ALE interpretability analysis based on machine learning MOFs adsorption of tetracycline antibiotics in water

Canhui Zhang, Sitan Wang*

College of Environmental Science and Engineering, Liaoning Technical University, Fuxin 123000, PR China

Abstract: With the large-scale use of tetracycline antibiotics (TCs), residues of TCs have been detected in various environmental systems. The adsorption of TCs in water by metal-organic framework materials has been proved to be an efficient removal method. In this study, machine learning (ML) method was used to train MOFs to adsorb TCs, and a ML model was obtained to accurately predict the adsorption of TCs by MOFs. Data sets for model training and evaluation were constructed by collecting data from literature. Six regression models, RF, Adaboost, GBDT, XGB, SVR and KNN, were trained to evaluate the model performance. According to the comparison of evaluation parameters, the model with high prediction performance for MOFs adsorption of TCs is XGB model. Accumulated Local Effects Plot (ALE) was used to visualize XGB model predictions. The results show that Zn and Zr as central metals have higher MOFs potential, while H_4TBAPy and H_4TCPP have higher MOFs advantage as organic ligands. When PS>9 nm, 200 $m^2\ g^{-1}$＜SA ＜1000 $m^2\ g^{-1}$, 0.25 $cm^3\ g^{-1}$＜PV＜0.5 $cm^3\ g^{-1}$ or PV>1.5 $cm^3\ g^{-1}$, the adsorption capacity of MOFs for TCs is higher, and PS should satisfy the following conditions first. This study accelerates the application of MOFs to TCs adsorption, provides a new way to screen and synthesize MOFs with high adsorption capacity for TCs.

1 Introduction

Tetracycline antibiotics have become the second largest antibiotic in the world due to their broad spectrum and low cost. Because TCs have good water solubility and mobility, various residues of TCs have been found in aquatic and terrestrial environments, resulting in environmental pollution and destroying ecosystem balance [1]. Currently, adsorption has become a widespread and effective way to remove organic pollutants in water and has shown good performance in removing TCs pollution from water.

Metal organic frameworks (MOFs) have been widely used in adsorption due to their adjustable pore topology, large specific surface area and abundant active sites. Theoretically, through different combinations, there are unlimited possibilities for MOFs.

Machine learning (ML) is able to decipher complex relationships between many variables, it provides visual insights for understanding and solving environmental problems. At present, ML has been used in screening functional materials and predicting gas adsorption efficiency of MOFs [2].

The purpose of this study is to propose a ML method to predict the adsorption performance of MOFs for TCs in aqueous environment. The basic data set of this study was constructed by collecting the structural characteristics, pore characteristics, experimental characteristics and adsorption capacity of MOFs. The ML model selects RF model, Adaboost model, GBDT model, XGB model, SVR model and KNN model as regression models respectively. By adjusting and training the hyperparameters of each model, the fitting effect of each model is evaluated, and the model with the best performance is selected as the regression model for prediction. By using the optimal model, we visually analyzed the variables affecting the adsorption capacity of MOFs, specifically using correlation analysis, feature importance evaluation and accumulative local effect point methods, and analyzed the influence degree of each feature in depth, and then clarified the relationship between each feature and the adsorption capacity of MOFs, and gave the application scope of the characteristics of high adsorption MOFs. This research introduces a new method to predict the adsorption capacity of MOFs to TCs. In addition, it provides new ideas for solving other environmental challenges.

2 Data and methods

2.1 Establish a database

In this study, we searched the papers on MOFs adsorption of TCs in water through Web of Since and other papers retrieval platforms. The data in the literature was extracted and outliers were eliminated. Finally, 2704 pieces of data were collected. MOFs central metal (CM), organic ligand (OL), modification method (ModM), specific surface area (SA/ $m^2\ g^{-1}$), pore diameter (PS/nm), pore volume (PV/ $cm^3\ g^{-1}$), adsorbate species (AD), initial concentration (C0/mg L^{-1}), solution volume (V/ml), adsorbent dose (m/mg), reaction time (t/min), reaction temperature (T/K)

wangsitan@lntu.edu.cn

and pH value (pH), adsorption capacity (q/mg g^{-1}) as collection characteristics.

Since there is missing data in the data set, data interpolation is used to fill in the missing data. For pH value and temperature T, fixed values of 7 and 298K were used for data interpolation; Subsequently, model interpolation was used for interpolation. The final random forest interpolation error was minimized, and thus this dataset was used as the basis for subsequent research. The discrete variables in the dataset are labeled and coded. At the same time, data sets are normalized.

2.2 Model Evaluation and Model Interpretation

RF model, Adaboost model, KNN model, SVR model, GBDT model and XGB model were used to model the dataset. 70% of the data was used for model training and tuning and 30% for model evaluation and selection. Model hyperparameters were tuned by ten-fold cross validation and grid search. The fitting coefficient R^2, mean absolute error MAE and root mean square error RMAE were chosen as performance metrics for model evaluation and the optimal model was selected.

A correlation analysis was performed on the dataset to analyze and compare the relationships between the variables. The importance of the characteristics was ranked to compare the degree of influence of the characteristics of each variable on the adsorption capacity. Finally, the effect of each variable on the adsorption capacity of MOFs is visualized using the optimal model and the cumulative local effect plot (ALE), and the range of applicability of each variable is given when the adsorption capacity is high.

3 Results and analysis

3.1 Model evaluation

Six models were constructed, and the dataset was divided into 7:3. The hyperparameters of the models were optimized by means of ten-fold cross validation and grid search, so that the models achieved the best results. The R^2 score, MAE and RMAE of each model were comprehensively compared. Figure 1 shows the fit and evaluation of each of the six models. XGB model has the best performance, R^2 = 0.96, MAE = 12.5, RMAE = 672.62.

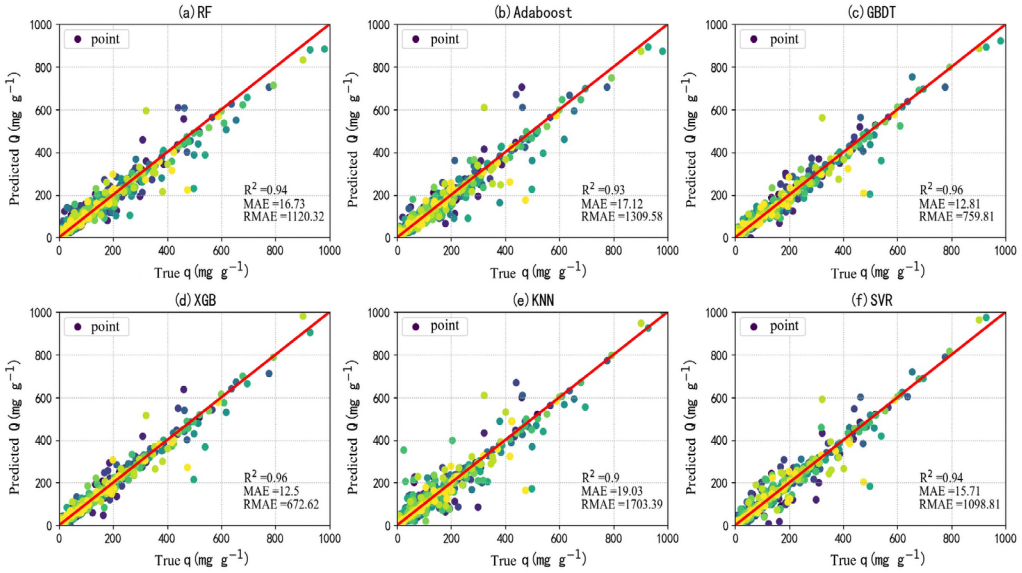

Figure 1. The effect of fitting each model

3.2 Correlation analysis and Feature importance

Pearson correlation coefficient is calculated to measure the correlation between quantitative variables in the dataset. Figure 2(a) shows the Pearson correlation matrix for quantitative variables, showing a strong positive correlation between SA and PV and a weak negative correlation between SA and PS. This can also explain the relationship between specific surface area, pore diameter and pore volume to some extent.

The XGB model quantifies the effect of each variable on adsorption capacity as shown in Figure 2(b). The initial concentration really calls the shots when it comes to adsorption capacity. This suggests that a steeper concentration gradient of TCs can really crank up the driving force, more than enough to overcome the resistance to mass transfer from the water to the solid adsorbent [3]. And for the characterization of the pore structure of MOFs, the order is given as PS>SA>PV. PS was arranged at the front end of MOFs, indicating that PS had the greatest influence on the adsorption of TCs by MOFs. Therefore, for the adsorption of TCs, the pore size of the adsorbent should be considered firstly, and a suitable range of pore size should be selected to obtain a better adsorbent material.

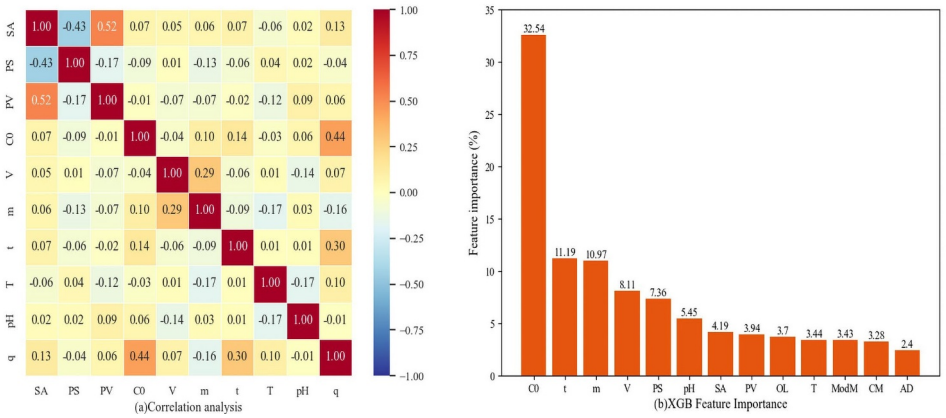

Figure 2. (a) Pearson correlation coefficient; (b)Feature importance

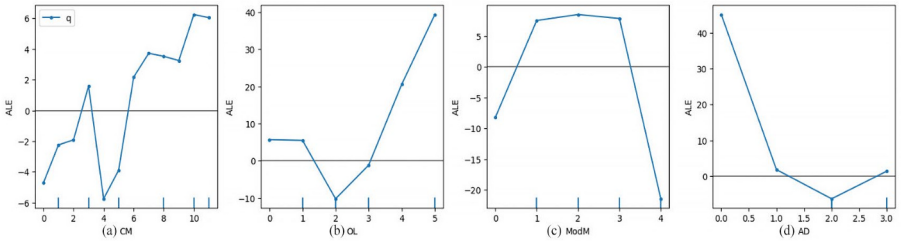

Figure 3. The ALE plot of discrete categorical variables(a)CM; (b)OL; (c)ModM; (d)AD,The horizontal coordinates are expressed in turn: (a)CM: Al, Bi, Co, Cr, Cu, Fe, Fe/Al, Fe/Co, Fe/Zr, Ti, Zn, Zr; (b)OL: $C_4H_4O_4$, $C_4H_6N_2$, H_2BDC, H_3BTC,H_4TBAPy, H_4TCPP; (c): Carbonized, Composite, Dopant, Doublemetal, Unmodified; (d):CTC, OTC, TC, TCH

3.3 Accumulative Local Effect Plot (ALE)

Accumulative local effects plots can depict the changes in the influence of features on the output values, thus visualizing the process of model prediction of the results, and thus discovering the range of each feature for high adsorption performance of MOFs.

The ALE plot of discrete categorical variables is shown in Figure 3. From Figure 3(a), it can be seen that monometallic MOFs with Zn and Zr as the center metals have a positive effect on the adsorption capacity. In addition, the adsorption capacity of bimetallic ionic MOFs was also relatively high, and it was shown that the synergistic effect between bimetallic ions and organic ligands could enhance the active adsorption effect [4]. Figure 3(b) shows the effect of organic ligands on adsorption capacity. MOFs with H_4TBAPy and H_4TCPP as organic ligands have greater advantages for adsorption of TCs [5]. The adsorption capacity of modified MOFs is higher than that of unmodified MOFs. When CTC is used as adsorbate, the adsorption capacity is the largest.

Figure 4 demonstrates the effect of pore structure characterization on adsorption capacity. From the figure, we can find that the adsorption capacity is larger when 200 m^2 g^{-1}<SA<1000 m^2 g^{-1}; And the adsorption level is basically consistent with the zero scale when SA>1500 m^2 g^{-1}. This may be related to the pore size; of course, ultra-high specific surface area goes hand in hand with very small pore size [6], which is also evident in Figure 4(b). When the pore size is too small, the adsorption capacity is at a low level, while when the pore size is larger than 9 nm, the adsorption capacity is higher than the average effect, which indicates that the pore size in this range has great advantages for adsorption [7]. This also indicates that for large molecules such as TCs, the larger the pore size, the more molecules can enter the pores and be trapped. In addition, the ALE of pore volume shows that when 0.25 cm^3 g^{-1}<PV<0.5 cm^3 g^{-1} and PV>1.5 cm^3 g^{-1}, the adsorption capacity is at a high level. At this time MOFs adsorption TCs molecules more favorable. Therefore, for designing MOFs with high adsorption capacity, the pore diameter should be satisfied to be larger than 9 nm first, and then the applicable range of specific surface area and pore volume should be considered.

Figure 5 shows the effect of experimental variables on adsorption capacity. From the figure, it can be seen that the adsorption capacity is higher when the initial concentration is greater than 50 mg L^{-1}, which indicates that a larger initial concentration can generate a larger concentration gradient force, thus favoring adsorption. And when the V is greater than 80 ml, the adsorption capacity is positively increasing at this time; the adsorbent dosage is negatively correlated with the adsorption capacity, and the adsorption capacity is larger when m<10 mg. From the ALE of t, it can be seen that the adsorption equilibrium is reached when t > 420 min. The ALE of T illustrates the time-absorption reaction of the adsorption process. Finally, MOFs have higher adsorption capacity when the pH value is between 3.3 and 5.4 or 6.5 and 7.5 due to the fact that, π-π interactions and van der Waals forces play a major role in adsorption. The adsorption of MOFs is stronger when the pH is neutral.

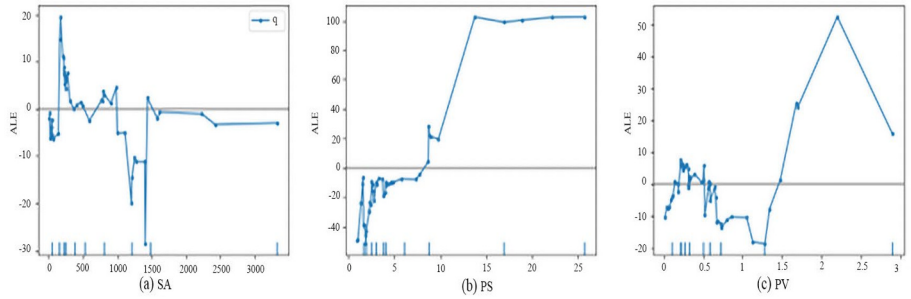

Figure 4. The ALE plot of pore structure characterization (a)SA; (b)PS; (c)PV

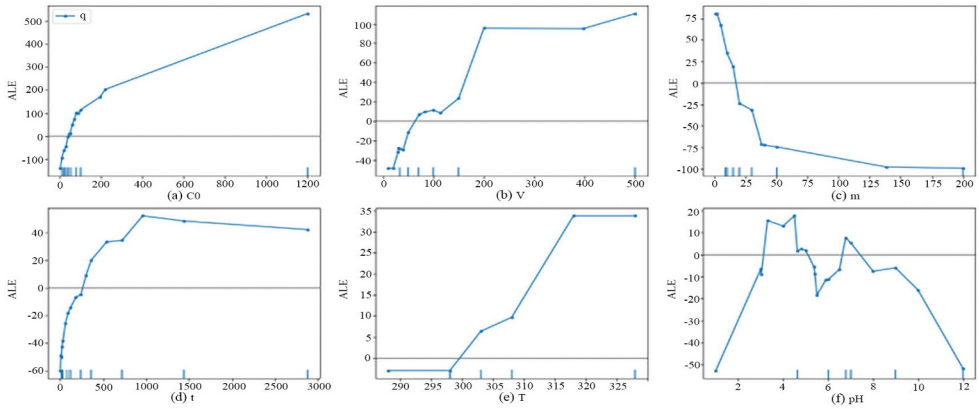

Figure 5. The ALE plot of experimental variables (a)C0; (b)V; (c)m; (d)t; (e)T; (f)pH

4 Conclusions

In this study, six machine learning methods were used to train and test the model based on the established data set of MOFs adsorbing TCs in water. The results show that XBG model has the best test performance. It is shown that C0 has that great influence on adsorption capacity through the feature importance, and for the characteristic of MOFs themselves, PS > SA > PV.

ALE was used to explain the variables and the results show that Zn and Zr as central metals, H_4TBAPy and H_4TCPP as organic ligands exhibit strong influence on the adsorption of TCs by MOFs. When PS>9 nm, 200 m^2 g^{-1} <SA<1000 m^2 g^{-1}, 0.25 cm^3 g^{-1}<PV<0.5 cm^3 g^{-1} or PV>1.5 cm^3 g^{-1}, MOFs exhibited higher adsorption capacity for tetracycline antibiotics. This provides a new idea for the synthesis of MOFs with high adsorption efficiency for tetracycline antibiotics.

Acknowledgment

This work was supported by the Basic Research Projects of Liaoning Provincial Education Department (LJ242410147014)

References

1. X.H. Liu, G.D. Zhang, Y. Liu, S.Y. Lu, P. Qin, X.C. Guo, B. Bi, L. Wang, B.D. Xi, F.C. Wu, W.L. Wang, T.T. Zhang, Occurrence and fate of antibiotics and antibiotic resistance genes in typical urban water of Beijing, China, Environ. Pollut. 246 (2019) 163-173.
2. M. Pardakhti, E. Moharreri, D. Wanik, S.L. Suib, R. Srivastava, Machine learning using combined structural and chemical descriptors for prediction of methane adsorption performance of metal organic frameworks (MOFs), ACS Comb. Sci. 19 (2017) 640-645.
3. Y. Kong, H. Lu, R. Wang, Q. Yang, B. Huang, Q. Zhou, W. Hu, J. Zou, Q. Chen, Adsorption characteristics of tetracycline hydrochloride and oxytetracycline by a MOF-525 (Co) metal organic framework, Colloids Surf., A 677 (2023) 132443.
4. R.Issa, F.A. Ibrahim, M. Al-Ghoul, M. Hmadeh, Controlled growth and composition of multivariate metal-organic frameworks-199 via a reaction-diffusion process, Nano Res. 14 (2021) 423-431.
5. C.Liang, Y. Tang, X. Zhang, H. Chai, Y. Huang, P. Feng, ZIF-mediated N-doped hollow porous carbon as a high performance adsorbent for tetracycline removal from water with wide pH range, Environ. Res. 182 (2020) 109059..
6. D.Y. Lee, D.V. Shinde, E.-K. Kim, W. Lee, I.-W. Oh, N.K. Shrestha, J.K. Lee, S.-H. Han, Supercapacitive property of metal–organic-frameworks with different pore dimensions and morphology, Microporous Mesoporous Mater. 171 (2013) 53-57.
7. X.D. Zhao, M.Q. Zheng, X.L. Gao, J. Zhang, E.B. Wang, Z.Q. Gao, The application of MOFs-based materials for antibacterials adsorption, Coord. Chem. Rev. 440 (2021) 213970.