for improving river suspended sediment predictions. *Applied Sciences*. 9(19): 4149.
- Ekmekcioğlu, Ö., Başakın, E. E. and Özger, M. 2022. Tree-based nonlinear ensemble technique to predict energy dissipation in stepped spillways. *European Journal of Environmental and Civil Engineering*. 26(8): 3547-3565.
- Hamed, K. H. 2009. Enhancing the effectiveness of prewhitening in trend analysis of hydrologic data. *Journal of hydrology*. 368(1-4): 143-155.
- Hamed, K. H. and Rao, A. R. 1998. A modified Mann-Kendall trend test for autocorrelated data. *Journal of hydrology*. 204(1-4): 182-196.
- Karami, H., DadrasAjirlou, Y., Jun, C., Bateni, S. M., Band, S. S., Mosavi, A. ... and Chau, K. W. 2022. A novel approach for estimation of sediment load in Dam reservoir with hybrid intelligent algorithms. *Frontiers in Environmental Science*. 10: 821079.
- Kendall, M. G. 1948. Rank correlation methods.
- Khalili, K., Tahoudi, M. N., Mirabbasi, R. and Ahmadi, F. 2016. Investigation of spatial and temporal variability of precipitation in Iran over the last half century. *Stochastic environmental research and risk assessment*. 30: 1205-1221.
- Khalilivandareh, S., Shahnazari, A. and Sarraf, A. 2022. Spatio-temporal variations of discharge and sediment in rivers flowing into the anzali lagoon. *Sustainability*. 14(1): 507.
- Kisi, O. and Shiri, J. 2012. River suspended sediment estimation by climatic variables implication: Comparative study among soft computing techniques. *Computers & Geosciences*. 43: 73-82.
- Kisi, O. and Yaseen, Z. M. 2019. The potential of hybrid evolutionary fuzzy intelligence model for suspended sediment concentration prediction. *Catena*. 174: 11-23.
- Kohavi, R. and John, G. H. 1997. Wrappers for feature subset selection. *Artificial intelligence*. 97(1-2): 273-324.
- Mann, H. B. 1945. Nonparametric tests against trend. *Econometrica: Journal of the econometric society*. 245-259.
- Mohammadi, B., Guan, Y., Moazenzadeh, R. and Safari, M. J. S. 2021. Implementation of hybrid particle swarm optimization-differential evolution algorithms coupled with multi-layer perceptron for suspended sediment load estimation. *Catena*. 198: 105024.
- Nourani, V. and Andalib, G. 2015. Daily and monthly suspended sediment load predictions using wavelet ساختار منقطع، مقاومت در برابر نوسان و قابلیت یادگیری روابط غیرخطی، گزینه‌ای مناسب برای پیش‌بینی بار رسوب معلق در مطالعات مهندسی منابع آب است. مدل GPR با وجود آنکه در فاز آموزش دقت کمتری داشت، در فاز آزمایش عملکرد خود را بهبود داد که می‌تواند به دلیل ساختار حساس آن به توزیع داده‌ها باشد. مدل-K-Star نیز گرچه در برخی بازه‌ها رفتار قابل قبولی داشت، اما در مجموع از دقت کمتری برخوردار بود و به نوسانات داده‌ها حساس‌تر عمل کرد. در مجموع، استفاده از مدل جنگل تصادفی می‌تواند به عنوان ابزاری کارآمد در پیش‌بینی بار رسوب معلق و مدیریت بهره‌برداری از مخازن مورد استفاده قرار گیرد و منجر به تصمیم‌گیری بهتر و کاهش هزینه‌های ناشی از رسوب‌گذاری شود.
- ### منابع
- اعلمی، م.، نورانی، و. و نظم آرا، ح. ۱۳۸۸. قابلیت شبکه‌های عصبی مصنوعی جهت مدل‌سازی چند ایستگاه بار معلق در مقایسه با روش منحنی سنج رسوب. دانش آب و خاک (دانش کشاورزی). ۱/۱۹: (۲): ۴۵-۵۵.
- بیات، م.، اخوان، ر.، حیدری مستعلی، س. و حمیدی، س. ک. ۱۴۰۱. مقایسه مدل‌های جنگل تصادفی، ماشین بردار تصمیم‌گیری و رگرسیون خطی چند متغیره در ارزیابی تنوع زیستی جنگل‌های هیرکانی. محیط‌شناسی. ۴۸ (۴): ۵۳۰-۵۱۳.
- ثاقبیان، س. م. ۱۴۰۰. تخمین بار معلق رسوبی با استفاده از روش‌های هوشمند تلفیقی با در نظر گرفتن عدم قطعیت مدل. آب و خاک (علوم و صنایع کشاورزی). ۳۵ (۴): ۴۸۸-۴۷۵.
- دنیادیده، م. و رستمی راوری، ا. ۱۳۹۵. ارزیابی معادلات رسوب در تخمین بار معلق رودخانه دالکی. کنفرانس هیدرولیک ایران. روشنگر، ک. و اخگر، س. ۱۳۹۸. بررسی پارامترهای هیدرولیکی تاثیرگذار بر آبستتگی پایین دست سازه‌های کنترل با استفاده از روش رگرسیون فرآیند گاوسی. مجله آبیاری و زهکشی ایران. ۱۳ (۶): ۱۸۶۸-۱۸۵۸.
- فضل‌الهی، ع. ۱۳۹۵. برآورد رسوب بار معلق رودخانه با بهره‌گیری از شبکه عصبی مصنوعی. همایش ملی علوم و مهندسی آبخیزداری ایران (توسعه مشارکتی در مدیریت حوزه‌های آبخیز).
- Abda, Z., Zerouali, B., Alqurashi, M., Chettih, M., Santos, C. A. G. and Hussein, E. E. 2021. Suspended sediment load simulation during flood events using intelligent systems: a case study on semiarid regions of Mediterranean Basin. *Water*. 13(24): 3539.
- Alizadeh Gharaei, M. S., Ramezani, Y. and Nazeri Tahroudi, M. 2024. Toward coupling of nonlinear support vector regression and crowd intelligence optimization algorithms in estimation of suspended sediment load. *Applied Water Science*. 14(9): 192.

- Geosciences. 14(3): 215.
- Sadeghian Agkandy, M., Rezaie, H., Khalili, K. and Ahmadie, F. 2024. Investigating the Performance of Kstar and GPR Algorithms in Modeling RDI Meteorological Drought Index (Case Study: East of Urmia Lake Basin). *Journal of Civil and Environmental Engineering*. 54(114): 142-151.
- Salih, S. Q., Sharafati, A., Khosravi, K., Faris, H., Kisi, O., Tao, H. ... and Yaseen, Z. M. 2020. River suspended sediment load prediction based on river discharge information: application of newly developed data mining models. *Hydrological Sciences Journal*. 65(4): 624-637.
- Samantaray, S., Sahoo, A., Satapathy, D. P., Oudah, A. Y. and Yaseen, Z. M. 2024. Suspended sediment load prediction using sparrow search algorithm-based support vector machine model. *Scientific Reports*. 14(1): 12889.
- Yilmaz, B., Aras, E., Kankal, M. and Nacar, S. 2019. Prediction of suspended sediment loading by means of hybrid artificial intelligence approaches. *Acta Geophysica*. 67: 1693-1705.
- based artificial intelligence approaches. *Journal of Mountain Science*. 12: 85-100.
- Olyaie, E., Banejad, H., Chau, K. W. and Melesse, A. M. 2015. A comparison of various artificial intelligence approaches performance for estimating suspended sediment load of river systems: a case study in United States. *Environmental monitoring and assessment*. 187: 1-22.
- Pronoos Sedighi, M., Ramezani, Y., Nazeri Tahroudi, M. and Taghian, M. 2023. Joint frequency analysis of river flow rate and suspended sediment load using conditional density of copula functions. *Acta Geophysica*. 71(1): 489-501.
- Rezaei, K. and Vadiati, M. 2020. A comparative study of artificial intelligence models for predicting monthly river suspended sediment load. *Journal of Water and Land Development*.
- Rezaei, K., Pradhan, B., Vadiati, M. and Nadiri, A. A. 2021. Suspended sediment load prediction using artificial intelligence techniques: comparison between four state-of-the-art artificial neural network techniques. *Arabian Journal of*

Estimation of Suspended Sediment Load Using Machine Learning Models (Case Study: Deh Molla Station, Zohreh River)

A. H. Ramezani Feriz¹, Y. Ramezani^{2*}, M. Nazeri Tahroudi³, D. Jhajharia⁴

Received: Jun.23, 2025

Accepted: Agu.04, 2025

Abstract

Accurate prediction of suspended sediment load, as one of the most important management parameters in watershed areas, plays a key role in the design of hydraulic structures, flood control, water resource protection, and reduction of damage caused by sedimentation. Given the complexities of this phenomenon and the influence of multiple factors on it, the use of accurate and reliable models is essential. In this study, the performance of three machine learning models, including Random Forest (RF), Gaussian Process Regression (GPR), and K-Star model, in estimating the suspended sediment load of the Zohreh River at the Deh Molla hydrometric station was investigated. To evaluate the accuracy of the models, statistical indices such as RMSE, R^2 , NSE, and KGE were used in both train and test phases. Additionally, scatter plots and Taylor diagrams were employed for graphical analysis of model performance and assessment of the certainty of predictions. The results indicated that the Random Forest model, with values of $R^2 = 0.97$, $NSE = 0.96$, $KGE = 0.88$, and $RMSE = 26982$ tons per day in the test phase, demonstrated the best performance and was the closest to the reference point in the Taylor diagram. The GPR model also showed relatively high accuracy, especially in predictions with mild fluctuations, ranking second. In contrast, the K-Star model exhibited weaker performance, particularly in estimating high suspended sediment load values, and was more sensitive to statistical fluctuations. The findings of this research showed that advanced machine learning models, especially flexible and robust models like RF, can be effectively used as tools for predicting the complex behavior of suspended sediment load in river engineering studies and play a significant role in management decision-making.

Keywords: Gaussian Process Regression, K-Star, Machine Learning Models, Random Forest, Simulation.

1- M.Sc. Student, Department of Water Engineering, University of Birjand, Birjand, Iran

2- Associate Professor, Department of Water Engineering, University of Birjand, Birjand, Iran

3- Assistant Professor, Department of Water Engineering, Lorestan University, Khorramabad, Iran

4- Professor, Department of Soil and Water Conservation Engineering, Central Agricultural University, Gangtok, India

(* - Corresponding Author Email: y.ramezani@birjand.ac.ir)