



Prediction and interpretation of photocatalytic NO removal on g-C₃N₄-based catalysts using machine learning

Jing Li^a, Xinyan Liu^b, Hong Wang^b, Yanjuan Sun^{a,*}, Fan Dong^{b,*}

^a School of Resources and Environment, University of Electronic Science and Technology of China, Chengdu 611731, China

^b Research Center for Carbon-Neutral Environmental & Energy Technology, Institute of Fundamental and Frontier Sciences, University of Electronic Science and Technology of China, Chengdu 611731, China

ARTICLE INFO

Article history:

Received 8 February 2023

Revised 28 April 2023

Accepted 21 May 2023

Available online 26 May 2023

Keywords:

Machine learning

g-C₃N₄-based catalysts

NO removal

Interpretability

Catalytic informatics

ABSTRACT

Predictive modeling of photocatalytic NO removal is highly desirable for efficient air pollution abatement. However, great challenges remain in precisely predicting photocatalytic performance and understanding interactions of diverse features in the catalytic systems. Herein, a dataset of g-C₃N₄-based catalysts with 255 data points was collected from peer-reviewed publications and machine learning (ML) model was proposed to predict the NO removal rate. The result shows that the Gradient Boosting Decision Tree (GBDT) demonstrated the greatest prediction accuracy with R^2 of 0.999 and 0.907 on the training and test data, respectively. The SHAP value and feature importance analysis revealed that the empirical categories for NO removal rate, in the order of importance, were catalyst characteristics > reaction process > preparation conditions. Moreover, the partial dependence plots broke the ML black box to further quantify the marginal contributions of the input features (e.g., doping ratio, flow rate, and pore volume) to the model output outcomes. This ML approach presents a pure data-driven, interpretable framework, which provides new insights into the influence of catalyst characteristics, reaction process, and preparation conditions on NO removal.

© 2023 Published by Elsevier B.V. on behalf of Chinese Chemical Society and Institute of Materia Medica, Chinese Academy of Medical Sciences.

A positive atmospheric quality is a prerequisite for ensuring human health and society development [1]. However, in recent years, fine particulate matter (PM_{2.5}) concentrations that still far exceed the guideline value of the World Health Organization (WHO) and growing ozone (O₃) pollution are causing respiratory diseases and ravaging crop growth in many regions of the world [2–4]. As a key precursor to the formation of PM_{2.5} and O₃, the persistently high concentration of nitrogen oxides (NO_x) in the atmosphere induces considerable challenges in solving pollution problems [5]. Accordingly, numerous photocatalysts have been developed to achieve efficient purification for the ppb level atmospheric NO_x [6]. From these works, it has been found that the NO removal rate is dependent on multiple factors, including the catalyst characteristics (e.g., doping ratio, band gap, and specific surface area), preparation conditions (e.g., precursor type, calcination temperature, calcination time, and heating rate), and reaction process (e.g., NO₂ generation, light intensity, and flow rate). While the catalyst design and experimental condition optimization are rather sophisticated [7–9], the dominant factor affecting the performance of NO removal rate

remains ambiguous. More significantly, an optimal and reasonable instruction scheme to achieve the best NO removal is still lacking [6,10].

Accordingly, it is necessary to systematically evaluate the effect of various factors on photocatalytic NO removal rate. An empirical method capable of predicting the photocatalytic NO removal performance is therefore beneficial for determining the optimal catalyst characteristics, preparation conditions, and reaction process. A robust model containing all possible factors can be used to emphasize the relative importance of each factor, which could strengthen the understanding of the catalytic reaction and contribute to the achievement of high photocatalytic NO removal under optimized conditions [11–13]. However, to our knowledge, this technique has yet to be developed and utilized on photocatalytic NO removal. Obviously, it is hard to achieve such a goal with sole efforts from experimentation [14]. On the other hand, data science and ML have grown tremendously in the past decades and have demonstrated themselves as one of the most powerful strategies in data mining and analysis. The ML method utilizes algorithms to gain knowledge from large, complex and multidimensional data and make fast and accurate predictions [13,15,16]. Success has been established in adopting ML methods (e.g., decision tree and neural network) to predict the NO_x removal efficiency of selective catalytic

* Corresponding authors.

E-mail addresses: sunyj@uestc.edu.cn (Y. Sun), dongfan@uestc.edu.cn (F. Dong).

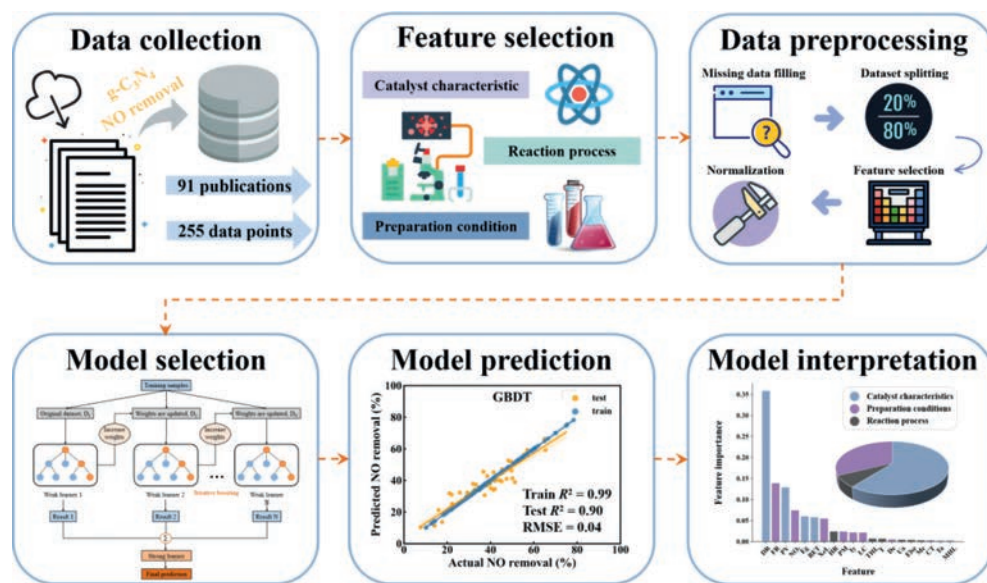


Fig. 1. Scheme of the workflow in this study and the detailed strategies of the machine learning framework to predict and interpret the photocatalytic NO removal rate by $g\text{-C}_3\text{N}_4$ -based catalysts.

reduction catalysts and NO decomposition conversions and yields of molecular sieves [17,18]. In addition, several studies have developed risk assessment methods for air pollution, using ML models to predict $\text{PM}_{2.5}$ based on meteorological conditions and land-use variables [19], as well as fine scale spatiotemporal estimation for ozone concentration prediction [20] and quantitative tracking of SiO_2 nanoparticles [21]. In complex air purification systems, the applied ML model has already demonstrated its capability to evaluate complicated interactions between dependent and independent variables [22]. Therefore, we expect it to be also effective in revealing the complex interrelationship between photocatalytic NO removal rate and multiple influencing factors [23,24].

Herein, we proposed a ML model to predict the photocatalytic NO removal rate. $g\text{-C}_3\text{N}_4$ -based catalysts were chosen as the model system due to their superior visible photocatalytic performance and wide application. 255 data points with 14 input features were focused and the Shapley additive explanation, feature importance, and partial dependence plot were utilized to explore the impact of input features on the target feature (NO removal rate) (Fig. 1). Considering the high dimensionality of the data, we developed three tree-based models (Random Forest (RF), Extreme Gradient Boost (XGB) and Gradient Boosting Decision Tree (GBDT)) to predict the photocatalytic purification rate of NO on $g\text{-C}_3\text{N}_4$ -based catalysts, achieving high prediction accuracies. Lastly, physical insights could be further extracted to guide the photocatalytic purification of NO.

For data collection, the experimental data on the photocatalytic NO removal by $g\text{-C}_3\text{N}_4$ -based catalysts were collected from 91 journal publications with 255 data points in the last decade, 52% of which came from our group. For data points that were not directly available, we extracted them from the figures using WebPlotDigitizer software [25]. The dataset did not include duplicate experiments to avoid data leakage [11]. Based on the study's openness, transparency and availability, we uploaded the raw data and ML methods on GitHub to make them publicly available to the community (https://github.com/rreality/C3N4_NO_removal.git).

The candidate features in developing the ML models for NO removal rate were categorized as: (i) Catalyst characteristics: doping ratio (wt%), specific surface area (m^2/g), pore volume (cm^3/g), and band gap (eV). (ii) Preparation conditions: precursors type of $g\text{-C}_3\text{N}_4$ (urea, thiourea, dicyandiamide, melamine), precursor masses (g), calcination temperature ($^\circ\text{C}$), calcination time (h), and heat-

ing rate ($^\circ\text{C}/\text{min}$). (iii) Reaction process: NO_2 generation (%), lamps (tungsten halogen lamp, Xe lamp, metal halide lamp, else lamp), load of catalyst (g), light intensity (W/m^2), and flow rate (L/min). (iv) Target feature: NO removal rate (%).

It should be noted that fluorescence lifetime, catalyst size thickness, and photoluminescence spectroscopy (PL) were also important factors affecting the NO removal rate. However, they are not included in this work due to the limited availability.

It is known that ML performance is strongly dependent on the quality of the data. The low-quality and flawed data will result in ML models producing meaningless output. It is therefore necessary to perform data cleaning and feature engineering on the dataset in order to obtain a meaningful ML model [11].

Raw data processing: KNN Imputer by Scikit-Learn (an open-source ML library for Python) [26] is a widely used method to fill in missing data. For input features like precursors type of $g\text{-C}_3\text{N}_4$ and lamps, which cannot be expressed numerically, we assigned them by one-hot encoding (0 or 1) [27].

Data set splitting: The preprocessed 255 data points were split into training and test sets randomly, with 80% of the total data points in the training set and the remaining 20% in test set.

Feature selection: A high degree of correlation between features can affect model's robustness and predictive performance. The highly correlated features need to be removed to improve the model's generalizability [28].

Normalization of Z-score: Normalization is mostly required to remove the effect of quantitative features on different scales. Based on the data structure, we used the Z-score normalization method [26].

According to the report of Yuan *et al.* [29], when the input feature set is between 5 and 15, the tree-based ML model is more suitable for datasets with 200–1000 data points, which is called a midsize dataset. Our dataset was 14 input features and 255 data points for this work, which can be termed a midsize dataset. Therefore, we focused on three tree-based ML models (Random Forest (RF), XGBoost (XGB), and Gradient Boosting Decision Tree (GBDT)) to analyze and predict the NO removal rate.

Hyperparameter tuning is the process of finding the hyperparameters configuration to achieve optimal performance [30]. The grid search method with five-fold cross-validation was used for hyperparameter tuning in this work [31].

In this work, the prediction of the NO removal rate was studied as a regression model rather than a classification model. The regression model performance was evaluated by R^2 (coefficient of determination), RMSE (root-mean-square error), and RE (relative error). Conceptually, higher R^2 , and lower RMSE and RE represent higher accuracy and better performance of the models [32]. The following equations were applied to calculate the model's performance:

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_{pi} - y_{ai})^2}{\sum_{i=1}^n (y_{ai} - \bar{y}_a)^2} \quad (1)$$

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_{pi} - y_{ai})^2} \quad (2)$$

$$\text{RE} = \frac{1}{n} \sum_{i=1}^n \frac{|y_{pi} - y_{ai}|}{y_{ai}} \quad (3)$$

Above, y_{pi} and y_{ai} represent the predicted and the actual values of the NO removal rate, respectively. \bar{y}_a is the average of NO removal rate of all instances; n and i are the total number of data points and the data point at any given instance.

Post-training model interpretation is a crucial step to examine whether the model predictions are consistent with the generalized experimental results [13]. There are convincing strategies (e.g., Shapley additive explanation, feature importance, and partial dependence plot) to explain black-box models [24].

Shapley additive explanations (SHAP): SHAP is a concept in the field of cooperative game theory that aims to measure the contribution of each player. Lundberg *et al.* [33] presented SHAP values explaining the model prediction method in their excellent work, which provided a high degree of interpretability to models. The marginal effects of each variable on NO removal rate were evaluated by the SHAP values.

Feature importance: Based on the relative influence of each variable on the ML model, the feature importance can be calculated in the range of 0–100%. A higher feature importance score means that the feature would have a greater impact on the model that is used to predict the output variables [34].

Partial dependence plot (PDP): PDP reveals dependence between the target variable (NO removal rate) and a set of input variables, marginalizing over the values of all other input features. The size of the set of input features is limited to univariate/bivariate due to the visual limitation [35]. The dependence relationships of important variables on NO removal rate were further examined by partial dependence plots.

255 data points originated from the published works were systematically collected, which contain the g-C₃N₄-based catalysts with good or bad performance. First, descriptive analysis in terms of maximum, minimum, mean value, and standard deviation values of input features and target feature (Table S1 in Supporting information) was performed to obtain a preliminary understanding of the raw data [36].

Fig. 2 explicitly shows data distributions of input variables and the target variable in the violin plot [32]. Noticeably, the doping ratio varied from 0 to 50 wt%, with an average value of 3.2 wt% (Fig. 2a). The mean values for microstructural characteristics including specific surface area (BET) and pore volume (PV) were 48.59 m²/g and 0.24 cm³/g (Figs. 2b and c), respectively. The mean value of the band gap was 2.62 eV, with a low standard deviation equal to 0.19 eV (Fig. 2d). The reported precursor masses varied from 0.5 g to 30 g, with a mean value of 9.8 g (Fig. 2e). The calcination process parameters including calcination temperature, calcination time, and heating rate range of 450–600 °C, 0–6 h, and 2–52 °C/min, respectively (Figs. 2f–h). The mean value of NO₂ generation on the g-C₃N₄-based catalysts was 20.6% with a standard deviation

Table 1

Comparative evaluation of three tree-based ML models using the data set.

Item	RF	XGB	GBDT
Training R^2	0.966	0.998	0.999
Test R^2	0.885	0.864	0.907
RMSE	0.048	0.052	0.043
RE	0.130	0.131	0.110

of 12.1% (Fig. 2i). The purification process including load of catalyst, light intensity, and flow rate ranged from 0.05 g to 0.4 g, 10 W/m² to 1800 W/m², and 0.027 L/min to 3 L/min, with the mean value of 0.16 g, 1310 W/m², and 2.1 L/min (Figs. 2j–l), respectively.

The reported NO removal rate in the dataset varied from 7.7% to 78.4%, with an average value of 37.8% (Fig. 2m). These data points reveal that how catalyst design and operating conditions might affect the purification performance. Notably, the broad distribution of each feature ensured that the model could learn from a variety of data, yielding model with higher robustness [37].

To obtain the best ML model, we trained three models and compared them based on R^2 , RMSE, and RE values. Fig. 3 shows the parity plots of actual and predicted values of the NO removal rate on the g-C₃N₄-based catalysts. The regression models all exhibit good predictive performance ($R^2 \geq 0.85$), which may be attributed to the availability and quality of the input data. For all three models, RF (Fig. 3a), XGB (Fig. 3b), and GBDT (Fig. 3c), the deviations between predicted and actual values were relatively small. As shown in Fig. 3, GBDT significantly outperformed RF and XGB in terms of prediction accuracy. The fit of the RF, XGB and GBDT models on the training data was 0.966, 0.998 and 0.999, expressed in R^2 . The R^2 were 0.885 (RF), 0.864 (XGB), and 0.907 (GBDT) for the test data, respectively. In addition, the RMSE and RE for RF on the test set were 0.048 and 0.130, as well as for XGB were 0.052 and 0.131, respectively. The GBDT showed the lowest RMSE (0.043) and RE (0.110) values compared to RF and XGB, which further indicates that GBDT has better prediction performance (Table 1).

As a result, the GBDT model shows superior predictive power for different experimental conditions on g-C₃N₄-based catalysts, thus the GBDT model was selected as the optimal model in this work. The high-quality prediction made by the GBDT model paves the way for interpreting the modeling results further.

Based on the GBDT model, we applied the SHAP value, feature importance, and PDP to determine the influence of the input variables, including catalyst characteristics, preparation conditions, and reaction process on the NO removal rate. Combining the SHAP value and feature importance, as shown in Fig. 4, the most important variables for predicting the NO removal rate can be revealed as doping ratio, flow rate, and pore volume, which are listed in a sequentially decreasing order. Except for the top three important features, the ranking of importance for the rest features showed slight differences between the SHAP value (Fig. 4a) and feature importance (Fig. 4b). Noticeably, precursors type of g-C₃N₄ feature assigned by the one-hot encoding, SHAP values presented in descending order of importance were urea (Ua), dicyandiamide (De), melamine (Me), and thiourea (Ta), while the feature importance was in the order of De, Ua, Me, and Ta. However, the feature importance score of dicyandiamide was 0.0047, while that of urea was 0.0042, thus their difference was negligible (Table S2 in Supporting information). Dong [38] and Liu *et al.* [39] consistently indicated that urea-derived g-C₃N₄ plays an essential role in enhancing photocatalytic performance because of the abundant mesopores and large specific surface area, which facilitates the adsorption of reactants and diffusion of products [40]. The result shows that urea and dicyandiamide as precursors of g-C₃N₄ can be selected preferentially. The remaining important features would be discussed and analyzed in later sections.

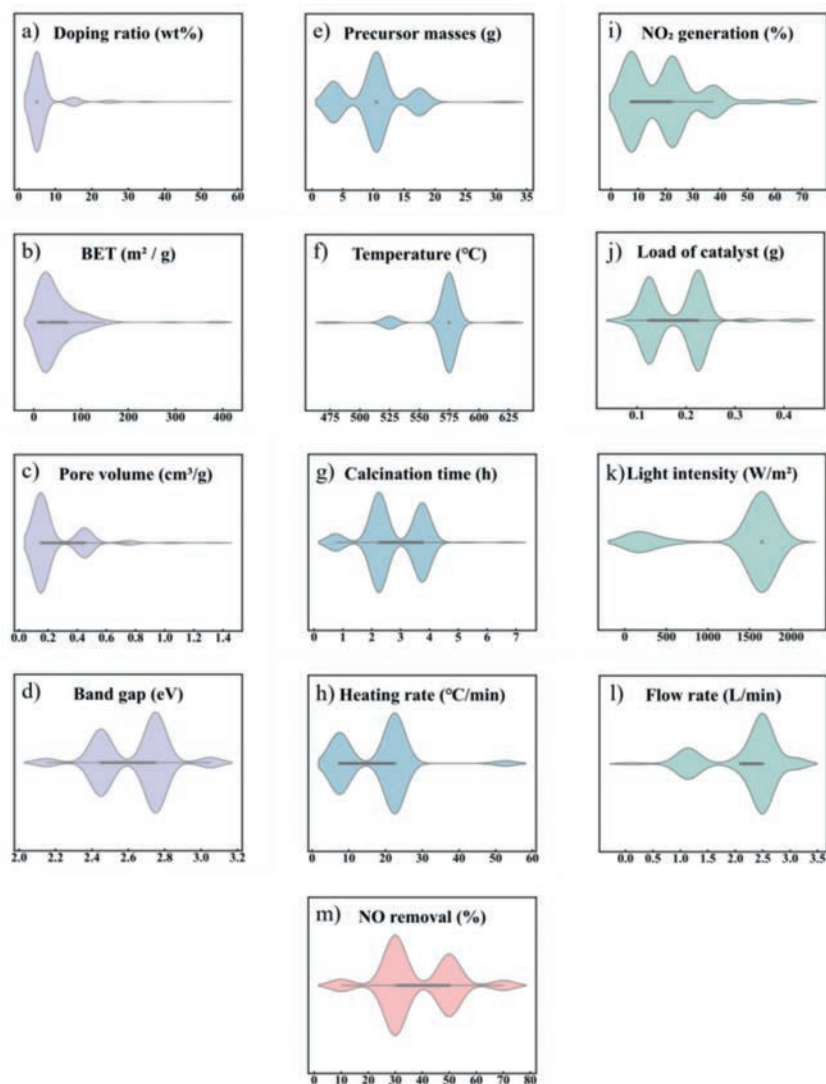


Fig. 2. Distribution of input features related to (a-d) catalyst characteristics (purple), (e-h) preparation conditions (blue), (i-l) reaction process (green), and (m) target variable (pink). Violin plot is a single-axis plot, and the width of each curve corresponds with the approximate frequency of data points in each region.

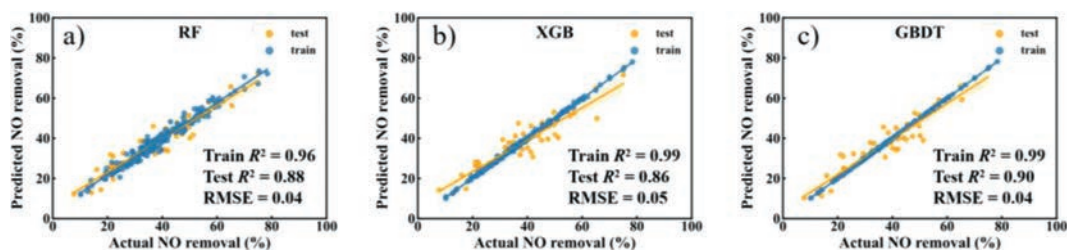


Fig. 3. Predicted NO removal rate vs actual values with (a) RF, (b) XGB, and (c) GBDT. The three models showed that tree-based models had good prediction performance on the training and test data set. The orange shades represent 95% confidence intervals of the regression line on the test points.

Grouping the features into a few categories, we found that "catalyst characteristics" is the most important category for NO removal, which accounted for 60.48% of the total importance. The sub-important strategies were "reaction process" and "preparation conditions", respectively (Fig. 4b). Thus, the combination of SHAP value and feature importance indicated that the priority of g-C₃N₄-based catalysts to achieve efficient NO removal may be: catalyst characteristics > reaction process > preparation conditions.

Based on the feature importance score (Table S2), the univariate partial dependence plots of the nine input features were presented

to further analyze how the important input features exactly affect the removal of NO.

As shown in Fig. 5a, elemental doping can facilitate NO removal. Specifically, the promotion effect of the doping ratio is significant in the range from 0 to 4 wt%, with little contribution to the NO removal rate when the doping ratio is higher than 4 wt%. Fe could be doped into g-C₃N₄ through the coordination between amido- and Fe, and the NO removal performance improved with the increase of Fe doping in 0 × 0.5 wt% but showed a decreasing trend after exceeding 0.5 wt% [41]. Fe doping could improve the

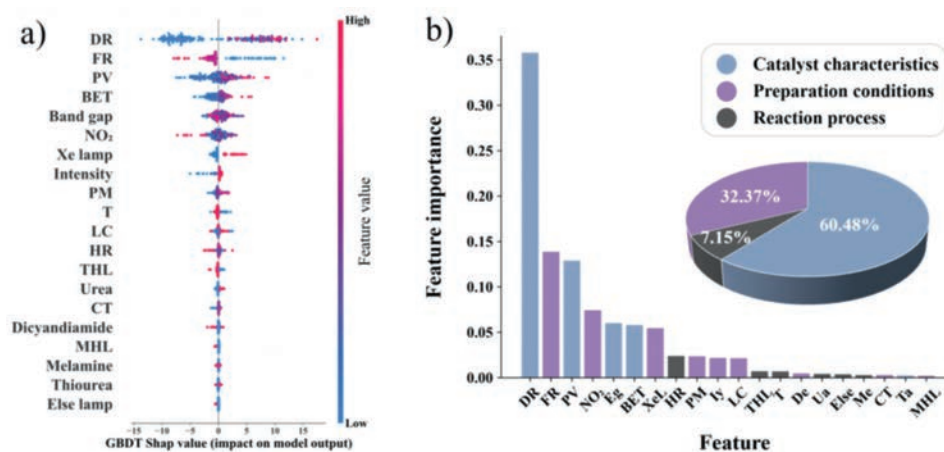


Fig. 4. (a) Shapley additive explanation method. Each dot represents an instance, the color represents the feature value and the X-axis position (SHAP value) represents the expected change in the predicted NO removal rate compared to the prediction when the feature took some baseline value. (b) Feature importance of each influential factor on the NO removal rate. Pie chart representing the average of the contributions for each of the feature classes (catalyst characteristics, preparation conditions, and reaction process) toward the target prediction. Note: DR: doping ratio (wt%), FL: flow rate (L/min), PV: pore volume (cm³/g), NO₂: NO₂ generation (%), Eg: band gap (eV), BET: specific surface area (m²/g), XeL: Xe lamp, HR: heating rate (°C/min), PM: precursor masses (g), Iy: intensity (W/m²), LC: load of catalyst (g), THL: tungsten halogen lamp, T: calcination temperature (°C), De: dicyandiamide, Ua: urea, Else: else lamp, Me: melamine, CT: calcination time (h), Ta: thiourea, and MHL: metal halide lamp.

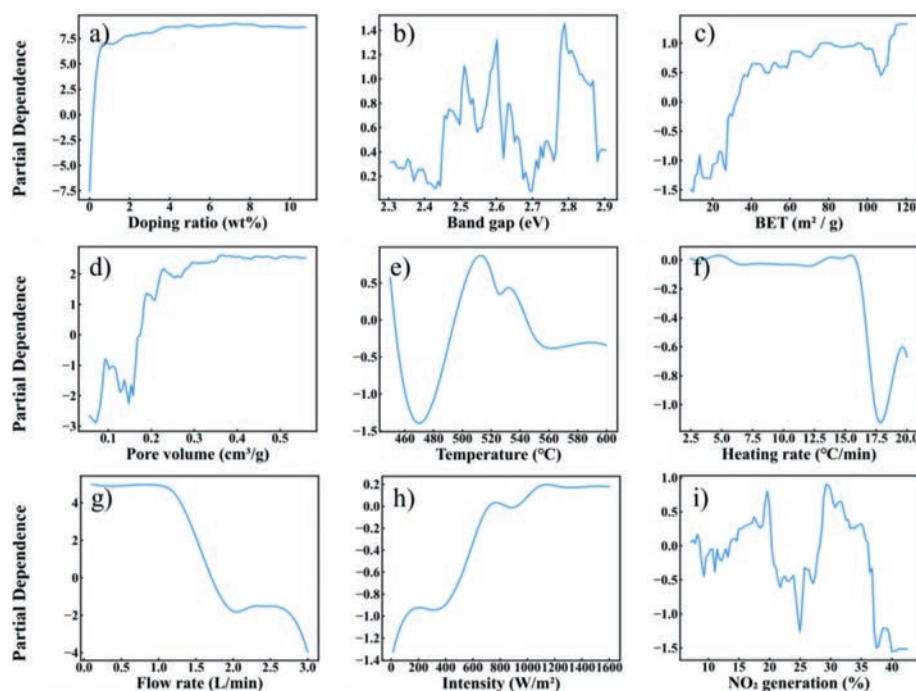


Fig. 5. (a-i) Plots of univariate partial dependence of NO removal rate on important variables.

transfer of photogenerated electrons and enhance the photocatalytic redox performance of g-C₃N₄. However, excessive Fe doping leads to the occupation of the adsorption sites of NO on the g-C₃N₄ surface, which would reduce the performance of g-C₃N₄ photocatalytic NO removal [42]. Typically, appropriate doping of elements could modify the energy band structure and promote charge separation [43,44]. In response, as can be seen from Fig. 5b, when the band gap of pure g-C₃N₄ (2.7 eV) is regulated to about 2.6 and 2.8 eV, the NO removal rate presents significant improvement. The narrowed band gap (about 2.6 eV) broadens the response range of light and enhances the absorption capacity of visible light, while the widened band gap (about 2.8 eV) holds the valence band or conduction band with stronger oxidation or reduction ability [7,45,46]. Both of the above could enhance the genera-

tion of reactive oxygen species (ROS), such as ·OH and ·O₂⁻, to promote efficient NO removal [44,47]. It is also worth mentioning that band gap which is too narrow would lead to carrier recombination, whereas the one that is too wide would reduce the light absorption efficiency [42]. Additionally, the removal of NO increases with the enhancement of BET and PV (Figs. 5c and d), which is attributed to the increased amount of active sites provided by larger microstructure [44,48].

Notably, the catalytic properties of g-C₃N₄ are also governed by synthesis process, such as calcination temperature and heating rate parameters [49,50]. The NO removal rate could be effectively improved by increasing calcination temperature to 520 °C (Fig. 5e). However, further increasing temperature to 550 °C did not significantly improve the NO removal rate. Li *et al.* [51] showed that urea

was not completely decomposed at the calcination temperature below 500 °C, and the $g\text{-C}_3\text{N}_4$ phase appeared at 500 °C. At calcination temperatures above 600 °C, urea decomposed completely and $g\text{-C}_3\text{N}_4$ volatilized completely. The PDP mean line was relatively higher than the average of NO removal rate over the heating rate range of 5.0–15.0 °C/min, which means the contribution of the heating rate is significant in the range from 5.0 °C/min to 15.0 °C/min (Fig. 5f). The slow heating rate (about 2 °C/min) favors the formation of $g\text{-C}_3\text{N}_4$ with a compact structure, complete lamellae, narrow crystalline spacing, and a good polymerization effect [52]. The rapid heating rate (about 15 °C/min) enables the formation of $g\text{-C}_3\text{N}_4$ with a porous structure, high specific surface area, complete CN skeleton, and more amino groups. The presence of loosely packed lightweight flakes results in a better crystal structure, which improves the adsorption and photocatalytic activity [42,52].

For an optimal photocatalyst, the proper reaction conditions are also crucial to enable efficient catalytic performance. It is commonly overlooked that the gas flow rate is a key factor affecting NO removal. As illustrated in Fig. 5g, the optimal gas flow rate for NO removal rate is below 1.0 L/min, as lower gas flow rate helps prolong the NO's duration on the catalyst surface. Fig. 5h displays an optimal light intensity of 1200 W/m² for the maximum NO removal rate, after which no significant increase for NO removal is observed. Therefore, the gas flow rate of 1.0 L/min and the light intensity of 1200 W/m² is sufficient for the ideal performance of the optimal photocatalyst. Especially, the generation of NO₂ [53], a toxic by-product, fluctuated on the NO removal rate (Fig. 5i). The result shows that NO removal rate hardly affects NO₂ production. The high NO removal rate may come from the increased selectivity of NO to NO₂, which is detrimental to the ecological environment. Therefore, researchers should not only focus on the NO removal rate but also reduce the generation of NO₂ toxic by-products and enhance green product selectivity.

To thoroughly investigate the co-effect of catalyst characteristics, preparation conditions, and reaction process on NO removal rate, the bivariate dependency plots were shown in Fig. 6. Since the BET was linearly correlated with PV in this dataset, the BET was used for bivariate PDP. Bivariate PDP analysis of NO removal on the doping ratio (the most critical factors) and band gap and BET was attempted (Figs. 6a and b). It can be seen that, compared to undoped catalysts, the dependence of NO removal rate is stronger on the catalyst characteristics (band gap and BET) for the doped catalysts. This suggested that the NO removal may be influenced by regulating the band structure and increasing the active site of catalysts due to the interactions between elemental doping and catalyst characteristics (band gap and BET) [54–56].

It could be observed that the dependence of the NO₂ generation and band gap on NO removal decreased with the increase in the flow rate (Figs. 6c and d). The fast gas flow rate and the short residence time in the reactor may seriously inhibit the adsorption and reaction of NO on the catalyst surface, leading to the reduction of photocatalytic efficiency [42,57]. The dependence variation of NO removal on the NO₂ generation and band gap was more evident at low flow rate, corroborating the critical role of the low flow (<1.0 L/min) rate, which was consistent with the existing conclusion in Fig. 5g. Therefore, the flow rate of 1.0 L/min was efficient for improving the NO removal of $g\text{-C}_3\text{N}_4$ -based catalysts (Figs. 6c and d).

The photocatalytic performance of $g\text{-C}_3\text{N}_4$ -based catalysts are closely connected with the preparation conditions [42,44,58]. Fig. 6e shows the synergistic effects of calcination temperature and heating rate on the NO removal. The dependence of the NO removal on the heating rate is decreased with the increase of calcination temperatures. With the increase of calcination temperature,

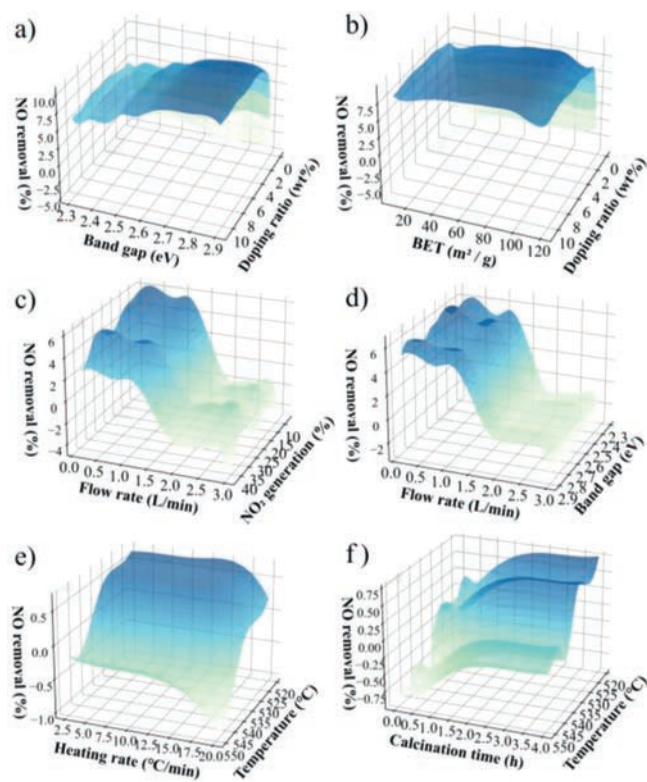


Fig. 6. (a–f) Plots of bivariate partial dependence of NO removal rate on any two important input variables and the interactions between the two variables.

the $g\text{-C}_3\text{N}_4$ structure may be distorted resulting in $n\text{-}\pi^*$ transitions [59]. In addition, the higher calcination temperatures may cause the decomposition of $g\text{-C}_3\text{N}_4$ [60]. Notably, the bivariate PDP (Fig. 6e) combined with univariate (Figs. 5e and f) indicated that calcination temperatures of 520 °C and heating rate of 15 °C/min could achieve excellent photocatalyst for ideal performance.

The interaction between calcination temperature and calcination time was displayed in Fig. 6f. The effects of calcination temperature on NO removal rate are more independent than calcination time at high calcination temperature (about 550 °C), and the dependence on calcination time becomes significant at lower calcination temperature (about 520 °C). In general, the intact $g\text{-C}_3\text{N}_4$ framework is not obtained due to insufficient calcination, leading to a lower specific surface area, which affects the adsorption of the catalyst [52].

Overall, post-training analysis of the model contributes to better understanding on the interactions between catalyst characteristics, preparation conditions, and reaction process, which could efficiently facilitate the catalyst design and process optimization for NO removal.

In this work, we established a data-driven, interpretable ML framework to predict NO removal rate on $g\text{-C}_3\text{N}_4$ -based catalysts. In contrast to traditional trial-and-error-based approaches, the insights obtained through the proposed model could offer further guidelines for designing the appropriate experimental schemes and selecting better catalysts for NO removal treatment. Overall, the above models showed the great predictive potential of ML-based approaches, where more accurate NO removal prediction might be realized with more available data. For instance, fluorescence lifetime, catalyst size thickness, PL and the data from density-functional theory (DFT) calculations can be regarded as additional input features for the NO removal models, which can help us further unveil the microscopic adsorption/desorption influences.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (Nos. 22172019, 22225606, 22176029) and Excellent Youth Foundation of Sichuan Scientific Committee Grant in China (No. 2021JDJQ0006).

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.ccl.2023.108596.

References

- [1] Y. Li, D. Guan, Y. Yu, et al., *Nat. Commun.* 10 (2019) 5473.
- [2] S. Rajagopalan, S.G. Al-Kindi, R.D. Brook, *J. Am. Coll. Cardiol.* 72 (2018) 2054–2070.
- [3] S.G. Al-Kindi, R.D. Brook, S. Biswal, et al., *Nat. Rev. Cardiol.* 17 (2020) 656–672.
- [4] L. Taylor, *BMJ* 374 (2021) n2349.
- [5] Y. Zeng, Y. Cao, X. Qiao, et al., *Sci. Total Environ.* 663 (2019) 329–337.
- [6] K. Li, H. Wang, J. Li, et al., *Environ. Chem. Lett.* 20 (2022) 2687–2708.
- [7] T. Xiong, W. Cen, Y. Zhang, et al., *ACS Catal.* 6 (2016) 2462–2472.
- [8] C. Liu, Y. Zhang, F. Dong, et al., *Appl. Catal. B: Environ.* 203 (2017) 465–474.
- [9] Y. Li, Y. Sun, W. Ho, et al., *Sci. Bull.* 63 (2018) 609–620.
- [10] F. Dong, T. Xiong, Y. Sun, et al., *Appl. Catal. B: Environ.* 219 (2017) 450–458.
- [11] T. Toyao, Z. Maeno, S. Takakusagi, et al., *ACS Catal.* 10 (2020) 2260–2297.
- [12] X. Liu, D. Lu, A. Zhang, et al., *Environ. Sci. Technol.* 56 (2022) 2124–2133.
- [13] S. Zhong, K. Zhang, M. Bagheri, et al., *Environ. Sci. Technol.* 55 (2021) 12741–12754.
- [14] A.J. Medford, M.R. Kunz, S.M. Ewing, et al., *ACS Catal.* 8 (2018) 7403–7429.
- [15] D. Xia, J. Chen, Z. Fu, et al., *Environ. Sci. Technol.* 56 (2022) 2115–2123.
- [16] S.M. Moosavi, K.M. Jablonka, B. Smit, et al., *J. Am. Chem. Soc.* 142 (2020) 20273–20287.
- [17] S. Bae, H. Lee, J. Shin, et al., *Chem. Mater.* 34 (2022) 7761–7773.
- [18] M. Sasaki, H. Hamada, Y. Kintaichi, et al., *Appl. Catal. A: Gen.* 132 (1995) 261–270.
- [19] X. Hu, J.H. Belle, X. Meng, et al., *Environ. Sci. Technol.* 51 (2017) 6936–6944.
- [20] X. Ren, Z. Mi, P.G. Georgopoulos, *Environ. Int.* 142 (2020) 105827.
- [21] X. Yang, X. Liu, A. Zhang, et al., *Nat. Commun.* 10 (2019) 1620.
- [22] P. Gupta, S.A. Christopher, *J. Geophys. Res. Atmos.* 114 (2009), doi:10.1029/2008JD011497.
- [23] J.R. Kitchin, *Nat. Catal.* 1 (2018) 230–232.
- [24] J.A. Esterhuizen, B.R. Goldsmith, S. Linic, *Nat. Catal.* 5 (2022) 175–184.
- [25] WebPlotDigitizer - Extract data from plots, images, and maps. <https://automeris.io/WebPlotDigitizer/>.
- [26] Scikit-Learn, Machine Learning in Python. <https://scikit-learn.org/stable/>.
- [27] J. Brownlee, Why One-Hot Encode Data in Machine Learning? <https://why-one-hot-encode-data-in-machine-learning/>.
- [28] S. Zhong, Y. Zhang, H. Zhang, *Environ. Sci. Technol.* 56 (2021) 681–692.
- [29] X. Yuan, M. Suvarna, S. Low, et al., *Environ. Sci. Technol.* 55 (2021) 11925–11936.
- [30] A.M. Żurański, J.I. Martinez Alvarado, B.J. Shields, et al., *Acc. Chem. Res.* 54 (2021) 1856–1865.
- [31] A.A. Orlov, D.Y. Demenko, C. Bignaud, et al., *Environ. Sci. Technol.* 55 (2021) 15542–15553.
- [32] K. Rahmani, A.H. Mamaghani, Z. Hashisho, et al., *J. Hazard. Mater.* 433 (2022) 128747.
- [33] S.M. Lundberg, S.I. Lee, A Unified Approach to Interpreting Model Predictions, 31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, U. S. A., 2007.
- [34] X. Zhu, Z. Wan, D.C. Tsang, et al., *Chem. Eng. J.* 406 (2021) 126782.
- [35] Molnar, Interpretable machine learning - A guide for making black box models explainable. <https://christophm.github.io/interpretable-ml-book/>.
- [36] X. Wang, L. Liu, W. Zhang, et al., *Environ. Sci. Technol.* 55 (2021) 7491–7500.
- [37] M. Khan, Z. Ullah, O. Mašek, et al., *Bioresour. Technol.* 355 (2022) 127215.
- [38] F. Dong, L. Wu, Y. Sun, et al., *J. Mater. Chem.* 21 (2011) 15171–15174.
- [39] J. Liu, T. Zhang, Z. Wang, et al., *J. Mater. Chem.* 21 (2011) 14398–14401.
- [40] F. Dong, Z. Ni, P. Li, et al., *J. Mater. Chem.* 39 (2015) 4737–4744.
- [41] G. Liu, G. Dong, Y. Zeng, et al., *Chin. J. Catal.* 41 (2020) 1564–1572.
- [42] W.J. Ong, L.L. Tan, Y.H. Ng, et al., *Chem. Rev.* 116 (2016) 7159–7329.
- [43] M. Zhou, G. Dong, F. Yu, et al., *Appl. Catal. B: Environ.* 256 (2019) 117825.
- [44] L. Jiang, X. Yuan, Y. Pan, et al., *Appl. Catal. B: Environ.* 217 (2017) 388–406.
- [45] I. Papailias, N. Todorova, T. Giannakopoulou, et al., *Appl. Catal. B: Environ.* 239 (2018) 16–26.
- [46] T. Giannakopoulou, I. Papailias, N. Todorova, et al., *Chem. Eng. J.* 310 (2017) 571–580.
- [47] X. Kang, G. Dong, T. Dong, A.C.S. Appl. Energy Mater. 6 (2023) 1025–1036.
- [48] J. Geng, L. Zhao, M. Wang, et al., *Environ. Sci. Nano.* 9 (2022) 742–750.
- [49] Y. Zheng, L. Lin, B. Wang, et al., *Angew. Chem. Int. Ed.* 54 (2015) 12868–12884.
- [50] X. Wang, S. Blechert, M. Antonietti, *ACS Catal.* 2 (2012) 1596–1606.
- [51] J. Li, M. Zhang, X. Li, et al., *Appl. Catal. B: Environ.* 212 (2017) 106–114.
- [52] W. Yang, L. Jia, P. Wu, et al., *J. Solid State Chem.* 304 (2021) 122545.
- [53] G. Dong, L. Yang, F. Wang, et al., *ACS Catal.* 6 (2016) 6511–6519.
- [54] T.Y. Ma, J. Ran, S. Dai, et al., *Angew. Chem.* 127 (2015) 4729–4733.
- [55] S. Guo, Y. Zhu, Y. Yan, et al., *Appl. Catal. B: Environ.* 185 (2016) 315–321.
- [56] Q. Han, C. Hu, F. Zhao, et al., *J. Mater. Chem. A* 3 (2015) 4612–4619.
- [57] K. Qi, J. Jing, G. Dong, et al., *Environ. Res.* 212 (2022) 113405.
- [58] Y. Li, Z. Jjiang, G. Dong, et al., *ChemSusChem* 15 (2022) e202200477.
- [59] H. Zhang, A. Yu, *J. Phys. Chem. C* 118 (2014) 11628–11635.
- [60] Z. Lin, X. Wang, *Angew. Chem.* 52 (2013) 1735–1738 *Int. Ed.*