



DATA-DRIVEN MODELING OF ORGANIC POLLUTANT ADSORPTION ONTO ADVANCED ADSORBENT MATERIALS UNDER VARIABLE PHYSICOCHEMICAL CONDITIONS

Ilona Israilova

PhD, Pedagogics, Tashkent state medical university, Uzbekistan, Email: ilona1israilova@gmail.com
ORCID ID:0000-0001-7275-4897

Received:-02/06/2025, Revised:-10/08/2026, Accepted:-12/09/2026, Published:-26/09/2026

ABSTRACT

Wastewater treatment is a persistent problem with organic pollutant contamination due to the significant role that the interacting material, pollutant, physicochemical, and operational factors play in the removal efficiency. This study has created a data-driven prediction model for advanced adsorbent material adsorption capacity under changing treatment conditions. An initial 6,172 observations were analyzed, with 101 duplicate observations removed, leaving 6,071 unique observations to be analyzed. Eight variable factors: specific surface area, pore volume, molecular characteristic, pH, temperature, initial concentration, flow rate, and column height were studied with respect to the adsorption capacity (Q). The regression approaches compared were: Multiple Linear Regression, Random Forest, XGBoost, Support Vector Regression, and K-Nearest Neighbors, with an 80:20 train-test split and 5-fold cross-validation. XGBoost demonstrated the best overall performance with $R^2 = 0.975$ and $RMSE = 12.529$ mg/g and the lowest MAE of 3.842 mg/g was obtained with Random Forest. Initial concentration is the most important feature in XGBoost, making up 65.404% of its importance, followed by column height and specific surface area. However, SHAP analysis revealed that the adsorption capacity increased with increasing concentration initially, and other variables had nonlinear and context-dependent effects. The results show the capability of interpretable machine learning models to predict the adsorption process accurately, optimize it, and minimize the number of experiments required for wastewater treatment. The outcomes also highlight the importance of considering both structural and operating parameters together.

KEYWORDS: *Organic pollutants; Adsorption capacity; XGBoost; Machine learning; SHAP*

1.INTRODUCTION

The organic contaminants from industrial effluents, pharmaceuticals, agrochemicals, dyes, petroleum-derived compounds, and other anthropogenic sources have emerged as one of the most predominant pollutants in the aquatic environment. These chemicals are often persistent, toxic, bioaccumulating, and not easily degraded, posing threats to aquatic life and human health. POCs and emerging contaminants are contaminants that stay in the water for a long time, necessitating the need for better treatment technologies (Ighalo et al., 2022). Conventional WWTS are often not able to completely remove pharmaceutical residues, which is a particular problem (De Andrade et al., 2018). Endocrine-disrupting compounds and other micropollutants, for example, remain detectable at low levels and have a detrimental impact on the ecology (Aris et al., 2020). Therefore, the removal of organic contaminants from wastewater has become an important priority in the environment.

Adsorption is extensively applied for the treatment of wastewater as contaminants can be moved from the solution to the surface or inner pores of solid adsorbent material. The advantages of adsorption are that it is simple to operate, has a relatively low energy requirement, is very effective and can be applied to a wide range of contaminants compared with many conventional treatment methods. The high affinity of advanced materials including graphene-based adsorbents, nanomaterials, metal-organic frameworks, activated carbon and modified biochar to various organic pollutants has been demonstrated (Awad et al., 2020; Baig et al., 2019). Metal-organic frameworks (MOFs) are exceptionally desirable in particular thanks to their tunable pore structure, high surface area, and chemical functionality (Huang et al., 2021; Jin et al., 2020). Moreover, due to its relatively low cost, porous structure and potential for regeneration and reuse, biochar is a promising adsorbent (Cheng et al., 2021; Dai et al., 2019).

The interactions between adsorbent characteristics, pollutant properties, physicochemical conditions and operational parameters determine the adsorption performance. The number and accessibility of the adsorption sites are determined by the specific surface area and the pore volume of the adsorbent; the diffusion, the affinity and the intermolecular interaction of the pollutant are determined by its molecular characteristics. Surface chemistry of an adsorbent and speciation of the pollutants may be altered by environmental factors like pH and temperature. Initial pollutant concentration has a direct impact on the concentration driving force for mass transfer while the flow rate and column height have an impact on residence time, breakthrough behavior, and overall treatment efficiency. It is known that such interacting variables have significant effects on the adsorption mechanisms and capacities over carbonaceous and advanced materials (Luo et al., 2022). Furthermore, machine learning modeling of biochar adsorption has revealed that the adsorption capacity relies on a complex function of the characteristics of the material and contaminant, instead of just one parameter (Jaffari et al., 2023).

With the growing number of adsorption datasets, there is a growing interest in using statistical and machine-learning techniques to predict treatment performance. Nonlinear adsorption relationships between two or more variables can be modeled using data-driven models, which cannot otherwise be captured using conventional empirical equations. Ensemble methods based on machine learning are very promising in predicting the efficiency of an adsorption process and identifying significant variables (Abbasi et al., 2022). In water management, AI can also aid in optimising the water treatment process, automation, and decision-making systems. Another recent study successfully applied machine learning in guiding adsorption experiments and predicting contaminant removal under different operating conditions (Jiang et al., 2024). Machine-learning models have also been used in the prediction of breakthrough behavior in granular activated carbon systems (Koyama et al., 2024).

In general, adsorption research has been conducted on one pollutant, one adsorbent or under restricted experimental conditions. These methods are not necessarily comprehensive when applied to the overall influence of structural, molecular, physico-chemical and operational parameters. While the removal capacity for the various advanced adsorbents has been demonstrated to be significant, the integrated predictive frameworks that incorporate a number of influencing factors are still rather limited. This requires the development of models that are physically meaningful and that assess material properties, pollutant properties and operating conditions all in one.

Thus, the aim of this study is to establish models to predict the capacity of organic pollutant adsorption under different material and physicochemical conditions using data. The aims are to describe the distributions of adsorption variables; to assess the relationships between adsorption capacity and predictor variables; to build and compare multiple models for predicting adsorption capacity; to find the most important variables determining the adsorption capacity; and to gain practical insights that help to optimize adsorption-based wastewater treatment systems.

2. MATERIALS AND METHODS

2.1 Study Design

This study used a quantitative data analysis design for model development to predict organic pollutant adsorption onto advanced adsorbent materials. The analytical framework integrated descriptive statistics, correlation analysis, machine learning regression, model validation, and explainable AI to predict the adsorption capacity under different material, physicochemical and operating conditions.

2.2 Dataset Description

A dataset with 6,172 numerical observations, eight predictor variables and one response variable. 101 duplicate records were located and deleted leaving 6,071 unique observations for modelling (Yang, 2025). There were no missing values in the data set. The adsorption capacity, Q (mg/g) was the target variable, and the specific surface area, pore volume, molecular characteristic, pH, temperature, initial concentration, flow rate and column height were used as predictors.

2.3 Study Variables

The adsorption capacity, Q (in mg/g) was selected as the dependent variable. Independent variables were specific surface area (SSA, m^2/g), pore volume (PV, cm^3/g), molecular characteristic (MC), pH, temperature (T , $^{\circ}\text{C}$), initial concentration (IC, mg/L), flow rate (FR, mL/min) and column height (CH, cm). These variables encompassed the following: adsorbent characteristics, pollutant characteristics, physicochemical conditions, and operating variables affecting the adsorption performance.

2.4 Data Preprocessing

The data set was examined for the presence of missing data, duplicate observations, invalid observations and extreme observations. To minimize information leakage, duplicate data were eliminated prior to the development of the models. Descriptive statistics and graphical methods were used to examine the distributions of the variables. Extreme values were kept if these were thought to be a valid experimental condition, as high adsorption capacities and concentrations can be indicative of data errors. For the models that are sensitive to scaling e.g. Support Vector Regression and K-Nearest Neighbours, the parameters were estimated from the training data only and applied to the test data.

2.5 Exploratory and Statistical Analysis

All variables were subjected to descriptive statistics such as mean, standard deviation, median, minimum and maximum. Variability and skewness were assessed using distribution plots. Pearson correlation was used for quantifying linear relationships between the predictors and adsorption capacity. Other data were explored using scatterplots to find possible nonlinear relationships between physicochemical parameters and Q .

2.6 Model Evaluation and Interpretation

Five regression models were tested: Multiple Linear Regression, Random Forest Regression, Extreme Gradient Boosting (XGBoost), Support Vector Regression, and K-Nearest Neighbors Regression. Multiple Linear Regression was used as a statistical baseline and the machine learning models were used to capture the non-linear relationships and interactions between the adsorption variables. The same 8 predictors were used throughout all the models to enable a performance comparison.

2.7 Model Training and Validation

The final dataset was split into an 80% training and 20% test set. For model fitting and hyperparameter optimization, the training set was used, and the testing set was an independent set for final evaluation. The training data set was subjected to 5-fold cross-validation to reduce bias in the model selection and to check the generalizability of the models. Hyperparameters of the used algorithms were optimized for the Random Forest, XGBoost, Support Vector Regression, and K-Nearest Neighbors.

2.8 Model Performance Evaluation

The coefficient of determination R^2 , mean absolute error, mean squared error and root mean squared error were used to assess the performance of the model. Good predictive performance was reflected in higher R^2 and lower MAE, MSE, RMSE values. The observed vs. predicted plots were also analyzed to evaluate the agreement between actual and predicted adsorption capacities.

2.9 Model Interpretation

The model with the highest performance was explained through feature-importance analysis and Shapley Additive Explanations (SHAP). The relative contribution of SSA, PV, MC, pH, temperature, initial concentration, flow rate and column height was ranked using feature importance. Next, the magnitude and direction of the individual predictor effects on the adsorption capacity were calculated using SHAP analysis. This allowed the identification of the most important material and physicochemical factors that control the adsorption of organic pollutants.

3. RESULTS

3.1 Dataset Characteristics and Descriptive Statistics

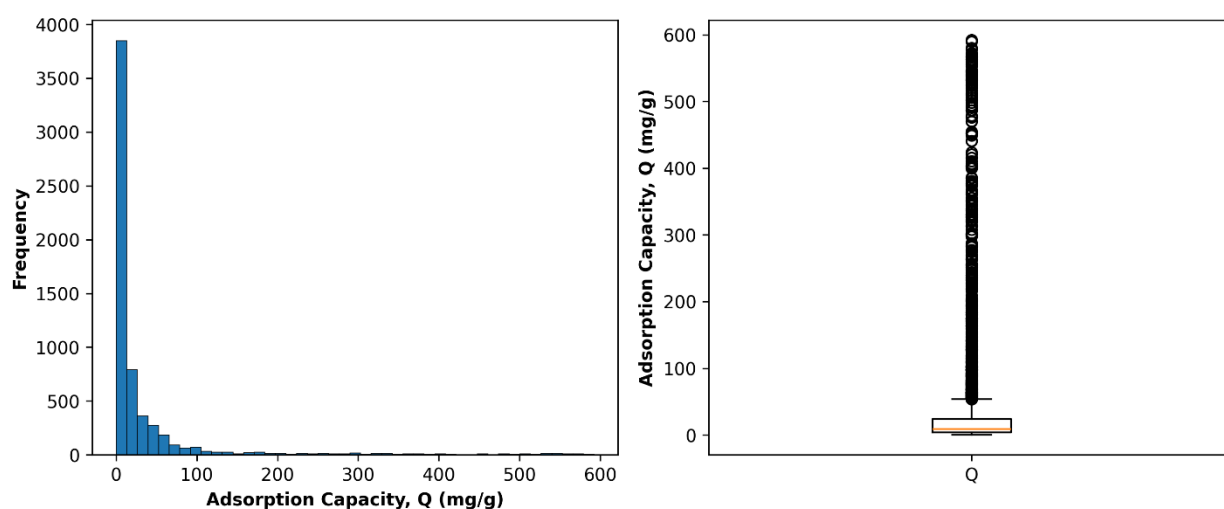
The final analytical dataset consisted of 6,071 unique observations that had no missing data. There was a significant difference in all material, physicochemical and operational parameters. Adsorption capacity (Q) ranged from 0.048 to 592.45 mg/g, with a mean of 33.12 ± 78.06 mg/g and a median of 8.91 mg/g. Specific surface area ranged from 0.43 to 2,311 m²/g and initial concentration varied from around 0.001 mg/L to 2,000 mg/L. Variations in the ranges of each variable were adequate to assess the adsorption behavior for a variety of experimental conditions. All variables used in the predictive analysis are summarized in Table 1, which presents descriptive statistics.

Table 1. Descriptive statistics of adsorption variables

Variable	Mean	SD	Median	Minimum	Maximum
SSA (m ² /g)	589.540	615.853	378.000	0.430	2311.000
PV (cm ³ /g)	0.648	0.312	0.590	0.165	2.600
MC	135.546	92.054	108.000	15.500	453.000
pH	5.367	1.045	5.000	1.000	11.000
T (°C)	25.671	4.989	25.000	4.500	60.000
IC (mg/L)	123.590	271.793	35.820	0.001	2000.000
FR (mL/min)	3.121	1.045	3.100	0.500	8.000
CH (cm)	9.499	1.205	9.500	1.000	15.000
Q (mg/g)	33.123	78.057	8.911	0.048	592.450

3.2 Distribution of Adsorption Capacity

The distribution of the observations of adsorption capacity showed a strong right skew, meaning that the majority of observations have been at relatively low Q values, and a smaller proportion of observations have been at much higher adsorption capacities. The mean adsorption capacity was significantly higher than the median value, which further supports the asymmetric distribution. The skewness value of Q is about 4.57, calculated. However, the high-capacity values were kept since they were considered to be possible adsorption conditions rather than definitely wrong measurements. Figure 1 shows the range of adsorption capacity in the data and its distribution.

**Figure 1. Distribution of adsorption capacity (Q) in the analyzed dataset**

3.3 Correlation among Adsorption Variables

Pearson correlation analysis showed that initial concentration was the most strongly positively linearly related to adsorption capacity ($r = 0.878$). Pore volume showed a weaker positive association with Q ($r = 0.204$), followed by SSA ($r = 0.102$), pH ($r = 0.090$), flow rate ($r = 0.063$), and temperature ($r = 0.061$). Column height showed a low negative correlation with Q ($r =$

-0.080) although MC had a weak negative correlation ($r = -0.047$). SSA and PV were the only predictors that intercorrelated moderately ($r = 0.510$). Table 2 indicates that the initial concentration was the most significant linear predictor of adsorption capacity.

Table 2. Pearson correlation matrix of adsorption variables

Variable	SSA	PV	MC	pH	T	IC	FR	CH	Q
SSA	1.000	0.510	0.118	0.399	0.068	0.053	-0.023	0.101	0.102
PV	0.510	1.000	-0.005	0.056	0.001	0.195	0.039	0.201	0.204
MC	0.118	-0.005	1.000	0.022	0.014	-0.065	-0.019	0.026	-0.047
pH	0.399	0.056	0.022	1.000	0.106	0.055	0.019	-0.044	0.090
T	0.068	0.001	0.014	0.106	1.000	0.065	-0.035	-0.007	0.061
IC	0.053	0.195	-0.065	0.055	0.065	1.000	0.077	-0.158	0.878
FR	-0.023	0.039	-0.019	0.019	-0.035	0.077	1.000	0.025	0.063
CH	0.101	0.201	0.026	-0.044	-0.007	-0.158	0.025	1.000	-0.080
Q	0.102	0.204	-0.047	0.090	0.061	0.878	0.063	-0.080	1.000

3.4 Relationships between Major Predictors and Adsorption Capacity

Graphical analysis verified that there was a considerable rise in adsorption capacity with a rise in initial concentration, but a certain fluctuation existed at similar concentrations. Conversely, the correlation of SSA, PV, pH, and temperature with Q was more nonlinear and heterogeneous. Its adsorption capacities were high over specific ranges of SSA and PV as opposed to simple monotonic variations. The same happened to pH and temperature with clusters of low and high Qs indicating that their influences were related to interactions with other adsorbent and operating characteristics. Figure 2 illustrates these contrasting predictor–response patterns and supports the use of nonlinear predictive algorithms.

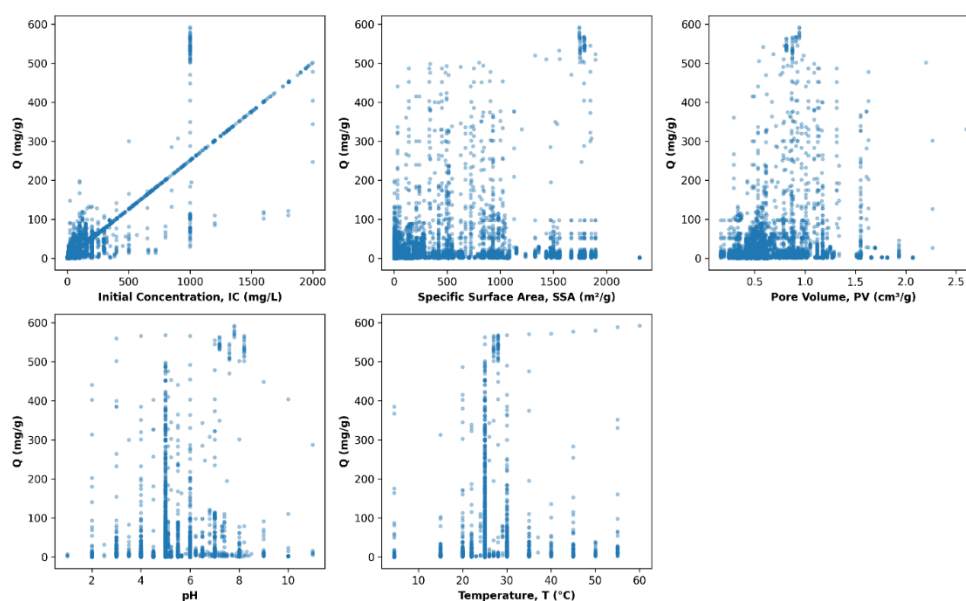


Figure 2. Relationships between major physicochemical predictors and adsorption capacity

3.5 Predictive Model Performance

The five regression models showed considerable differences in the prediction of the adsorption capacity. XGBoost had the best overall predictive performance with an R-squared of about 0.975 (~97.48% of the variance explained) and RMSE of 12.529 mg/g on the independent test dataset. Random Forest exhibited similar performance ($R^2 = 0.972$; RMSE = 13.257 mg/g) and had the lowest MAE of 3.842 mg/g. KNN and SVR also gave a good predictive accuracy with R^2 value of 0.959 and 0.950 respectively. Multiple Linear Regression was found to be far less successful ($R^2 = 0.792$), which means that the adsorption capacity was not well described by purely linear relationships. Table 3 indicates that the nonlinear ensemble models were the most predictive.

Table 3. Performance comparison of adsorption-capacity prediction models

Model	R^2	MAE (mg/g)	MSE	RMSE (mg/g)
XGBoost	0.975	5.096	156.976	12.529
Random Forest	0.972	3.842	175.753	13.257
K-Nearest Neighbors	0.959	6.962	258.122	16.066
Support Vector Regression	0.950	6.980	309.009	17.579
Multiple Linear Regression	0.792	14.078	1295.577	35.994

3.6 Observed versus Predicted Adsorption Capacity

Based on the highest R^2 value and lowest overall RMSE, XGBoost was used as the key predictive model. The observed and predicted adsorption capacities were tightly clustered around the line of perfect agreement over most of the range of response. The model was able to successfully mimic both the low and high adsorption capacities, but there were a few mismatches with individual high adsorption observations. Close agreement between experimental and XGBoost-predicted Q values is observed in Figure 3 which also shows an R^2 of 0.975.

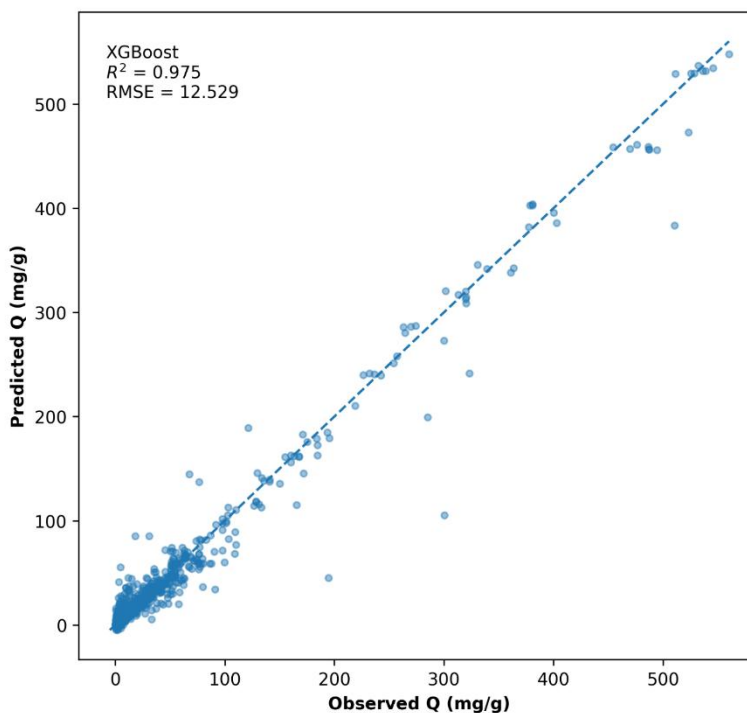


Figure 3. Observed versus XGBoost-predicted adsorption capacity

3.7 Feature Importance Analysis

XGBoost feature-importance analysis revealed that the initial concentration is the strongest predictive variable and explains around 65.40% of the importance of the model. Column height was second with 14.73 percent, specific surface area was third with 7.63 percent and molecular characteristic was fourth with 4.83 percent. The proportions contributed by pore volume, pH, flow rate and temperature were smaller. Notably, the fact that column height was relatively high with relatively weak Pearson correlation with Q indicates that its predictive value was largely nonlinear or dependent on interactions.

Table 4 presents the ranked contributions of the eight input variables to XGBoost predictions.

Table 4. Ranked XGBoost feature importance for adsorption-capacity prediction

Rank	Variable	Importance (%)
1	Initial concentration (IC)	65.404
2	Column height (CH)	14.729
3	Specific surface area (SSA)	7.630
4	Molecular characteristic (MC)	4.825
5	Pore volume (PV)	2.571
6	pH	2.306
7	Flow rate (FR)	1.355
8	Temperature (T)	1.181

3.8 Explainable Machine-Learning Analysis

SHAP analysis also sheds light on the direction and magnitude of the contributions of predictors. The concentration had the largest range of SHAP values and was definitely the most dominant impact on the model output. The SHAP values were mostly positive, with larger values corresponding to higher IC values, signifying significant increases in adsorption capacity. The effects of PV, MC, column height, pH, temperature, and flow rate were also generally less and more sensitive than SSA, but were still meaningful and positive at selected values. The data in Figure 4 showed that the material structure and operating conditions had both positive and negative contributions to the height of the column, supporting the fact that there are nonlinear effects that are not accounted for with simple correlation with respect to Q.

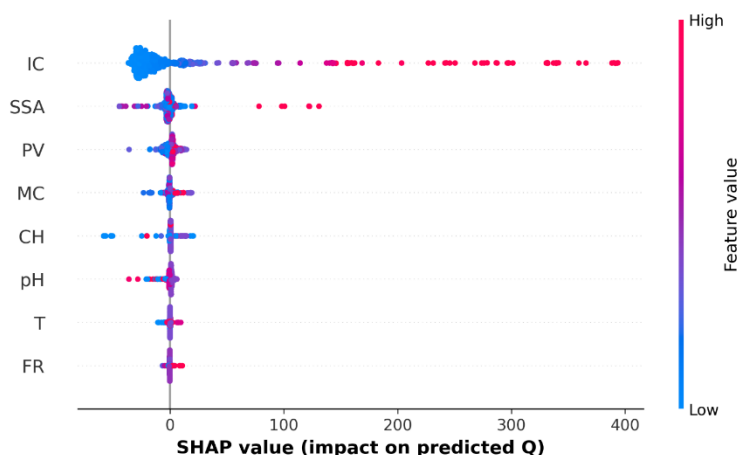


Figure 4. SHAP summary plot showing the magnitude and direction of predictor effects on XGBoost adsorption-capacity predictions

4. Discussion

The present study proved the ability of an integrated data-driven model containing the characteristics of the adsorbent, properties of the pollutant, physicochemical conditions, and the operational variables to predict the adsorption capacity. XGBoost had the highest predictive performance for all metrics, with an R^2 of 0.975 and RMSE of 12.529 mg/g, and Random Forest made equally good predictions. These results strengthen the use of machine learning for understanding complex adsorption systems with multiple variables that interact nonlinearly (Taoufik et al., 2022; Wang et al., 2024).

Specific surface area and pore volume, in general, were associated with the adsorption capacity as linear functions, but with less intensity than initial concentration. Under the proper conditions, the more available adsorption sites, the more pollutant uptake that can be achieved. SSA was among the most influential predictors in the XGBoost model. The adsorption behaviour also affected by pore volume by its effect on the accessibility and diffusion in the structure of the adsorbent. These mechanisms have also been observed for modified and functionalized biochars, where the pore structure and chemical modifications play a significant role in the adsorption ability (Qiu et al., 2022; Qu et al., 2023). Advanced graphene-based materials also have high surface area and surface properties (Thakur & Kandasubramanian, 2019).

The initial concentration was the most significant predictor, with the highest features contribution as well as the highest correlation to the adsorption capacity. As pollutant concentration increases, the driving force for mass transfer would increase, and more adsorption sites would be occupied until saturation is reached. The effect of molecular properties, pH, and temperature was relatively less significant but still important, as it influences the speciation of the pollutants, the surface charge, intermolecular interactions, and the energy of adsorption. This physicochemical dependence has been reported widely in biochar adsorption systems (Rajput et al., 2025; Tong et al., 2019). Machine learning research on the adsorption of pharmaceuticals has also demonstrated that joint molecular and environmental properties are helpful to interpret adsorption mechanisms (Nguyen et al., 2022; Sigmund et al., 2020). Simple linear relationships with flow rate and column height were also less important but were informative in the nonlinear models. Bed depth influences the contact with the adsorbent and the number of active sites, while the flow rate affects the residence time and the mass transfer opportunity. This type of behavior reinforces independent vs. jointly evaluating operational variables.

The performance degradation of Multiple Linear Regression in comparison to XGBoost and Random Forest clearly shows that the adsorption behavior was influenced by nonlinearity and higher-order interactions. Ensemble tree methods are more flexible in dealing with threshold effects, variable interactions and heterogeneous response patterns than linear regression. Other benefits of machine learning methods for sorption prediction in carbonaceous materials and for biochar adsorption-capacity estimation have been reported. Both feature-importance and SHAP analyses consistently showed that initial concentration was the most important factor for the predicted adsorption capacity, followed by the column height, SSA and molecular properties. High concentrations at the start generally raised predicted Q , while the other variables had mixed positive and negative effects based on their value and interaction. This validates the adsorption performance as a multifactor coupling effect as opposed to single-factor effects, which aligns with the machine-learning approach for the analysis of complex wastewater adsorption systems (Yang et al., 2025).

The results are consistent with previous work that has shown that high removal efficiencies can be obtained via a combination of optimal surface area, porosity, and chemical functionality of the adsorbents. MOFs have demonstrated excellent adsorption capacity for pharmaceuticals, dyes and other environmental pollutants (Rasheed, Bilal, et al., 2020; Rasheed, Hassan, et al., 2020; Uddin et al., 2021). Based on these findings, the present results show that predictive modelling can be used to measure the combined structural and process contribution.

The proposed framework can be used for screening of adsorbents, optimization of operating conditions and initial prediction of adsorption performance prior to in-depth experimentation. These data-driven techniques could minimize the amount of experimentation required and help design more efficient wastewater treatment systems by identifying variables that are

significant and likely operating ranges for the treatment system. Future work should include more adsorbent classes and/or more pollutant-specific descriptors, regeneration properties, adsorption kinetics, and externally validated, larger datasets to further strengthen the generalizability and mechanistic interpretation of the models.

5. Conclusion

This study showed that data-driven adsorptive modeling is effective for predicting the adsorption capacity of organic pollutants under different adsorptive, physicochemical and operational conditions. The analysis involved 6,071 unique observations and eight predictor variables and revealed that the nonlinear machine-learning models significantly outperformed conventional linear regression. For predictive performance, XGBoost was the best overall algorithm evaluated with an R^2 of 0.975 and RMSE of 12.529 mg/g followed by the Random Forest algorithm which also had very competitive results. The most important factors affecting the adsorption capacity were the initial concentration of pollutants, column height, specific surface area, and molecular properties. SHAP analysis also showed that higher initial concentrations had a positive influence on the prediction of adsorption capacity for the most part, while the other variables had more context-dependent and nonlinear effects. The results support the findings that adsorption performance is caused by the synergistic effect of material properties, pollutant characteristics and treatment conditions, and not by single factors. The developed framework thus be used to aid in quick design estimation, decrease experimental efforts, and aid in optimizing adsorption wastewater treatment systems. The lack of identity categories for adsorbent and pollutant, and the limited number of descriptors, however, prevent wider generalization. More diverse materials and pollutants, molecular descriptors and externally validated datasets should be included in future studies for a more robust, interpretable and applicable model.

References

1. Abbasi, E., Moghaddam, M. R. A., & Kowsari, E. (2022). A systematic and critical review on development of machine learning based-ensemble models for prediction of adsorption process efficiency. *Journal of Cleaner Production*, 379, 134588.
2. Aris, A. Z., Hir, Z. A. M., & Razak, M. R. (2020). Metal-organic frameworks (MOFs) for the adsorptive removal of selected endocrine disrupting compounds (EDCs) from aqueous solution: A review. *Applied Materials Today*, 21, 100796.
3. Awad, A. M., Jalab, R., Benamor, A., Nasser, M. S., Ba-Abbad, M. M., El-Naas, M., & Mohammad, A. W. (2020). Adsorption of organic pollutants by nanomaterial-based adsorbents: An overview. *Journal of Molecular Liquids*, 301, 112335.
4. Baig, N., Sajid, M., & Saleh, T. A. (2019). Graphene-based adsorbents for the removal of toxic organic pollutants: A review. *Journal of Environmental Management*, 244, 370–382.
5. Cheng, N., Wang, B., Wu, P., Lee, X., Xing, Y., Chen, M., & Gao, B. (2021). Adsorption of emerging contaminants from water and wastewater by modified biochar: A review. *Environmental Pollution*, 273, 116448.
6. Dai, Y., Zhang, N., Xing, C., Cui, Q., & Sun, Q. (2019). The adsorption, regeneration and engineering applications of biochar for removal organic pollutants: A review. *Chemosphere*, 223, 12–27.
7. De Andrade, J. R., Oliveira, M. F., Da Silva, M. G. C., & Vieira, M. G. A. (2018). Adsorption of Pharmaceuticals from Water and Wastewater Using Nonconventional Low-Cost Materials: A Review. *Industrial & Engineering Chemistry Research*, 57(9), 3103–3127. <https://doi.org/10.1021/acs.iecr.7b05137>
8. Huang, L., Shen, R., & Shuai, Q. (2021). Adsorptive removal of pharmaceuticals from water using metal-organic frameworks: A review. *Journal of Environmental Management*, 277, 111389.

9. Ighalo, J. O., Yap, P.-S., Iwuozor, K. O., Aniagor, C. O., Liu, T., Dulta, K., Iwuchukwu, F. U., & Rangabhashiyam, S. (2022). Adsorption of persistent organic pollutants (POPs) from the aqueous environment by nano-adsorbents: A review. *Environmental Research*, 212, 113123.
10. Jaffari, Z. H., Jeong, H., Shin, J., Kwak, J., Son, C., Lee, Y.-G., Kim, S., Chon, K., & Cho, K. H. (2023). Machine-learning-based prediction and optimization of emerging contaminants' adsorption capacity on biochar materials. *Chemical Engineering Journal*, 466, 143073.
11. Jiang, S., Hou, Y., Man, Z., Wang, C., Shi, X., Shang, J., & Cheng, X. (2024). Guiding experiment with Machine Learning: A case study of biochar adsorption of Ciprofloxacin. *Separation and Purification Technology*, 334, 126023.
12. Jin, E., Lee, S., Kang, E., Kim, Y., & Choe, W. (2020). Metal-organic frameworks as advanced adsorbents for pharmaceutical and personal care products. *Coordination Chemistry Reviews*, 425, 213526.
13. Koyama, Y., Fasaee, M. A., Berglund, E. Z., & Knappe, D. R. (2024). Machine learning models to predict early breakthrough of recalcitrant organic micropollutants in granular activated carbon adsorbers. *Environmental Science & Technology*, 10–1021.
14. Luo, Z., Yao, B., Yang, X., Wang, L., Xu, Z., Yan, X., Tian, L., Zhou, H., & Zhou, Y. (2022). Novel insights into the adsorption of organic contaminants by biochar: A review. *Chemosphere*, 287, 132113.
15. Nguyen, X. C., Ly, Q. V., Nguyen, T. T. H., Ngo, H. T. T., Hu, Y., & Zhang, Z. (2022). Potential application of machine learning for exploring adsorption mechanisms of pharmaceuticals onto biochars. *Chemosphere*, 287, 132203.
16. Qiu, B., Shao, Q., Shi, J., Yang, C., & Chu, H. (2022). Application of biochar for the adsorption of organic pollutants from wastewater: Modification strategies, mechanisms and challenges. *Separation and Purification Technology*, 300, 121925.
17. Qu, J., Meng, Q., Peng, W., Shi, J., Dong, Z., Li, Z., Hu, Q., Zhang, G., Wang, L., & Ma, S. (2023). Application of functionalized biochar for adsorption of organic pollutants from environmental media: Synthesis strategies, removal mechanisms and outlook. *Journal of Cleaner Production*, 423, 138690.
18. Rajput, P., Yadav, S., Liu, C., & Balasubramanian, P. (2025). Predicting biochar adsorption capacity for methylene blue removal using machine learning. *Journal of Water Process Engineering*, 69, 106749.
19. Rasheed, T., Bilal, M., Hassan, A. A., Nabeel, F., Bharagava, R. N., Ferreira, L. F. R., Tran, H. N., & Iqbal, H. M. (2020). Environmental threatening concern and efficient removal of pharmaceutically active compounds using metal-organic frameworks as adsorbents. *Environmental Research*, 185, 109436.
20. Rasheed, T., Hassan, A. A., Bilal, M., Hussain, T., & Rizwan, K. (2020). Metal-organic frameworks based adsorbents: A review from removal perspective of various environmental contaminants from wastewater. *Chemosphere*, 259, 127369.
21. Sigmund, G., Gharasoo, M., Hüffer, T., & Hofmann, T. (2020). Deep learning neural network approach for predicting the sorption of ionizable and polar organic pollutants to a wide range of carbonaceous materials. *Environmental Science & Technology*, 54(7), 4583–4591.
22. Taoufik, N., Boumya, W., Achak, M., Chennouk, H., Dewil, R., & Barka, N. (2022). The state of art on the prediction of efficiency and modeling of the processes of pollutants removal based on machine learning. *Science of The Total Environment*, 807, 150554.
23. Thakur, K., & Kandasubramanian, B. (2019). Graphene and Graphene Oxide-Based Composites for Removal of Organic Pollutants: A Review. *Journal of Chemical & Engineering Data*, 64(3), 833–867. <https://doi.org/10.1021/acs.jced.8b01057>
24. Tong, Y., McNamara, P. J., & Mayer, B. K. (2019). Adsorption of organic micropollutants onto biochar: A

-
- review of relevant kinetics, mechanisms and equilibrium. *Environmental Science: Water Research & Technology*, 5(5), 821–838.
25. Uddin, M. J., Ampia, R. E., & Lee, W. (2021). Adsorptive removal of dyes from wastewater using a metal-organic framework: A review. *Chemosphere*, 284, 131314.
26. Wang, Z., Wang, Q., Yang, F., Wang, C., Yang, M., & Yu, J. (2024). How machine learning boosts the understanding of organic pollutant adsorption on carbonaceous materials: A comprehensive review with statistical insights. *Separation and Purification Technology*, 350, 127790.
27. Yang, Z. (2025). Adsorption Dataset. 1. <https://doi.org/10.17632/3ncczsrzh8.1>
28. Yang, Z., Liu, Y., Wang, Z., Yang, R., Lei, J., Zhang, Y., Zhang, A., Liu, Z., & Li, Z. (2025). Machine learning-assisted simulated annealing for deciphering multi-factor coupling mechanisms in complex wastewater adsorption systems with experimental validation. *Environmental Research*, 122826.