ICA2025OSLO
14th IWA Specialised Conference on
Instrumentation, Control and Automation

# Prediction of Heavy Metal Adsorption Performance by Biochar Based on Machine Learning

Shuoyang Wang[1], Fangchao Zhao[1], Dong Chen[1]

[1] School of Environmental and Municipal Engineering, Qingdao University of Technology, Jialingjiang Dong 777, Qingdao 266520, China

**Corresponding author: Dong Chen, chendong_cau@163.com**

**Abstract:** Heavy metal pollution in water bodies has become a severe environmental issue globally. Biochar, as a low-cost adsorbent material, is widely used for the removal of heavy metals from water. However, traditional studies often focus on only a few physicochemical properties of biochar and external environmental conditions, neglecting the complex interactions of multiple factors in water bodies. Machine learning, as a data-driven mathematical approach, can effectively analyze the nonlinear relationships among various characteristic variables that influence adsorption performance.This study employs an automated machine learning framework to construct models and thoroughly investigates the comprehensive effects of the physical and chemical properties of biochar and external environmental conditions on its adsorption performance. The results indicate that all automated machine learning frameworks accurately capture the relationships between input features and output features, with $H_2O$ AutoML identified as the optimal model, achieving an $R^2$ of 0.918. Furthermore, feature importance analysis reveals that the concentration of heavy metals in water has the most significant impact on the adsorption performance of biochar, followed by the chemical properties of biochar, particularly the H/C ratio, highlighting its core role in the adsorption performance of biochar.This study is the first to apply automated machine learning models to predict the adsorption performance of biochar for heavy metals in water bodies, pioneering a new path for the application of machine learning in the environmental field.

**Keywords**: Automated Machine Learning; Biochar; Heavy metal; Feature importance; machine learning

## 1 Introduction

Heavy metal pollution in water bodies has become a serious environmental issue worldwide, posing significant threats to ecosystems and human health[1]. Biochar, as a promising adsorbent material, is widely applied for the removal of heavy metals from water. However, traditional studies often focus only on a limited number of physicochemical properties of biochar and external environmental conditions, neglecting the complex interactions of multiple factors in water bodies. This limitation results in insufficient predictive capability and adaptability of traditional methods in practical applications.Fortunately, with the rapid development of artificial intelligence technologies, particularly the emergence of machine learning, new solutions have been provided to overcome these challenges. Machine learning, as a data-driven mathematical approach, can effectively analyze the nonlinear relationships among various characteristic variables that influence adsorption performance without requiring extensive assumptions about the relationships between variables[2,3]. Building on this advantage, this study employs an automated machine learning framework to construct models, comparing model performance to identify the optimal model. Furthermore, through feature importance analysis and partial dependence plots, this research delves into the comprehensive effects of the physical and chemical properties of biochar and external environmental conditions on its adsorption performance.

## 2 Method

In this experiment, we utilized four different machine learning frameworks: TPOT, FLAML, AutoGluon, and $H_2O$ AutoML (as shown in Figure 2). To further analyze the specific impacts of each feature variable on the adsorption performance of biochar, we employed the SHAP (Shapley Additive Explanations) additive explanation method, combined with Partial Dependence Plots (PDP) to reveal the interactions between features and their contributions to the adsorption performance of biochar. The SHAP method is a powerful tool for feature importance analysis, as it quantifies the contribution of each feature to the model's prediction results, providing a transparent explanation of the decision-making process. At the same time, PDP help us understand the relationship between individual features and model predictions, particularly how a feature influences adsorption performance across different value ranges. The combination of these

methods allows us to more intuitively identify the factors that play a critical role in the process of biochar adsorbing heavy metals.

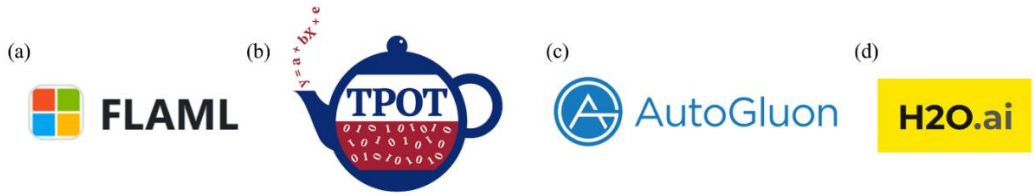


**Figure 2** Four different automated machine learning frameworks. (a) FLAML. (b) TPOT. (c) AutoGluon. (d)$H_2O$ AutoML.

**3 Model Performance Comparison**

Figure 3.1 illustrates the performance of the four automated machine learning frameworks at varying training times, ranging from 60 seconds to 2400 seconds. The results show distinct performance characteristics for each framework at different training durations. Specifically, TPOT achieved its best performance at a training time of 300 seconds, with an $R^2$ value of 0.912. In contrast, FLAML required a longer training time to reach optimal predictive performance; at 1500 seconds of training, FLAML's $R^2$ value was 0.913, after which its performance remained relatively stable. AutoGluon's performance differed from the aforementioned two frameworks, as it also reached its optimal performance at a training time of 300 seconds, with an $R^2$ value of 0.911. However, extending the training time further did not improve AutoGluon's predictive performance, which remained consistent. Among all the models, $H_2O$ AutoML demonstrated the best performance, achieving an $R^2$ value of 0.918 at a training time of 1800 seconds, surpassing the other three frameworks.

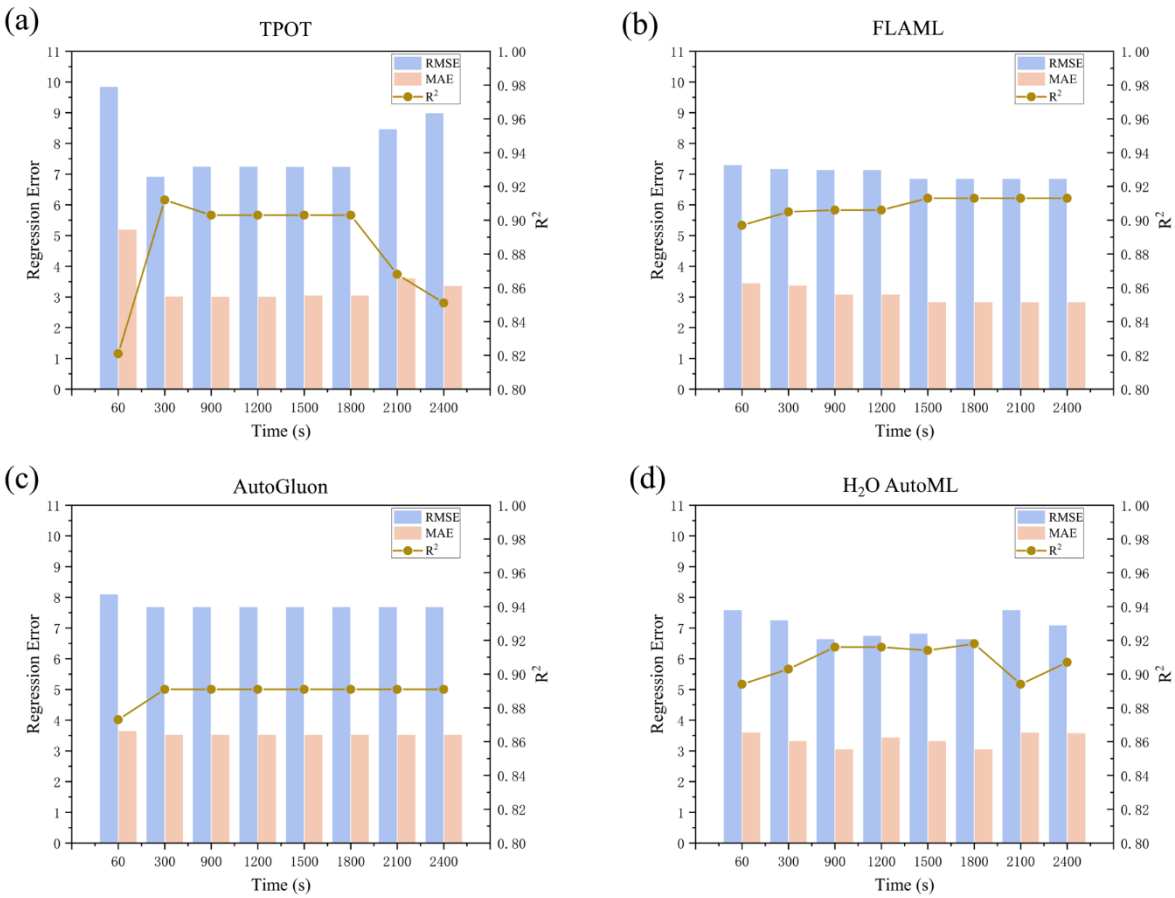


**Figure 3.1** Performance of four automated machine learning frameworks at different training times. (a) TPOT. (b) FLAML. (c) AutoGluon. (d)$H_2O$ AutoML.

## 4 Feature Importance Analysis

Figure 4.1 presents the ranking of feature importance for the predictive model. The top three key factors affecting the adsorption performance of biochar are the concentration of heavy metals in the solution, the stirring rate, and the amount of biochar added. Among these, the concentration of heavy metals in the solution has the most significant impact on biochar's adsorption performance. This is likely because a higher concentration of heavy metals leads to stronger adsorption forces, allowing metal ions to more readily occupy the active sites on the biochar surface, thereby significantly influencing adsorption performance. The stirring rate is the second most important feature. Its impact on the adsorption process primarily lies in its ability to enhance the contact between the solution and biochar, allowing heavy metal ions to diffuse more effectively to the biochar surface, thus improving adsorption efficiency. The amount of biochar added ranks third in importance. Increasing the amount of biochar can provide more adsorption sites, thereby enhancing the adsorption effect. However, excessive amounts of biochar may lead to a dilution effect in the solution, potentially reducing adsorption efficiency, highlighting the need to find a balance.

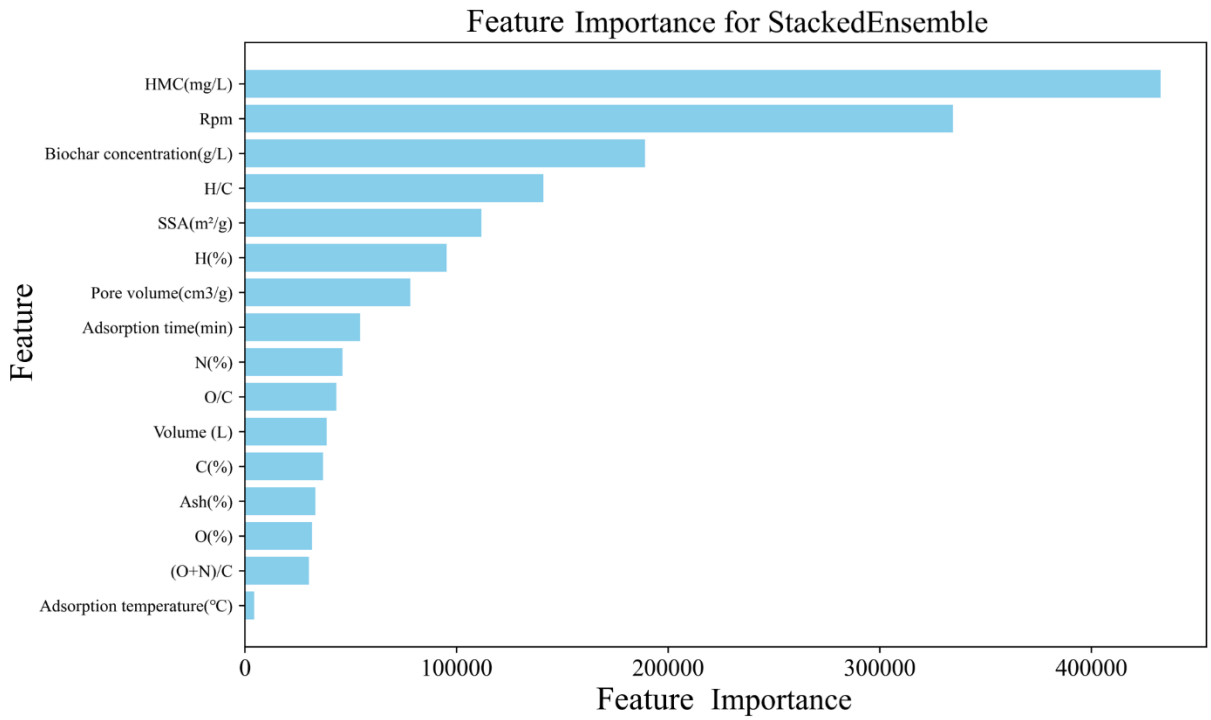


**Figure 4.1** Feature Importance Analysis

Automated machine learning frameworks, by providing a variety of model selections and automatic hyperparameter tuning, can effectively analyze the nonlinear relationships between features, significantly reducing experimental time and costs while enhancing research efficiency. These frameworks overcome the limitations of traditional material preparation processes and notably optimize the prediction and evaluation of adsorbent performance. Through automation, researchers can more quickly explore and validate different models and parameter combinations, thus identifying the optimal adsorbent formulations and operating conditions for more efficient removal of heavy metals or other applications. This approach not only improves the reproducibility and reliability of research but also provides strong support for innovation in the field of materials science.

**REFERENCES**

[1] Predicting the Bioaccessibility of Soil Cd, Pb, and As with AdvancedMachine Learning for Continental-Scale Soil Environmental CriteriaDetermination in China[Z].

[2] Tian Y, Yang X, Chen N, et al. Data-driven interpretable analysis for polysaccharide yield prediction[J/OL]. Environmental Science and Ecotechnology, 2024, 19: 100321. DOI:10.1016/j.ese.2023.100321.

[3] Chen H, Cao X, Wang J, et al. Machine Learning Accelerating the Condition Screening of Ceftriaxone Sodium Anaerobic Co-Metabolic Degradation[J/OL]. ACS ES&T Engineering, 2024, 4(4): 947-955. DOI:10.1021/acsestengg.3c00518.