

UC Santa Barbara

UC Santa Barbara Previously Published Works

Title

Interpretable-generative machine learning approaches for predicting simultaneous removal of organic pollutants and heavy metals from water by adsorbent materials

Permalink

<https://escholarship.org/uc/item/306730bd>

Journal

Journal of Cleaner Production, 532

ISSN

0959-6526

Authors

Zhai, Mudi

Wu, Zhaozhong

Fu, Bomin

et al.

Publication Date

2025-11-01

DOI

10.1016/j.jclepro.2025.146980

Copyright Information

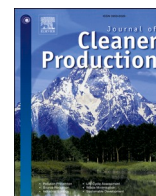
This work is made available under the terms of a Creative Commons Attribution License, available at <https://creativecommons.org/licenses/by/4.0/>

Peer reviewed



Contents lists available at ScienceDirect

Journal of Cleaner Production

journal homepage: www.elsevier.com/locate/jclepro

Interpretable-generative machine learning approaches for predicting simultaneous removal of organic pollutants and heavy metals from water by adsorbent materials

Mudi Zhai^{a,b,1}, Zhaozhong Wu^{c,1}, Bomin Fu^{d,*}, Jingzhang Sun^e, Mohamad Sleiman^f, Frederic C. Meunier^g, Junsen Wang^a, Weijie Wang^a, Tianrun Wang^a, Haoran Duan^b, Zisheng Ai^h, Jose Luis Valverdeⁱ, Anne Giroir-Fendler^g, Jean-Marc Chovelon^g, Arturo A. Keller^j, Hongtao Wang^{a,k,**}

^a College of Environmental Science and Engineering, Key Laboratory of Yangtze River Water Environment, Ministry of Education, Tongji University, Siping Rd 1239, Shanghai, 200092, China

^b School of Civil and Environmental Engineering, University of New South Wales, Sydney, NSW, 2052, Australia

^c School of Computer Science, Northwestern Polytechnical University, Xi'an, 710129, China

^d Key Laboratory of Agro-Forestry Environmental Processes and Ecological Regulation of Hainan Province, School of Environmental Science and Engineering, Hainan University, Haikou, 570228, China

^e School of Cyberspace Security, Hainan University, Haikou, Hainan, China

^f Université Clermont Auvergne, CNRS, ICCF, F-63000, Clermont-ferrand, France

^g Université Claude Bernard Lyon 1, IRCELYON, CNRS UMR 5256, 2 avenue Albert-Einstein, Villeurbanne, F-69626, France

^h Department of Medical Statistics, Tongji University School of Medicine, Shanghai, China

ⁱ Department of Chemical Engineering, University of Castilla La Mancha, Avda. Camilo José Cela 12, Ciudad Real, 13071, Spain

^j Bren School of Environmental Science and Management, University of California, Santa Barbara, CA, 93106, United States

^k Shanghai Institute of Pollution Control and Ecological Security, Shanghai, 200092, China

* Corresponding author. Key Laboratory of Agro-Forestry Environmental Processes and Ecological Regulation of Hainan Province, School of Environmental Science and Engineering, Hainan University, Haikou, 570228, China.

** Corresponding author. College of Environmental Science and Engineering, Key Laboratory of Yangtze River Water Environment, Ministry of Education, Tongji University, Siping Rd 1239, Shanghai, 200092, China.

E-mail addresses: fubomin@hainanu.edu.cn (B. Fu), hongtao@tongji.edu.cn (H. Wang).

¹ Co-first authors: these authors contributed equally to this work.

<https://doi.org/10.1016/j.jclepro.2025.146980>

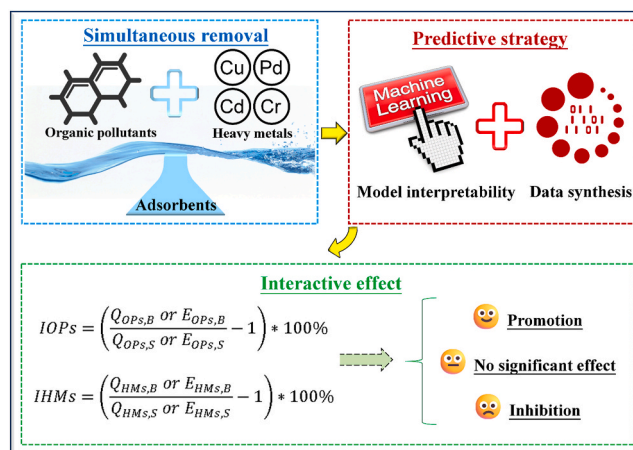
Available online 2 November 2025

0959-6526/© 2025 Elsevier Ltd. All rights are reserved, including those for text and data mining, AI training, and similar technologies.

HIGHLIGHTS

- Adsorbents capable of simultaneously removing organic pollutants and heavy metals.
- Simultaneous removal mainly includes inhibition, promotion and no significant effect.
- CatBoost model shows its better predictive performance and model interpretability.
- WGAN method assists the stability and classification ability of CatBoost model.

GRAPHICAL ABSTRACT



ARTICLE INFO

Keywords:

Combined contamination
Ensemble learning model
Shapley additive explanation
Feature importance
Generative adversarial network

ABSTRACT

This study constructed a dataset containing commonly used adsorbents for the simultaneous removal of organic pollutants (OPs) and heavy metals (HMs) from aqueous phases, and evaluated machine learning models to predict the interactions between OPs and HMs on the adsorbents. Data synthesis using Wasserstein generative adversarial networks (WGAN) and feature importance interpretation via Shapley additive methods were analyzed. The results revealed that categorical boosting (CatBoost) emerged as the best machine learning model due to its excellent adsorption prediction performance and model interpretability. The interactive effects were classified into “inhibition”, “no significant effect” and “promotion” categories. The original dataset was computationally augmented with 1500 reliable synthetic data by WGAN, improving the precision for “promotion” category of CatBoost from 0.79 to 0.93. Adsorbent surface area and solution pH were the primary factors causing the “promotion” of adsorptive behavior. Bridging and electrostatic interaction mechanisms may play important roles in promoting simultaneous adsorption. The strategy of combining interpretable machine learning and data synthesis with the WGAN method, not only provides new insights into the prediction of the composite adsorption behavior of OPs and HMs by adsorbents in the aqueous phase, but also addresses the problem of insufficient experimental data to improve the prediction performance.

1. Introduction

The combined pollution of organic pollutants (OPs) and heavy metals (HMs) is routinely found in medical, livestock, industrial and municipal wastewater (Zhai et al., 2023). The complexes formed by OPs and HMs are generally more stable and more difficult to remove by sewage treatment plants and the natural environment. These OP-HM mixtures may pose greater risks than individual effects in water systems, the so-called “cocktail” effect. Thus, the development of effective removal technologies is essential to minimize the hazards induced by OP-HM combinations. Adsorption technology can concentrate and separate pollutants through attractive forces between adsorbents and adsorbates. It offers the advantages of high efficiency and economy, easy operation and no secondary pollution, as well as the adsorbents can be recovered and recycled, which makes it a promising means of simultaneously removing OPs and HMs from water (Zhai et al., 2024). Adsorption behavior is generally evaluated by batch equilibrium experiments under various operating conditions (Al-Ghouti and Da’ana, 2020). However, this evaluation process is time-consuming and costly (Taoufik et al., 2022).

With the rapid development of artificial intelligence, machine learning techniques provide new tools for improving the efficiency of processing large datasets (Lin and Zhang, 2025; Wei et al., 2025; Xu et al., 2024). Earlier studies have confirmed the effectiveness of machine learning in predicting the adsorption performance of OPs or HMs on

different adsorbents (Palansooriya et al., 2022; Zhu et al., 2022). A gradient boosting decision tree model based on machine learning showed satisfactory performance ($R^2 > 0.90$) in predicting the adsorption of tetracycline antibiotics by biomass charcoal (Zhou et al., 2023). Leng et al. constructed a hybrid machine learning model for biomass production applications that yielded an R^2 as high as 0.996 for the prediction of adsorption of HMs by biochar (Leng et al., 2025). Nevertheless, most machine learning studies still focus on a single system, and there is a lack of systematic studies on the prediction of adsorption behaviors of adsorbent materials for pollutant mixtures of OPs and HMs. Moreover, if experimental data are scarce, the predictive ability of the machine learning model can hardly be guaranteed.

Several approaches have been developed, such as synthetic data strategy, transfer learning, active learning and label propagation, to solve the problem of data scarcity (Koç et al., 2025). Among those, generative adversarial networks (GANs) have become a research hotspot because of their outstanding performance in generating synthetic data. GANs were first proposed in 2014, and they can learn from a set of training data and generate synthetic data that strongly resemble to original data (Chakraborty et al., 2024). The performance of machine learning models can be improved by supplementing the training dataset with synthetic data generated by GANs. Currently, GAN-related research has been applied in many fields, such as biomedicine, computer vision, financial markets, information security, image processing, audio recognition and generation (Lv et al., 2024), but there is a lack of

attention in the environmental protection field, especially in the context of material-based wastewater treatment.

The performance of machine learning models was evaluated in the present work, to determine the best model for removal prediction. A Wasserstein-based GAN (WGAN) approach was employed to augment the original dataset involving the removal of OPs and HMs from aqueous phases by typical adsorbents, thereby further optimizing the model. In addition, feature importance calculations and Shapley additive explanation (SHAP) analysis were performed to reveal the main conditions and related mechanisms affecting the adsorption behavior. This study provides new insights for the treatment of wastewater contaminated with both OPs and HMs, as well as for the rational selection and design of environmentally functional adsorbent materials.

2. Materials and methods

2.1. Dataset construction

A total of 63 papers on related topics, including 1787 data points were collected from the Web of Science to construct a dataset. The related adsorbent materials included biochars (BC), activated carbons (AC), resins (RES), metal-organic frameworks (MOF), cyclodextrins (CD), and cellulose (CEL). The HMs considered were Cu^{2+} , Cd^{2+} , Pd^{2+} , and Cr^{6+} . There were 24 different OPs, including antibiotics, non-steroidal anti-inflammatory drugs, pesticides, chemical by-products and environmental hormones. The missing data patterns in the dataset are summarized in Table S1, and detailed descriptions of data pre-processing are provided in Text S1.

Three types of features were considered as input parameters: adsorbent characteristics (specific surface area (SSA), elemental composition (C%, H%, O%, N%), pH_{pzc}, total pore volume and pore diameter), environmental characteristics (initial pH, temperature,

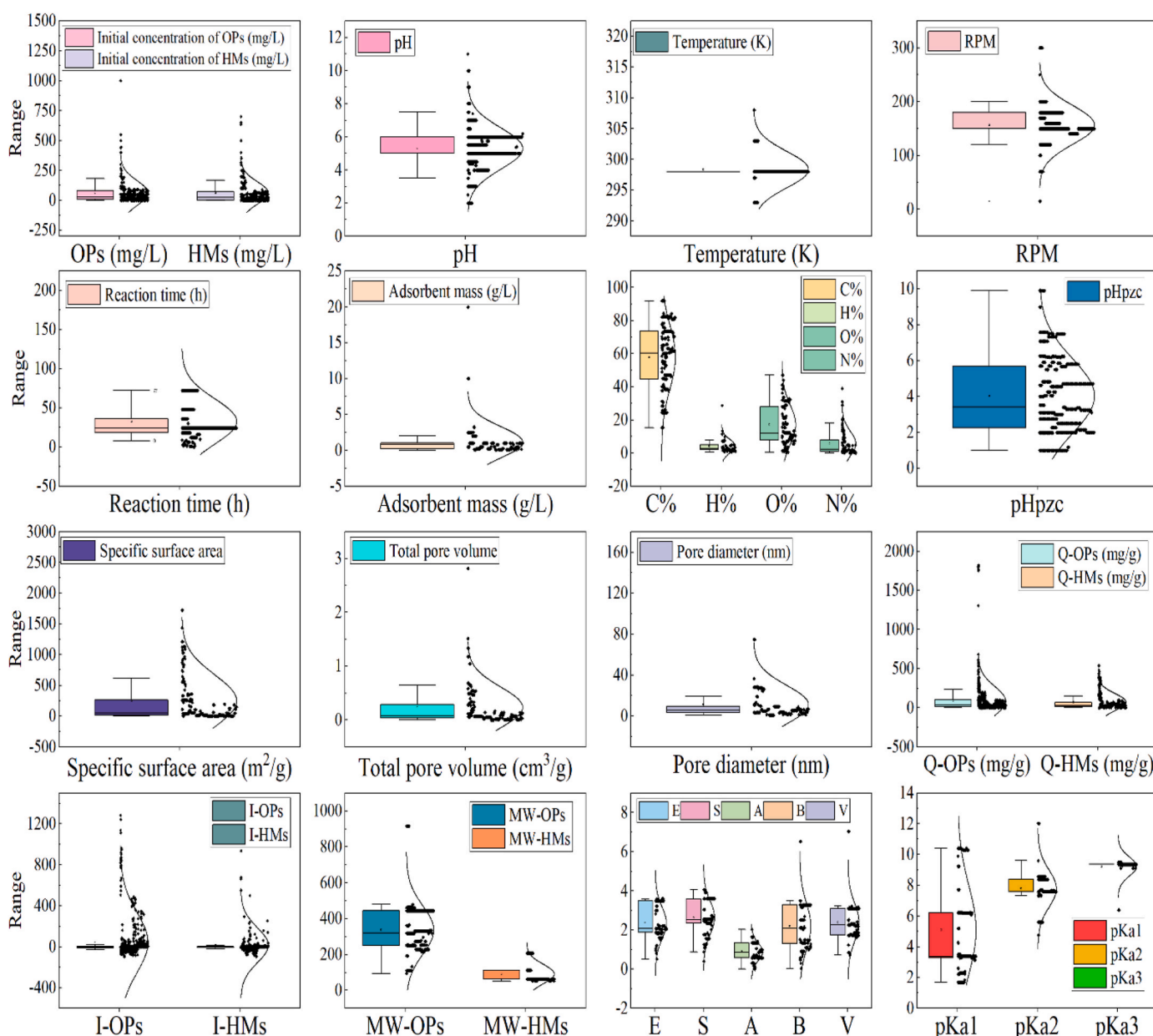


Fig. 1. Box plot of the collected data (E: molar refractivity, S: polarizability, A: hydrogen bond acidity, B: hydrogen bond basicity, V: molar volume).

revolutions per minute (RPM), adsorbent mass, reaction time, initial concentration of OPs and HMs), and pollutant characteristics (OPs and HMs)). The characteristics of OPs were molecular weight (MW-OPs), hydrogen bond acidity (A), hydrogen bond basicity (B), polarizability (S), molar refractivity (E), molar volume (V) and dissociation constant (pKa). The characteristics of HMs consisted in the molecular weight (MW-HMs), electronegativity (X), ionic radius (r), and valence. The adsorption capacity for OPs (QOPs), the adsorption capacity for HMs (QHMs), the influence of HMs on OP adsorption (IOPs), and the influence of OPs on HM adsorption (IHMs) were set as output parameters. The characteristic distributions of input and output parameters are described in Text S2 and in Fig. 1 and S1.

The adsorption capacity (Q) and removal efficiency (U) were calculated via equations (1) and (2), respectively:

$$Q = \frac{(C_0 - C_e)V}{m} \quad (1)$$

$$U = \frac{(C_0 - C_e)}{C_0} \quad (2)$$

where C_0 is the initial pollutant concentration, C_e is the equilibrium concentration, V is the reaction volume, and m is the mass of the adsorbent.

To quantify the interactive effects of OPs and HMs, two percentage indices of IOPs and IHMs were employed and obtained via equations (3) and (4), respectively:

$$IOPs = \left(\frac{Q_{OPs,B} \text{ or } U_{OPs,B}}{Q_{OPs,S} \text{ or } U_{OPs,S}} - 1 \right) * 100\% \quad (3)$$

$$IHMs = \left(\frac{Q_{HMs,B} \text{ or } U_{HMs,B}}{Q_{HMs,S} \text{ or } U_{HMs,S}} - 1 \right) * 100\% \quad (4)$$

where $Q_{OPs,B}$ and $Q_{OPs,S}$ represent the adsorption capacities of OPs in binary and single systems, respectively, whereas $Q_{HMs,B}$ and $Q_{HMs,S}$ embody the adsorption capacities of HMs in binary and single systems, respectively; IOPs indicate the effect of HMs on OP adsorption in a binary system, whereas the IHMs show the effect of OPs on HM adsorption. Positive values of IOPs or IHMs indicate that the adsorption of OPs or HMs onto the adsorbent is promoted by the other pollutant. In contrast, negative values indicate adsorption inhibition by the other pollutant. Nil values of IOPs or IHMs indicate no effect.

2.2. Development and evaluation of machine learning models

In this study, models that have been well performed for adsorption prediction in recent years were selected, including multiple linear regression (MLR), five ensemble learning models, including extreme gradient boosting (XGBoost), light gradient boosting machine (LGBM), categorical boosting (CatBoost), stacking and random forest (RF), as well as two deep learning (DL) models, including back propagation neural network (BPNN) and one dimensional convolutional neural network (1D-CNN) (Abbasi et al., 2022). A brief description of these models can be found in Text S3. To avoid biases caused by scale effects, all the input parameters were standardized using Z score normalization. The dataset was randomly assigned into two parts: 80 % of it was arranged as a training set for the purpose of model development, and the remaining 20 % was used as a validation set to evaluate the model.

Bayesian optimization (Fig. S19 and S20) and five-fold cross-validation were employed to tune the hyperparameters of the models (Tables S17 and S18). Regarding the regression tasks for predicting the adsorption capacity, the coefficient of determination (R^2) and root mean square error (RMSE) were used to measure the predictive performance. For the classification tasks predicting interactive effects, model performance was evaluated in terms of accuracy, precision, recall, and F1 score. The calculation processes of the above evaluation indicators are

shown in Text S4.

2.3. Interpretability analysis of machine learning models

Recursive feature elimination (RFE) was employed to select the best features for the machine learning models. Based on the calculated importance of features in the tree, RFE with cross-validation removes the least important features each time, and can automatically find the number of features that make the model perform optimally. SHAP values of the features were used to analyze the specific impact of features on the output parameters. To further explore feature relationships, a partial dependency plot (PDP) was applied to reveal the nature of dependence between the inputs and outputs. Detailed information on the SHAP and PDP methods has been described in previous studies (Greenwell, 2017; Vimbi et al., 2024; Zhang et al., 2025a, 2025b).

2.4. WGAN method for data synthesis

WGAN is an unsupervised learning model that trains two neural networks, including a generator and a discriminator in an adversarial manner to generate synthetic data. The Wasserstein distance can be regarded as the minimum “cost” required to transfer one “item” to another between two distributions, based on the following formula (Jiang et al., 2024):

$$W(p_g, p_r) = \inf_{\gamma \in \Pi(p_g, p_r)} \mathbb{E}_{(x,y) \sim \gamma} [\|x - y\|] \quad (5)$$

where p_g is the generated distribution, p_r is the real distribution, and $\Pi(p_g, p_r)$ is the set of all possible joint distributions $\gamma(x, y)$. The goal of the WGAN is to minimize the Wasserstein distance between the generated distribution and the true distribution. The original and generated data were mapped into a two-dimensional space using t-distributed stochastic neighbor embedding (t-SNE) to explore the similarity of the data distributions (Van der Maaten and Hinton, 2008).

3. Results and discussion

3.1. Performance comparison of machine learning models

The predictive performances of eight machine learning models for adsorption capacity of QOPs and QHMs are shown in Fig. 2 and Tables S2 and S3. MLR model performed the worst in terms of the prediction of QOPs and QHMs, with testing set R^2 values of 0.60 and 0.63, respectively. Fig. S2 and S3 reveal a significant deviation between the actual and predicted values via MLR model, with errors increasing as the

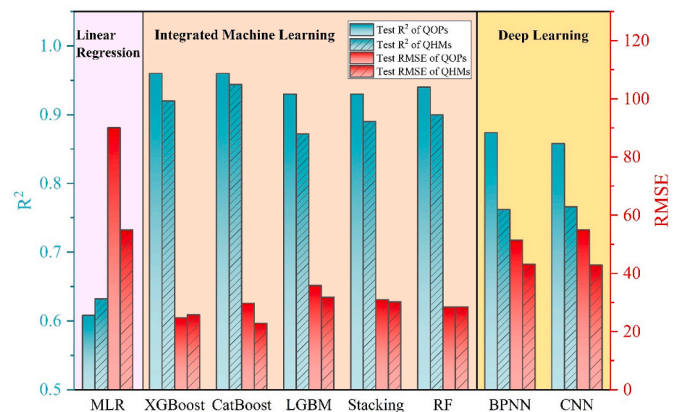


Fig. 2. Predictive performance of eight machine learning models for adsorption capacity of QOPs and QHMs. R^2 and RMSE are the average results from five-fold cross-validation of the models.

prediction values increased. This finding indicates that MLR model struggled to capture the nonlinear relationships within the data. BPNN and 1D-CNN models also did not achieve satisfactory predictive performance. The R^2 values of the DL models for the test sets of QOPs and QHMs were less than 0.85 and 0.75, respectively. Similar to those of MLR model, the errors of ANN and 1D-CNN models increased significantly with increasing prediction values. This may be due to the inability of small sample data to meet the requirements of DL models for capturing complex data structures.

In contrast, the ensemble learning models exhibited better performance in predicting QHMs and QOPs, with average R^2 values from five-fold cross-validation for both the training and test sets exceeding 0.85, with a smaller error between the actual and predicted values (Fig. S4 and S5). The predictive performance of the models for QOPs was generally better than that for QHMs. Regarding the QOP prediction, XGBoost and CatBoost achieved the highest R^2 values (0.96 for both), even though XGBoost had a slightly lower RMSE. For the prediction of QHMs, CatBoost significantly outperformed the other models. As illustrated in Fig. 3, the predicted values of CatBoost model for QOPs and QHMs closely matched the actual values, with residuals distributed near the zero line, indicating a strong predictive ability. This superior performance of CatBoost may be related to the dataset size and its efficient categorical feature handling method. One-hot encoding for many OP categories can easily lead to a dimensional explosion of features. CatBoost employs an integrated categorical feature handling method that can automatically and efficiently process categorical features. This method effectively prevents feature dimensionality explosion in a small dataset while retaining the intrinsic information of categorical features, thereby yielding superior results on the smaller QHM dataset (530) compared to the QOP dataset (1,015) (Hancock and Khoshgoftaar, 2020).

3.2. Feature analysis of adsorbent removal ability

Owing to the increased complexity caused by the inclusion of numerous variables, the key features influencing the prediction of QOPs and QHMs were selected using feature importance and the RFE. The feature importance of the two ensemble learning models (XGBoost and CatBoost) that have demonstrated better predictive performance was examined. Fig. S6 shows the top ten feature importance rankings of XGBoost for QOPs and QHMs. XGBoost model heavily relied on the activated carbon feature in adsorbent category, which accounted for 93 % of the overall feature importance for QOPs. In addition, it mainly

relied on the atomic mass of HMs (76 %), leading to the neglect of other features in QHMs. This over-reliance on certain features may be due to the coverage-based selection criterion of XGBoost, which tends to select features with high frequency in nodes, thereby overestimating their importance (Nguyen et al., 2021).

In contrast, the CatBoost model had a more balanced distribution of feature importance, as shown in Fig. 4. Except for the adsorbent feature contributing 44 % of the total feature importance, the environmental features, including initial OP concentration and adsorbent mass also contributed 48 % for QOPs. Although the proportion of environmental features of QHMs (40 %) was close to that of QOPs, the former exhibited a stronger pollutant feature dependence of 43 %, with electronegativity (x) being the most important feature accounting for 37 %. The visualization of RFE process for CatBoost model is illustrated in Fig. S7. The prediction performance of QOPs and QHMs was found to be degraded after the removal of eight and three features, respectively. Therefore, the feature importance of the model was calculated after removing these features.

Feature significance provides insight into the relationships between features and output variables, but is still insufficient to reveal their specific effects. To address this, SHAP and PDP analyses were employed to explore the deeper impact on features of QOPs and QHMs based on CatBoost model, and the results are shown in Fig. S8 and 9. The initial pollutant concentrations of OPs and HMs exhibited similar trends in their impacts on QOPs and QHMs. Both the SHAP values and the predicted QOPs and QHMs increased markedly with increasing initial pollutant concentrations. An increase in initial concentration led to the occupation of otherwise unsaturated sites at the surface of the adsorbent, thereby leading to an increase in adsorption capacity. However, as the concentration continued to rise, the sorption sites on the adsorbent surface became gradually saturated, causing the capacity to reach an asymptote value (Aziz et al., 2022). A similar trend was observed for the concentrations of coexisting pollutants. The predicted values of QOPs increased with increasing initial HM concentration in the range of 0–120 mg/L, but leveled off above 120 mg/L. For the QHMs, the adsorption no longer increased after the initial OP concentration exceeded 110 mg/L. In general, appropriate initial concentrations of OPs and HMs may promote mutual adsorption through a bridging effect. Once the initial concentration continues to increase, especially beyond the saturation point of the adsorbent, the adsorption sites may become clogged, allowing the competitive effect to outweigh the bridging effect, resulting in a weakening of the “promotion” in the adsorption process (Zhai et al., 2023).

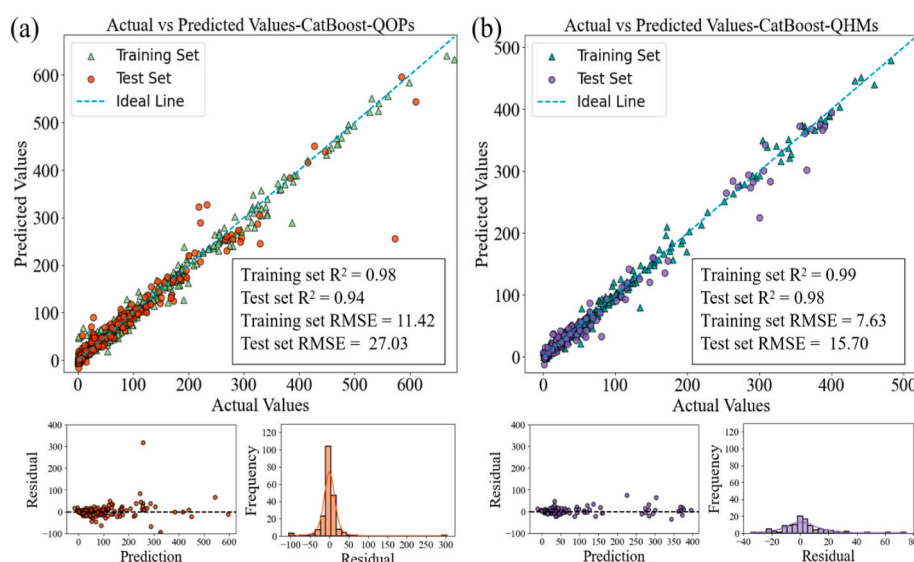


Fig. 3. Actual and predicted values of CatBoost model for QOPs (a) and QHMs (b).

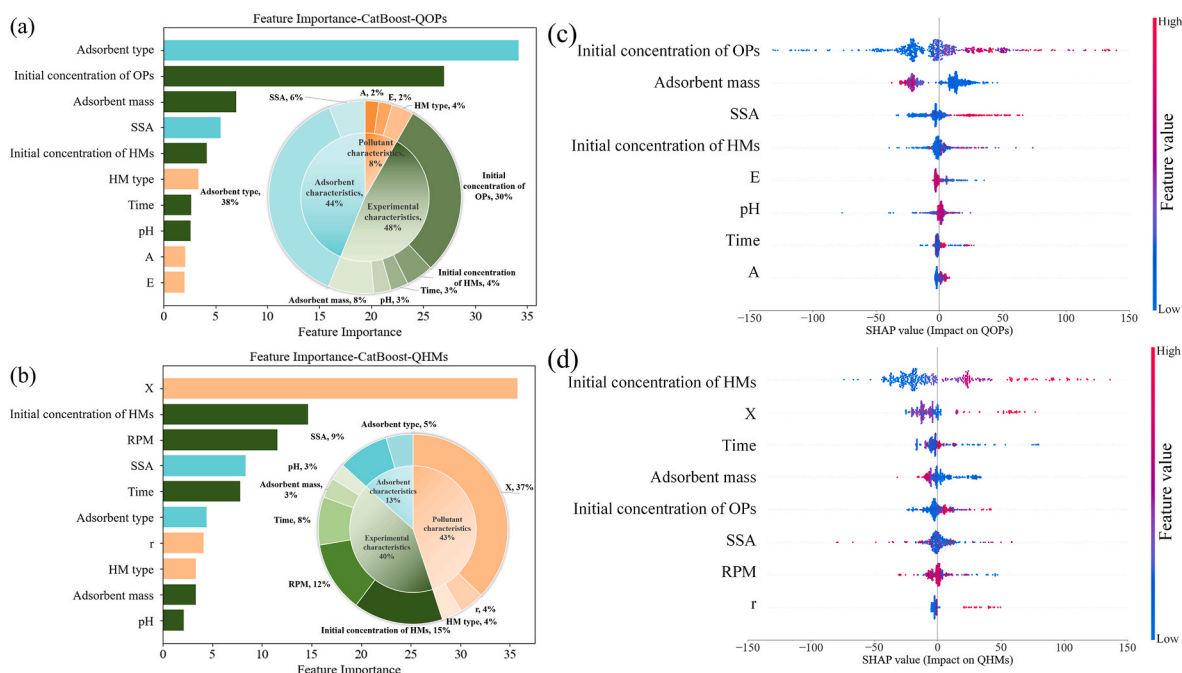


Fig. 4. Feature importances of CatBoost and SHAP scatter plots for QOPs (a, c) and QHMs (b, d).

The SSA of adsorbent had a positive effect on the adsorption capacity. The larger the SSA of adsorbent is, the greater the number of sorption sites per unit mass of adsorbent, and the greater the adsorption capacity for QOPs. However, the adsorbent mass had a significant negative effect on both QOPs and QHMs, causing the adsorption capacity to decrease with increasing adsorbent mass. Although a higher adsorbent mass improves the pollutant removal efficiency, it leads to a decrease in pollutant adsorption per unit mass of adsorbent. The effects of pH on QOPs showed multiplicity: both the SHAP and the predicted values of QOPs increased rapidly with pH in the range of pH 2.5–6.5, but decreased gradually when pH exceeded 6.5. On the contrary, pH had a significant positive effect on QHMs, which increased with increasing pH. The OPs considered in this study are dissociable and exist in various ionic forms depending on the pH of the aqueous solution, which makes the effect of pH on OP adsorption very complicated. 90 % of the HMs are cations, and thus most of the electrostatic attraction between adsorbents, OPs and HMs increases with pH. Higher electronegativity demonstrated a positive effect on the adsorption process as it not only reflected the electron affinity of HMs, but also enhanced the electrostatic interactions with OPs and adsorbent surfaces, which facilitated adsorption. The RPM also gained relatively high importance for QHMs, but the trend of its effect on the PDP and SHAP values was not regular. The effect of adsorption time on QOPs was also not regular. In addition, the effects of E and A on QOPs and QHMs were weak.

3.3. Transformation of prediction task for interactive effect

To reveal the interactive effects between OPs and HMs within binary systems, IOPs and IHMs were quantified. Given the outstanding performance in both adsorption prediction and feature analysis, CatBoost model was chosen to predict the performance of adsorbent materials for IOPs and IHMs. The five-fold cross-validation results of CatBoost model on the training and test sets are shown in Table S4. The predictive ability of IOPs and IHMs was significantly weaker than that of QOPs and QHMs, especially on the test set, where the performance was poor with an average R^2 of less than 0.5. Compared with the adsorption capacity of a single system, the interactive effects in a binary system are usually more complex, and it is difficult for machine learning models to capture and predict the intrinsic relationship, resulting in very poor predictive

stability of IOPs and IHMs.

To address the poor stability of predictions, the numerical prediction task was transformed into a classification task by assigning a certain range and a clear meaning to IOPs and IHMs. When IOPs or IHMs were greater than 5%, OPs/HMs were significantly promoted in a binary system, and the “Promotion” category was used; between –5% and 5%, it indicated that they were not significantly affected, so they were assigned to the “no significant effect” category; and for IOPs or IHMs lower than –5%, the combined adsorption was significantly inhibited, so they were assigned to the “inhibition” category. The adsorption effect classification of HMs on OPs is designated as ICOP in a binary system, and ICHM denotes the adsorption effect of OPs on HMs.

The results of CatBoost model on the classification task are shown in Table S5, and a threshold-sensitivity study as shown in Tables S15 and S16. The predictive stability of CatBoost model for both ICOPs and ICHMs was significantly improved, with an average evaluation metric of more than 0.7 on the test set. The predictive performance for ICOPs was slightly better than that for ICHMs. The accuracy, precision recall and F1 score of the model for ICOPs are 0.83, 0.83, 0.82 and 0.82, respectively, compared with values of 0.75, 0.75, 0.74 and 0.74 for ICHMs.

The prediction performance of CatBoost for three categories of interactive effects is shown in Table 1. For ICOPs, the prediction ability of “no significant effect” (e.g., F1 score = 0.76) was weaker than that of the other two categories (F1 scores of 0.85 for both “inhibition” and “promotion”). The predictions of “no significant effect” (F1 score = 0.75) and “promotion” (F1 score = 0.72) in ICHMs were weaker than that of “inhibition” (F1 score = 0.82). In particular, the recall rate of the

Table 1
Five-fold cross-validation average results for classification task of CatBoost model.

Model	Metrics (testing set)	Inhibition	No significant effect	Promotion
ICOPs	Precision	0.88 ± 0.012	0.77 ± 0.013	0.83 ± 0.011
	Recall	0.83 ± 0.014	0.75 ± 0.015	0.88 ± 0.012
	F1 Score	0.85 ± 0.011	0.76 ± 0.013	0.85 ± 0.011
ICHMs	Precision	0.81 ± 0.012	0.75 ± 0.013	0.79 ± 0.012
	Recall	0.84 ± 0.011	0.75 ± 0.014	0.68 ± 0.015
	F1 Score	0.82 ± 0.012	0.75 ± 0.012	0.72 ± 0.013

test set for “promotion” category of ICHMs was 0.68, which was significantly lower than the precision of 0.79. This suggests that the model is overly conservative in predicting the “promotion” category and missed many samples that belong to the “promotion” category. The difference in the prediction performance of various categories may be caused by the imbalance and limited size of the dataset.

3.4. WGAN-based dataset expansion strategy

The comparison of traditional GAN and WGAN on the prediction performance is shown in Table S6. The accuracy of the GAN-generated data in the range of 500–1500 data fluctuated between 0.39 and 0.58, which is significantly lower than that of the WGAN (0.75–0.90). Traditional GAN often faces gradient vanishing and pattern collapse problems during training, and the difference between generated and real data measured by Jensen-Shannon divergence is not sensitive enough in high-dimensional spaces, leading to unstable training and lower quality of generated data (Park et al., 2022). In contrast, WGAN provides smoother and more meaningful gradients due to the better mathematical properties of the incorporated Wasserstein distance, thus significantly improving the stability of model training, mitigating the phenomenon of

pattern collapse, and improving the diversity and quality of the generated data (Gao et al., 2020).

To further optimize the generation performance of WGAN, the variation of Wasserstein distance with the number of training rounds in ICOPs and ICHMs was evaluated. Fig. 5 illustrates the changes in the Wasserstein distance during the WGAN training process. In the initial phase (0–2000 iterations), the discriminator was able to easily distinguish between the virtual and original data because of their significant differences, resulting in a large Wasserstein loss. As the training progressed (approximately 2000–4000 iterations), the generator gradually learned to produce more realistic data, making it difficult for the discriminator to distinguish between real and generated data, and leading to a decreasing trend in Wasserstein loss. When it reached the range of 4000–7000 iterations, the Wasserstein loss increased again to a high value and then gradually decreased, indicating that the adversarial training between the generator and discriminator was ongoing. After approximately 8000 iterations, the Wasserstein loss gradually stabilized at a lower level, suggesting that the generator had learned to generate relatively realistic data, and that the discriminator had difficulty in distinguishing between real and generated data. At this point, a balanced state between the generator and discriminator was achieved

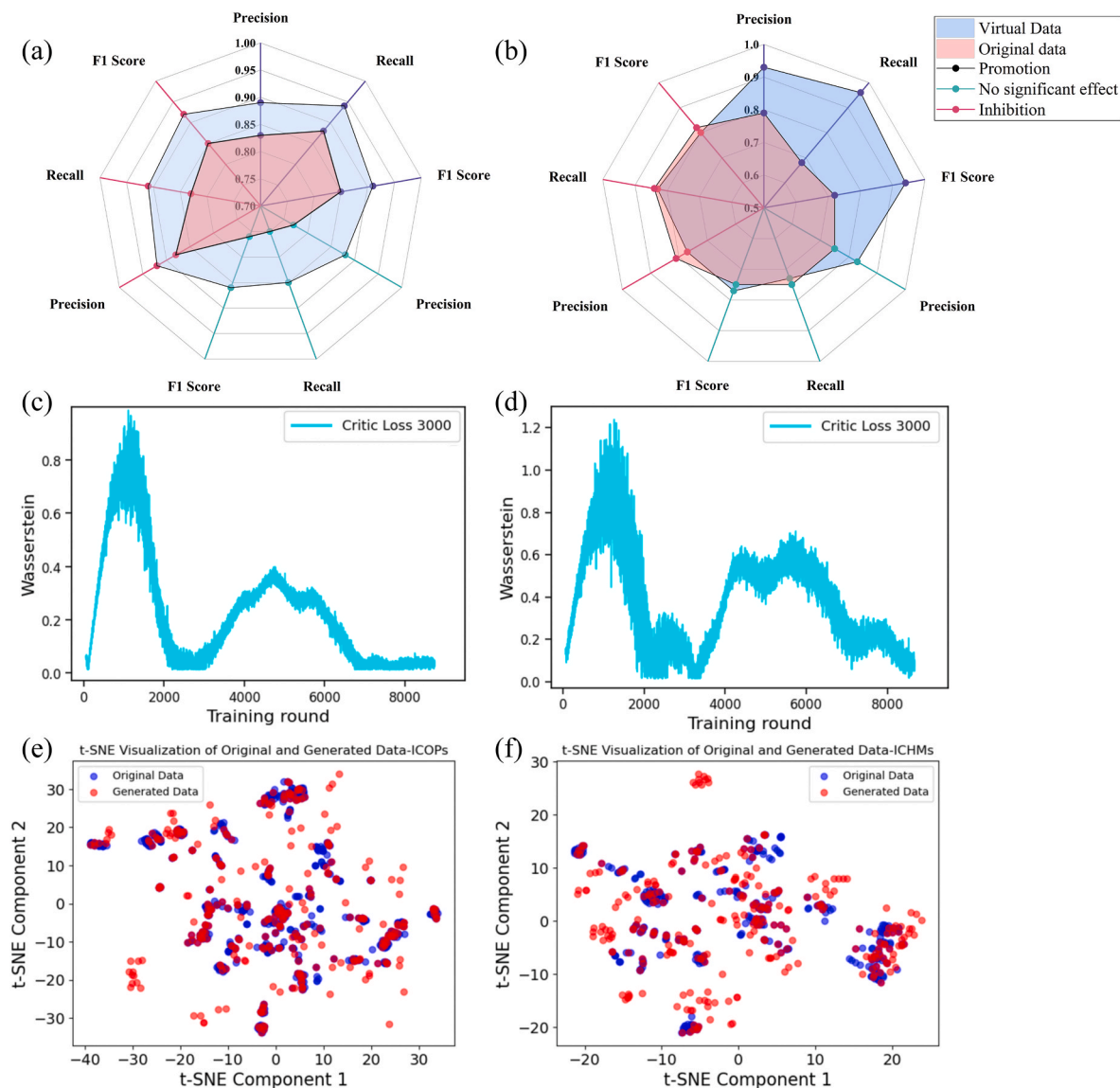


Fig. 5. Model performance radar chart for real original and virtual data for ICOPs (a) and ICHMs (b); variation of Wasserstein distance with the number of WGAN training rounds for ICOPs (c) and ICHMs (d); t-SNE distributions of generated and original data for ICOPs (e) and ICHMs (f).

(Xiong et al., 2020).

After completing the iterative training, the properties of the data generated by WGAN were verified with those of the original data to ensure the quality of the synthetic data. The t-SNE visualization reveals

that the generated data (red points) effectively covered the blue points represented as original data in the space, indicating that the WGAN is capable of generating data with a distribution similar to the original data. Additionally, the generated data filled the areas where the original

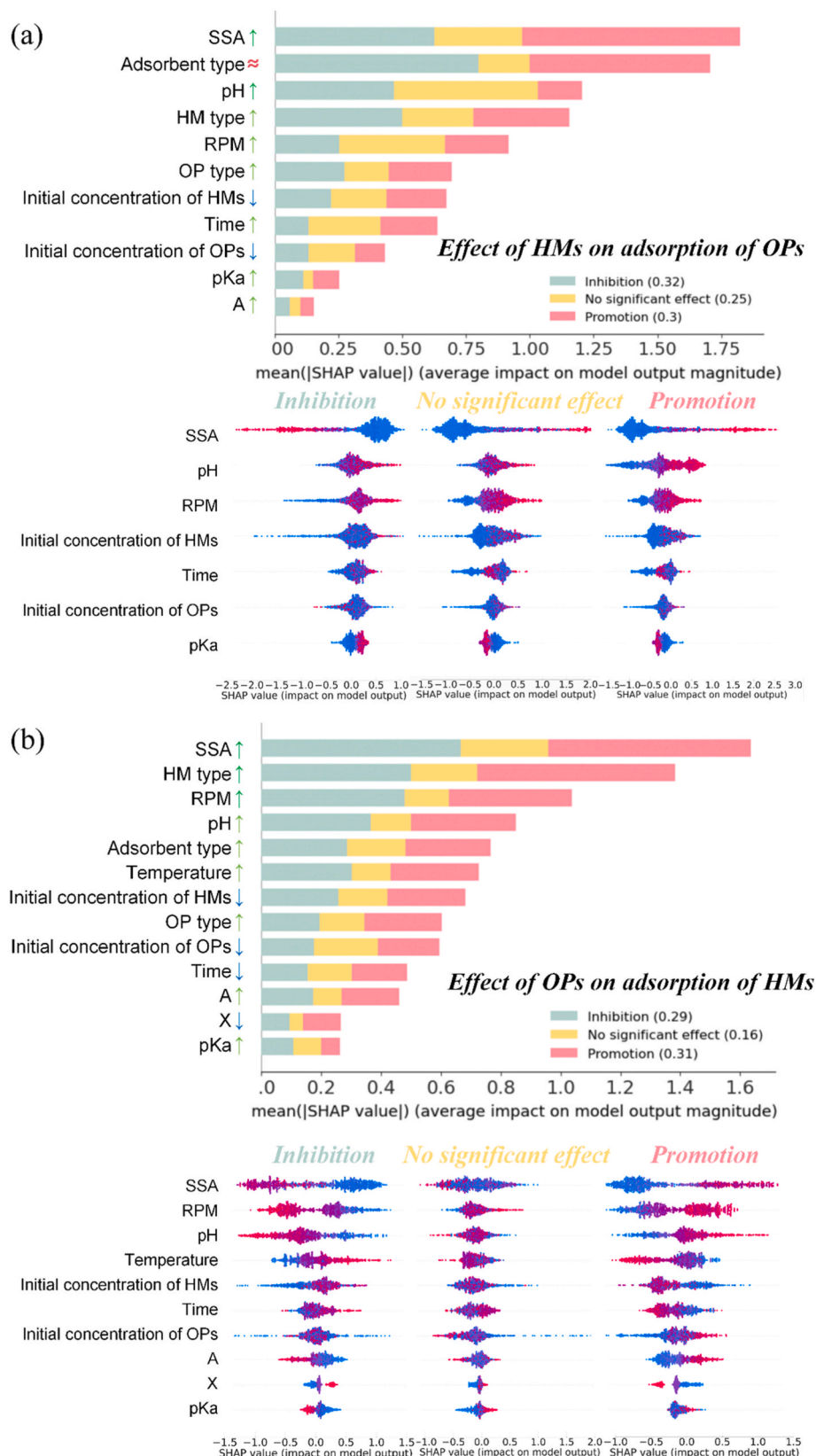


Fig. 6. Average absolute values and scatter plots of SHAPs on “inhibition”, “no significant effect” and “promotion” categories for ICOPs (a) and ICHMs (b).

data were not distributed, demonstrating that the WGAN can generate new data distributions based on the original data distribution, thereby increasing the diversity of the data (Liu et al., 2024; Zhai et al., 2025). Additionally, we conducted quantitative checks including descriptive statistics and Kolmogorov-Smirnov tests as shown in Table S11–14, Fig. S17 and S18. For ICHM, feature descriptive statistics and two-sample KS tests showed no differences at the 5 % level, with only minor changes in means and variances. For ICOP, several features differed from the original, including initial concentrations of OP and HM, RPM, time, adsorbent mass, specific surface area, predicted Q-HM and Q-OP, and some differences were notably reduced, such as adsorbent mass by approximately 83 % and Q-OP by approximately 36 %. Overall, the augmented data covered the main regions of the original samples but modified some marginal distributions.

The synthetic data randomly generated by WGAN were combined with the original data to construct an expanded dataset. The effect of dataset containing different numbers of synthetic data on the performance of the optimal machine learning model (CatBoost) was explored. As shown in Tables S7 and S8, the incorporation of synthetic data from 500 to 1500 resulted in an improvement of the average F1 scores of CatBoost for ICOPs and ICHMs from 0.80 to 0.75 to 0.90 and 0.84, respectively. This indicates that the incorporation of sufficient WGAN-generated data as the training set boosted the performance of the machine learning model. In contrast, the CatBoost model performance showed a decreasing trend when adding between 1500 and 2500 synthetic data. This may be because the distribution of synthetic data generated by weighted clipping has difficulty matching the real data, resulting in training bias caused by excessive and diverse synthetic data, thereby reducing the accuracy of the machine learning model (Lee et al., 2021). Therefore, 1500 was selected as the most suitable number of synthetic data supplements for expanding the dataset, as it helped CatBoost to perform better in all the metrics.

Tables S9 and 10 demonstrate the CatBoost model performance before and after dataset expansion. The predictive ability of the model in each category (“inhibition”, “no significant effect”, “promotion”) in the ICOPs was improved, especially the precision, recall and F1 score for the “no significant effect” were improved from 0.77, 0.75 and 0.76, respectively, to 0.88, 0.84 and 0.86 after adding the 1500 synthetic data generated by WGAN. Regarding the ICHMs, although the “inhibition” and “no significant effect” categories did not result in a significant improvement for model performance after expanding the dataset, the precision, recall, and F1 score in the “promotion” category significantly improved from 0.79, 0.68, and 0.72 to 0.93, 0.96, and 0.94, respectively. This shows that the data synthesis based on WGAN can serve to solve the class imbalance problem in the dataset and further enhance the stability and classification ability of CatBoost model (Liu et al., 2019).

3.5. Feature analysis for interactive effects of OPs and HMs onto adsorbents

Fig. S10 illustrates the RFE process based on 5-fold cross-validation for ICOPs and ICHMs of CatBoost model. The model performance decreased after 14 and 12 features were removed from ICOPs and ICHMs, respectively. The number of features removed by RFE was greater than that of the single system. The feature importance of three categories of interactive effects was quantified using mean absolute SHAP value and is shown in Fig. 6, S11 and S12. The “promotion” and “inhibition” categories had similar proportions of feature importance for ICOPs and ICHMs, whereas the proportion of “no significant effect” features was smaller than those of the other two categories. With the exception of the “no significant effect” category in ICOPs, the feature with largest proportion in the three categories of ICOPs and ICHMs was SSA. Compared to the single system, SSA was more important for predicting the interactive effect of the binary system, and was at the top of the list of feature importance for ICOPs and ICHMs. As SSA increased, the SHAP values of ICHMs decreased in the “inhibitory” category, but

increased in the “promotional” category. ICOPs showed a similar trend overall, but with both positive and negative SHAP values at lower SSA. The value of SSA is proportional to the number of adsorption sites. Fewer adsorption sites indicated that OPs and HMs competed fiercely for adsorption onto the adsorbent, thus inhibiting adsorption. With the increase of SSA, more OPs or HMs can be adsorbed at more adsorption sites, which further promotes adsorption through bridging effect (Zhai et al., 2023). As shown in Fig. 7a, the surface of AC is governed by an electron polarizable cloud that may interact favorably with pollutants via non- or low-polar groups, dispersive and van der Waals forces.

Compared with the single system, pH was also more important in the binary system and ranked third and fourth in the feature importance of ICOPs and ICHMs, respectively. In particular, pH was the feature with highest percentage in “no significant effect” category of ICOPs. The SHAP values of “no significant effect” and “promotion” categories for ICOPs showed decreasing and increasing performance with increasing pH, respectively. This suggests that HMs have a weak effect on OP adsorption under acidic conditions, and promote adsorption in alkaline environments. pH is closely correlated with electrostatic interactions during the adsorption process. For example, under the influence of pH, protons are exchanged between AC and aqueous media, generating positively and negatively charged AC derivatives (Fig. 7b), which provide electrostatic interaction sites for pollutants. OPs were deprotonated with increasing pH, causing them to exist as more negatively charged anions in solution. This leads to enhanced electrostatic attraction between anionic OPs and cationic HMs (e.g., Cu^{2+} , Pb^{2+} , and Cd^{2+}), which increases the potential for complexation between OPs and HMs. Moreover, cationic HMs generally show stronger adsorption under neutral or slightly alkaline conditions, allowing more HMs to be immobilized onto the adsorbent surface, and providing more adsorption sites for OPs (Zhai et al., 2024). This could be further confirmed by the influence of pKa on OP adsorption. When pKa was increased, the SHAP values of ICOPs increased in the “inhibition” category but decreased in the “promotion” category, indicating that pKa negatively affects the adsorption capacity of adsorbent material.

The two-dimensional PDP for the initial concentrations of ICOPs and ICHMs are shown in Fig. S13 and S14. The initial pollutant concentration was less important in the binary system compared to the single system, and slightly more important in HMs than in OPs. The maximum predicted probability of “no significant effect” category in ICOPs and ICHMs occurred at high OP and low HM concentrations. This is consistent with the conventional understanding that the adsorption process of a pollutant at a high concentration is less affected by the presence of another pollutant at a low concentration. Since the initial concentration of HMs was more important than other factors, the trend of impacts in the two-dimensional PDP was found to depended mostly on this concentration. As the initial concentration of HMs increased, the predicted probability of “inhibition” category in ICOPs increased, whereas the probability of “promotion” category showed a tendency of increasing and then decreasing. Similarly, the predicted probability of “promotion” category for ICHMs increased and then decreased with increasing HM concentration. This may be related to the difference in chemical properties between HMs and OPs. The adsorption of HMs onto adsorbent materials may be stronger than OPs, so that the initial HM concentration dominates the interactive effect in binary systems.

Fig. S15 and S16 show the SHAP value distributions for the features of adsorbents, OPs and HMs. There was no significant difference in the ranking of the importance for adsorbent features between single and binary systems. Compared with that of ICOPs, the range of SHAP values for the six adsorbents of ICHMs was smaller and not significantly different. Especially in the “promotion” category, AC had the lowest mean SHAP value, and the other adsorbents had similar mean values less than 0.8. RES, MOF, and AC had greater mean SHAP values than BC, CD, and CEL in the “promotion” category of ICOPs, whereas only BC and CEL had values greater than zero in the “inhibition” category. This indicates RES, MOF and AC have positive effects on OP adsorption in the presence

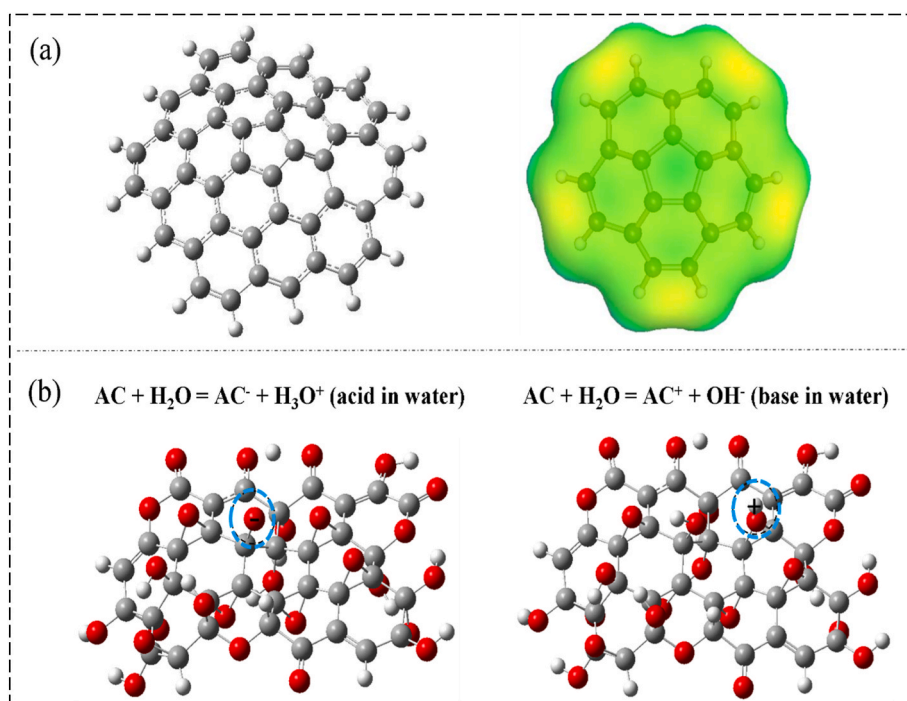


Fig. 7. Polarized charge distribution at the surface (a) and single-charged derivatives (b) of activated carbon (AC). The black, white and red models represent carbon, hydrogen and oxygen atoms, respectively. AC^- and AC^+ mean they carry negative and positive charges during proton exchange. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

of HMs, which may be related to the SSA of the adsorbent. MOF and AC tend to have larger SSA, which is consistent with the SHAP analysis. In the “no significant effect” category of adsorbents, MOF and CD had greater mean values, indicating that their positive effects were not significant compared with those of the other adsorbents. This may be related to the difference in the adsorption mechanism of OPs and HMs on CD and MOF. For example, the cavity structure inside the CD makes it very suitable for adsorbing OPs through hydrophobic or π - π interactions, whereas the functional groups on its surface can adsorb HMs by electrostatic or coordination interactions. This difference in adsorption mechanism leads to a decrease in competition and bridging between OPs and HMs, resulting in the “no significant effect” (Usman et al., 2021; Zhao et al., 2015). The ordering of HMs in ICOPs and ICHMs was similar, both exhibiting as $Cd^{2+} > Cr^{6+} > Cu^{2+} > Pb^{2+}$. The distribution differences in ICHMs were greater than those in ICOPs in terms of OPs. In addition, atrazine and bisphenol A had SHAP values less than 0 for “inhibition” category but greater than 0 for “promotion” category, suggesting that they facilitated the adsorption of adsorbent materials.

4. Conclusion

The coexistence of OPs and HMs in aquatic environments poses significant risks, threatening these ecosystems and public health. Although adsorption processes have demonstrated effectiveness in controlling this combined contamination, several challenges remain, including time-consuming development of adsorbent materials, suboptimal adsorption efficiency, and incomplete understanding of adsorption mechanisms for combined OP-HM pollution in the aqueous phase.

In this study, a dataset containing common adsorbent materials for the removal of OPs and HMs from water was constructed. Machine learning models were applied to predict the adsorption capacity of adsorbents and the interactive effects of OPs and HMs in the binary system. CatBoost was determined to be the best prediction model due to its excellent adsorption capacity prediction performance and good model interpretability. Moreover, the model performance of CatBoost was further significantly improved by predicting task transformation

combined with dataset expansion based on the WGAN method. The “inhibition”, “no significant effect” and “promotion” were defined as the main interactive effect categories of OPs and HMs onto adsorbents. SSA and solution pH were found to be important parameters influencing the “promotion” of combined adsorption. Mechanisms such as bridging and electrostatic interactions may dominate the process by which adsorption is promoted.

This work provides new insights for the investigation of adsorption behavior in complex pollution environments in the aqueous phase. It can predict the adsorption capacity and interaction effects in a multi-component adsorption system by minimizing the number of experiments, and can also predict the promotion and inhibition behavior in the interaction effects through model feature feedback. Furthermore, the strategy based on the combination of interpretable machine learning and data synthesis with WGAN method, not only reduces the cost and time of experimental process, but also addresses the issue of insufficient data. Its application in the field of material-based treatment of water and wastewater can help to rapidly optimize the design of environmentally functional adsorbent materials, and select appropriate working conditions, thereby improving the combined removal efficiency of OPs and HMs from contaminated wastewater.

CRediT authorship contribution statement

Mudi Zhai: Writing – original draft, Visualization, Software, Methodology, Investigation, Formal analysis, Data curation. **Zhaozhong Wu:** Visualization, Validation, Software, Methodology, Data curation. **Bomin Fu:** Writing – review & editing, Validation, Supervision, Resources, Conceptualization. **Jingzhang Sun:** Writing – review & editing, Validation. **Mohamad Sleiman:** Writing – review & editing, Validation. **Frederic C. Meunier:** Writing – review & editing, Validation. **Junsen Wang:** Validation, Data curation. **Weijie Wang:** Validation, Data curation. **Tianrun Wang:** Validation, Data curation. **Haoran Duan:** Writing – review & editing. **Zisheng Ai:** Writing – review & editing. **Jose Luis Valverde:** Writing – review & editing. **Anne Giroir-Fendler:** Writing – review & editing. **Jean-Marc Chovelon:** Writing – review &

editing. **Arturo A. Keller:** Writing – review & editing. **Hongtao Wang:** Writing – review & editing, Supervision, Resources, Project administration, Funding acquisition.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This study was funded by the National Natural Science Foundation of China (No. 42407061, 52270188 and 72348002). The authors are very grateful to Professor Peng Wang from Sun Yat-Sen University for his helpful comments on the manuscript.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.jclepro.2025.146980>.

Data availability

Data will be made available on request.

References

- 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. *J. Clean. Prod.* 379, 134588.
- Al-Ghouti, M.A., Da'ana, D.A., 2020. Guidelines for the use and interpretation of adsorption isotherm models: a review. *J. Hazard Mater.* 393, 122383.
- Aziz, F.F.A., Jalil, A.A., Hassan, N.S., Fauzi, A.A., Azami, M.S., Jusoh, N.W.C., Jusoh, R., 2022. A review on synergistic coexisting pollutants for efficient photocatalytic reaction in wastewater remediation. *Environ. Res.* 209, 112748.
- Chakraborty, T., Ks, U.R., Naik, S.M., Panja, M., Manvitha, B., 2024. Ten years of generative adversarial nets (GANs): a survey of the state-of-the-art. *Mach. Learn. Sci. Technol.* 5, 011001.
- Gao, X., Deng, F., Yue, X., 2020. Data augmentation in fault diagnosis based on the Wasserstein generative adversarial network with gradient penalty. *Neurocomputing* 396, 487–494.
- Greenwell, B.M., 2017. Pdp: an R Package for Constructing Partial Dependence Plots.
- Hancock, J.T., Khoshgoftaar, T.M., 2020. CatBoost for big data: an interdisciplinary review. *J. Big Data* 7, 94.
- Jiang, Y., Li, J., Yang, X., Yuan, R., 2024. Applications of generative adversarial networks in materials science. *Materials. Genome. Eng. Adva.* 2, e30.
- Koç, C., Özyurt, F., Iantovics, L.B., 2025. Survey on Latest Advances in Natural Language Processing Applications of Generative Adversarial Networks, vol. 15. Wiley Interdiscip Rev Data Min Knowl Discov, e70004.
- Lee, S., Kim, J., Lee, G., Hong, J., Bae, J.H., Lim, K.J., 2021. Prediction of aquatic ecosystem health indices through machine learning models using the WGAN-based data augmentation method. *Sustainability* 13, 10435.
- Leng, L., Zheng, H., Shen, T., Wu, Z., Xiong, T., Liu, S., Cao, J., Peng, H., Zhan, H., Li, H., 2025. Engineering Biochar from biomass pyrolysis for effective adsorption of heavy metal: an innovative machine learning approach. *Sep. Purif. Technol.* 361, 131592.
- Lin, C., Zhang, H., 2025. Polymer biodegradation in aquatic environments: a machine learning model informed by meta-analysis of structure-biodegradation relationships. *Environ. Sci. Technol.* 59, 1253–1263.
- Liu, R., Zayed, T., Xiao, R., 2024. Advanced acoustic leak detection in water distribution networks using integrated generative model. *Water Res.* 254, 121434.
- Liu, Y., Zhou, Y., Liu, X., Dong, F., Wang, C., Wang, Z., 2019. Wasserstein GAN-based small-sample augmentation for new-generation artificial intelligence: a case study of cancer-staging data in biology. *Engineering* 5, 156–163.
- Lv, Y., Duan, J., Li, X., 2024. A survey on modeling for behaviors of complex intelligent systems based on generative adversarial networks. *Comput. Sci. Rev.* 52, 100635.
- Nguyen, D.H., Le, X.H., Heo, J.-Y., Bae, D.-H., 2021. Development of an extreme gradient boosting model integrated with evolutionary algorithms for hourly water level prediction. *IEEE Access* 9, 125853–125867.
- Palansooriya, K.N., Li, J., Dissanayake, P.D., Suvarna, M., Li, L., Yuan, X., Sarkar, B., Tsang, D.C.W., Rinklebe, J., Wang, X., 2022. Prediction of soil heavy metal immobilization by biochar using machine learning. *Environ. Sci. Technol.* 56, 4187–4198.
- Park, C., Lee, J., Kim, Y., Park, J.-G., Kim, H., Hong, D., 2022. An enhanced AI-based network intrusion detection system using generative adversarial networks. *IEEE Internet Things J.* 10, 2330–2345.
- 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. *Sci. Total Environ.* 807, 150554.
- Usman, M., Ahmed, A., Ji, Z., Yu, B., Shen, Y., Cong, H., 2021. Environmentally friendly fabrication of new β -Cyclodextrin/ZrO₂ nanocomposite for simultaneous removal of Pb (II) and BPA from water. *Sci. Total Environ.* 784, 147207.
- Van der Maaten, L., Hinton, G.E., 2008. Visualizing data using t-SNE. *J. Mach. Learn. Res.* 9, 2579–2605.
- Vimbi, V., Shaffi, N., Mahmud, M., 2024. Interpreting artificial intelligence models: a systematic review on the application of LIME and SHAP in Alzheimer's disease detection. *Brain Inform* 11, 10.
- Wei, J., Shen, T., Wang, K., Liu, J., Wang, S., Hu, W., 2025. Transfer learning framework for the wind pressure prediction of high-rise building surfaces using wind tunnel experiments and machine learning. *Build. Environ.* 271, 112620.
- Xiong, X., Hongkai, J., Li, X., Niu, M., 2020. A Wasserstein gradient-penalty generative adversarial network with deep auto-encoder for bearing intelligent fault diagnosis. *Meas. Sci. Technol.* 31, 045006.
- Xu, Q., Li, J., Fan, Y., Gao, Z., Wang, Z., Xu, L., Wang, S., Liu, J., 2024. Intelligent prediction framework for axial compressive capacity of FRP-RACFST columns. *Mater. Today Commun.* 41, 110999.
- Zhai, M., Fu, B., Wu, Z., Wang, J., Wang, W., Wang, H., 2025. Data-driven machine learning improves prediction of sulfonamide antibiotic adsorption by biochar in aqueous phase. *Bioresour. Technol.*, 132773.
- Zhai, M., Fu, B., Zhai, Y., Wang, W., Maroney, A., Keller, A.A., Wang, H., Chovelon, J.-M., 2023. Simultaneous removal of pharmaceuticals and heavy metals from aqueous phase via adsorptive strategy: a critical review. *Water Res.*, 119924.
- Zhai, M., Fu, B., Zhang, F., Tian, X., Wang, H., Chovelon, J.-M., 2024. Simultaneous removal of non-steroidal anti-inflammatory drugs and cadmium in aqueous phase by activated carbon: behaviors, mechanisms and application potentials. *J. Environ. Chem. Eng.* 12, 112013.
- Zhang, Y., Peng, J., Wang, Z., Xi, M., Liu, J., Xu, L., 2025a. Machine learning-assisted sustainable mix design of waste glass powder concrete with strength–cost–CO₂ emissions trade-offs. *Buildings* 15, 2640.
- Zhang, Y., Wang, H., Liu, J., Liu, F., Lv, X., 2025b. Intelligent structural design of composite concrete-encased steel columns based on hybrid machine learning and multiobjective optimization. *Struct. Concr.* 1–24.
- Zhao, F., Repo, E., Yin, D., Meng, Y., Jafari, S., Sillanpää, M., 2015. EDTA-cross-linked β -cyclodextrin: an environmentally friendly bifunctional adsorbent for simultaneous adsorption of metals and cationic dyes. *Environ. Sci. Technol.* 49, 10570–10580.
- Zhou, B.-Q., Yang, R.-C., Hui-Ping, L.L., Wang, Y.-J., Zhang, C.-Y., Xiao, Z.-J., He, Z.-Q., Pang, W.-H., 2023. Numeric and nonnumeric information input to predict adsorption amount, capacity and kinetics of tetracyclines by biochar via machine learning. *Chem. Eng. J.* 471, 144636.
- Zhu, X., He, M., Sun, Y., Xu, Z., Wan, Z., Hou, D., Alessi, D.S., Tsang, D.C.W., 2022. Insights into the adsorption of pharmaceuticals and personal care products (PPCPs) on biochar and activated carbon with the aid of machine learning. *J. Hazard Mater.* 423, 127060.