

Research Paper

Application of Feature Selection Methods in Predicting Phosphate Adsorption from Aqueous Solutions Using Machine Learning Algorithms

Laleh Divband Hafshejani^{1*}, Hamid Abdolabadi², Abd Ali Naseri³

¹ Department of Environmental Engineering, Faculty of Water and Environmental Engineering, Shahid Chamran University of Ahvaz, Ahvaz, Iran

² Department of Environmental Engineering, Faculty of Water and Environmental Engineering, Shahid Chamran University of Ahvaz, Ahvaz, Iran

³ Department of Irrigation and Drainage Engineering, Faculty of Water and Environmental Engineering, Shahid Chamran University of Ahvaz, Ahvaz, Iran



10.22125/iwe.2025.518646.1876.

Received:
April 23, 2025
Accepted:
November 3, 2025
Available online:
December 25, 2025

Keywords:
Hydrochar, Contact time, Initial concentration, Algorithm

Abstract

Optimization using artificial intelligence methods is an effective approach to improving the performance of systems and processes. These methods enable the identification of more effective parameters and their optimization to enhance phosphate adsorption efficiency.

The present study focused on developing predictive machine learning algorithms with a dimensionality reduction approach. To develop the predictive model, experimental data on phosphate adsorption by sugarcane bagasse hydrochar were obtained through laboratory-scale adsorption experiments. Five independent input variables, including initial pollutant concentration, contact time, adsorbent mass, solution temperature, and pH, were considered in the training process. Additionally, phosphate adsorption efficiency was considered as the output. Among the applied algorithms, the Extra Trees Regressor (ET) demonstrated relatively better performance in predicting phosphate adsorption efficiency, with an R^2 value of 0.922, as well as lower RMSE (0.074) and MAE (0.048) values. Based on the results, the two input factors with the greatest impact on phosphate adsorption effectiveness were contact time and initial phosphate concentration. Furthermore, the adsorbent mass was identified as the parameter with the least impact.

1. Introduction

Phosphorus (P) is an essential nutrient for all living organisms. However, rapid industrialization and population growth have led to excessive phosphorus discharge through human-related activities such as phosphate rock mining, fertilizer application, animal production, food processing, and domestic waste generation (Zhu and Ma, 2020). The accumulation of phosphorus in soils and sediments contributes to internal pollution, while excessive concentrations in surface waters cause eutrophication, algal blooms, oxygen depletion, and aquatic ecosystem degradation. These processes reduce water transparency, alter species composition, and threaten water supply and recreational uses.

* **Corresponding Author:** Laleh Divband Hafshejani

Address: Department of Environmental Engineering, Faculty of Water and Environmental Engineering, Shahid Chamran University of Ahvaz, Ahvaz, Iran. **Email:** l.divband@scu.ac.ir **Tel:** +989132387133

Various physicochemical methods have been employed for phosphorus removal from wastewater, including electro dialysis, reverse osmosis, nanofiltration, ultrafiltration, and adsorption. Among these, adsorption stands out as a proven and economical technique due to its operational simplicity, high efficiency, and reusability of materials (Crini, 2008; Divband Hafshejani et al., 2022). Recent studies have focused on developing innovative adsorbents with superior performance for pollutant removal (Hafshejani et al., 2016; Rahdar et al., 2021). However, adsorption efficiency depends on several factors such as adsorbent type, dosage, solution pH, temperature, contact time, and initial pollutant concentration (Almanassra et al., 2021). Machine learning (ML) approaches have recently emerged as promising tools to overcome the limitations of traditional adsorption models by enabling accurate prediction and optimization of adsorption performance under diverse operational conditions (Awolusi et al., 2019; Wu et al., 2023). ML models can extract complex nonlinear relationships between adsorbent properties and pollutant removal efficiency, reducing the need for laborious experiments (Fu et al., 2024). Moreover, feature selection—a critical step in ML preprocessing—enhances model accuracy and interpretability by identifying the most relevant input parameters (Büyükkеçeci and Okur, 2022; El Touati et al., 2024). Despite growing interest in ML-based adsorption modeling, limited research has focused on sustainable adsorbents such as sugarcane bagasse hydrochar. Therefore, this study aims to develop an optimized phosphate adsorption model using feature selection techniques integrated with advanced ML algorithms, offering a novel approach for improving adsorption performance.

2. Materials and Methods

In this study, the adsorption of phosphate from aqueous solutions using sugarcane bagasse biochar was investigated through data-driven modeling and optimization. Experimental data were collected from previous studies, focusing on key operational parameters such as initial solution pH, contact time, process temperature, initial phosphate concentration, and adsorbent dosage. These parameters were selected due to their known influence on adsorption efficiency and water quality improvement.

After data collection, the dataset was divided into training and testing subsets to enable proper evaluation of the predictive models. All variables were standardized to ensure that differences in scale did not bias the model performance. The relationships between input variables and the target output were analyzed using statistical correlation measures to identify the most influential parameters in the adsorption process.

Several machine learning algorithms were applied to simulate and predict phosphate adsorption, including Decision Tree, Random Forest, and Extreme Gradient Boosting (XGBoost). These models were chosen for their ability to capture complex nonlinear relationships and provide reliable predictions without requiring explicit mathematical formulas. Model hyperparameters were tuned through iterative training to maximize prediction accuracy while avoiding overfitting.

Feature selection techniques were applied to identify the most significant variables affecting phosphate adsorption. By reducing the number of input variables, model performance improved in terms of both accuracy and computational efficiency. The final models were trained using the selected features, and their performance was evaluated using metrics such as R^2 , mean absolute error (MAE), and root mean squared error (RMSE).

The workflow of this study included data preparation, model training, feature importance analysis, model retraining with selected features, and final performance assessment. This approach ensured that the predictive models not only provided accurate phosphate adsorption predictions but also highlighted the critical factors driving adsorption efficiency.

3. Results

Descriptive statistics showed that phosphate removal ranged from 1.7% to 93.3%, with a mean of 55.9%, indicating moderate efficiency under typical conditions. Contact time and pH displayed wide variations, allowing for comprehensive analysis of adsorption behavior. Correlation analysis revealed that initial phosphate concentration had the strongest negative correlation with removal efficiency ($r = -0.8$), confirming its dominant role. Conversely, contact time ($r = 0.6$) and temperature ($r = 0.4$) showed positive associations with removal rate, whereas pH and adsorbent dosage exhibited weaker effects.

Feature selection results using the three approaches consistently ranked contact time and initial concentration as the most influential variables, followed by temperature and pH, while adsorbent dosage had negligible influence. These findings align with previous studies indicating that phosphate uptake is controlled primarily by concentration gradient and reaction kinetics (Ye et al., 2016; Nouaa et al., 2024).

4. Discussion and Conclusion

The integration of feature selection techniques significantly enhanced the interpretability and efficiency of the machine learning models. Among the tested algorithms, the Extra Trees Regressor (ET) achieved the highest predictive accuracy ($R^2 = 0.922$, RMSE = 0.074, MAE = 0.048), outperforming both XGBoost and Decision Tree models. The superior performance of ET is attributed to its ability to capture nonlinear relationships through ensemble randomization while mitigating overfitting.

The analysis confirmed that initial phosphate concentration and contact time are the key factors controlling adsorption efficiency, while temperature exerts a secondary positive influence, suggesting an endothermic process. The weak impact of pH and adsorbent dosage within the studied range highlights the stability of the hydrochar surface chemistry and its moderate sensitivity to environmental changes.

Overall, the study demonstrates that coupling feature selection with ensemble ML algorithms provides an effective framework for predicting and optimizing adsorption processes. This data-driven approach reduces the need for extensive laboratory experiments and supports the rational design of sustainable adsorbents for nutrient recovery and water purification. Future research should focus on expanding datasets with real wastewater samples and exploring hybrid models integrating ML with adsorption isotherm and kinetic theories for deeper mechanistic insight.

5. Six important references

- 1) Büyükkeçeci, S., & Okur, A. (2022). Feature selection methods in ML-based environmental modeling. *Applied Soft Computing*.
- 2) Crini, G. (2008). Recent developments in polysaccharide-based materials for adsorption of water contaminants. *Progress in Polymer Science*.
- 3) Divband Hafshejani, L., Naseri, A., et al. (2022). Removal of phosphate from aqueous solutions using hydrochar derived from sugarcane bagasse. *Journal of Environmental Chemical Engineering*.
- 4) Fu, J., Feng, L., et al. (2024). Advances in ML models for adsorption performance prediction of sustainable materials. *Chemical Engineering Journal*.
- 5) Wu, Y., Li, J., et al. (2023). Machine learning-driven prediction of phosphate adsorption capacity of biochar. *Journal of Environmental Management*.
- 6) Zhu, Y., & Ma, L. (2020). Phosphorus flows and recycling in global agricultural systems. *Environmental Science & Technology*.

Conflict of Interest

Authors declared no conflict of interest.

کاربرد روش‌های انتخاب ویژگی در پیش‌بینی جذب فسفات از محلول‌های آبی توسط الگوریتم‌های یادگیری ماشین

لاله دیوبند هفشجانی^{۱*}، حمید عبدل آبادی^۲، عبدعلی ناصری^۳

تاریخ ارسال: ۱۴۰۴/۰۲/۰۳

تاریخ پذیرش: ۱۴۰۴/۰۸/۱۲

مقاله پژوهشی

چکیده

بهینه‌سازی با استفاده از روش‌های هوش مصنوعی، یک رویکرد موثر برای بهبود عملکرد سامانه‌ها و فرایندها است. این روش‌ها امکان یافتن پارامترهای مؤثرتر و برنامه‌ریزی بر روی آن‌ها جهت ارتقای راندمان جذب فسفات را فراهم می‌آورند. پژوهش حاضر بر توسعه الگوریتم‌های یادگیری ماشین پیش‌بینی‌کننده با رویکرد کاهش ابعاد متمرکز بود. به‌منظور توسعه مدل پیش‌بینی، داده‌های تجربی جذب فسفات به‌وسیله هیدروچار باگاس نیشکر از طریق تنظیم جذب در مقیاس آزمایشگاهی به‌دست آمد. پنج متغیر ورودی مستقل، شامل غلظت اولیه آلاینده، زمان تماس، جرم جاذب، دمای محلول و pH، در فرآیند آموزش در نظر گرفته شدند. علاوه بر این، راندمان جذب فسفات به عنوان خروجی در نظر گرفته شد. از میان الگوریتم‌های به کار گرفته شده، رگرسیون درختان اضافی (ET) با R^2 برابر با ۰/۹۲۲ و همچنین با مقادیر پایین‌تر RMSE (۰/۰۷۴) و MAE (۰/۰۴۸) عملکرد نسبتاً بهتری در پیش‌بینی راندمان جذب فسفات ارائه کرد. بر اساس نتایج، دو عامل ورودی که بیشترین تأثیر را بر اثربخشی جذب فسفات دارند زمان تماس و غلظت اولیه فسفات هستند. علاوه بر این، مقدار جاذب به عنوان پارامتری با کمترین تأثیر شناخته شد.

واژه‌های کلیدی: هیدروچار؛ زمان تماس؛ غلظت اولیه؛ الگوریتم.

^۱ گروه مهندسی محیط زیست، دانشکده مهندسی آب و محیط زیست، دانشگاه شهید چمران اهواز، اهواز، ایران (*نویسنده مسئول):

(l.divband@scu.ac.ir)

^۲ گروه مهندسی محیط زیست، دانشکده مهندسی آب و محیط زیست، دانشگاه شهید چمران اهواز، اهواز، ایران. ایمیل: h.abdolabadi@scu.ac.ir

^۳ گروه آبیاری و زهکشی، دانشکده مهندسی آب و محیط زیست، دانشگاه شهید چمران اهواز، اهواز، ایران. ایمیل: abdalinaseri@scu.ac.ir



مقدمه

فسفر یکی از مواد مغذی ضروری برای حیات است. با پیشرفت سریع صنعت و افزایش جمعیت جهانی، مقدار قابل توجهی فسفر در فرآیندهای انسانی مانند استخراج سنگ فسفات، استفاده از کودها، تولید محصولات دامی و غذایی، مصرف انسانی و تولید فضولات آزاد می‌شود (Zhu & Ma, 2020). این فسفر آزاد شده می‌تواند در خاک و رسوبات انباشته شود و منجر به تجمع بیش از حد فسفر و آلودگی داخلی گردد، در حالی که مقادیر اضافی فسفر ممکن است در آب‌های سطحی باقی بماند و به اتروفیکاسیون و کاهش کیفیت آب منجر شود. اتروفیکاسیون به معنای ورود بیش از حد مواد مغذی به یک سیستم آبی است که معمولاً تغییرات نامطلوبی در اکوسیستم‌های آبی ایجاد می‌کند، از جمله شکوفه‌های سمی جلبکی، کاهش شفافیت آب، کاهش اکسیژن ناشی از تجزیه مواد آلی، تغییر در ترکیب گونه‌ها و افزایش مرگ و میر ماهیان. اتروفیکاسیون همچنین می‌تواند خدمات آبی مورد نیاز انسان مانند تامین آب آشامیدنی و استفاده‌های تفریحی از منابع آبی را به خطر بیندازد.

روش‌های مختلفی از جمله الکترودیالیز، اسمز معکوس، نانوفیلتراسیون، اولترافیلتراسیون و جذب سطحی برای حذف فسفر از محلول‌های آبی وجود دارد. جذب سطحی به طور خاص یک فناوری اثبات شده و شناخته شده برای تصفیه آب به دلیل مزایای فنی و اقتصادی آن است (Crini, 2008; Divband Hafshejani et al., 2022). پیشرفت‌های اخیر در زمینه جذب برای حذف آلاینده‌های مختلف عمدتاً بر تهیه و ارزیابی مواد جدید با ظرفیت‌های جذب برجسته متمرکز شده است (Hafshejani et al., 2021; Rahdar et al., 2016). اثربخشی جذب فسفات به عوامل متعددی از جمله نوع ماده جاذب مورد استفاده، میزان ماده مورد استفاده، دمای آب، زمان تماس، pH آب، غلظت فسفات و غیره... بستگی دارد (Almanassra et al., 2021).

پیش‌بینی جذب آلاینده‌ها از آب با استفاده از جاذب‌های مختلف باعث می‌شود که روش‌های تصفیه بهینه‌تر و با کارایی

بالاتری اجرا شوند. این امر باعث می‌شود تا کیفیت آب تصفیه شده بهبود یابد و آلاینده‌ها به طور موثرتری حذف شوند. در سال‌های اخیر، روش‌های هوش مصنوعی این امکان را فراهم کرده‌اند که برخی از محدودیت‌های مدل‌های جذب سنتی، به‌ویژه در زمینه پیش‌بینی‌های دقیق‌تر تحت شرایط عملیاتی مختلف، برطرف شوند. این روش‌ها به دلیل توانایی در تحلیل و درک تعاملات پیچیده، به عنوان رویکردی موثر برای مدل‌سازی و بهینه‌سازی تجربی شناخته شده‌اند (Awolusi et al., 2019). از جمله روش‌های هوش مصنوعی که به عنوان یک ابزار قدرتمند برای پیش‌بینی جذب آلاینده‌ها توسط جاذب‌های مختلف در حال ظهور است، یادگیری ماشین است (Fu et al., 2024; Wu et al., 2023). این رویکرد از داده‌های تجربی موجود برای شناسایی الگوها و توسعه مدل‌هایی استفاده می‌کند که می‌توانند کارایی جذب جاذب‌های جدید را برای آلاینده‌های مختلف پیش‌بینی کنند (Wu et al., 2023; Zhang et al., 2023). مدل‌های یادگیری ماشین می‌توانند روابط پیچیده‌ای بین خواص جاذب و جذب آلاینده را ثبت کنند، که به طور بالقوه منجر به پیش‌بینی‌های دقیق‌تر نسبت به روش‌های سنتی می‌شود (Fu et al., 2024). با استفاده از یادگیری ماشین برای پیش‌بینی جذب، محققان می‌توانند به طور بالقوه نیاز به آزمایش‌های آزمایشگاهی زمان‌بر و پرهزینه را کاهش دهند (Fu et al., 2024). مدل‌های یادگیری ماشینی را می‌توان برای شناسایی پارامترهای طراحی بهینه برای جاذب‌های جدید با قابلیت‌های جذب آلاینده افزایش یافته استفاده کرد (Zhang et al., 2023). به طور کلی، یادگیری ماشین یک رویکرد امیدوارکننده برای پیشرفت زمینه جذب آلاینده‌ها و توسعه مواد جاذب موثرتر ارائه می‌دهد. در این راستا، انتخاب ویژگی‌ها به عنوان یکی از مراحل کلیدی پیش‌پردازش داده‌ها به شمار می‌رود. انتخاب ویژگی با شناسایی مرتبط‌ترین ویژگی‌ها و حذف موارد اضافی یا نامربوط، دقت و کارایی مدل‌های یادگیری ماشین را افزایش می‌دهد. این منجر به تعمیم بهتر و کاهش بیش از حد برآزش می‌شود (Büyükkeçeci & Okur, 2022). با کاهش تعداد ویژگی‌ها، هزینه محاسباتی مورد نیاز

غلظت اولیه فسفات (۵۰ تا ۱۰۰ میلی گرم بر لیتر) و مقدار جاذب مورد استفاده (۲ تا ۱۰ گرم) بوده‌اند. راندمان جذب فسفات به عنوان پارامتر خروجی و هدف اصلی مدل سازی انتخاب شده است. این داده‌ها مستقیماً از مطالعه انجام شده توسط دیوبند هفشجانی و همکاران استخراج شده‌اند (Divband Hafshejani et al., 2022) که به عنوان مبنایی برای تحلیل و توسعه مدل‌های پیش‌بینی کننده در این پژوهش مورد استفاده قرار گرفته‌اند.

$$R = \frac{C_0 - C_f}{C_0} \times 100 \quad (1)$$

در روابط بالا، C_0 : غلظت اولیه فسفات (mg/l)، C_f : غلظت نهایی فسفات (mg/l)، m : جرم جاذب (g) و R : راندمان جذب فسفات (%) می‌باشد.

سپس داده‌ها ۷۰ درصد داده‌ها برای آموزش و ۳۰ درصد برای تست به کار برده شدند. تمامی داده‌ها (ورودی و خروجی) با استفاده از رابطه (۲) در محدوده ۰ و ۱ نرمال سازی شدند:

$$x_i^* = \frac{(x_i - x_{min})}{(x_{max} - x_{min})} \quad (2)$$

در رابطه (۱)، x_i^* مقدار نرمال شده متغیر اولیه x_i و x_{min} و x_{max} به ترتیب حداقل و حداکثر مقادیر x_i هستند.

همبستگی بین متغیرهای ورودی و متغیر هدف با استفاده از ضریب همبستگی پیرسون (رابطه ۳) اندازه‌گیری شد.

$$r_{xy} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}} \quad (3)$$

در رابطه بالا \bar{x} و \bar{y} میانگین متغیرهای x و y می‌باشد.

الگوریتم‌های هوش مصنوعی

الگوریتم تقویت گرادیان (XGBoost) مخفف Extreme Gradient Boosting است و یک الگوریتم کارآمد و قدرتمند است که برای رگرسیون و طبقه بندی استفاده می‌شود. این الگوریتم بر روی درختان تصمیم ساخته شده است و از فنی به نام تقویت گرادیان استفاده می‌کند که با افزودن درخت‌های تصمیم جدید که خطاهای قبلی را اصلاح می‌کند، عملکرد مدل را بهبود می‌بخشد. الگوریتم تقویت گرادیان به دلیل سرعت، مقیاس پذیری، و

برای برداشت داده، آموزش و پیش‌بینی مدل کاهش می‌یابد و علاوه بر این مدل‌هایی با ویژگی‌های کمتر و مرتبط‌تر قابل تفسیر هستند (El Touati et al., 2024; Suresh et al., 2022).

روش‌های متعددی برای انتخاب ویژگی‌ها وجود دارد که هر یک دارای مزایا و معایب خاص خود هستند. از جمله این روش‌ها می‌توان به انتخاب ویژگی مبتنی بر فیلتر و انتخاب ویژگی مبتنی بر پوششی اشاره کرد. با استفاده از این فن‌ها، می‌توان ویژگی‌هایی را که بیشترین تأثیر را بر مدل دارند شناسایی کرده و ویژگی‌های کم‌اهمیت‌تر را حذف نمود. این اقدام به بهبود دقت و کارایی مدل‌های یادگیری ماشین در زمینه جذب آلاینده‌ها کمک می‌کند. مرور ادبیات نشان می‌دهد که یادگیری ماشین به طور فزاینده‌ای در مدل سازی فرآیندهای جذب آلاینده‌ها مورد توجه قرار گرفته است و توانایی آن در تحلیل داده‌های پیچیده و پیش‌بینی عملکرد جاذب‌ها، مزیت قابل توجهی نسبت به روش‌های سنتی دارد. با این حال، بیشتر مطالعات موجود بر استفاده از جاذب‌های متداول مانند کربن فعال یا زئولیت متمرکز بوده‌اند و کمتر به مواد جاذب نوین و پایدار مانند هیدروچار باگاس نیشکر پرداخته‌اند. همچنین، استفاده ترکیبی از روش‌های انتخاب ویژگی و الگوریتم‌های یادگیری ماشین برای بهینه‌سازی جذب فسفات با این ماده جاذب تاکنون گزارش نشده است. این شکاف تحقیقاتی، ضرورت مطالعه حاضر را برجسته می‌کند که با بهره‌گیری از هیدروچار باگاس نیشکر و ترکیب روش‌های انتخاب ویژگی و یادگیری ماشین، رویکردی نوین برای بهینه‌سازی فرآیند جذب فسفات ارائه می‌دهد.

مواد و روش‌ها

آماده سازی داده‌ها

در این مطالعه، برای شبیه‌سازی و بهینه‌سازی فرآیند جذب فسفات با استفاده از هیدروچار باگاس نیشکر، از الگوریتم‌های پیشرفته یادگیری ماشین بهره گرفته شده است. داده‌های مورد استفاده شامل متغیرهای ورودی کلیدی نظیر pH اولیه محلول (۲ تا ۱۱)، زمان تماس (۵ تا ۲۴۰ دقیقه)، دمای فرآیند (۲۰ تا ۴۰ درجه سانتی‌گراد)،



کاهش ناهمگونی (impurity decrease) یا میزان مشارکت در کاهش خطای پیش‌بینی محاسبه می‌شود. این روش برای داده‌های با روابط غیرخطی و پیچیده بسیار مناسب است.

اطلاعات متقابل: این معیار آماری میزان وابستگی میان یک ویژگی و متغیر هدف را اندازه‌گیری می‌کند، حتی اگر این رابطه غیرخطی باشد. ویژگی‌هایی که اطلاعات بیشتری درباره متغیر هدف ارائه می‌دهند، اهمیت بالاتری دارند. این روش به‌ویژه در شناسایی ویژگی‌های کلیدی برای مدل‌های یادگیری ماشین مفید است.

اهمیت جایگشتی: در این روش، مقادیر هر ویژگی به‌طور تصادفی جابه‌جا می‌شوند و تغییرات در دقت مدل اندازه‌گیری می‌شود. هر چه کاهش دقت بیشتر باشد، اهمیت آن ویژگی بالاتر است. این روش برای مدل‌های پیچیده و غیرخطی کاربرد دارد و وابستگی کمی به مقیاس داده‌ها دارد.

با ترکیب نتایج این سه روش، می‌توان وزن نسبی و نقش هر ویژگی در پیش‌بینی متغیر هدف را به‌صورت جامع و معتبر مشخص کرد و مقایسه‌ای میان اهمیت ویژگی‌ها ارائه داد.

معیارهای ارزیابی

در پژوهش حاضر از دو مجموعه داده برای پیش‌بینی جذب استفاده گردید. در مرحله اول از همه پارامترها (pH زمان تماس، دما، غلظت اولیه فسفات و مقدار جاذب) به عنوان ورودی استفاده شد. در مرحله بعد پیش‌بینی تنها به کمک پارامترهای موثر که از مرحله قبل تعیین شده بودند. در هر مرحله عملکرد الگوریتم‌ها با استفاده از پارامترهای آماری مانند ضریب تعیین (R^2)، ریشه میانگین مربعات خطا (RMSE) و میانگین خطای مطلق (MAE) (روابط ۴-۶) مورد ارزیابی قرار گرفت تا الگوریتم موفق برای توسعه مشخص گردد.

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i^{exp} - y_i^{pred})^2}{\sum_{i=1}^n (y_i^{exp} - y_{ave}^{exp})^2} \quad (4)$$

توانایی مدیریت داده‌های از دست رفته شناخته شده است (Abdi & Mazloom, 2022). رگرسیون درختان اضافی (ET) مخفف Extra Trees Regressor یک روش یادگیری ماشین گروهی است که برای رگرسیون استفاده می‌شود، شبیه به جنگل‌های تصادفی اما با یک لایه تصادفی اضافه شده است.

این کار با برآزش چندین درخت تصمیم تصادفی بر روی زیرمجموعه‌های مختلف داده‌ها، بهبود دقت کلی مدل و کاهش بیش از حد برآزش کار می‌کند (Abdi & Mazloom, 2022). درخت تصمیم (DT) مخفف Decision Tree به عنوان یک ابزار یادگیری ماشین برای حل مسائل طبقه بندی و رگرسیون می‌باشد که با ساختاری شبیه به فلوچارت عمل می‌کنند. آن‌ها از گره‌های داخلی برای نمایش ویژگی‌های داده، شاخه‌ها برای نمایش تصمیمات و برگ‌ها برای نمایش نتایج استفاده می‌کنند. اما چالش اصلی درختان تصمیم‌گیری، تمایل بالای آن‌ها به بیش برآزش است.

انتخاب ویژگی برای مدل‌های هوش مصنوعی

به منظور افزایش دقت و اطمینان در فرایند انتخاب ویژگی، از چندین روش مختلف استفاده شد. هر یک از این روش‌ها بر مبنای معیارها و فرضیات متفاوتی برای سنجش اهمیت متغیرها عمل می‌کنند. استفاده هم‌زمان از چند روش انتخاب ویژگی موجب می‌شود ویژگی‌های نهایی با اتکا بر شواهد متنوع و مستقل تعیین شوند و نقش واقعی آنها در پیش‌بینی مدل با دقت بیشتری شناسایی گردد، نه صرفاً به دلیل ماهیت خاص یک الگوریتم. در این مطالعه، اهمیت ویژگی‌ها برای پیش‌بینی متغیر هدف با استفاده از سه روش جنگل تصادفی (Random Forest)، اطلاعات (Mutual Information) و اهمیت جایگشتی (Permutation Importance) ارزیابی شد.

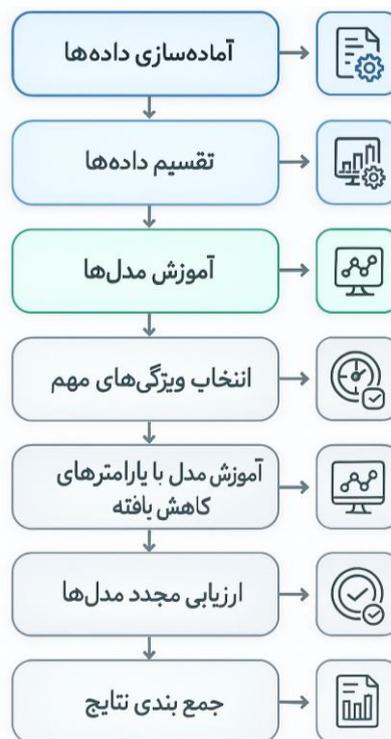
جنگل تصادفی: این روش مبتنی بر الگوریتم درخت تصمیم است که چندین درخت مستقل ایجاد می‌کند و پیش‌بینی نهایی را با استفاده از میانگین یا رأی‌گیری ترکیبی درخت‌ها ارائه می‌دهد. اهمیت هر ویژگی در جنگل تصادفی بر اساس

ریشه میانگین مربعات خطا برابر با ۰ است. این مقدار نشان می‌دهد که تفاوت بین مقادیر پیش‌بینی شده و مقادیر واقعی صفر است. هرچه ریشه میانگین مربعات خطا کوچک‌تر باشد، مدل دقیق‌تر است. مقدار بهینه میانگین خطای مطلق نیز برابر با صفر است. این مقدار نشان می‌دهد که اختلاف میانگین مطلق بین مقادیر پیش‌بینی شده و مقادیر واقعی صفر است. هرچه میانگین خطای مطلق کوچک‌تر باشد، مدل بهتر عمل کرده است. فلوجارت مراحل مختلف انجام تحقیق در شکل (۱) نشان داده شده است.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (Y_i^{exp} - Y_i^{pred})^2} \quad (5)$$

$$MAE = \frac{\sum_{i=1}^n |Y_i^{exp} - Y_i^{pred}|}{n} \quad (6)$$

در روابط بالا Y_i^{pred} و Y_i^{exp} به ترتیب مقادیر آزمایشی و پیش‌بینی شده را دارند. Y_{ave}^{exp} میانگین داده‌های آزمایشی و n تعداد اندازه‌گیری‌ها است. مقدار بهینه ضریب تعیین برابر با ۱ است. این مقدار نشان می‌دهد که مدل تمامی واریانس‌های موجود در داده‌ها را به خوبی توضیح می‌دهد. مقدار ضریب تعیین بین ۰ تا ۱ قرار می‌گیرد، و هرچه به ۱ نزدیک‌تر باشد، مدل بهتر عمل کرده است. مقدار بهینه



شکل (۱): فلوجارت مراحل انجام تحقیق حاضر

متعددی از جمله زمان تماس، pH، غلظت اولیه آلاینده، مقدار جاذب و دما قرار دارد. میانگین درصد جذب برابر با ۵۵/۹۴ درصد بوده که نشان‌دهنده عملکرد نسبتاً مطلوب جاذب در شرایط معمولی است. با این حال، اختلاف قابل توجه میان مقادیر حداقل (۱/۷ درصد) و حداکثر (۹۳/۲۸ درصد) حاکی از آن است که راندمان جذب به شدت وابسته به شرایط عملیاتی بهینه می‌باشد.

نتایج و بحث

اطلاعات آماری مربوط به درصد جذب فسفات و پارامتر مختلف مؤثر بر آن در جدول (۱) نشان داده شده است. این اطلاعات شامل میانگین، حداکثر، حداقل و انحراف معیار هر پارامتر است. داده‌های جدول (۱) و شکل (۲) نشان می‌دهند که نتایج آماری حاصل از داده‌های آزمایشگاهی بیانگر آن است که فرآیند جذب مورد مطالعه تحت تأثیر پارامترهای



می‌یابد. این رفتار به‌طور معمول ناشی از محدودیت تعداد مکان‌های فعال جاذب است؛ به‌طوری‌که با افزایش غلظت اولیه، بخش بیشتری از این مکان‌ها توسط یون‌های فسفات اشباع شده و ظرفیت جذب مؤثر کاهش می‌یابد (Thawornchaisit et al., 2021)

. علاوه بر این، در سطوح مختلف pH، فسفات به اشکال متفاوتی مانند H_2PO_4^- ، HPO_4^{2-} و PO_4^{3-} وجود دارد که هرکدام تمایل متفاوتی برای اتصال به جاذب دارند. تغییر غلظت اولیه ممکن است تعادل شیمیایی محلول و pH را تغییر داده و در نتیجه بر گونه‌های فسفات و در نهایت بر راندمان حذف تأثیر بگذارد (Manyatshe et al., 2022).

مقدار جاذب در محدوده ۰/۱۰ تا ۰/۵ گرم بر لیتر بررسی شده است. این گستره نشان می‌دهد که اثر مقدار جاذب بر راندمان حذف به‌طور کامل ارزیابی شده است. معمولاً افزایش مقدار جاذب موجب افزایش تعداد سایت‌های فعال و سطح تماس بیشتر میان فاز جامد و مایع می‌شود، اما در غلظت‌های بسیار بالا ممکن است به‌دلیل هم‌پوشانی ذرات و ایجاد توده، ظرفیت ویژه جاذب کاهش یابد (da Silva et al., 2022; Tee et al., 2022).

دما در بازه ۲۲ تا ۴۰ درجه سانتی‌گراد با میانگین ۲۴/۳۲ درجه سانتی‌گراد بررسی شد. تغییرات محدود دما نشان می‌دهد که این پارامتر اثر نسبتاً خفیفی بر فرآیند داشته است. با این حال، ضریب همبستگی مثبت ۰/۴ بین دما و درصد جذب حاکی از آن است که فرآیند تا حدودی اندوترمیک بوده و افزایش دما موجب افزایش نرخ نفوذ یون‌های فسفات به سطح جاذب می‌شود

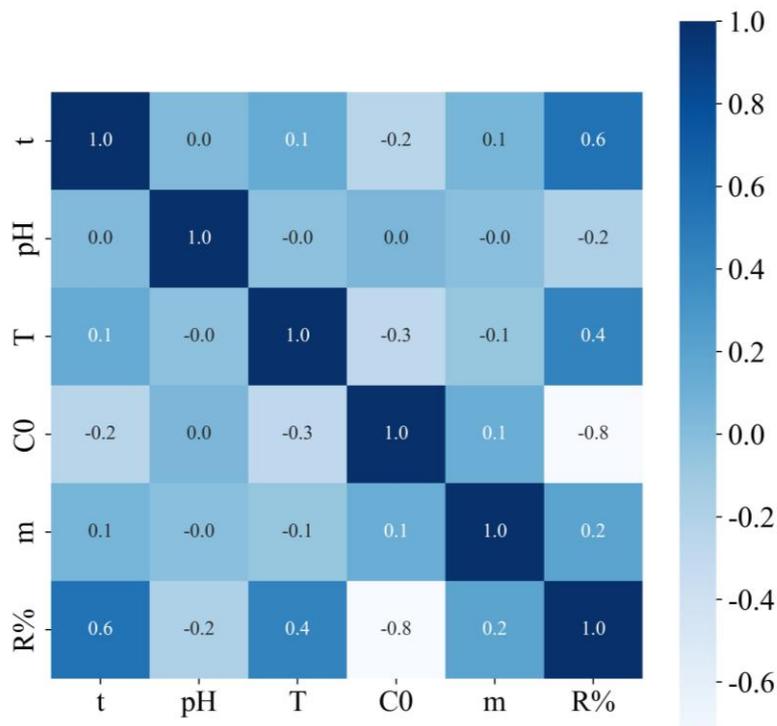
زمان تماس با میانگین حدود ۱۶۰ دقیقه و دامنه تغییر ۵ تا ۲۴۰ دقیقه، نقش مهمی در دستیابی به تعادل جذب ایفا کرده است. افزایش زمان تماس معمولاً باعث افزایش راندمان جذب تا رسیدن به نقطه اشباع جاذب می‌شود که نشانگر ماهیت تدریجی فرآیند است. زمان رسیدن به تعادل بسته به نوع جاذب و شرایط محیطی متفاوت است. جاذب‌های با ساختار متخلخل یا دارای جایگاه‌های فعال بالا (مانند هیدروکسیدهای دوگانه لایه‌ای) معمولاً زمان تعادل کوتاه‌تری (۴۰ تا ۶۰ دقیقه) دارند (Yang et al., 2014). در حالی که جاذب‌های پیچیده‌تر یا دارای مکانیزم جذب کندتر (مانند هیدروکسید نیکل-آلمینیوم-زیرکونیوم) به زمان طولانی‌تری (تا ۲۴ ساعت) نیاز دارند (Ogata et al., 2021).

مقدار pH با میانگین نزدیک به ۶ و دامنه وسیع بین ۲ تا ۱۱ نشان‌دهنده بررسی جامع رفتار جاذب در محیط‌های اسیدی و بازی است. تغییر pH می‌تواند بر بار سطحی جاذب، برهم‌کنش‌های الکترواستاتیکی و گونه‌های شیمیایی آلاینده تأثیرگذار باشد (Nouaa et al., 2024; Ye et al., 2016)؛ از این رو، این پارامتر یکی از عوامل کلیدی در تعیین کارایی فرآیند جذب محسوب می‌شود.

غلظت اولیه آلاینده با میانگین ۸۲/۲۵ میلی‌گرم بر لیتر و دامنه ۲۰ تا ۱۰۰ میلی‌گرم بر لیتر نشان می‌دهد که جاذب در غلظت‌های نسبتاً بالا مورد ارزیابی قرار گرفته است. مطابق شکل (۲)، میزان جذب فسفات توسط هیدروچار باگاس نیشکر رابطه‌ی قوی اما معکوسی با غلظت اولیه فسفات دارد (ضریب همبستگی برابر با ۰/۸-) بدین معنا که با افزایش غلظت اولیه فسفات، راندمان جذب کاهش

جدول (۱): مشخصه‌های آماری داده‌های تحقیق حاضر

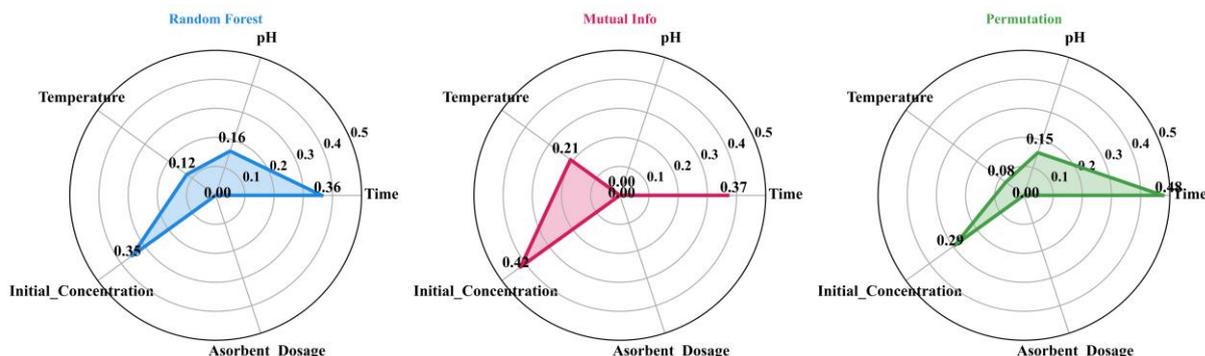
پارامتر	زمان (دقیقه)	pH	دما (سانتی‌گراد)	غلظت اولیه (میلی‌گرم بر لیتر)	مقدار جاذب	درصد جذب
میانگین	۱۶۰/۳۲	۶/۰۹۶	۲۴/۳۲	۸۲/۲۵	۰/۰۶۴	۵۵/۹۴
حداکثر	۲۴۰	۱۱	۴۰	۱۰۰	۰/۵	۹۳/۲۸
حداقل	۵	۲	۲۲	۲۰	۰/۰۱	۱/۷۰
انحراف معیار	۵۴/۷۵	۱/۳۵	۶/۱۳	۲۷/۲۲	۰/۰۸	۲۳/۶۸



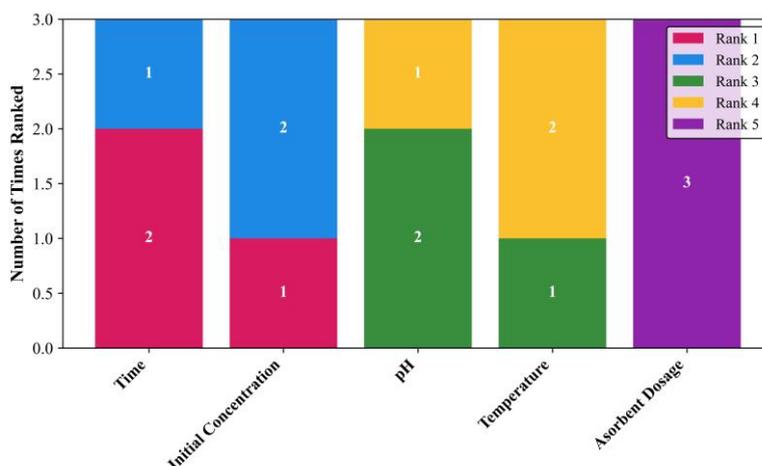
شکل (۲): ماتریس همبستگی بین پارامترهای مختلف موثر بر فرایند جذب فسفات به وسیله هیدروچار باگاس نیشکر

منفی pH به تغییر بار سطحی جاذب و رقابت میان یون‌های OH^- و فسفات برای مکان‌های جاذب مربوط می‌شود. به‌طور کلی، نتایج شکل (۲) و جدول (۱) نشان می‌دهد که غلظت اولیه فسفات مهم‌ترین عامل تعیین‌کننده راندمان حذف در شرایط آزمایشی حاضر است، در حالی که زمان تماس و دما نیز اثر مثبت و معنی‌داری بر بهبود کارایی جاذب دارند. از سوی دیگر، تأثیر pH و جرم جاذب در محدوده‌های بررسی‌شده کمتر محسوس است. این یافته‌ها بیانگر آن است که فرآیند جذب فسفات توسط هیدروچار باگاس نیشکر به‌شدت وابسته به شرایط عملیاتی بوده و دستیابی به درصد حذف بالاتر از ۹۰ درصد تنها در شرایط بهینه حاصل می‌شود.

دومین فاکتور مؤثر بر جذب فسفات، زمان تماس است (ضریب همبستگی برابر ۰/۶). این همبستگی مثبت بیانگر آن است که افزایش زمان تماس موجب افزایش میزان جذب فسفات می‌شود، اما این رابطه کاملاً خطی نیست. در مراحل اولیه تماس، سرعت جذب بالا بوده و سپس با نزدیک شدن به زمان تعادل، شیب افزایش کاهش می‌یابد. این پدیده نشان‌دهنده کنترل سینتیکی فرآیند در مراحل آغازین و محدودیت نفوذ در مراحل نهایی است (Zhang et al., 2012). در نهایت، دو پارامتر جرم جاذب و pH به ترتیب دارای رابطه مستقیم و معکوس ضعیفی با راندمان جذب فسفات هستند. اثر مثبت جرم جاذب ناشی از افزایش سطح مؤثر جذب و تعداد جایگاه‌های فعال است، در حالی که تأثیر ود.



شکل (۳): نمودار راداری اهمیت نسبی ویژگی‌ها (زمان، pH، دما، غلظت اولیه، دوز جاذب) با سه روش جنگل تصادفی، اطلاعات متقابل و اهمیت جایگشتی



شکل (۴): فراوانی رتبه‌های ۱-۵ ویژگی‌ها در سه روش انتخاب ویژگی

رویکرد جایگشتی، نیز زمان تماس را به‌عنوان مهم‌ترین پارامتر و سپس غلظت اولیه را معرفی می‌کند، در حالی که جرم جاذب همچنان بی‌تأثیر باقی می‌ماند. همگرایی در هر سه روش (زمان تماس و غلظت اولیه) تأیید می‌کند که پارامترهای کلیدی در فرآیند جذب فسفات هستند. در حالی که جرم جاذب تأثیر قابل‌توجهی ندارد و دما و pH نقش متوسطی ایفا می‌کند. این نتایج با مطالعات تجربی و مدل‌سازی‌های آماری و یادگیری ماشین نیز همخوانی دارد؛ به‌ویژه مدل‌های شبکه عصبی و رگرسیون نشان داده‌اند که افزایش زمان تماس و غلظت اولیه، کارایی جذب را به طور

نتایج تجزیه و تحلیل انتخاب ویژگی برای الگوریتم‌های هوش مصنوعی

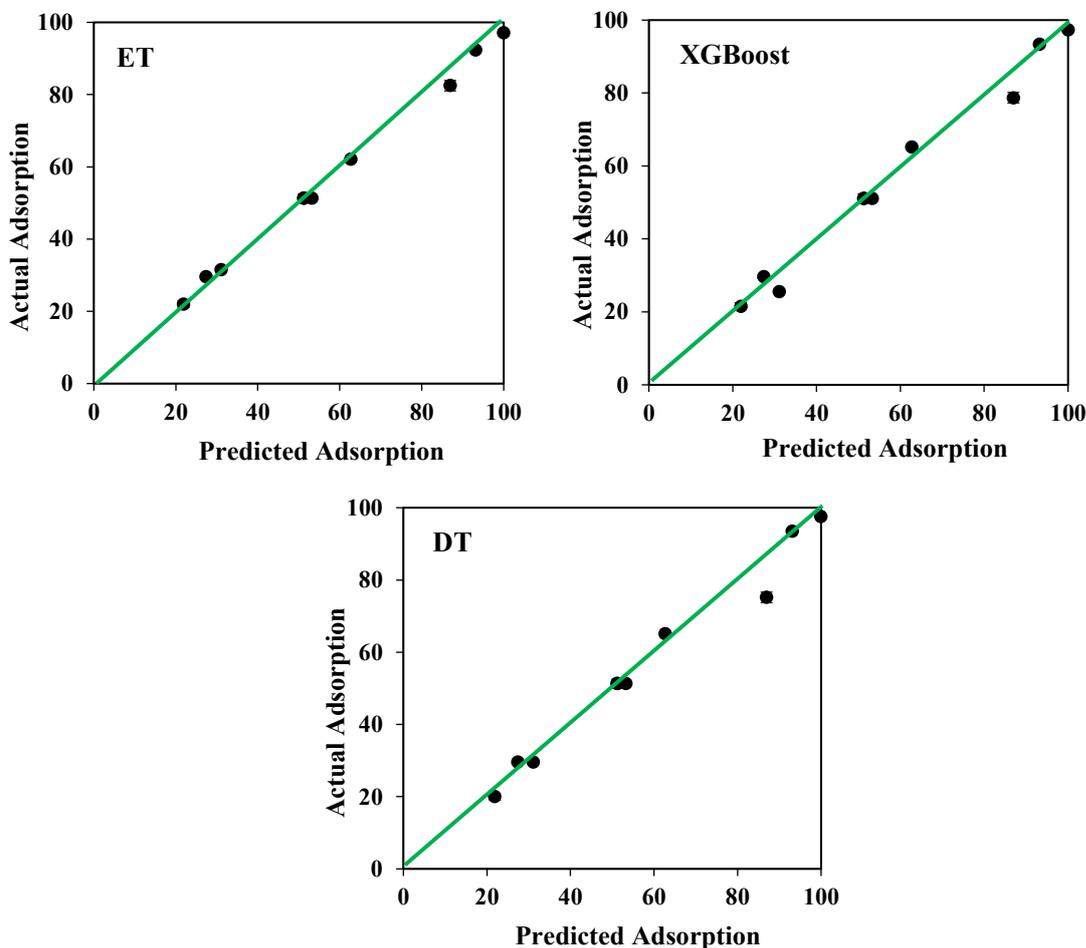
شکل (۳) سه روش مختلف برای رتبه‌بندی پارامترهای مؤثر بر جذب فسفات را نشان می‌دهد که شامل روش جنگل تصادفی، اطلاعات متقابل و اهمیت جایگشتی است. در روش جنگل تصادفی، اهمیت ویژگی‌ها بر اساس الگوریتم جنگل تصادفی محاسبه شده و نشان می‌دهد که زمان تماس و غلظت اولیه بیشترین تأثیر را دارند، در حالی که جرم جاذب بی‌تأثیر و دو پارامتر دیگر کم اهمیت هستند. روش استفاده از اطلاعات متقابل، نقش غلظت اولیه و دما را برجسته می‌کند، همچنان با اهمیت صفر برای جرم جاذب، روش

فسفات غیرخطی باشد، که منجر به عملکرد ضعیف‌تر DT می‌شود. DT به نوبت موجود در داده‌های آموزشی حساس است و نوبت می‌تواند منجر به پیش‌بینی‌های نادرست شود (Wu et al., 2023). مشابه با نتایج این تحقیق، تحقیقات اخیر نشان می‌دهد که مدل‌های یادگیری ماشین (ML) - به ویژه الگوریتم‌های مبتنی بر درخت و شبکه‌های عصبی مصنوعی (ANNs) - در پیش‌بینی جذب فسفات در مواد جاذب مختلف، دقت بالایی دارند (Elshishini et al., 2025; Iftikhar et al., 2025; Zhang & Pan, 2014). مطالعه‌ای با بهره‌گیری از الگوریتم‌های یادگیری ماشین از جمله Random Forest و CatBoost به شبیه‌سازی و پیش‌بینی ظرفیت جذب فسفر توسط انواع مختلف هیدروچار پرداخته شد. نتایج نشان می‌دهد که مدل CatBoost عملکرد بسیار دقیقی دارد و مقدار هیدروچار و غلظت اولیه فسفر هیدروچار به عنوان مهم‌ترین پارامترهای مؤثر شناسایی شده‌اند (Lyu et al., 2024).

معناداری افزایش می‌دهد (Nouaa et al., 2024; Ye et al., 2016).

ارزیابی دقت الگوریتم‌های یادگیری ماشین در پیش-بینی جذب فسفات

نتایج پیش‌بینی جذب فسفات بر روی هیدروچار باگاس نیشکر با استفاده از ۵ پارامتر (pH، زمان تماس، دما، غلظت اولیه فسفات و مقدار جاذب) به عنوان ورودی در شکل (۳) و جدول (۲) نشان داده شده است. در میان الگوریتم‌های مختلف مورد بررسی، ET با R^2 برابر با ۰/۹۲۲، که بالاترین مقدار در بین الگوریتم‌ها است، و همچنین با مقادیر پایین‌تر RMSE (۰/۰۷۴) و MAE (۰/۰۴۸)، بیشترین کارایی را در پیش‌بینی راندمان جذب فسفات داشت. پایین‌ترین دقت پیش‌بینی با R^2 (۰/۷۴۳)، RMSE (۰/۱۳۴) و MAE (۰/۰۸)، مربوط به الگوریتم DT بود. انتخاب الگوریتم مناسب برای پیش‌بینی به عوامل مختلفی مانند ماهیت داده‌ها، پیچیدگی مدل و دقت مورد نیاز بستگی دارد. الگوریتم ET به طور ذاتی برای مدل‌سازی روابط پیچیده بین متغیرها مناسب است. در این مورد، ET می‌تواند به طور مؤثر تعاملات بین ۵ پارامتر ورودی و راندمان جذب فسفات را مدل‌سازی کند. علاوه بر این ET، با استفاده از داده‌های آموزشی، الگویی را برای پیش‌بینی راندمان جذب فسفات یاد می‌گیرد. این قابلیت به الگوریتم ET اجازه می‌دهد تا به طور دقیق‌تر به تغییرات در پارامترهای ورودی پاسخ دهد. دلایل احتمالی عملکرد ضعیف DT ناتوانی آن در مدل‌سازی روابط غیرخطی بین متغیرها ضعیف است. در این مورد، ممکن است رابطه بین پارامترهای ورودی و راندمان جذب



شکل (۵): نتایج الگوریتم‌های یادگیری ماشین برای پیش‌بینی جذب فسفات توسط هیدروچار باگاس نیشکر

جدول (۲): مقایسه عملکرد الگوریتم‌های یادگیری ماشین برای پیش‌بینی جذب فسفات توسط هیدروچار باگاس نیشکر

معیارهای ارزیابی	مدل		
R ²	RMSE	MAE	
۰/۹۲۲	۰/۰۷۴	۰/۰۴۸	ET
۰/۷۶۱	۰/۱۲۹	۰/۰۷۷	XGBoost
۰/۷۴۳	۰/۱۳۴	۰/۰۸۰	DT

می‌گذارد، زیرا به مولکول‌های جذب کننده اجازه می‌دهد تا با سطح جاذب برهم کنش داشته باشند. در فرایندهای جذب زمان تماس کافی برای رسیدن به تعادل و دستیابی به حداکثر جذب ضروری است (Pannu et al., 2024). غلظت اولیه نقش حیاتی در تعیین ظرفیت جذب یک جاذب با تأثیرگذاری بر نیروی محرکه برای انتقال جرم دارد که به

ارزیابی انتخاب ویژگی در بهینه‌سازی پیش‌بینی جذب فسفات

بر اساس مجموع روش‌های انتخاب ویژگی، زمان تماس، غلظت اولیه، pH، دما و جرم جاذب به ترتیب نقش مهمی در کارایی و اثربخشی فرآیند جذب دارند. زمان تماس یکی از پارامترهای مهمی است که بر فرآیند جذب تاثیر

می‌گذارند. با درک و کنترل این پارامترها می‌توان عملکرد فرآیند جذب را برای کاربردهای مختلف افزایش داد. مقایسه نتایج الگوریتم موفق در مرحله قبل از میان الگوریتم‌های مورد مطالعه با ورودی‌های مختلف براساس درجه اهمیت برای پیش‌بینی جذب فسفات توسط هیدروچار باگاس نیشکر در جدول ۳ نشان داده شده است. براساس این نتایج، پیش‌بینی جذب فسفات توسط هیدروچار باگاس نیشکر با استفاده از دو پارامتر غلظت اولیه و زمان تماس با R^2 برابر با ۰/۷۲۸) نسبت به استفاده از چهار پارامتر غلظت اولیه، زمان تماس، جرم جاذب و pH با R^2 برابر با ۰/۶۶۱) دقت بیشتری دارد. این امر به دلیل کاهش پیچیدگی مدل و تأثیر بیشتر این دو پارامتر نسبت به سایر پارامترها است. غلظت اولیه به عنوان یکی از عوامل کلیدی تعیین‌کننده میزان آلودگی و زمان تماس نیز به عنوان فاکتور مهم در فرآیند جذب عمل می‌کند، در حالی که اضافه کردن جرم جاذب و pH ممکن است به افزایش دقت مدل کمک نکند و باعث افزایش پیچیدگی محاسباتی شود. نتایج حاصل از روش‌های تعیین ویژگی در تعیین پارامترهای پر اهمیت‌تر (غلظت اولیه و زمان تماس) با نتایج حاصل از پارامترهای با اهمیت از طریق الگوریتم‌های مورد مطالعه (شکل ۳) و همچنین نتایج ماتریس همبستگی پیرسون (شکل ۲) مطابقت داشت. (Wu et al., 2023).

نوبه خود بر فرآیند جذب تا رسیدن به تعادل تأثیر می‌گذارد. غلظت اولیه بالاتر معمولاً منجر به نیروی محرکه قوی‌تر برای انتقال جرم می‌شود که منجر به افزایش ظرفیت جذب می‌شود. این نیروی محرکه برای غلبه بر مقاومت‌های انتقال جرم مختلف که بین فازهای آبی و جامد در طول فرآیند جذب وجود دارد، ضروری است. علاوه بر این، غلظت اولیه برای تخمین نرخ جذب اولیه و ثابت سرعت فرآیند جذب بسیار مهم است و اهمیت آن در بهینه‌سازی سینتیک جذب نشان داده شده است (Tran, 2023).

pH محلول با تغییر بار سطحی و یونیزاسیون جاذب، ظرفیت جذب را تحت تأثیر قرار می‌دهد و در نتیجه بر تعامل بین جاذب و آلاینده تأثیر می‌گذارد (Vujčić et al., 2023). دما نیز می‌تواند بر میزان جذب و ظرفیت تعادلی جاذب تأثیر بگذارد. به طور کلی، فرآیندهای جذب می‌توانند گرمازا یا گرماگیر باشند و تغییرات دما می‌تواند کارایی جذب را افزایش یا مهار کند (Zhou et al., 2019). جرم جاذب مورد استفاده در فرآیند بر تعداد مکان‌های جذب در دسترس تأثیر می‌گذارد. جرم جاذب بزرگ‌تر، ظرفیت جذب و کارایی را با ایجاد مکان‌های بیشتر، برای چسباندن ماده جاذب افزایش می‌دهد. قابل ذکر است میزان جذب به دلیل افزایش سطح در دسترس برای جذب، تأثیر مستقیمی بر ظرفیت جذب دارد (Guo et al., 2022). این ویژگی‌ها برای بهینه‌سازی فرآیند جذب ضروری هستند، زیرا در مجموع بر کارایی و اثربخشی سیستم تأثیر

جدول (۳): نتایج الگوریتم ET با ورودی‌های مختلف براساس روش حذف ویژگی بازگشتی برای پیش‌بینی جذب فسفات توسط هیدروچار باگاس نیشکر

R^2	ترکیب پارامترهای مختلف
۰/۶۷۱	زمان تماس
۰/۷۲۸	غلظت اولیه + زمان تماس
۰/۶۱۷	غلظت اولیه + زمان تماس + دوز جاذب
۰/۶۶۱	غلظت اولیه + زمان تماس + دوز جاذب + اسیدیته

نتیجه‌گیری

از مهم‌ترین دستاوردهای این مطالعه می‌توان به استفاده هم‌زمان از چند مدل داده‌محور و سه رویکرد متمایز انتخاب ویژگی، به‌منظور مقایسه و شناسایی عوامل کلیدی مؤثر در جذب، اشاره کرد که موجب افزایش قابلیت اعتماد و تفسیرپذیری مدل‌ها گردید. با این حال، محدود بودن داده‌های آزمایشگاهی و عدم در نظر گرفتن اثرات فیزیکی-شیمیایی در مقیاس مولکولی از جمله کاستی‌های پژوهش به‌شمار می‌رود. بنابراین، توسعه مدل‌های ترکیبی داده‌محور و فیزیکی، استفاده از داده‌های میدانی و چندمنبعی، و اعتبارسنجی در شرایط واقعی می‌تواند در پژوهش‌های آینده موجب ارتقای دقت و کارایی پیش‌بینی‌ها گردد. به‌طور کلی، یافته‌های این تحقیق نشان داد که بهره‌گیری از الگوریتم‌های یادگیری ماشین همراه با انتخاب ویژگی می‌تواند رویکردی مؤثر برای طراحی بهینه جاذب‌ها و پیش‌بینی عملکرد فرآیندهای حذف آلاینده‌ها از محلول‌های آبی باشد.

در این پژوهش، با بهره‌گیری از الگوریتم‌های یادگیری ماشین شامل ET، DT و XGBoost، تلاش شد تا راندمان جذب فسفات توسط هیدروچار به‌صورت دقیق مدل‌سازی و پیش‌بینی شود. برای بهبود عملکرد مدل‌ها و افزایش تفسیرپذیری، سه روش انتخاب ویژگی مبتنی بر ضریب پیرسون، حذف بازگشتی ویژگی‌ها و اهمیت ویژگی بر پایه مدل مورد استفاده قرار گرفت. نتایج نشان داد که ترکیب روش‌های انتخاب ویژگی با مدل‌های یادگیری ماشین موجب افزایش چشمگیر دقت پیش‌بینی و کاهش پیچیدگی محاسباتی گردید. در میان مدل‌ها، الگوریتم اکسترا تری بهترین عملکرد را از نظر شاخص‌های آماری ارائه داد و تحلیل اهمیت متغیرها بیانگر آن بود که زمان تماس و غلظت اولیه فسفات بیشترین نقش را در فرآیند جذب داشته‌اند، در حالی که مقدار جاذب و pH تأثیر کمتری داشته‌اند.

منابع

- Abdi, J., & Mazloom, G. (2022). Machine learning approaches for predicting arsenic adsorption from water using porous metal-organic frameworks. *Scientific Reports*, 12(1), 16458 .
- Almanassra, I. W., Mckay, G., Kochkodan, V., Atieh, M. A., & Al-Ansari, T. (۲۰۲۱). A state of the art review on phosphate removal from water by biochars. *Chemical Engineering Journal*, 409, 128211 .
- Awolusi, T. F., Oke, O. L., Akinkurolere, O. O., & Atoyebi, O. D. (2019). Comparison of response surface methodology and hybrid-training approach of artificial neural network in modelling the properties of concrete containing steel fibre extracted from waste tyres. *Cogent engineering*, 6(1), 1649852 .
- Büyükeçeci, M., & Okur, M. C. (2022). A comprehensive review of feature selection and feature selection stability in machine learning. *Gazi University Journal of Science*, 36(4), 1506-1520 .
- Crini, G. (2008). Kinetic and equilibrium studies on the removal of cationic dyes from aqueous solution by adsorption onto a cyclodextrin polymer. *Dyes and Pigments*, 77(2), 415-426 .
- da Silva Bruckmann, F., Schnorr, C. E., da Rosa Salles, T., Nunes, F. B., Baumann, L., Müller, E. I., Silva, L. F., Dotto, G. L., & Bohn Rhoden, C. R. (2022). Highly efficient adsorption of tetracycline using chitosan-based magnetic adsorbent. *Polymers*, 14(22), 4854 .
- Divband Hafshejani, L., Naseri, A. A., Moradzadeh, M., Daneshvar, E., & Bhatnagar, A. (2022). Applications of soft computing techniques for prediction of pollutant removal by environmentally friendly adsorbents (case study: the nitrate adsorption on modified hydrochar). *Water Science & Technology*, 86(5), 1066-1082 .



- El Touati, Y., Slimane, J. B., & Saidani, T. (2024). Adaptive Method for Feature Selection in the Machine Learning Context. *Engineering, Technology & Applied Science Research*, 14(3), 14295-14300 .
- Elshishini, H. M., El-Subruiti, G. M., Ghatass, Z. F., Farag, N. H., & Eltaweil, A. S. (2025). Assessing decision-based machine learning algorithms for predicting adsorption efficiency: a detailed study of MnFe LDH functionalized La (OH)₃@ AC chitosan beads. *Journal of Water Process Engineering*, 69, 106678 .
- Fu, W., Feng, M., Guo, C., Zhou, J., Zhang, X., Lv, S., Huo, Y., & Wang, F. (2024). Machine learning-driven prediction of phosphorus removal performance of metal-modified biochar and optimization of preparation processes considering water quality management objectives. *Bioresource Technology*, 130861 .
- Guo, Y., Ali, W., Schneider, A., Salma, A., Mayer-Gall, T., Gutmann, J. S., & Fernandez Lahore, H. M. (2022). Megaporous monolithic adsorbents for bioproduct recovery as prepared on the basis of nonwoven fabrics. *Electrophoresis*, 43(13-14), 1387-1398 .
- Hafshejani, L. D., Hooshmand, A., Naseri, A. A., Mohammadi, A. S., Abbasi, F., & Bhatnagar, A. (2016). Removal of nitrate from aqueous solution by modified sugarcane bagasse biochar. *Ecological Engineering*, 95, 101-111 .
- Iftikhar, S., Ishtiaq, R., Zahra, N., Ruba, F., Lam, S.-M., Abbas, A., & Jaffari, Z. H. (2025). Probabilistic prediction of phosphate ion adsorption onto biochar materials using a large dataset and online deployment. *Chemosphere*, 370, 144031 .
- Lyu, H., Xu, Z., Zhong, J., Gao, W., Liu, J., & Duan, M. (2024). Machine learning-driven prediction of phosphorus adsorption capacity of Biochar :insights for adsorbent design and process optimization. *Journal of environmental management*, 369, 122405 .
- Manyatshe, A., Cele, Z. E., Balogun, M. O., Nkambule, T. T., & Msagati, T. A. (2022). Chitosan modified sugarcane bagasse biochar for the adsorption of inorganic phosphate ions from aqueous solution. *Journal of Environmental Chemical Engineering*, 10(5), 108243 .
- Nouaa, S., Aziam, R., Benhiti, R., Carja, G., Zerbet, M., & Chiban, M. (2024). Exploiting RSM and ANN modeling methods to optimize phosphate ions removal using LDH/alginate composite beads. *Journal of Water Process Engineering*, 68, 106333 .
- Ogata, F., Nagai, N., Iijima, S., Toda, M., Otani, M., Nakamura, T., & Kawasaki, N. (2021). Exploiting the different parameters on the adsorption of phosphate ions and its subsequent recovery using complex nickel–aluminum–zirconium hydroxide. *Chemical and Pharmaceutical Bulletin*, 69(8), 789-795 .
- Pannu, M. W., Huang, A., & Plumlee, M. H. (2024). Variable PFAS removal by adsorbent media with sufficient prediction of breakthrough despite reduced contact time at pilot scale. *Water Environment Research*, 96(5), e11035 .
- Rahdar, S., Rahdar, A., Sattari, M., Hafshejani, L. D., Tolkou, A. K., & Kyzas, G. Z. (2021). Barium/cobalt@ polyethylene glycol nanocomposites for dye removal from aqueous solutions. *Polymers*, 13(7), 1161 .
- Suresh, S., Newton, D. T., Everett IV, T. H., Lin, G., & Duerstock, B. S. (2022). Feature selection techniques for a machine learning model to detect autonomic dysreflexia. *Frontiers in Neuroinformatics*, 16, 901428 .



- Tee, G. T., Gok, X. Y., & Yong, W. F. (2022). Adsorption of pollutants in wastewater via biosorbents, nanoparticles and magnetic biosorbents: A review. *Environmental Research*, 212, 113248 .
- Thawornchaisit, U., Onlamai, T., Phurkphong, N., & Sukharom, R. (2021). Sugarcane Bagasse-derived Hydrochar: Modification with Cations to Enhance Phosphate Removal: 10.32526/enrj/19/202100036. *Environment and Natural Resources Journal*, 19(5), 371-380 .
- Tran, H. N. (2023). Applying linear forms of pseudo-second-order kinetic model for feasibly identifying errors in the initial periods of time-dependent adsorption datasets. *Water*, 15(6), 1231 .
- Vujić, M., Vasiljević, S., Nikić, J., Kordić, B., Agbaba, J., & Tubić, A. (2023). Sorption Behavior of Organic Pollutants on Biodegradable and Nondegradable Microplastics: pH Effects. *Applied Sciences*, 13(23), 12835 .
- Wu, Y., Li, Y., Jiang, Z., Xu, Z., Yang, M., Ding, J., & Zhang, C. (2023). Machine Learning Prediction of Phosphate Adsorption on Six Different Metal-Containing Adsorbents. *ACS ES&T Engineering*, 3(8), 1135-1146 .
- Yang, K., Yan, L.-g., Yang, Y.-m., Yu, S.-j., Shan, R.-r., Yu, H.-q., Zhu, B.-c., & Du, B. (2014). Adsorptive removal of phosphate by Mg–Al and Zn–Al layered double hydroxides: kinetics, isotherms and mechanisms. *Separation and Purification Technology*, 124, 36-42 .
- Ye, J., Cong, X., Zhang, P., Zeng, G., Hoffmann, E., Wu, Y., Zhang, H., & Fang, W. (2016). Operational parameter impact and back propagation artificial neural network modeling for phosphate adsorption onto acid-activated neutralized red mud. *Journal of molecular liquids*, 216, 35-41 .
- Zhang, J., Shan, W., Ge, J., Shen, Z., Lei, Y., & Wang, W. (2012). Kinetic and equilibrium studies of liquid-phase adsorption of phosphate on modified sugarcane bagasse. *Journal of Environmental Engineering*, 138(3), 252-258 .
- Zhang, W., Huang, W., Tan, J., Huang, D., Ma, J., & Wu, B. (2023). Modeling, optimization and understanding of adsorption process for pollutant removal via machine learning: Recent progress and future perspectives. *Chemosphere*, 311, 137044 .
- Zhang, Y., & Pan, B. (2014). Modeling batch and column phosphate removal by hydrated ferric oxide-based nanocomposite using response surface methodology and artificial neural network. *Chemical Engineering Journal*, 249, 111-120 .
- Zhou, Y., Lu, J., Zhou, Y., & Liu, Y. (2019). Recent advances for dyes removal using novel adsorbents: a review. *Environmental pollution*, 252, 352-365 .
- Zhu, X., & Ma, J. (2020). Recent advances in the determination of phosphate in environmental water samples: Insights from practical perspectives. *TRAC Trends in Analytical Chemistry*, 127, 115908 .