Evaluating the Performance of Machine Learning Models for Predicting Suspended Sediment Load (case study: Taleghan watershed, Iran)

Zeinab Mohammadi-Raigani [a], Hamid Gholami [b*], Mojtaba Mohammadi [c]
[a] *Postdoctoral researcher, Department of Natural Resources Engineering, Faculty of Natural Resources and Agriculture, University of Hormozgan, Bandar-Abbas, Hormozgan, Iran.*
[b] *Professor, Department of Natural Resources Engineering, Faculty of Natural Resources and Agriculture, University of Hormozgan, Bandar-Abbas, Hormozgan, Iran.*
[c] *Assistant professor, Department of Desert Management and Control, Faculty of Environmental Sciences, Planning and Sustainable Development, University of Saravan, Saravan, Sistan va Baluchestan, Iran.*

Research Full Paper
Article History (Received: 2024/11/17 Accepted: 2024/12/28)

Extended abstract

1- Introduction

Accurate prediction of suspended sediment load (SSL) in rivers is crucial for sustainable water resources management. Sediment, as a significant factor in degrading aquatic ecosystems, reducing the lifespan of hydraulic structures, and deteriorating water quality, has been a major concern. Given the complexity of hydrological processes and the interplay of various factors affecting sediment production and transport, accurate modeling of this phenomenon has always been challenging. In recent years, advancements in machine learning techniques have enabled the development of highly accurate predictive models. This study aims to enhance the understanding of sediment dynamics in watersheds and provide more precise tools for SSL prediction by applying machine learning models to daily SSL forecasting in the Taleghan watershed. The Taleghan watershed, characterized by its mountainous topography and specific hydrological conditions, was selected for this study due to its significance in supplying water to the Tehran region. Accurate SSL prediction in this watershed is of paramount importance.

2- Material and Methods

For this study, hydrological data including discharge, precipitation, snow index, and suspended sediment load were collected from the Gelink Taleghan hydrometric station over the period 2000 to 2018. The collected data, after undergoing quality checks and outlier removal, were prepared for training and testing machine learning models. Data preprocessing involved standardization, normalization, and outlier removal to enhance model accuracy. Principal Component Analysis (PCA) was employed to reduce data dimensionality and select the most significant input variables for the model. This method aids in identifying new linear combinations of original variables that explain the maximum variance in the data. By using PCA, the number of input variables in the model can be reduced, thereby decreasing model training time and complexity. Six machine learning models, including xgbTree, Cubist, qrnn, Ctree, Cforest, and LASSO, were utilized in this study. These models were chosen due to their ability to model nonlinear and complex relationships between variables. To evaluate model performance, various statistical metrics such as Root Mean Squared Error (RMSE), Nash-Sutcliffe Efficiency (NSE), and Mean Absolute Error (MAE) were employed. Additionally, time series plots, scatter plots, and Taylor diagrams were used for visual assessment of model performance.

3- Results

Results indicated that xgbTree, Cubist, and qrnn models outperformed other models in predicting SSL. These models effectively simulated daily variations in SSL. Overall, the xgbTree model demonstrated the best

* **Corresponding Author: hadesertt64@gmail.com**

[DOI: 10.61186/jeer.15.1.83]

performance in SSL prediction. A comparison of the models' predictive accuracy revealed that ML algorithms can successfully predict daily SSL, particularly xgbTree (RMSE = 63.301, NSE = 0.97), Cubist (RMSE = 46.330, NSE = 0.96), and qrnn (RMSE = 85.349, NSE = 0.96), which exhibited the lowest prediction errors and highest efficiency metrics. Taylor diagrams further confirmed that xgbTree, Cubist, and qrnn models demonstrated a better fit to the observed data.

Sensitivity analysis revealed that discharge and precipitation had the most significant influence on SSL variations. Additionally, employing PCA to reduce data dimensionality and select key variables played a crucial role in enhancing model performance. The findings of this study demonstrate that machine learning models can serve as powerful tools for predicting SSL in watersheds. Furthermore, the results indicated that machine learning models are capable of identifying complex and nonlinear patterns in hydrological data that cannot be detected using traditional modeling techniques. This is particularly significant in scenarios where climate change and human activities impact hydrological processes.

4- Discussion & Conclusions

This study employed machine learning models to forecast daily suspended sediment load (SSL) in the Taleghan watershed. Results indicated that xgbTree, Cubist, and qrnn models exhibited superior performance in predicting SSL. These models can serve as powerful management tools for assessing the impacts of climate change and human activities on sedimentation processes, and for planning and managing water resources sustainably. The findings of this study demonstrate that machine learning models can be effective alternatives to traditional modeling methods for SSL prediction. These models can capture complex nonlinear relationships between input and output variables, leading to more accurate forecasts. However, to enhance predictive accuracy, further research is needed using higher spatial and temporal resolution data, as well as considering additional factors influencing sedimentation processes. For future research, it is recommended to explore hybrid models combining machine learning and physical models. Additionally, deep learning techniques can be employed to extract more complex features from data. Furthermore, investigating the impacts of climate change on sedimentation processes and developing predictive models under various climate scenarios is a crucial topic for future research.

Key Words: Suspended sediment load, Machine learning models, Principal Component Analysis, Taleghan watershed**.**

Cite this article: Mohammadi-Raigani, Z., Gholami, H., & Mohammadi, M. (2025). Evaluating the Performance of Machine Learning Models for Predicting Suspended Sediment Load (case study: Taleghan watershed, Iran). *Journal of* Environmental *Erosion Research. 2025; 15 (1):83-104*. http//doi.org/10.61186/jeer.15.1.83

DOI: http//doi.org/10.61186/jeer.15.1.83
Published by Hormozgan University Press.
URL: http://magazine.hormozgan.ac.ir

[DOI: 10.61186/jeer.15.1.83]

ارزیابی عملکرد مدل‌های یادگیری ماشین در پیش‌بینی بار رسوب معلق (مطالعه موردی: حوضه آبخیز طالقان)

زینب محمدی رایگانی: *پژوهشگر فوق دکتری گروه مهندسی منابع طبیعی، دانشگاه هرمزگان، بندرعباس*
حمید غلامی *: *استاد گروه مهندسی منابع طبیعی، دانشگاه هرمزگان، بندرعباس*
مجتبی محمدی: *استادیار گروه مدیریت و کنترل بیابان، دانشگاه سراوان، سراوان*

نوع مقاله: پژوهشی
تاریخچهٔ مقاله (تاریخ دریافت: ۱۴۰۳/۰۸/۲۷ تاریخ پذیرش: ۱۴۰۳/۱۰/۰۸)
DOI: http//doi.org/10.61186/jeer.15.1.83

چکیده

مدل‌سازی و پیش‌بینی بار رسوب معلق (SSL) یک موضوع مهم در مدیریت یکپارچه محیطی و منابع آب است، زیرا رسوب بر کیفیت آب و زیستگاه‌های آبی تأثیر می‌گذارد. از سوی دیگر، کمی‌سازی و درک تعاملات غیرخطی در دینامیک رسوب همواره به عنوان یک چالش اساسی مطرح بوده است. هدف از این مطالعه پیش‌بینی بار رسوب معلق روزانه در حوضه طالقان در شمال غرب تهران با بکارگیری و مقایسه عملکرد شش مدل یادگیری ماشین (ML) شامل Cforest، Ctree، Cubist، LASSO، qrnn و xgbTree بود. داده‌های ده متغیر ورودی (دبی، رسوب روزانه، بارش، بارش تجمعی، شاخص برف، دبی و رسوب با تاخیر زمانی یک روزه (t-1)) از سال ۲۰۰۰ تا ۲۰۱۸ برای آموزش و آزمایش (به ترتیب ۷۵ درصد آموزش و ۲۵ درصد آزمایش)، مدل‌های ML استفاده شد. همچنین تکنیک آنالیز مؤلفه‌های اصلی (PCA) برای کاهش تعداد متغیرهای ورودی بکار گرفته شد. عملکرد مدل‌ها برای پیش‌بینی SSL با استفاده از چندین معیار کمی و گرافیکی، از جمله ریشه میانگین مربعات خطا (RMSE)، ضریب نش-ساتکلیف (NSE) و میانگین خطای مطلق (MAE)، نمودار تیلور، نمودار تغییرات زمانی و نمودار پراکندگی ارزیابی شد. مقایسه دقت پیش‌بینی مدل‌ها نشان داد که الگوریتم‌های ML می‌توانند SSL روزانه را به‌طور رضایت‌بخش پیش‌بینی کنند، به‌ویژه مدل‌های xgbTree (RMSE= ۳۰۱/۶۳؛ NSE= ۰/۹۷)، Cubist (RMSE= ۳۳۰/۴۶؛ NSE= ۰/۹۶) و qrnn (RMSE= ۳۴۹/۸۵؛ NSE= ۰/۹۶)، که کمترین خطای پیش‌بینی و بالاترین معیارهای کارایی را نشان دادند. علاوه بر این، نمودار تیلور تأیید کرد که مدل‌های xgbTree، Cubist و qrnn بهترین تطابق بین مقادیر مشاهده شده و پیش‌بینی‌شده را برای پارامترهای هیدرولیکی مختلف به دست آوردند. این نتایج حاکی از آن است که عملکرد مدل‌های یادگیری ماشین به عنوان یک روش مناسب برای پیش‌بینی و تحلیل دینامیک رسوب در حوضه‌های آبخیز است.

۱. واژگان کلیدی: مدل‌سازی پیش‌بینی، دینامیک رسوب، کیفیت آب، الگوریتم‌های پیشرفته.

* **نویسندهٔ مسئول:** hadesertt64@gmail.com

[DOI: 10.61186/jeer.15.1.83]

۱ـ مقدمه

فرسایش آب بخشی جدایی ناپذیر از چرخه زمین شناسی و ژئومورفولوژیکی سیستم زمین است (Pourghasemi و همکاران، ۲۰۱۷). دینامیک (حمل و نقل) رسوب به عنوان یکی از پیامدهای منفی فرسایش خاک می‌تواند باعث مسائل و نگرانی‌های زیست محیطی مانند آسیب به اکوسیستم‌های آبی، کاهش کیفیت آب‌های سطحی و زیرزمینی، و تغییرات در مخزن سدها و مورفولوژی رودخانه شود (Ren و همکاران ۲۰۲۰: Raigani و همکاران ۲۰۱۹؛ Shojaeezadeh و همکاران ۲۰۱۸). بار رسوب معلق (SSL)[1] در حوضه‌های آبخیز یکی از مهم‌ترین پارامترهای هیدرولیکی و هیدرولوژیکی است که می‌تواند بر عملکرد سازه‌های هیدرولیکی و پروژه‌های انتقال آب تأثیر بگذارد. علاوه بر این، رسوبات منتقل شده به مخازن می‌تواند ظرفیت مخزن را کاهش داده و بر سیاست‌های عملیاتی مانند تامین آب، تولید انرژی و آبیاری تاثیر بگذارد (Darabi و همکاران، ۲۰۲۱). بنابراین، پیش‌بینی SSL در رودخانه‌ها نقش کلیدی در مدیریت حوضه و منابع آب پایدار، ریخت‌شناسی رودخانه و بهره‌برداری از سازه‌های هیدرولیکی ایفا می‌کند (Nosrati و همکاران، ۲۰۲۱؛ Haghighi و همکاران، ۲۰۱۹: Yang همکاران ۲۰۰۹). ناهمگونی فضایی ویژگی‌های مختلف فیزیکی، هیدرومترولوژیکی و ژئومورفولوژیک حوضه‌های رودخانه‌ای و رابطه غیرخطی بین این متغیرها و فرآیند رسوب‌گذاری، مانع بزرگی در پیش‌بینی دقیق بار و غلظت رسوب بوده است (Melesse و همکاران، ۲۰۱۱). در نتیجه، تخمین و پیش‌بینی بار رسوب در یک سیستم رودخانه‌ای نیازمند مدل‌های قوی است که می‌توانند روابط غیرخطی و مقادیر گمشده را مدیریت کنند (Khosravi و همکاران، ۲۰۱۸؛ Choubin و همکاران، ۲۰۱۲). در دهه‌های اخیر، توسعه تکنیک‌های هوش مصنوعی (Cobaner و همکاران، ۲۰۰۹) و داده‌کاوی به‌عنوان پیش‌بینی‌کننده پدیده‌های هیدرولوژیکی، تحول بزرگی در پیش‌بینی‌ها ایجاد کرده است (Lafdani و همکاران، ۲۰۱۳: Li و همکاران ۲۰۲۲). این مدل‌ها می‌توانند رویدادهای آینده را به طور موثر از طریق شناسایی الگوهای داده در سوابق تاریخی پیش بینی کنند (Khosravi و همکاران، ۲۰۱۸).

محققان مختلف از مدل‌های یادگیری ماشین برای مطالعات هیدرولوژیکی از جمله پیش‌بینی سری‌های زمانی رواناب یا جریان (Tingsanchali و Gautam، ۲۰۰۰: و وانگ، ۲۰۰۶)، مدل‌سازی انتقال رسوب (Khan و همکاران، ۲۰۱۸)، شبیه سازی جریان رودخانه (Bou-Fakhreddine و همکاران، ۲۰۱۸؛ Diop و همکاران، ۲۰۱۸؛)، مدیریت سطح آب (Yang و همکاران، ۱۹۹۸)، بهینه سازی عملیات مخزن (Solomatine و Torres، ۱۹۹۶)، مدیریت کیفیت آب (Wen و Lee، ۱۹۹۸)، تخمین پارامترهای کیفیت آب (Melesse و همکاران، ۲۰۰۸)، شبیه سازی آب‌های زیرزمینی (Nadiri و همکاران، ۲۰۱۷؛ Yu و همکاران، ۲۰۱۸)، تخمین تبخیر و تبخیر و تعرق (Ghorbani و همکاران، ۲۰۱۷؛ Tao و همکاران، ۲۰۱۸)، و پیش بینی رسوب (Abrahart و White، ۲۰۰۱؛ Nagy و همکاران، ۲۰۰۲؛ Alp و Cigizoglu، ۲۰۰۷) استفاده کرده‌اند. در سال‌های اخیر مطالعات زیادی برای پیش‌بینی SSL با استفاده از تکنیک‌های هوش مصنوعی و داده‌کاوی انجام شده است (Mohammadi-Raigani و همکاران، ۲۰۲۴؛ Rahul و همکاران، ۲۰۲۱؛ Choubin و همکاران، ۲۰۱۸؛ Liu و همکاران، ۲۰۱۳؛ Lafdani و همکاران، ۲۰۱۳؛ Rajaee و همکاران، ۲۰۱۳). به عنوان مثال، Melesse و همکاران (۲۰۱۱)، یک رویکرد مدل‌سازی مبتنی بر ANN برای تخمین SSL در سه رودخانه اصلی در ایالات متحده توسعه و نشان دادند که پیش‌بینی‌های ANN بهتر از نتایج حاصل از میانگین متحرک یکپارچه (ARIMA)، رگرسیون خطی چندگانه (MLR) و رگرسیون غیرخطی چندگانه (MNLR) بود. Chiang و Tsai (۲۰۱۱) گزارش دادند که مدل‌های SVM در برآورد بارهای رسوب معلق در حوضه رودخانه کائوپینگ تایوان از مدل‌های ANN بهتر عمل کردند. به طور مشابه، Çimen (۲۰۱۶)، عملکرد SVM را برای پیش‌بینی SSL دو رودخانه در ایالات متحده بررسی کرد و دریافت که یک مدل SVM می‌تواند SSL را بدون تولید مقادیر منفی دبی رسوب پیش‌بینی کند. Kisi و همکاران، (۲۰۱۲) ، الگوریتم‌های برنامه‌ریزی ژنتیکی (GP)، ANFIS، ANN و SVM را برای پیش بینی SSL روزانه در دو ایستگاه در رودخانه

[1] Suspended sediment load

[DOI: 10.61186/jeer.15.1.83]

کامبرلند در ایالات متحده مقایسه کرد و نشان داد که GP نسبت به مدل‌های دیگر برتری دارد. Khosravi و همکاران، (۲۰۱۸) ، در پژوهشی ارزیابی مدل‌های داده‌کاوی مستقل (به عنوان مثال، کاهش خطای درخت هرس (REPT)، M5P و یادگیری مبتنی بر نمونه ((IBK) و مدل‌های ترکیبی، (bagging-M5P، کمیته تصادفیREPT–(RCREPT) و زیرفضای تصادفی REPT (RS-REPT))) برای پیش‌بینی بار رسوب معلق (SSL) ناشی از ذوب یخبندان در حوضه آند در شیلی بررسی کردند. نتایج آن‌ها نشان داد که مدل‌های ترکیبی بهتر از مدل‌های فردی عمل کردند. bagging-M5P بهترین قابلیت پیش بینی را داشت در حالی که REPT ضعیف‌ترین را داشت. Hasanpour و همکاران (۲۰۱۹) ، از یک مدل ترکیبی فازی C-means (-FCM SVR) برای ارزیابی SSL استفاده کردند که عملکرد پیش‌بینی بهتری را نسبت به SRC، ANN، ANFIS و SVRنشان داد. Mehri و همکاران، (۲۰۲۱)، در این مطالعه با استفاده از الگوریتم‌های داده‌کاوی شامل چهار روش هوشمند ANFIS-PSO، ANFIS-GA، ANFIS و GMDH برای پیش‌بینی توزیع غلظت رسوب استفاده کردند. همچنین در این پژوهش مشخص شد که روش ANFIS-PSO می‌تواند روش مناسب و دقیق‌تری نسبت به سایر روش‌ها باشد.

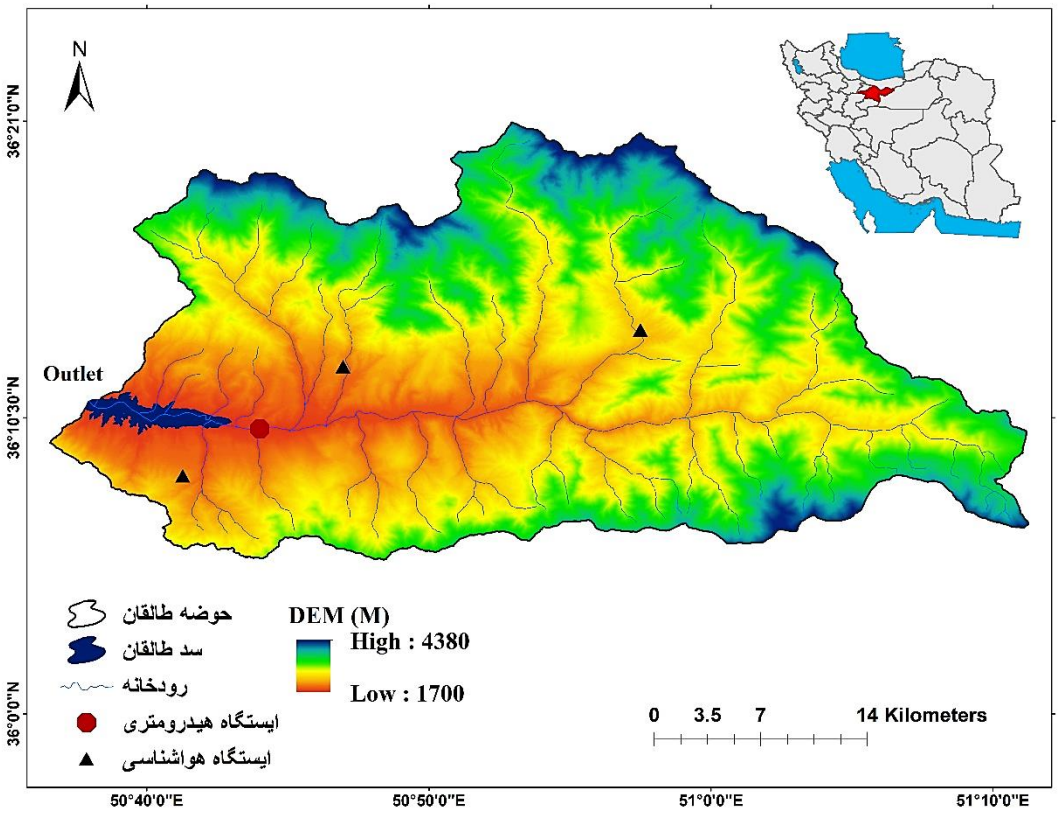
با وجود پیشرفت‌های قابل توجه در مدل‌سازی هیدرولوژیکی، پیش‌بینی دقیق بار رسوب معلق (SSL) همچنان یک چالش است، به ویژه در حوضه‌های آبخیز پیچیده‌ای مانند حوضه آبخیز طالقان در ایران. تحقیقات موجود اغلب به مدل‌های تجربی سنتی یا مدل‌های مبتنی بر فیزیک متکی هستند که می‌توانند توسط الزامات داده، پیچیدگی‌های کالیبراسیون پارامترها و فرضیات مربوط به فرآیندهای زیربنایی محدود شوند. در حالی که برخی از مطالعات، مدل‌های یادگیری ماشین را برای پیش‌بینی SSL بررسی کرده‌اند، شکافی در تحقیقات برای ارزیابی و مقایسه سیستماتیک عملکرد طیف متنوعی از الگوریتم‌های پیشرفته یادگیری ماشین، مانند CTree، CForest، QRNN، LASSO، Cubist و XGBoost، در این زمینه خاص وجود دارد. علاوه بر این، تحقیقات محدودی بر درک بین دقت پیش‌بینی و تفسیرپذیری مدل در مدل‌سازی SSL متمرکز شده است، که برای آگاه کردن تصمیمات مدیریتی بسیار مهم است. این مطالعه با ارزیابی دقیق و مقایسه عملکرد این مدل‌ها در پیش‌بینی SSL در حوضه آبخیز طالقان، با در نظر گرفتن جنبه‌های دقت و تفسیرپذیری، قصد دارد به این شکاف‌ها بپردازد. این امر بینش‌های ارزشمندی را در مورد مناسب بودن رویکردهای مختلف یادگیری ماشین برای پیش‌بینی SSL در محیط‌های نیمه‌خشک مشابه ارائه می‌دهد. از این رو، با توجه به توانایی بالای مدل‌های یادگیری، هدف اصلی این مقاله بررسی کارایی و دقت شش مدل یادگیری ماشین (xgbTree، Cubist، qrnn، Ctree، Cforest وLASSO) در پیش‌بینی SSL روزانه در رودخانه می‌باشد. برای این منظور، رودخانه طالقان در ایران به عنوان مطالعه موردی انتخاب شده و ۲۰۳ داده بار رسوب معلق روزانه از دوره ۲۰۰۰-۲۰۱۸ در نظر گرفته شده است. در این مطالعه از روش آماری آنالیز مولفه‌های اصلی (PCA) برای انتخاب ورودی مهم در پیش‌بینی SSL روزانه در رودخانه طالقان استفاده گردید. عملکرد مدل‌های یادگیری ماشین با استفاده از معیارهای آماری مختلف، نمودار تیلور و نمودارهای بصری به‌منظور شناسایی بهترین مدل‌های پیش‌بینی ارزیابی شدند. لازم به ذکر است، طبق دانش نویسندگان، هیچ اثری در ادبیات مطالعات داخلی ایران گزارش نشده است که از xgbTree، Cubist، qrnn، Ctree و LASSO برای پیش‌بینی SSL روزانه استفاده کند.

۲ـ منطقهٔ مورد مطالعه

حوضه آبخیز طالقان رود در شمال غرب تهران در استان البرز به مساحت ۱۳۵۲ کیلومتر مربع واقع شده است. منطقه مورد مطالعه در قسمت بالایی سد طالقان (۳۶ درجه و ۴ دقیقه تا ۳۶ درجه و ۲۱ دقیقه شمالی و ۵۰ درجه و ۳۸ دقیقه تا ۵۱ درجه و ۱۲ دقیقه شرقی) واقع شده است (شکل ۱). ارتفاع توپوگرافی حوضه آبخیز رودخانه طالقان بین ۱۷۰۰ متر تا ۴۳۸۰ متر از سطح متوسط دریا متغیر است و میانگین وزنی آن ۲۷۵۳ متر است که منطقه عمدتاً کوهستانی آن را برجسته می‌کند. مهمترین ویژگی منطقه مورد مطالعه، ارتفاع زیاد و شیب تند آن است که بیش از ۴۰ درصد از حوزه آبخیز طالقان دارای شیب‌هایی به این بزرگی

[DOI: 10.61186/jeer.15.1.83]

و جهت گیری اصلی شرقی-غربی است. طول رودخانه طالقان ۳۹ کیلومتر است که فصل طغیان آن بیشتر در ماه‌های بهار اتفاق می‌افتد. بیشترین و کمترین میانگین بارندگی سالانه به ترتیب ۸۱۴ و ۴۵۴ میلی‌متر و میانگین بارندگی سالانه در حوضه تقریباً ۶۰۰ میلی‌متر است (Hosseini و همکاران، ۲۰۱۲). بر اساس داده‌های ایستگاه زیدشت، میانگین دما در منطقه مورد مطالعه ۸ درجه سانتی‌گراد است. حداکثر مطلق دما در ماه مرداد ۳۷ درجه سانتی‌گراد و حداقل دما ۱۸ درجه سانتی‌گراد است. مرتع بیشترین کاربری منطقه مورد مطالعه را دارد و تقریباً ۸۱/۷ درصد از محدوده مورد مطالعه را در بر می‌گیرد.



شکل ۱: موقعیت جغرافیایی حوضهٔ آبخیز رودخانه طالقان و ایستگاه‌های مورد استفاده

۳ـ مواد و روش

۳ـ۱ـ آنالیز آماری و مجموعه داده‌های مورد استفاده برای مدل‌سازی

مجموعه داده‌های مورد استفاده در این پژوهش برای پیش‌بینی بار رسوب معلق (SSL) در حوضه آبخیز رودخانه طالقان شامل: داده‌های دبی، رسوب روزانه، بارش، بارش تجمعی، شاخص برف، دبی و رسوب با تاخیر زمانی یک روزه (t-1)) از سال ۲۰۰۰ تا ۲۰۱۸ برای آموزش و آزمایش (به ترتیب ۷۵ درصد آموزش و ۲۵ درصد آزمون)، مدل‌های ML استفاده شد. سری زمانی داده‌های دبی روزانه و رسوب معلق از ایستگاه هیدرومتری گلینک واقع در خروجی حوزه آبخیز طالقان (شکل ۱)، از سازمان آب منطقه‌ای استان تهران و البرز جمع‌آوری و برای ساخت مدل‌های پیش بینی مورد استفاده قرار گرفت. داده‌های بارندگی حوضه آبخیز طالقان از سه ایستگاه باران سنجی زیدشت (۵۰ درجه و ۶۹ دقیقه شرقی و ۳۶ درجه و ۱۴ دقیقه شمالی)، دیزان (۵۰ درجه و ۹۶ دقیقه شرقی و ۳۶ درجه و ۲۳ دقیقه شمالی) و جزینان (۵۰ درجه و ۷۸ دقیقه شرقی و ۳۶ درجه و ۲۰ دقیقه شمالی) جمع آوری شد. همچنین شاخص برف از طریق آنالیز تصاویر ماهواره‌ای لندست در نرم افزار GIS بدست آمد. شکل (۱) موقعیت ایستگاه‌های اندازه‌گیری را در حوضه آبخیز رودخانه طالقان نشان می‌دهد.

[DOI: 10.61186/jeer.15.1.83]

اطلاعات آماری روزانه دبی رودخانه (Q)، بار رسوب معلق (SSL)، مجموعه داده‌های بارندگی و شاخص برف مورد استفاده در مطالعه حاضر در جدول (۱) آورده شده است. پارامترهای آماری برای مجموعه داده‌ها شامل میانگین، انحراف معیار (Sd)، حداکثر و حداقل (جدول ۱) است. هنگام تقسیم داده‌ها به زیر مجموعه‌های آموزشی و آزمون، بررسی داده‌هایی که جامعه آماری مشابهی را ارائه می‌دهند ضروری است (Masters، ۱۹۹۳). لازم به ذکر است که این مدل‌ها زمانی بهترین عملکرد را دارند که فراتر از محدوده داده‌های مورد استفاده برای آموزش مدل برون‌یابی نشوند (Tokar و Johnson، ۱۹۹۹). مقدار انحراف معیار (sd) دبی روزانه (Q) و دبی با تاخیر یک روز ((Q (t-1) در هر دو مجموعه داده آموزش و آزمون برابر بود، در حالی‌که این مقدار، در SSL و و بار رسوب با تاخیر یک روز (SSL-1) برای دوره آموزش کمتر از دوره آزمون بود. اما در شاخص برف (Snow index) بالعکس بود. طبق جدول ۱، میانگین Q، (Q (t-1، SSL و SSL-1 در دوره آزمون بیشتر از دوره آموزش است. اما در شاخص برف (Snow index) میانگین در دوره آموزش بیشتر است.

جدول ۱: خلاصه آماری داده¬های آموزشی، آزمایشی و تمام مجموعه داده‌ها

پارامترهای آماری	حداقل	حداکثر	میانگین	انحراف استاندارد
Training set (152)				
Q (m^3s^{-1})	۱/۲۷	۷۸/۲۸	۱۲/۵۱	۱۳/۹۹
SSL (ton/day)	۳/۲۳	۱۲۳۵۰/۴۷	۷۷۰/۴۹	۱۶۱۳/۱۸
Q (t-1)	۰/۰۰	۷۸/۲۸	۱۲/۴۹	۱۴/۰۱
SSL-1	۰/۰۰	۱۲۳۵۰/۴۷	۷۷۰/۴۹	۱۶۱۳/۱۸
Rainfall (mm)- Zidasht station	۰/۰۰	۲۶/۰۰	۰/۹۱	۳/۳۵
Cumulative precipitation (mm)	۰/۰۰	۳۳/۰۰	۴/۸۹	۷/۵۰
Rainfall (mm) – Jazinan station	۰/۰۰	۲۵/۰۰	۱/۴۰	۴/۲۹
Cumulative Rainfall (mm)	۰/۰۰	۶۱/۰۰	۷/۰۷	۱۰/۹۱
Rainfall (mm)- Dizan station	۰/۰۰	۳۹/۰۰	۲/۳۲	۶/۲۰
Cumulative Rainfall (mm)	۰/۰۰	۵۱/۰۰	۱۲/۳۲	۱۳/۶۰
Snow index	۰/۰۰	۷۹/۵۵	۲۳/۴۴	۲۷/۰۱
Testing set (51)				
Q (m^3s^{-1})	۲/۲۲	۵۳/۴۶	۱۳/۲۹	۱۳/۹۷
SSL (ton/day)	۷/۲۱	۷۰۳۸/۶۲	۹۶۵/۳۷	۱۷۶۹/۲۵
Q (t-1)	۲/۲۲	۵۳/۴۶	۱۳/۲۶	۱۳/۹۹
SSL-1	۷/۲۱	۷۰۳۸/۶۲	۹۶۵/۴۸	۱۷۶۹/۱۹
Rainfall (mm)- Zidasht station	۰/۰۰	۳۲/۴۰	۲/۵۸	۷/۱۳
Cumulative precipitation (mm)	۰/۰۰	۵۹/۴۰	۱۲/۶۲	۱۵/۰۹
Rainfall (mm) – Jazinan station	۰/۰۰	۳۲/۴۰	۲/۵۸	۷/۱۳
Cumulative Rainfall (mm)	۰/۰۰	۵۹/۴۰	۱۲/۶۲	۱۵/۰۹
Rainfall (mm)- Dizan station	۰/۰۰	۵۸/۰۰	۲/۳۹	۸/۷۷
Cumulative Rainfall (mm)	۰/۰۰	۸۲/۰۰	۱۲/۸۸	۲۰/۹۱
Snow index	۰/۰۰	۷۰/۰۰	۱۳/۹۳	۲۰/۲۲
All data set (203)	۲/۲۲	۵۳/۴۶	۱۳/۲۹	۱۳/۹۷

۲-۳- یادگیری ماشین بینی SSL

در این پژوهش از مدل‌های Ctree، Cforest، LASSO، xgbTree، qrnn و Cubist برای پیاده‌سازی مدل‌ها با استفاده از بسته Caret نسخه 4.3.2 نرم‌افزار R، استفاده شد.

[DOI: 10.61186/jeer.15.1.83]

Q (m^3s^{-1})	۱/۲۷	۷۸/۲۸۱	۱۲/۶۵	۱۳/۹۴
SSL (ton/day)	۳/۲۳	۱۲۳۵۰/۷۴	۸۱۵/۷۳	۱۶۴۸/۵۱
Q (t-1)	۰/۰۰	۷۸/۲۸۱	۱۲/۶۴	۱۳/۹۵
SSL (t-1)	۰/۰۰	۱۲۳۵۰/۷۴	۸۱۵/۶۸	۱۶۴۸/۵۴
Rainfall (mm)- Zidasht station	۰/۰۰	۳۲/۴	۱/۳۳	۴/۶۳
Cumulative precipitation (mm)	۰/۰۰	۵۹/۴	۶/۸۵	۱۰/۴۶
Rainfall (mm) – Jazinan station	۰/۰۰	۳۲/۴	۱/۶۹	۵/۱۵
Cumulative Rainfall (mm)	۰/۰۰	۶۱	۸/۴۷	۱۲/۲۸
Rainfall (mm)- Dizan station	۰/۰۰	۵۸	۲/۳۴	۶/۹۰
Cumulative Rainfall (mm)	۰/۰۰	۸۲	۱۲/۴۱	۱۵/۶۹
Snow index	۰/۰۰	۷۹/۵۴۷	۲۰/۹۴	۲۵/۷۳

ماهیت متنوع طیف وسیعی به وظیفه این مدل‌ها، از مزایا را بسته تحلیلی ارائه

می‌دهد.Ctree، با چارچوب آزمون فرضیه خود، تفسیرپذیری بیشتری را در مقایسه با درخت‌های تصمیم سنتی فراهم می‌کند و امکان بینش در مورد اهمیت آماری تقسیم‌ها را فراهم می‌آورد (Hothornو همکاران، ۲۰۱۵). CForest از این قدرت بهره می‌برد و عملکرد پیش‌بینی را از طریق جمع‌آوری بهبود می‌بخشد (Luo و همکاران، ۲۰۲۰). qrnnها به‌ویژه زمانی ارزشمند هستند که تمرکز فراتر از پیش‌بینی‌های نقطه‌ای به درک کل توزیع شرطی متغیر هدف گسترش یابد و ارزیابی ریسک قوی را ممکن سازد (Li و همکاران، ۲۰۲۳). LASSO، با کوچک کردن ضرایب و انجام انتخاب متغیر، مدل‌ها را ساده می‌کند و تعمیم را به‌ویژه در مجموعه داده‌های با ابعاد بالا بهبود می‌بخشد (Muthukrishnan و همکاران، ۲۰۱۶). مدل‌های Cubist در ثبت روابط و تعاملات غیرخطی برتری دارند و دقت بهبود یافته‌ای را در مقایسه با مدل‌های کاملاً خطی ارائه می‌دهند (Houborg و همکاران، ۲۰۱۸). XGBoost، که به دلیل کارایی و عملکرد خود شناخته شده است، اغلب به دلیل چارچوب تقویت گرادیان و تکنیک‌های منظم‌سازی خود، به نتایج پیشرفته در وظایف مختلف پیش‌بینی دست می‌یابد (Ma و همکاران، ۲۰۲۲). انتخاب مدل مناسب به نیازهای خاص تحقیق، از جمله سطح تفسیرپذیری مورد نظر، ماهیت داده‌ها و پیچیدگی روابط مدل‌سازی شده بستگی دارد.

مزیت دیگر استفاده از مجموعه‌ای متنوع از مدل‌ها، پتانسیل مقایسه مدل و ایجاد گروه است. مقایسه عملکرد مدل‌های مختلف، مانند CForest و XGBoost، می‌تواند بینش‌هایی را در مورد داده‌ها و روابط پیش‌بینی کننده زیربنایی آشکار کند. علاوه بر این، ترکیب پیش‌بینی‌ها از چندین مدل از طریق روش‌های گروهی اغلب می‌تواند منجر به بهبود دقت و استحکام کلی در مقایسه با تکیه بر یک مدل واحد شود. این رویکرد از نقاط قوت هر مدل به صورت جداگانه استفاده می‌کند و در عین حال نقاط ضعف آنها را کاهش می‌دهد. به عنوان مثال، ترکیب تفسیرپذیری CTree با قدرت پیش‌بینی XGBoost می‌تواند تجزیه و تحلیل جامع‌تر و روشنگرانه‌تری ارائه دهد. بنابراین، در نظر گرفتن طیف وسیعی از مدل‌ها نه تنها انعطاف‌پذیری را ارائه می‌دهد، بلکه راه‌هایی را برای استراتژی‌های مدل‌سازی پیچیده‌تر نیز باز می‌کند.

۳ـ۳ـ انتخاب ورودی‌های مناسب برای مدل سازی SSL

انتخاب ویژگی یا متغیرهای ورودی یکی از مراحل کلیدی در کاربرد الگوریتم‌های ML است (Gholami و همکاران، ۲۰۲۱). تحلیل مؤلفه اصلی (PCA) روشی مؤثر برای شناسایی ورودی مدل‌ها و کاهش تعداد پارامترهای ورودی مورد نیاز است (Lu و همکاران، ۲۰۱۹). در نتیجه، PCA را می‌توان برای کاهش پیچیدگی متغیرهای ورودی، زمانی که حجم زیادی از اطلاعات وجود دارد و به‌منظور تفسیر بهتر متغیرها استفاده کرد (Noori و همکاران، ۲۰۱۰). PCA با توصیف حداکثر مقدار واریانس مشترک در یک ماتریس همبستگی با استفاده از کمترین تعداد مفاهیم گویا، به صرفه‌جویی بدست می‌یابد. معیار Kaiser-Meyer-

[DOI: 10.61186/jeer.15.1.83]

Melkin (KMO)برای بررسی کفایت داده‌ها با استفاده از رابطه (۱) بدست می آید. KMO اندازه گیری نسبت واریانس بین متغیرهایی است که ممکن است واریانس مشترک باشد (Darabi و همکاران ۲۰۲۱).

رابطهٔ ۱
$$KMO = \frac{\sum(correlation)^2}{\sum(correlation)^2+\sum(partial\ correlation)^2}$$

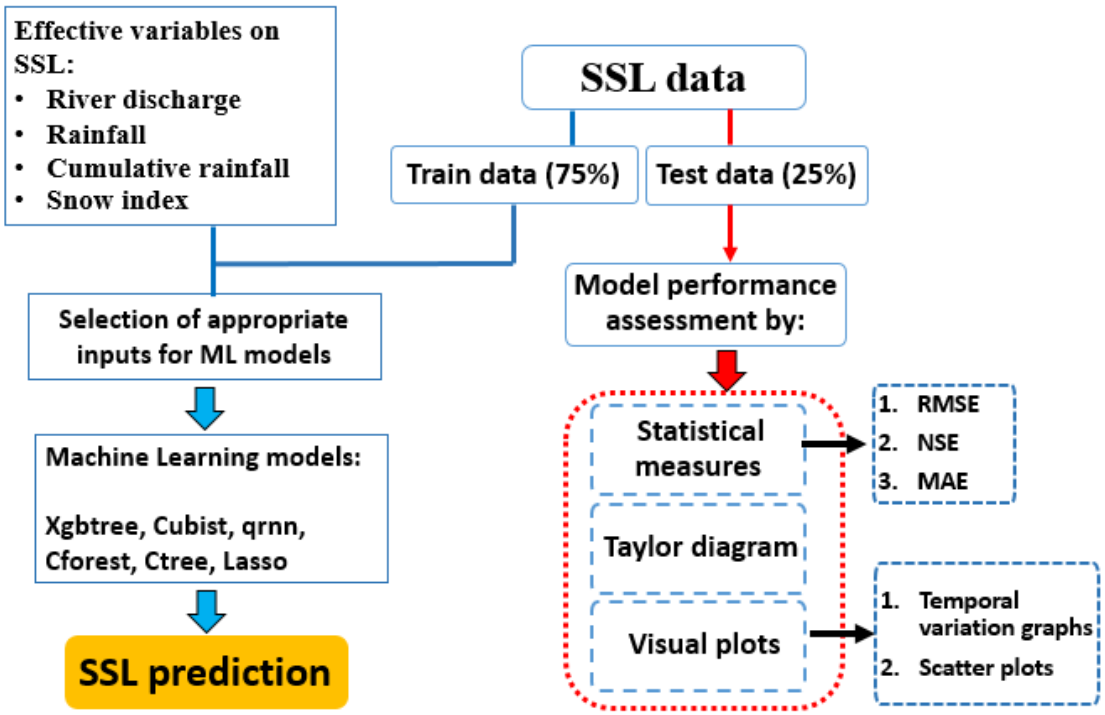
طبق ادبیات، حداقل مقدار KMO باید ۰/۵ باشد. در این مطالعه، KMO مقدار ۰/۶۵ بود. همبستگی بین متغیرها باید بررسی شود تا از مشکلات چند خطی جلوگیری شود (Lu و همکاران، ۲۰۱۹). در این مطالعه تمامی مقادیر همبستگی زیر آستانه (۰/۹) بوده و در نتیجه مشکلی از چند خطی وجود نداشت. مروری جامع از روش مؤلفه اصلی توسط Nosrati و همکاران (a۲۰۱۸) ارائه شده است. تمام تجزیه و تحلیل‌های آماری با استفاده از STATISTICA V.16.0 (StatSoft, 2016) انجام شد.

۴-۳- معیارهای ارزیابی عملکرد مدل‌ها

ارزیابی مدل نباید بر اساس یک شاخص واحد با توجه به ماهیت تصادفی متغیرهای هیدرولوژیکی باشد (Dawson و همکاران، ۲۰۰۷: Idrees و همکاران، ۲۰۲۱: Khosravi و همکاران، ۲۰۱۸). در نتیجه، در این مطالعه از سه شاخص رایج برای ارزیابی عملکرد و مقایسه مدل‌های ML استفاده شد. که شامل دو معیار انحراف (شامل ریشه میانگین مربعات خطا (RMSE)، میانگین خطای مطلق (MAE) و یک معیار اندازه گیری بهره‌وری (ضریب کارایی نش- ساتکلیف[1] (NSE) شد. همچنین به منظور ارزیابی و مقایسه عملکرد مدل‌های پیش بینی علاوه بر معیارهای آماری کمی، برای به دست آوردن درک بصری عملکرد مدل‌های سطحی برای پیش‌بینی SSL در منطقه مورد مطالعه، ترکیبی از نمودارهای تیلور (Taylor، ۲۰۰۱)، نمودارهای خطی و نمودارهای پراکندگی استفاده گردید. مراحل اصلی برای مدل سازی زمانی SSL رودخانه توسط مدل‌های یادگیری ماشین در شکل ۲ ارائه شده است.

[1] Nash-Sutcliffe (NSE)

[DOI: 10.61186/jeer.15.1.83]



شکل ۲: نمودار جریانی تحقیق

۴-نتایج و بحث

۴-۱- متغیرهای ورودی مدل¬های پیش بینی SSL

پس از حل معادله (KMO)، ۱۰ مقدار ویژه به دست آمد. ویژگی‌های مولفه‌ها (PC) شامل مقادیر ویژه، نسبت واریانس و نسبت واریانس تجمعی در جدول ۲، ارائه شده است. با توجه به جدول ۲، نتایج PCA نشان داد که سه مولفه اصلی[1] اول (PC1–PC3) با مقادیر ویژه > ۱، نزدیک به ۷۰ درصد از نسبت کل واریانس متغیرهای ورودی را نشان می‌دهند (جدول ۲). علاوه بر این، بردارهای ویژه که ضرایب تشکیل PC را ارزیابی می‌کنند، از طریق کاربرد PCA به دست می‌آیند (جدول ۳).

در جدول (۳)، مؤثرترین متغیرها در تشکیل مولفه‌ها با فونت درشت نشان داده شده است. به طور کلی مشخص است که دبی (Q) و بار رسوب معلق (SSL) با تاخیر یک روز (t-1) بیشترین تأثیر را بر PC1 دارند که بیش از ۳۶/۲۷ درصد نسبت‌های واریانس متغیرهای ورودی را شامل می‌شود. همچنین متغیر بارندگی ایستگاه زیدشت و جزینان بیشترین تأثیر را بر مولفه دوم (PC2) دارد که بیش از ۲۱/۶۸ درصد نسبت های واریانس متغیرهای ورودی را شامل می شود. علاوه بر این، مولفه اصلی سوم (PC3) مربوط به کوچک‌ترین مقدار ویژه منتخب (۱/۱۷) تقریبا ۱۱/۷۱ درصد از واریانس کل را تشکیل می‌دهد. تحت تأثیر دبی روزانه، بارش تجمعی ایستگاه زیدشت و جزینان قرار می‌گیرد. سایر اطلاعات مولفه‌ها را می‌توان از این جدول به دست آورد (جدول ۳). در این مقاله، سه مولفه اول به عنوان ورودی مدل‌های یادگیری ماشین بر اساس مدل PCA (جدول ۳) انتخاب شدند. به‌طور کلی از میان ده متغیر ورودی سه فاکتور شاخص برف، بارش

[1] principal components

[DOI: 10.61186/jeer.15.1.83]

ایستگاه جزینان و بارش تجمعی ایستگاه جزینان بر روی هیچ مولفه‌ای اثرگذار نبوده و در نتیجه حذف شدند. نتایج مراحل آموزش و آزمون مدل‌های یادگیری ماشین توسط متغیرهای انتخابی با روش PCA در جدول ۴ آورده شده است.

جدول ۲. آمار توصیفی مولفه‌های اصلی ایجاد شده

PCs	Eigenvalue	Variance proportion	Cumulative variance proportion
PC 1	۳/۶۳	۳۶/۲۷	۳۶/۲۷
PC 2	۲/۱۷	۲۱/۶۸	۵۷/۹۵
PC 3	۱/۱۷	۱۱/۷۱	۶۹/۶۶
PC 4	۰/۹۱۲	۹/۱۲۳	۷۸/۷۸۳
PC 5	۰/۷۷۹	۷/۷۹۰	۸۶/۵۷۳
PC 6	۰/۴۴۲	۴/۴۲۱	۹۰/۹۹۴
PC 7	۰/۳۸۱	۳/۸۱۱	۹۴/۸۰۵
PC 8	۰/۳۱۶	۳/۱۶۰	۹۷/۹۶۴
PC 9	۰/۱۵۴	۱/۵۴۵	۹۹/۵۰۹
PC 10	۰/۰۴۹	۰/۴۹۱	۱۰۰

جدول ۳. مقادیر تحلیل مؤلفه اصلی (PCA) متغیرها و مقادیر ویژه ماتریس همبستگی

Parameter	PC 1	PC 2	PC 3	Communalities
Q (m^3s^{-1})	۰/۶۴۵	۰/۵۳۲	۰/۰۷۹	۰/۷۱
Q (t-1)	**۰/۹۳۱**	-۰/۱۴۵	۰/۲۱۹	۰/۹۴
SSL (ton/day) (t-1)	**۰/۸۸۴**	-۰/۱۶۵	۰/۲۸۱	۰/۸۹
Rainfall (mm)- Zidasht station	-۰/۰۶۹	۰/۸۳۸	۰/۲۹۳	۰/۷۹
Cumulative Rainfall (mm)	۰/۰۰۸	۰/۱۷۳	**۰/۸۷۵**	۰/۸۰
Rainfall (mm) - Jazinan station	۰/۰۳۲	۰/۸۰۳	۰/۲۴۴	۰/۷۱
Cumulative Rainfall (mm)	۰/۱۸۱	۰/۲۲۸	۰/۷۸۸	۰/۷۰
Rainfall (mm)- Dizan station	۰/۰۵۰	۰/۷۱۳	۰/۱۳۷	۰/۵۳
Cumulative Rainfall (mm)	۰/۱۰۷	۰/۲۵۵	۰/۷۴۹	۰/۶۴
Snow index	-۰/۴۹۳	-۰/۱۴۲	۰/۱۰۰	۰/۲۷
Eigenvalue	۳/۶۳	۲/۱۷	۱/۱۷	
% Total variance (Variance proportion)	۳۶/۲۷	۲۱/۶۸	۱۱/۷۱	
Cumulative % variance (Cumulative variance proportion)	۳۶/۲۷	۵۷/۹۵	۶۹/۶۶	

[DOI: 10.61186/jeer.15.1.83]

۴-۲- ارزیابی عملکرد مدل‌های پیش‌بینی‌کننده SSL

پژوهش حاضر عملکرد شش مدل ML را برای پیش‌بینی SSL در حوضه طالقان اعمال و مقایسه گردید. این مدل‌های ML شامل مدل‌های Ctree، Cforest، Cubist، LASSO، qrnn و xgbTree بودند. مدل‌ها با استفاده از شاخص‌های آماری از جمله RMSE، NSE و MAE در دوره‌های آموزشی و آزمایشی ارزیابی و مقایسه شدند. مقادیر این آمار به دست آمده توسط مدل‌ها در جدول (۴) آورده شده است. طبق جدول (۴)، مقادیر RMSE، NSE و MAE و برای Cforest به ترتیب در محدوده ۴۷۱/۱۱ تا ۶۱۹/۹۷، ۶۴/۷۶- تا ۶/۹۳ و ۰/۸۵ تا ۰/۹۳ قرار دارند. پارامترهای آماری ذکر شده برای مدل Ctree به ترتیب در محدوده ۵۰۲/۸۲ تا ۶۳۴/۴۸، ۱۳/۵۲- تا ۰ و ۰/۸۴ تا ۰/۹۲ بوده است. برای مدل Cubist، به ترتیب از ۱۲۵/۳۴ تا ۳۳۰/۴۶، ۵۱/۸۸- تا ۲۱/۷۲- و ۰/۹۶ تا ۰/۹۹، برای مدل LASSO، به ترتیب ۵۷۵/۴۷ تا ۵۹۳/۸۸، ۱۰۰/۳۰- تا ۰ و ۰/۸۷ تا ۰/۸۸، برای مدل qrnn، به ترتیب از ۱۸۴/۹۶ تا ۳۴۹/۸۵، ۸۹/۷۶- تا ۳۴/۴۱- و ۰/۹۶ تا ۰/۹۹ در نهایت، برای مدل xgbTree پارامترهای آماری ذکر شده به ترتیب در بازه های ۱۹/۷۵ تا ۳۰۱/۶۳، ۵۶/۵۹- تا ۱ و ۰/۹۷ تا ۱ است.

با توجه به تمام معیارهای ارزیابی، از این جدول به وضوح می‌توان دریافت که مدل xgbTree، Cubist و qrnn بسیار بهتر از مدل‌های Ctree، Cforest و LASSO عمل می‌کند. در نمودار تغییرات زمانی مدل xgbTree (R^2=۱)، Cubist (R^2=۱) و qrnn (R^2=۰/۹۹) (شکل ۳-a-b-c؛ جدول: ۴)، ما انحراف قابل توجهی از مقادیر رسوب ثبت شده در طول سری آموزشی مشاهده نمی‌کنیم. در مقابل، مدل‌های Ctree، Cforest و LASSO (مقدار R^2 به‌ترتیب برابر با ۰/۸۴، ۰/۸۵ و ۰/۸۸) به طور مداوم رکوردهای رسوب مشاهده‌شده را دست کم یا بیشتر برآورد می‌کنند (شکل ۳- d-e-f). به طور کلی، این مدل‌ها تمایل به دست کم گرفتن مقادیر بالای SSL دارند، که ممکن است به غلظت بالای رسوب موجود در دوره آموزش نسبت داده شود. همانطور که از جدول (۱) مشهود است، دوره آموزشی به طور کلی مقادیر SSL بالاتری را در مقایسه با دوره آزمایشی شامل می‌شود.

از سوی دیگر، تمام مقادیر RMSE برای دوره آزمایش از مدل xgbTree، Cubist و qrnn بسیار کمتر از مدل‌های Ctree، Cforest و LASSO بودند و مقادیر R^2 و NSE مربوطه بالاتر بودند (جدول: ۴). در این مورد، مقادیرRMSE و R^2 برای مدل xgbTree به ترتیب ۳۰۱/۶۳ و ۰/۹۸ بود که دقیق‌ترین پیش بینی بار رسوب معلق را برآورد کرد. علاوه بر این، مدلLASSO ضعیف‌ترین پیش‌بینی (RMSE= ۵۹۳/۸۸ وR^2 =۰/۹۲) را نسبت به سایر مدل‌ها داشت. نتایج محاسبات نشان می‌دهد که بهترین مقادیر RMSE بدست‌آمده توسط مدل‌های xgbTree، Cubist، qrnn، Ctree، Cforest و LASSO به ترتیب، ۳۰۱/۶۳، ۳۳۰/۴۶، ۳۴۹/۸۵، ۵۰۲/۸۲، ۴۷۱/۱۱ و ۵۹۳/۸۸ است. بهترین مقادیر MAE به ترتیب ۵۶/۵۹-، ۵۱/۸۸-، ۸۹/۷۶-، ۱۳/۵۲-، ۶۴/۷۶- و ۱۰۰/۳۰- هستند. و بهترین مقادیر NSE به ترتیب ۰/۹۷، ۰/۹۶، ۰/۹۶، ۰/۹۲، ۰/۹۳ و ۰/۸۸ هستند (جدول: ۴).

با توجه به تمام معیارهای ارزیابی، از این جدول به وضوح می‌توان دریافت که مدل xgbTree، Cubist و qrnn بسیار بهتر از مدل‌های Ctree، Cforest و LASSO عمل می‌کند. در نمودار تغییرات زمانی مدل xgbTree (R^2=۱)، Cubist(R^2=۱) و qrnn (R^2=۰/۹۹) (شکل ۳-a-b-c؛ جدول: ۴)، ما انحراف قابل توجهی از مقادیر رسوب ثبت شده در طول سری آموزشی مشاهده نمی‌کنیم. در مقابل، مدل‌های Ctree، Cforest و LASSO (مقدار R^2 به‌ترتیب برابر با

[DOI: 10.61186/jeer.15.1.83]

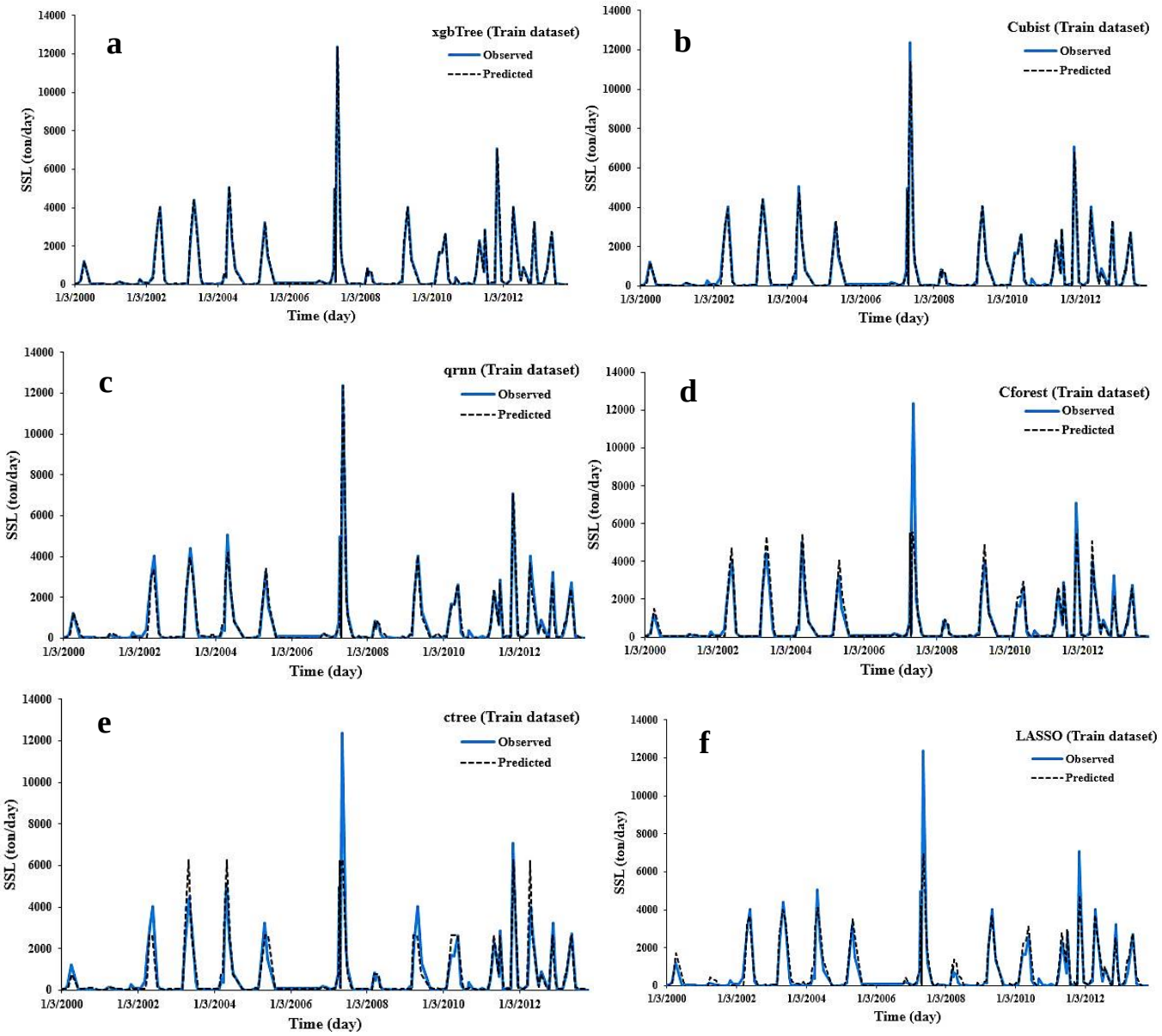
۰/۸۴، ۰/۸۵ و ۰/۸۸) به طور مداوم رکوردهای رسوب مشاهده‌شده را دست کم یا بیشتر برآورد می‌کنند (شکل ۳-d-e-f). به طور کلی، این مدل‌ها تمایل به دست کم گرفتن مقادیر بالای SSL دارند، که ممکن است به غلظت بالای رسوب موجود در دوره آموزش نسبت داده شود. همانطور که از جدول (۱) مشهود است، دوره آموزشی به طور کلی مقادیر SSL بالاتری را در مقایسه با دوره آزمایشی شامل می‌شود.

جدول ۴. نتایج ارزیابی عملکرد مدل‌ها برای هر دو مجموعه داده

Model	Train (152)				Test (51)			
	RMSE	NSE	MAE	R^2	RMSE	NSE	MAE	R^2
Cforest	۶۱۹/۹۷	۰/۸۵	۶/۹۳	۰/۸۵	۴۷۱/۱۱	۰/۹۳	-۶۴/۷۶	۰/۹۳
Ctree	۶۳۴/۴۸	۰/۸۴	۰/۰۰	۰/۸۴	۵۰۲/۸۲	۰/۹۲	-۱۳/۵۲	۰/۹۲
Cubist	۱۲۵/۳۴	۰/۹۹	-۲۱/۷۲	۱/۰۰	۳۳۰/۴۶	۰/۹۶	-۵۱/۸۸	۰/۹۷
LASSO	۵۷۵/۴۷	۰/۸۷	۰/۰۰	۰/۸۸	۵۹۳/۸۸	۰/۸۸	-۱۰۰/۳۰	۰/۹۲
qrnn	۱۸۴/۹۶	۰/۹۹	-۳۴/۴۱	۰/۹۹	۳۴۹/۸۵	۰/۹۶	-۸۹/۷۶	۰/۹۷
xgbTree	۱۹/۷۵	۱/۰۰	۰/۰۰	۱/۰۰	۳۰۱/۶۳	۰/۹۷	-۵۶/۵۹	۰/۹۸

از سوی دیگر، تمام مقادیر RMSE برای دوره آزمایش از مدل xgbTree، Cubist و qrnn بسیار کمتر از مدل‌های Ctree ، Cforest و LASSO بودند و مقادیر R^2 و NSE مربوطه بالاتر بودند (جدول: ۴). در این مورد، مقادیرRMSE و R^2 برای مدل xgbTree به ترتیب ۳۰۱/۶۳ و ۰/۹۸ بود که دقیق‌ترین پیش بینی بار رسوب معلق را برآورد کرد. علاوه بر این، مدلLASSO ضعیف‌ترین پیش‌بینی (RMSE= ۵۹۳/۸۸ وR^2 =۰/۹۲) را نسبت به سایر مدل‌ها داشت. نتایج محاسبات نشان می‌دهد که بهترین مقادیر RMSE بدست‌آمده توسط مدل‌های xgbTree، Cubist، qrnn، Ctree، Cforest و LASSO به ترتیب، ۳۰۱/۶۳، ۳۳۰/۴۶، ۳۴۹/۸۵، ۵۰۲/۸۲، ۴۷۱/۱۱ و ۵۹۳/۸۸ است. بهترین مقادیر MAE به ترتیب ۵۶/۵۹-، ۵۱/۸۸-، ۸۹/۷۶-، ۱۳/۵۲-، ۶۴/۷۶- و ۱۰۰/۳۰- هستند. و بهترین مقادیر NSE به ترتیب ۰/۹۷، ۰/۹۶، ۰/۹۲، ۰/۹۳ و ۰/۸۸ هستند (جدول: ۴).

[DOI: 10.61186/jeer.15.1.83]



شکل ۳. مقایسه بین SSL مشاهده شده و پیش بینی شده توسط مدل‌های ML برای مجموعه داده‌های آموزشی در ایستگاه سنجش گلینک در حوضه رودخانه طالقان.

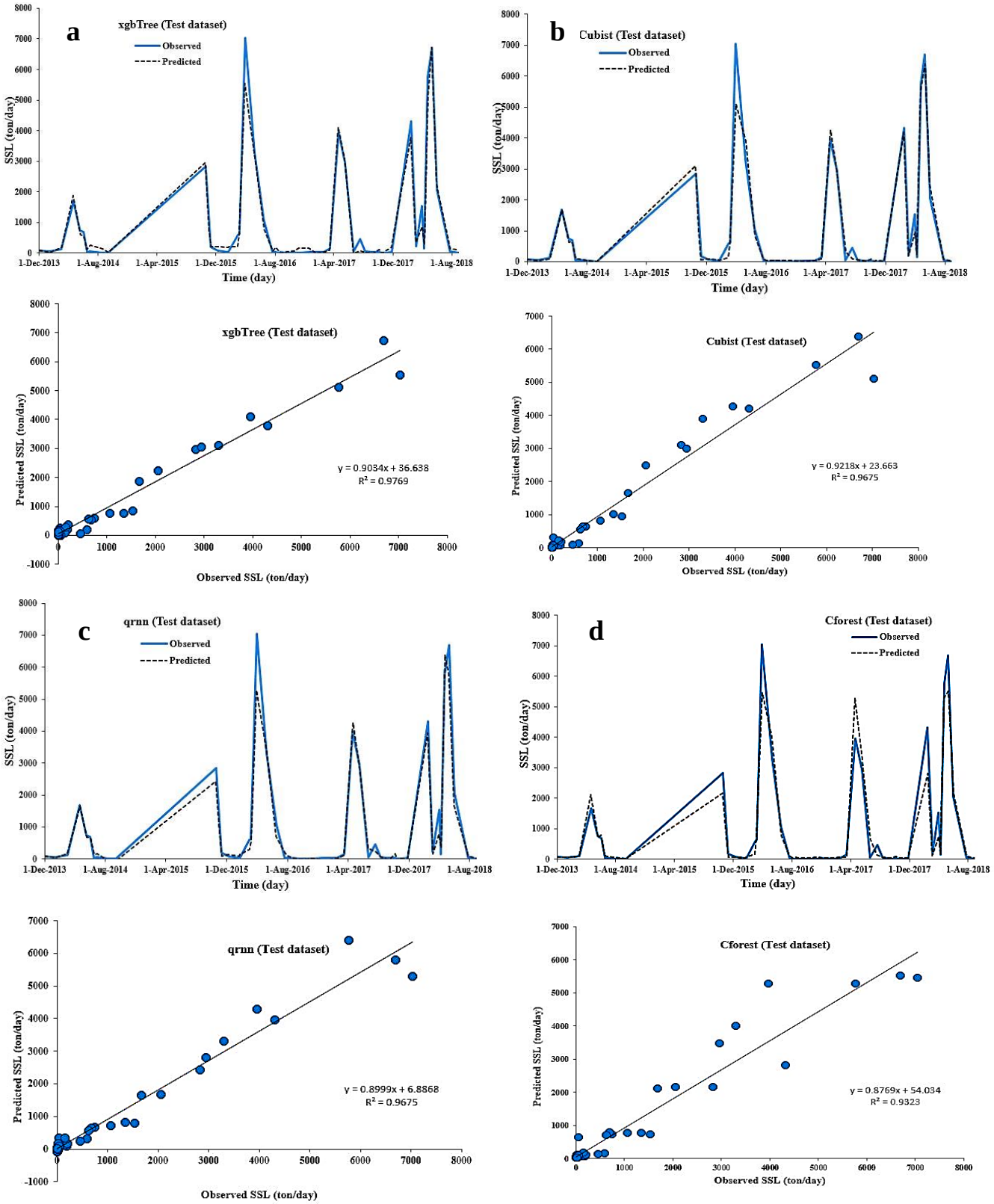
شکل (۴) مقادیر SSL مشاهده شده و پیش‌بینی‌شده به‌دست‌آمده از مدل‌های xgbTree، Cubist، qrnn، Ctree، Cforest و LASSO را برای دوره آزمایشی در قالب نمودارهای تغییرات زمانی (سمت بالا) و نمودارهای پراکندگی (سمت پایین) نشان می‌دهد. نمودارها به وضوح تطابق نزدیکی را بین مقادیر SSL مشاهده شده و تخمینی برای مدل‌های xgbTree ، Cubist و qrnn نشان می‌دهند (شکل ۴-a-b-c). این مدل‌ها با موفقیت هم قله‌های بالا و هم پایین را ثبت می‌کنند، اگرچه گاهاً تخمین‌های دست کم گرفتن مقادیر رسوب مشهود است. نمودار پراکندگی متناظر همچنین نشان می‌دهد که مقادیر پیش بینی شده SSL به خوبی در اطراف خط رگرسیون با مقدار بالای ضریب تعیین پراکنده شده است (شکل ۴-a-b-c). به عبارت دیگر، نتایج xgbTree، Cubist و qrnn در نمودارهای پراکندگی به خط مستقیم ۴۵ درجه نزدیک‌تر از سایر مدل‌ها بود که نشان‌دهنده سطح بالای قابلیت پیش‌بینی این مدل‌ها در مطالعه حاضر است. در حالی که مدل‌های Ctree، Cforest و LASSO نتایج ضعیفی ارائه کردند و مدل‌ها به طور قابل‌توجهی پیک‌ها را دست کم و در

[DOI: 10.61186/jeer.15.1.83]

مواردی بیشتر برآورد کردند. پیش بینی بیش از حد/کمتر از مقادیر روزانه SSL مدل‌های Ctree، Cforest و LASSO به وضوح از نمودارهای هیدروگراف و نمودار پراکندگی دیده می‌شود (شکل۴-d-e-f).

شکل ۴ (d-f) نشان می‌دهد که مدل‌های Cforest (R^2= ۰/۹۳۲) و Ctree (R^2= ۰/۹۱۹) مقادیر پایین و متوسط SSL را بیش از یا کمتر از حد پیش‌بینی می‌کنند، در حالی که آن‌ها به طور مداوم مقادیر بالای SSL را کمتر پیش‌بینی می‌کنند. برعکس، مدل LASSO (R^2= ۰/۹۲۳) به طور مداوم مقادیر SSL بالا و متوسط را دست کم برآورد می‌کند (شکل ۴-e). در بین مدل‌ها، مدل LASSO (RMSE= ۵۹۳/۸۸) ضعیف‌ترین پیش‌بینی‌ها را ایجاد کرد که نتایج جدول (۴) را تأیید می‌کند. این را می‌توان بیشتر در نمودارهای تغییرات زمانی و نمودارهای پراکندگی SSL مشاهده شده و پیش بینی شده در شکل۴ (a)-(f) مشاهده کرد. همچنین، انحراف قابل توجهی از خط رگرسیون برای همه مدل‌ها به سمت روش‌های بیش از حد/کم تخمین قابل تشخیص است، به جز نتایج به‌دست‌آمده با استفاده از مدل xgbTree، Cubist و qrnn، که نتایج به‌دست‌آمده با استفاده از معیارهای ارزیابی مدل ارائه شده در جدول (۴) مطابقت دارد.

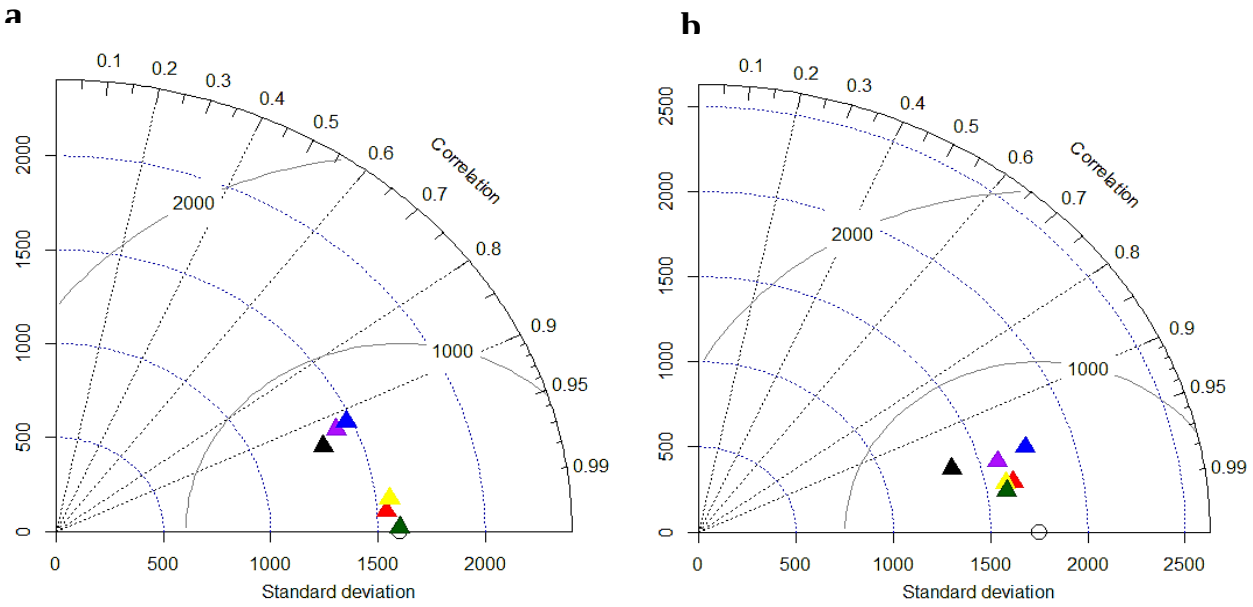
[DOI: 10.61186/jeer.15.1.83]



شکل ۴. مقادیر SSL مشاهده شده و پیش بینی شده، به دست آمده از مدل‌های ML برای مجموعه داده آزمایشی، همراه با نمودارهای پراکندگی متناظر آن‌ها در ایستگاه گلینک در حوضه آبخیز رودخانه طالقان.

[DOI: 10.61186/jeer.15.1.83]

برای تحلیل بیشتر عملکرد مدل‌ها، آن‌ها را با استفاده از نمودارهای تیلور ارزیابی کردیم (شکل ۵). نمودارهای تیلور، همانطور که توسط (Taylor، ۲۰۰۱) پیشنهاد شد، برای ارزیابی دقت مدل‌های ML بر اساس مجموعه داده‌های آموزشی (شکل ۵-a) و مجموعه داده‌های آزمایشی (شکل ۵-b) ساخته شدند. با توجه به نمودار تیلور برای مجموعه داده آموزشی، ضریب همبستگی (r) برای مدل‌های xgbTree، Cubist، qrnn، Ctree، Cforest و LASSO به ترتیب، ۱، ۱، ۰/۹۹، ۰/۹۲، ۰/۹۲ و ۰/۹۴ است. و مقادیر RMSE به ترتیب ۱۹/۷۵، ۱۲۵/۳۴، ۱۸۴/۹۶، ۶۳۴/۴۸، ۶۱۹/۹۷ و ۵۷۵/۴۷ هستند. نمودارهای تیلور (شکل ۵) برای مجموعه داده‌های آموزشی و آزمایشی نشان می‌دهد که مدل xgbTree (نزدیک به میدان مشاهده شده) نسبت به مدل‌های دیگر، دارای ضریب همبستگی بالاتر (Train=1؛ Test=0/99) و RMSE کمتر (Train=19.75؛ Test=301.63) است. جایی که xgbTree دارای انحراف استاندارد نزدیک به مشاهده شده است، در حالی که مدل‌های دیگر دارای انحراف استاندارد کمتر از مشاهده شده هستند. سپس مدل‌های Cubist و qrnn نسبت به مدل‌های Ctree، Cforest و LASSO تطابق بهتری دارند و به مقادیر مشاهده‌شده نزدیک‌تر از بقیه هستند این را می‌توان در شکل (۵) نیز مشاهده کرد. این نشان می‌دهد که مدل‌های Ctree، Cforest وLASSO نمی‌توانند به اندازه کافی نوسانات در داده‌های مشاهده‌شده را پیش‌بینی کنند. به طور خلاصه، نمودارهای تیلور (شکل ۵) عملکرد کمی بهتر مدل xgbTree و در ادامه مدل‌های Cubist و qrnn را در پیش‌بینی SSL تأیید می‌کنند، همانطور که توسط نمودارهای تغییرات زمانی و نمودارهای پراکندگی مشخص شده است. به طور کلی، مدل xgbTree دقیق‌ترین مدل است و به دنبال آن مدل‌های LASSO> Cforest >Ctree >qrnn >Cubist قرار دارد.



شکل ۵. نمودارهای تیلور برای ارزیابی عملکرد مدل‌های ML؛ a) مجموعه داده‌های آموزشی، و b) مجموعه داده‌های آزمایشی. نقاط بنفش، آبی، قرمز، سیاه، زرد و سبز به ترتیب نشان دهنده مدل‌های Cforest، Ctree، Cubist، LASSO، qrnn و xgbTree هستند.

[DOI: 10.61186/jeer.15.1.83]

۵ـ نتیجه‌گیری

این مطالعه عملکرد شش مدل یادگیری ماشین (xgbTree، Cubist، qrnn، Ctree، Cforest و LASSO) را در پیش‌بینی بار رسوب معلق (SSL) در حوضه آبخیز طالقان ایران ارزیابی کرد. یافته‌ها نشان می‌دهد که مدل‌های xgbTree، Cubist و qrnn عملکرد پیش‌بینی بهتری نسبت به مدل‌های Ctree، Cforest و LASSO در تخمین SSL روزانه دارند. این برتری می‌تواند به دلیل توانایی ذاتی این مدل‌ها در یادگیری روابط پیچیده و غیرخطی بین متغیرهای ورودی (دبی، بار رسوب معلق با تأخیر زمانی، شاخص برف و بارندگی) و SSL باشد. به طور خاص، xgbTree با استفاده از الگوریتم تقویت گرادیان، Cubist با ترکیب قوانین و مدل‌های خطی، و qrnn با تمرکز بر پیش‌بینی چندک‌ها به جای میانگین، می‌توانند الگوهای غیرخطی موجود در داده‌های SSL را بهتر از مدل‌های خطی یا مبتنی بر درخت تصمیم ساده مانند CTree، Cforest و LASSO شناسایی و مدل‌سازی کنند.

استفاده از آنالیز مؤلفه‌های اصلی (PCA) برای انتخاب متغیرهای ورودی، روشی کارآمد برای کاهش ابعاد داده‌ها و انتخاب مؤثرترین متغیرها ارائه می‌دهد. این روش در مقایسه با روش‌های آزمون و خطا، زمان و هزینه محاسباتی را کاهش می‌دهد و به شناسایی ترکیبی از متغیرها که بیشترین اطلاعات را برای پیش‌بینی SSL دارند، کمک می‌کند. با این حال، مقایسه نتایج PCA با سایر تکنیک‌های انتخاب متغیر مانند آزمون گاما و انتخاب رو به جلو (FS) می‌تواند موضوع تحقیقات آینده باشد تا بهترین روش انتخاب متغیر برای مدل‌سازی SSL مشخص شود.

با توجه به تأثیر پارامترهای مختلف آب و هوایی بر SSL، پیش‌بینی SSL در شرایط متغیر آب و هوایی از اهمیت بالایی برخوردار است. مطالعات آینده باید بر پیش‌بینی SSL تحت سناریوهای مختلف تغییر اقلیم و بررسی تأثیر متغیرهایی مانند شدت بارش بر SSL تمرکز کنند. گسترش مدل برای پیش‌بینی SSL ماهانه و مبتنی بر رویداد و همچنین استفاده از داده‌های ماهواره‌ای به عنوان اطلاعات مکمل، به ویژه در مناطقی با داده‌های زمینی محدود، می‌تواند دقت پیش‌بینی را بهبود بخشد. این یافته‌ها مبنایی برای توسعه مدل‌های پیش‌بینی SSL دقیق‌تر و مدیریت بهینه منابع آب در حوضه‌های آبخیز مشابه ارائه می‌دهد.

۶ـ قدردانی و تشکر

این مقاله خروجی پروژه پسادکتری با عنوان " ارایه یک مدل پیش بینی بار رسوب معلق رودخانه بر پایه الگوریتم‌های یادگیری (مطالعه موردی: آبخیز طالقان) " با شماره ۴۰۰۴۹۸۶ تحت حمایت بنیاد ملی علم ایران (INSF) می باشد.

منابع:

1. Abrahart, R.J., White, S.M., 2001. Modeling sediment transfer in Malawi: comparing backpropagation neural network solutions against a multiple linear regression benchmark using small data sets. Physics and Chemistry of the Earth B 26 (1), 19–24.
2. Alp, M., Cigizoglu, H.K., 2007. Suspended sediment estimation by feed forward back propagation method using hydro meteorological data. Environmental Modelling & Software 22 (1), 2–13.
3. Bou-Fakhreddine, B., Mougharbel, I., Faye, A., Abou Chakra, S., Pollet, Y., 2018. Daily river flow prediction based on Two-Phase Constructive Fuzzy Systems Modeling: A case of hydrological – meteorological measurements asymmetry. J. Hydrol. 558, 255–265. https://doi.org/10.1016/j.jhydrol.2018.01.035.

[DOI: 10.61186/jeer.15.1.83]

4. Chiang, J.-L., Tsai, Y.-S., (2011). Suspended sediment load estimate using support vector machines in Kaoping river basin, in: Consumer Electronics, Communications and Networks (CECNet), 2011 International Conference On. pp. 1750–1753.
5. Choubin, Bahram, Darabi, Hamid, Rahmati, Omid, Sajedi-Hosseini, Farzaneh, & Kløve, Bjørn. (2018). River suspended sediment modelling using the CART model: A comparative study of machine learning techniques. Science of The Total Environment, 615, 272-281. doi: https://doi.org/10.1016/j.scitotenv.2017.09.293.
6. Çimen, M., 2016. Estimation of daily suspended sediments using support vector machines. Hydrol. Sci. J. 6667.
7. Cobaner, M., Unal, B., Kisi, O., (2009). Suspended sediment concentration estimation by an adaptive neuro-fuzzy and neural network approaches using hydrometeorological data. J. Hydrol. 367, 52–61.
8. Darabi, H., Mohamadi, S., Karimidastenaei, S., Kisi, S., Ehteram, M., ELShafie, A., Torabi Haghighi, A., 2021. Prediction of daily suspended sediment load (SSL) using new optimization algorithms and soft computing models. Soft Computing (2021) 25:7609–7626, https://doi.org/10.1007/s00500-021-05721-5.
9. Dawson, Christian W, Abrahart, Robert J, & See, Linda M. (2007). HydroTest: a web-based toolbox of evaluation metrics for the standardised assessment of hydrological forecasts. Environmental Modelling & Software, 22(7), 1034-1052.
10. Diop, L., Bodian, A., Djaman, K., Yaseen, Z.M., Deo, R.C., El-shafie, A., Brown, L.C., (2018). The influence of climatic inputs on stream-flow pattern forecasting: case study of Upper Senegal River. Environ. Earth Sci. 77, 182. https://doi.org/10.1007/s12665-018-7376-8.
11. Gholami, Hamid, Mohammadifar, Aliakbar, Golzari, Shahram, Kaskaoutis, Dimitris G, & Collins, Adrian L. (2021 b). Using the Boruta algorithm and deep learning models for mapping land susceptibility to atmospheric dust emissions in Iran. Aeolian Research, 50, 100682.
12. Ghorbani, M.A., Deo, R.C., Yaseen, Z.M., H. Kashani, M., Mohammadi, B., 2017. Pan evaporation prediction using a hybrid multilayer perceptron-firefly algorithm (MLP-FFA) model: case study in North Iran. Theor. Appl. Climatol. https://doi.org/10.1007/s00704- 017-2244-0
13. Haghighi AT, Darabi H, Shahedi K, Solaimani K, Kløve B (2019) A scenario-based approach for assessing the hydrological impacts of land use and climate change in the Marboreh Watershed. Iran. Environ Model Assess 25:1–17.
14. Hassanpour, F., et al., 2019. Development of the FCM-SVR hybrid model for estimating the suspended sediment load. KSCE Journal of Civil Engineering, 23 (6), 2514–2523. doi:10.1007/s12205-019-1693-7.
15. Hosseini, Majid, Ghafouri, A., Amin, M., Tabatabaei, Mohsen, Goodarzi, Massoud, & Abdeh Kolahchi, Abdolnabi. (2012). Effects of Land Use Changes on Water Balance in Taleghan Catchment, Iran. Journal of Agricultural Science and Technology, 14, 1159-1172.
16. Hothorn, T., Hornik, K., & Zeileis, A. (2015). ctree: Conditional inference trees. The comprehensive R archive network, 8, 1-34.
17. Houborg, R., & McCabe, M. F. (2018). A hybrid training approach for leaf area index estimation via Cubist and random forests machine-learning. ISPRS Journal of Photogrammetry and Remote Sensing, 135, 173-188.
18. Idrees, M. B., Jehanzaib, M., Kim, D., Kim, T.D., 2021. Comprehensive evaluation of machine learning models for suspended sediment load inflow prediction in a reservoir. Stochastic Environmental Research and Risk Assessment

[DOI: 10.61186/jeer.15.1.83]

https://doi.org/10.1007/s00477-021-01982-6(0123456789().,-volV)(0123456789,-().volV).
19. Khan, A.I., Topping, B.H.V., Bahreininejad, A., 1993. Parallel training of neural networks for finite element mesh generation. In: Topping, B.H.V., Khan, A.I. (Eds.), Neural Networks and Combinatorial Optimization in Civil&Structural Engineering. Civil-Comp Press, Edinburgh, pp. 81–94.
20. Khosravi, Khabat, Mao, Luca, Kisi, Ozgur, Yaseen, Zaher Mundher, & Shahid, Shamsuddin. (2018). Quantifying hourly suspended sediment load using data mining models: Case study of a glacierized Andean catchment in Chile. Journal of Hydrology, 567, 165-179. doi: https://doi.org/10.1016/j.jhydrol.2018.10.015.
21. Kisi, Ozgur, Dailr, Ali Hosseinzadeh, Cimen, Mesut, & Shiri, Jalal. (2012). Suspended sediment modeling using genetic programming and soft computing techniques. Journal of Hydrology, 450-451, 48-58. doi: https://doi.org/10.1016/j.jhydrol.2012.05.031.
22. Lafdani, E Kakaei, Nia, A Moghaddam, & Ahmadi, A. (2013). Daily suspended sediment load prediction using artificial neural networks and support vector machines. Journal of Hydrology, 478, 50-62.
23. Li, Y., Wang, Z., Han, R., Shi, S., Li, J., Shang, R., ... & Gu, Y. (2023). Quantum recurrent neural networks for sequential learning. Neural Networks, 166, 148-161.
24. Liu, Qian-Jin, Shi, Zhi-Hua, Fang, Nu-Fang, Zhu, Hua-De, & Ai, Lei. (2013). Modeling the daily suspended sediment concentration in a hyperconcentrated river on the Loess Plateau, China, using the Wavelet–ANN approach. Geomorphology, 186, 181-190. doi: https://doi.org/10.1016/j.geomorph.2013.01.012.
25. Lu H, Meng Y, Yan K, Gao Z (2019) Kernel principal component analysis combining rotation forest method for linearly inseparable data. Cogn Syst Res 53:111–122
26. Luo, Y., Xue, Y., Liu, W., Song, H., Huang, Y., Tang, G., ... & Sun, Z. (2022). Development of diagnostic algorithm using machine learning for distinguishing between active tuberculosis and latent tuberculosis infection. *BMC Infectious Diseases*, *22*(1), 965.
27. Ma, J., Yu, Z., Qu, Y., Xu, J., & Cao, Y. (2020). Application of the XGBoost machine learning method in PM2. 5 prediction: a case study of Shanghai. *Aerosol and Air Quality Research*, *20*(1), 128-138.
28. Masters, Timothy. (1993). Practical neural network recipes in C++: Morgan Kaufmann.
29. Mehri, Y., Nasrabadi, M., Omid, M.H., (2021). Prediction of suspended sediment distributions using data mining algorithms. Ain Shams Engineering Journal, https://doi.org/10.1016/j.asej.2021.02.034.
30. Melesse, A. M., Ahmad, S., McClain, M. E., Wang, X., & Lim, Y. H. (2011). Suspended sediment load prediction of river systems: An artificial neural network approach. Agricultural Water Management, 98(5), 855-866. doi: https://doi.org/10.1016/j.agwat.2010.12.012.
31. Melesse, A., Jayachandran, K., Zhang, K., 2008. Modeling coastal eutrophication at Florida Bay using neural networks. Journal of Coastal Research 24, 190–196.
32. MixSIR model. CATENA 164, 32–43.
33. Muthukrishnan, R., & Rohini, R. (2016, October). LASSO: A feature selection technique in predictive modeling for machine learning. In *2016 IEEE international conference on advances in computer applications (ICACA)* (pp. 18-20). Ieee.
34. Nadiri, A.A., Gharekhani, M., Khatibi, R., Sadeghfam, S., Moghaddam, A.A., (2017). Groundwater vulnerability indices conditioned by Supervised Intelligence Committee Machine (SICM). Sci. Total Environ. 574, 691–706. https://doi.org/10.1016/j.scitotenv.2016.09.093.

[DOI: 10.61186/jeer.15.1.83]

35. Nagy, H.M., Watanabe, K., Hirano, M., 2002. Prediction of sediment load concentration in rivers using artificial neural network model. Journal of Hydraulic Engineering 128 (6), 588–595.
36. Noori, R., Sabahi, M.S., Karbassi, A.R., Baghvand, A., Taati Zadeh, H., 2010d. Multivariate statistical analysis of surface water quality based on correlations and variations in the data set. Desalination 260, 129–136.
37. Nosrati, K., Collins, A.L., Madankan, M., 2018a. Fingerprinting sub-basin spatial sediment sources using different multivariate statistical techniques and the Modified
38. Nosrati, Kazem, Mohammadi-Raigani, Zeinab, Haddadchi, Arman, & Collins, Adrian L. (2021). Elucidating intra-storm variations in suspended sediment sources using a Bayesian fingerprinting approach. Journal of Hydrology, 596, 126115. doi: https://doi.org/10.1016/j.jhydrol.2021.126115.
39. Pourghasemi, H.R., Yousefi, S., Kornejady, A., Cerda, A., (2017). Performance assessment of individual and ensemble data-mining techniques for gully erosion modeling. Sci. Total Environ. 609, 764–775.
40. Rahul, Atul Kumar, Shivhare, Nikita, Kumar, Shashi, Dwivedi, Sumita, & Dikshit, P. K. S. (2021). Modelling of Daily Suspended Sediment Concentration Using FFBPNN and SVM Algorithms.
41. Raigani, Zeinab Mohammadi, Nosrati, Kazem, & Collins, Adrian L. (2019). Fingerprinting sub-basin spatial sediment sources in a large Iranian catchment under dry-land cultivation and rangeland farming: Combining geochemical tracers and weathering indices. Journal of Hydrology: Regional Studies, 24, 100613. doi: https://doi.org/10.1016/j.ejrh.2019.100613.
42. Rajaee, T., Nourani, V., Zounemat-Kermani, M., Kisi, O., (2011). River suspended sediment load prediction: application of ANN and wavelet conjunction model. J. Hydrol. Eng. 16 (8), 613–627.
43. Ren J, Zhao M, Zhang W, Xu Q, Yuan J, Dong B (2020) Impact of the construction of cascade reservoirs on suspended sediment peak transport variation during flood events in the Three Gorges Reservoir. CATENA 188:104409.
44. Shojaeezadeh SA, Nikoo MR, McNamara JP, AghaKouchak A, Sadegh M (2018). Stochastic modeling of suspended sediment load in alluvial rivers. Adv Water Resour 119:188–196.
45. Solomatine, D.P., Torres, L.A., 1996. Neural network approximation of a hydrodynamicmodel. In: Optimizing Reservoir Operation, Proceedings of the Second International Conference on Hydroinformatics , Zurich, pp. 201–206.
46. Tao, H., Diop, L., Bodian, A., Djaman, K., Ndiaye, P.M., Yaseen, Z.M., 2018. Reference evapotranspiration prediction using hybridized fuzzy model with firefly algorithm: Regional case study in Burkina Faso. Agric. Water Manag.
47. Taylor, Karl E. (2001). Summarizing multiple aspects of model performance in a single diagram. Journal of Geophysical Research: Atmospheres, 106(D7), 7183-7192. doi: https://doi.org/10.1029/2000JD900719
48. Tokar, A. Sezin, & Johnson, Peggy A. (1999). Rainfall-Runoff Modeling Using Artificial Neural Networks. Journal of Hydrologic Engineering, 4(3), 232-239. doi: doi:10.1061/(ASCE)1084-0699(1999)4:3(232)
49. Wen, C.G., Lee, C.S., (1998). A neural network approach to multi-objective optimization for water quality management in a river basin. Water Resources Research 34 (3), 427–436.
50. Yang CT, Marsooli R, Aalami MT (2009). Evaluation of total load sediment transport formulas using ANN. Int J Sedim Res 24(3):274–286.

[DOI: 10.61186/jeer.15.1.83]

51. Yang, C.C., Prasher, S.O., Tan, C.S., 1998. An artificial neural network model for water table management systems. In: Drainage in the 21st Century: Food Production and the Environment. Proceedings of the Seventh International Drainage Symposium, Orlando, Florida, USA, 8–10 March 1998, ASAE, St. Joseph, MI.
52. Yu, H., Wen, X., Feng, Q., Deo, R.C., Si, J., Wu, M., 2018. Comparative Study of Hybrid- Wavelet Artificial Intelligence Models for Monthly Groundwater Depth Forecasting in Extreme Arid Regions, Northwest China. Water Resour. Manag. https://doi.org/10.1007/s11269-017-1811-6.

[DOI: 10.61186/jeer.15.1.83]