

Application of machine learning algorithms in predicting the photocatalytic degradation of perfluorooctanoic acid

Amir H. Navidpour^a, Ahmad Hosseinzadeh^a, Zhenguo Huang^a, Donghao Li^b, John L. Zhou^{a,*}

^a Centre for Green Technology, School of Civil and Environmental Engineering, University of Technology Sydney, 15 Broadway, Ultimo, NSW 2007, Australia

^b Department of Chemistry, Yanbian University, Park Road 977, Yanji 133002, Jilin Province, China

Corresponding author:

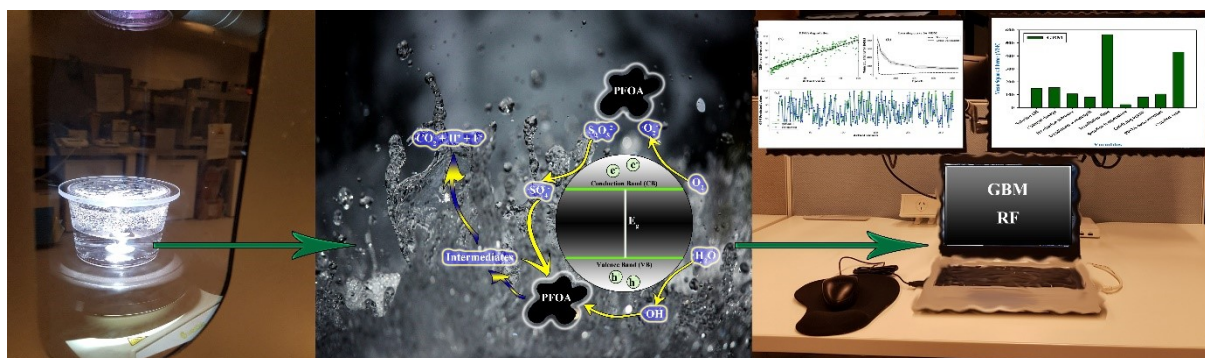
Prof John Zhou, junliang.zhou@uts.edu.au

ABSTRACT

Perfluorooctanoic acid (PFOA) is used in a variety of industries and highly persistent in the environment, with potential human health risks. Photocatalysis has been extensively used for the decomposition of various organic pollutants, yet its simulation and modelling are challenging. This research aimed to establish different machine learning (ML) algorithms which can simulate and predict the photocatalytic degradation of PFOA. The published results were used to estimate and predict the photocatalytic degradation of PFOA. Statistical criteria including the coefficient of determination (R^2), mean absolute error (MAE), and mean squared error (MSE) were considered in assessing the best method of modeling. Among seven ML algorithms pre-screened, Adaptive Boosting (AdaBoost), Gradient Boosting Machine (GBM) and Random Forest (RF) showed the best performance and were chosen for deep modelling and analysis. Grid search was used to optimize the models developed by AdaBoost, GBM, and RF; and permutation variable importance (PVI) was used to analyze the relative importance of different variables. Based on the modeling results, GBM model ($R^2 = 0.878$, MSE = 106.660, MAE = 6.009) and RF model ($R^2 = 0.867$, MSE = 107.500, MAE = 6.796) showed superior performances compared with AdaBoost model ($R^2 = 0.574$, MSE = 388.369, MAE = 16.480). Furthermore, the PVI results suggested that the GBM model provided the best outcome, with the light irradiation time, type of catalyst, dosage of catalyst, solution pH, irradiation intensity, initial PFOA concentration, oxidizing agents (peroxymonosulfate, ammonium persulfate, and sodium persulfate), irradiation wavelength, and solution temperature as the most important process variables in decreasing order.

Keywords: Gradient boosting machine; Machine Learning; Modeling; Perfluorooctanoic acid; Photocatalysis

Graphic abstract



Highlights

- Machine learning models were developed to predict PFOA photocatalytic degradation.
- The GBM and RF models were more robust than AdaBoost model.
- The best modeling performance was achieved by GBM based on PVI analysis.
- PVI analysis showed high importance of irradiation time, catalyst type, catalyst dosage and pH in declining order.

1. Introduction

Per- and polyfluoroalkyl substances (PFAS) possess important properties as repellents (of oil, dirt, and water), surfactants, and frictions reducers. Thus, they are widely used in a variety of consumer products such as carpets and household products to improve stain-proof and water-proof properties. They are also employed for several other purposes including membranes, lubricants (for machinery), firefighting foams, and adjuvants in pesticides.^[1,2] Due to their toxicity and persistence in the environment, bioaccumulation, and probable health impacts, PFAS have recently received increased worldwide attention.^[3] It has been reported that human exposure to even trace level of PFAS can lead to their bioaccumulation in the blood, with correlations between exposure to PFAS and immunotoxicity.^[4] Notably, perfluorooctanoic acid (PFOA, $C_7F_{15}COOH$) and perfluorooctane sulfonic acid (PFOS, $C_8F_{17}SO_3H$) are the most frequently reported PFAS in the environment.^[3]

Several processes have been used for PFAS removal such as nanofiltration membranes,^[5] ion exchange resins,^[6] hybrid membrane filtration,^[7] biodegradation,^[8] and adsorption.^[9] In adsorption, PFAS are only transferred from one phase to another whereas redox techniques could be used for decomposition of PFAS.^[10] Photocatalytic degradation is widely used for PFOA treatment, resulting in the formation of various intermediates including PFHpA, PFHxA, PFPeA, PFBA, PFPrA and TFA^[11,12] which are less toxic than PFOA.^[13]

A variety of metal oxides have been used for the photocatalytic decomposition of PFOA. TiO_2 , as the most frequently used photocatalyst, has shown insignificant efficiency for the degradation of PFOA. In comparison, In_2O_3 and $\beta-Ga_2O_3$ have shown good performances^[14] which are mainly related to their wide band gap energies (4.8 eV and 3.6 eV, respectively)^[14,15] and appropriate surface properties. Such different performance is mainly due to much lower position of the conduction band of TiO_2 (-4.21 eV) than that of $\beta-Ga_2O_3$ (-2.95 eV), relative to the vacuum energy level. Therefore, the reductive potential of the photoinduced electrons of $\beta-$

Ga_2O_3 is remarkably higher than that of TiO_2 .^[16] ZnO is a substantial alternative to TiO_2 which has been widely used for photocatalytic degradation of various organic pollutants.^[17-20] In addition to its sufficient electron mobility, abundance and chemical inertness, ZnO absorbs a larger fraction of light (compared to TiO_2).^[19] Besides, ZnO has shown higher efficiencies than TiO_2 in some cases.^[21,22] CeO_2 is another metal oxide which has found applications in photodegradation of PFOA.^[23] Other than metal oxides, some other materials including BiOCl (bismuth oxychloride), BN (boron nitride), NiAl-LDHs (layered double hydroxides), $\text{Bi}_3\text{O}(\text{OH})(\text{PO}_4)_2$ (bismuth oxyhydroxyphosphate or BOHP), and BiFeO_3 (bismuth ferrite or BFO) have also been used recently for the photocatalytic decomposition of PFOA. It is worth mentioning that unlike hydroxyl radicals ($\cdot\text{OH}$) which have been inert to PFOA in some reports,^[24] persulfates ($\text{S}_2\text{O}_8^{2-}$) and sulfate radicals ($\cdot\text{SO}_4^-$) are highly efficient in the photodecomposition of PFOA. Notably, sulfate radicals and persulfates could be produced from oxidizing agents including ammonium persulfate (APS), sodium persulfate (NaPS), and peroxymonosulfate (PMS).^[25-27]

In spite of unique features of photocatalysis, evaluating the feasibility and efficiency of semiconductors in photocatalytic applications is often a challenging task, by conducting lengthy and carefully designed experiments under controlled conditions. Often, many hours or days are needed to determine the degradation efficiency of organic pollutants. In addition, some semiconductors are very expensive. For instance, although In_2O_3 and $\beta\text{-Ga}_2\text{O}_3$ have been widely researched for the degradation of PFOA, they are not cost effective. Furthermore, many variables are involved in photocatalysis such as light wavelength and intensity, solution pH, pollutant concentration, and catalyst amount, which all need to be carefully assessed. Overall, the rapid estimation of degradation efficiency of organic pollutants in photocatalytic reactions is a major challenge. As a result, modeling, simulation and prediction of photocatalytic

efficiency for organic pollutant degradation are a highly valuable and complimentary tool to experimental research.

The prediction of the photocatalytic degradation of PFOA has not been studied yet. Moreover, it is highly valuable to use modelling to support the photocatalytic experiments by evaluating the relative importance of process variables which saves experimental time, costs, and energy consumption by optimizing and reducing the number of experiments. Mathematical modelling has been widely used for photocatalytic applications in both air and water.^[28-31] Acceptable performances could be achieved by mathematical models. For example, Zekri et al. (2013) used mathematical modeling for estimating the degradation performance of a photocatalytic reactor, with advantages of no need to introduce adjustment variables and a minimum number of experimental data.^[29] On the downside, applying these models could be accompanied by limitation of the number of effective factors^[32] whereas there are several effective process variables in photocatalysis, as discussed previously. In addition, **increased consumption of computational resources** is needed to apply these models for complex systems with numerous variables.^[33,34] More importantly, Afzal et al. (2021) suggested that the predictions of mathematical models could be acceptable on the condition of well understanding the impacts of underlying assumptions.^[35] Due to their potential performances and lack of need to consider the scientific background of processes and mechanisms to predict the correlations, machine learning (ML) processes are being widely used for modeling the correlations among various independent and dependent parameters in different applications.^[36] They have been highly promising to optimize the quality control in complex manufacturing environments in which the causes of problems are hardly detected. The identification of implicit relationships in a dataset is a main advantage of ML procedures. The ability of handling high dimensional problems and data is another advantage of ML procedures.^[37] Notably, ML approaches are superior to mathematical models in modeling the complicated correlations among numerous

process variables. There are several types of ML processes though with different performances in variety of applications.^[36] It means that each method of ML has unique advantages and disadvantages^[37] depending highly on its application. As mentioned, PFOA is one of the most frequently observed PFAS in the environment.^[3] Despite its vital importance, the application of ML processes in predicting the photocatalytic degradation of PFOA as an emerging persistent organic pollutant has not been up to now. To bridge this knowledge gap, a systematic investigation was carried out to develop the most promising ML algorithms for such a purpose followed by analyzing the relative importance of numerous process variables for the first time, to the best of our knowledge. In this research, different ML algorithms including Multiple Linear Regression (MLR), Random Forest (RF), Ridge Regression (RR), Multilayer Perceptron (MLP), Gradient Boosting Machine (GBM), Adaptive Boosting (AdaBoost), and Support Vector Machine (SVM) were used to nominate a potential method to estimate the photocatalytic decomposition of PFOA over various photocatalysts. The performance of ML algorithms is highly dependent on the hyperparameters. Thus, tuning all hyperparameters is essential to develop an optimized model.^[38] The efficiency of PFOA photocatalysis is dependent on various parameters. Among those, nine key variables including solution pH, solution temperature, catalyst dosage, light irradiation intensity, irradiation wavelength, irradiation duration, initial PFOA concentration, type of catalyst, and oxidizing agents (PMS, APS, and NaPS) were used to model the process carefully. Eventually, the performance of the developed models was assessed based on both the outcomes and the relative importance of process variables.

2. Materials and methods

2.1. Data collection and processing

The published results were used to estimate and predict the percentage of PFOA photocatalytic degradation. A detailed literature review was conducted by considering several factors, e.g. reporting solution pH and light irradiation intensity (in term of wattage), and using a single/distinct irradiation wavelength in the UV region (200-400 nm). Notably, a single/distinct irradiation wavelength is not generally used in the visible region (400-800 nm). Furthermore, compared to UV light illumination, visible light irradiation has much lower effect on the photocatalytic degradation of PFOA. For instance, using Ga₂O₃ assisted by PMS, the photodegradation of PFOA under UV light (254 nm) and visible light (400-800 nm) was 100% and <10%, respectively, although the irradiation intensity of visible light was higher than that of UV light.^[26] In another research, using TiO₂ assisted by PMS, PFOA was thoroughly degraded after 2 h and 9 h under UV and visible irradiation, respectively, whereas the irradiation intensity of visible light was approximately 224 times higher than that of UV light.^[27] Therefore, in order to consider the effect of irradiation wavelength in the visible region (in addition to UV region), the mean wavelength value (i.e. 600 nm) was used in the dataset.

Due to their potential performances, PMS, APS, and NaPS were selected to evaluate the effects of oxidizing agents. Overall, a comprehensive dataset containing 1343 datapoints was compiled by considering 18 different categories of catalysts including ZnO, ZnO-rGO, Ga₂O₃, TiO₂, TiO₂-MWCNT, modified TiO₂ by loading metals, metals-doped TiO₂, TiO₂-Sb₂O₃, BiFeO₃-rGO, In₂O₃, BiOCl, CeO₂, NiAl-LDHs, CeO₂@NiAl-LDHs, BN, BiPO₄, BOHP, and BOHP-CS^[11,12,23,25-27,39-52] where the effects of several key process variables on the photocatalytic decomposition of PFOA have been reported. Plot Digitizer was used to extract the datapoints from those publications. The following equation was used to normalize all of the input data in the range of 0 to 1 for data processing:^[53]

$$x_i \text{ normalised proportion} = \frac{x_i - \text{minimum value of data}}{\text{maximum value of data} - \text{minimum value of data}} \quad (1)$$

where x_i is a data point. It is notable that data normalization was performed to reduce computational complexity and to prevent over training for the input data (not for output data).

2.2. Selection of ML procedures and modelling generality

As the performances of various models vary in different applications, appropriate selection of the procedures for modeling is critical. Besides, Occam's Razor principle states that "a model should be as simple as possible, and as complex as needed". Accordingly, RR, LR, MLP, AdaBoost, RF, SVR, and GBM algorithms using default hyper-parameters were pre-screened from *Scikit-learn* library. To assess the performances of the mentioned methods, R^2 (Eq. 2) and mean squared error (MSE) (Eq. 3) were applied.^[36,54] The datasets were divided into train (80%) and test (20%) datasets. Five-fold cross validation was applied to assess the validity of the models developed to prevent overfitting and devastating the data. The test dataset was employed to check the prediction strength of the models by unseen data points. A grid search was carried out to tune the hyperparameters for each of the models. The outcomes of the grid search were employed in all phases of modeling.

$$R^2 = 1 - \frac{\sum_{i=1}^N (y_{prd,i} - y_{Act,i})}{\sum_{i=1}^N (y_{prd,i} - y_m)} \quad (2)$$

$$MSE = \frac{1}{N} \sum_{i=1}^N (y_{prd,i} - y_{Act,i})^2 \quad (3)$$

where N is the total number of datapoints, y_m is the average value of actual PFOA photocatalytic degradation (%), $y_{Act,i}$ is the actual value of PFOA photocatalytic degradation (%), and $y_{prd,i}$ is the predicted value of PFOA photocatalytic degradation (%).

2.3. Gradient boosting machine

GBM is among powerful ML methods developed by Friedman that enables data modeling and analysis for either classification problems or regression.^[55,56] GBM is an ensemble learning algorithm where a series of individual models (generally decision trees) are employed to create

the final model. Like neural networks that employ gradient descent for weight optimization, the loss function is minimized by the gradient in GBM.^[57] The learning process fits new models to yield a more accurate prediction of the response in GBMs. Maximizing the correlation between the new constructed base-learners and the negative gradient of the loss function (correlated with the total ensemble) is the main principle of this algorithm. Although nomination of the loss function is optional, considering the classic squared-error loss as the error function eventuates in consecutive error-fitting in the learning process. So far, numerous loss functions have been derived, and such a high flexibility could result in the promising customization of GBMs to different specific data-driven tasks. It allows plenty of flexibility in the model design, resulting in a trial-and-error process when it comes to selecting the most suitable loss function. GBMs have been successfully used in practical applications to address a variety of ML challenges.^[58] Available elements in GBMs include additive model, strong and weak (base) learners, and loss function. Notably, the weak learners which are the initial decision trees are more powerful than the random estimate in prediction. The strong learners with their remarkable power in prediction are composed of various weak learners. Boosting the weak learners to the strong learners is required for analyzing and modeling the processes in GBMs. For such a purpose, training decision trees are used in serial, gradual and additive approaches. While maintaining the current base learners, new weak learners could be added to the model to reduce either loss function or total error of model.^[59,60] The best situation of the hyperparameters (in grids) was determined by a grid search for estimation of the photocatalytic degradation efficiency of PFOA in GBM.^[61] Notably, there were five key hyperparameters in this analysis: the minimum number of samples in each leaf (*min_samples_leaf*), the number of gradients boosted trees (*n_estimator*), the maximum depth of GBM trees (*max_depth*), the minimum number of samples for splitting an internal node (*min_samples_split*), and the maximum number of features for the best split (*max_features*). All parameters were adapted in

the ranges as (2, 3, 4, 5, 6, and 7), (100-1000), (1, 2, 3, 4, and 5), (2, 3, 4, 5, 6, and 7), and (2, 3, 4, 5, 6, and 7), respectively. In addition, the generality of the GBM model was assessed within a grid search as discussed in section 2.2.

2.4. Random forest

RF is among powerful ML methods that enables modelling both regression and classification cases. With RF algorithm, a variety of decision trees are generated, as regression functions, where the average of decision trees represents the final proportion of the response variable.^[62]

RF is a combination of tree predictors, whereby each tree is dependent on values of a random vector distributed equally and independently across all trees in the forest. Increasing the number of trees in the forest causes the generalization error to converge to a limit. An individual tree classifier's generalization error relies on the strength of individual trees in the forest as well as the correlation between them. Each node is split by using a random selection of features which generates error rates that are favorably comparable to those obtained from AdaBoost. However, they are more robust in terms of noise. The response to the increased number of features, in splitting, is evaluated by strength, error, and correlation monitored by internal estimates.^[63] Parallel ensembling is used in an RF classifier to fit numerous decision tree classifiers, in parallel, on various dataset samples. Overwhelming voting or average is used to yield the final result, therefore reducing over-fitting and improving the accuracy of prediction. Thus, single decision tree-based models are less accurate than multiple decision tree-based models. A combination of bootstrap aggregation and random feature selection is used to develop a series of decision trees, and applicable to both regression or classification problems.^[57] The generality of the RF model was assessed within a grid search as discussed in section 2.2. The maximum number of features for the best split (*max_features*), the minimum number of samples in each leaf (*min_samples_leaf*), the number of gradients boosted trees

(*n_estimator*), and the minimum number of samples for splitting an internal node (*min_samples_split*), i.e. the hyperparameters, were adapted in the ranges as (2, 3, 4, 5, 6, 7 and 8), (1, 2, 3, 4, 5, 6, 7 and 8), (100-1000), and (0.5, 1, 2, 3, 4, 5 and 6), respectively.

2.5. AdaBoost

The AdaBoost approach can be used for both regression and classification cases.^[64] AdaBoost is an ensemble learning process where an iterative process is used to improve poor classifiers by learning from the relevant errors. Sequential ensembling is employed in AdaBoost, whereas parallel ensembling is employed in RF. To achieve a powerful classifier with high accuracy, a lot of poor performing classifiers are combined. Although the remarkably improved efficiency of the classifier introduces AdaBoost as an adaptive classifier, it could yield overfits. Overall, AdaBoost is sensitive to either outliers or noisy data, but it is highly promising for binary classification problems in which the performance of decision trees should be boosted.^[57] In this research, the development of the AdaBoost model has been generalized based on the condition discussed in section 2.2. Optimization and tuning the hyperparameters were conducted by a grid search to identify the best loss function. There were two key hyperparameters in this analysis: learning rate (0.1, 0.5, 1, 2, 3, 4 and 5) and various gradients boosted trees (*n_estimator*) in the range of 20-500. It is notable that goodness of fitness was assessed by the learning curve for all GBM, RF, and AdaBoost models.

2.6. Evaluation of variable importance

Permutation variable importance (PVI) approach is a method for inspection of fitted models from tabular data.^[63] In general, PVI takes advantage of its generality, ease and rate of calculation, and analyzing either interactive or individual effects of variables.^[65-67] In such an approach, random alteration of the input is used for consideration of the model errors in

prediction of the output. Thus, a feature is more important when there are more errors.^[68] With regard to the errors, the feature importance was measured by MSE. It is notable that input variables importance in photocatalytic degradation of PFOA over different photocatalysts was evaluated by PVI approach for all models (GBM, RF, and AdaBoost).

2.7. Comparison of model performance and strength

The strengths and performances of different ML-based models in simulating PFOA photodegradation over different photocatalysts were compared by considering key statistical indices such as MSE, MAE, and R^2 which were calculated using the test datasets. MAE was calculated based on equation (4):^[36]

$$MAE = 1 - \frac{\sum_{i=1}^n |y_i - z_i|}{n} \quad (4)$$

where n , y_i , and z_i are the overall number of data points, predicted and actual values, consecutively.

3. Results and discussion

3.1. Selection of ML algorithms

The performances of seven ML algorithms used for the estimation of PFOA photocatalytic degradation were evaluated (Table 1). According to the statistical indices and in order to compare performance and PVI of different algorithms, the first three models with the highest prediction strengths, i.e. GBM, RF and AdaBoost, were nominated to deeply model and analyze the photocatalytic degradation of PFOA over various photocatalysts. The high efficiencies of these three algorithms have been reported in other applications such as the prediction of H_2 production by fermentation process.^[36]

Table 1. Performance comparison of seven ML algorithms for photocatalytic decomposition of PFOA by different photocatalysts.

	GBM	RR	AdaBoost	RF	LR	SVR	MLP
Train MSE	171.14	608.13	444.61	54.86	618.57	627.97	484.82
Test MSE	189.18	693.48	496.30	178.14	644.62	618.97	552.07
Total-Train R^2	0.80	0.26	0.458	0.93	0.278	0.267	0.409
Total-Test R^2	0.78	0.28	0.485	0.81	0.211	0.242	0.428

3.2. AdaBoost

A grid search, with each loss function, was used to tune the various hyperparameters conditions in order to nominate the most suitable loss function. The values of tuned hyperparameters along with the relevant values of R^2 and MSE for various phases of modeling, i.e. exponential, square, and linear, are shown in Table 2. As evident, square phase resulted in the best loss function. The corresponding learning rate and $n_estimator$ were 2 and 40, respectively. The values of MSE for both validation and training phases, i.e. 401.877 and 343.997, and R^2 for both validation and training phases, i.e. 0.515 and 0.587, were calculated accordingly. The model performance in PFOA photocatalytic degradation over several photocatalysts was evaluated by providing a test dataset including one fifth of the hidden data points. As evident in Table 2 and Fig. 1, a medium strength of prediction (57%) with an MSE of 388.369 was provided by this model in the test phase. Figure 1(b) illustrates learning conditions of training and validation for the developed AdaBoost model in various epochs.

Table 2. Outcomes of AdaBoost model by using several loss functions with adapted hyperparameters.

	Grid search		R^2	R^2	R^2	R^2	MSE	MSE	MSE	MSE
	learning rate	n-estimator	Test	Total-Train	Validation	Train	Test	Total-Train	Validation	Train
Square	2	40	0.574	0.607	0.515	0.587	388.369	327.102	401.877	343.997
Exponential	5	20	0.444	0.477	0.430	0.496	465.523	445.783	429.341	484.286
Linear	2	160	0.430	0.465	0.445	0.474	485.895	453.803	466.045	444.846

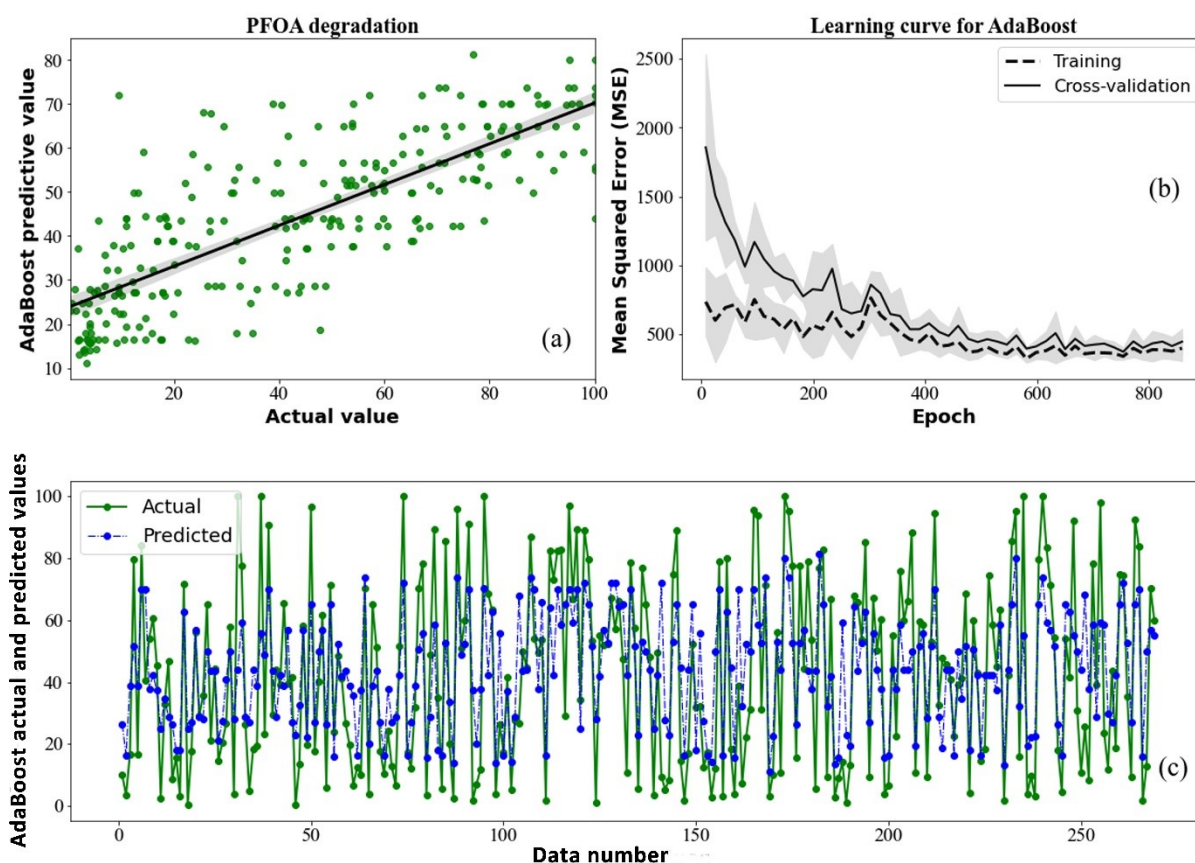


Figure 1. (a) Coefficient of correlation (in test phase), (b) learning curve, and (c) strength of prediction (in test phase) of the model developed by AdaBoost.

There is a significant difference between the prediction strength of this model for estimating photocatalytic degradation of tetracycline antibiotics^[33] and that of PFOA (this research). Firstly, another pollutant was considered in the other modeling, so the targets are different. Secondly, another type of catalysts (MOFs) was considered in their research. Under these circumstances, similar outcomes might not be expected since they are both of crucial importance. More importantly, different model performances have been reported in different applications although a similar modeling approach was used.^[53,69]

3.3. Gradient boosting machine

Tuning all hyperparameters was performed in a grid search to evolve a GBM model. The best situations of the hyperparameters were obtained consisting of 3 (*min_samples_leaf*), 100

($n_estimator$), 5 (max_depth), 6 ($min_samples_split$), and 6 ($max_features$). Both cross-validation and training phases were conducted under the mentioned conditions. The corresponding MSE values were 114.061 and 42.750; and R^2 values were 0.864 and 0.950 for these phases, respectively. In addition, MSE values of 146.90 and 51.485, and R^2 values of 0.840 and 0.937 were derived for the test and total training phases, respectively, signifying the remarkable potential of this model with a prediction strength of 84% in the photocatalytic degradation of PFOA over various photocatalysts (total training includes both cross-validation and train phases). The promising strength of prediction obtained by using this model in the test phase can be evidenced in Fig. 2.

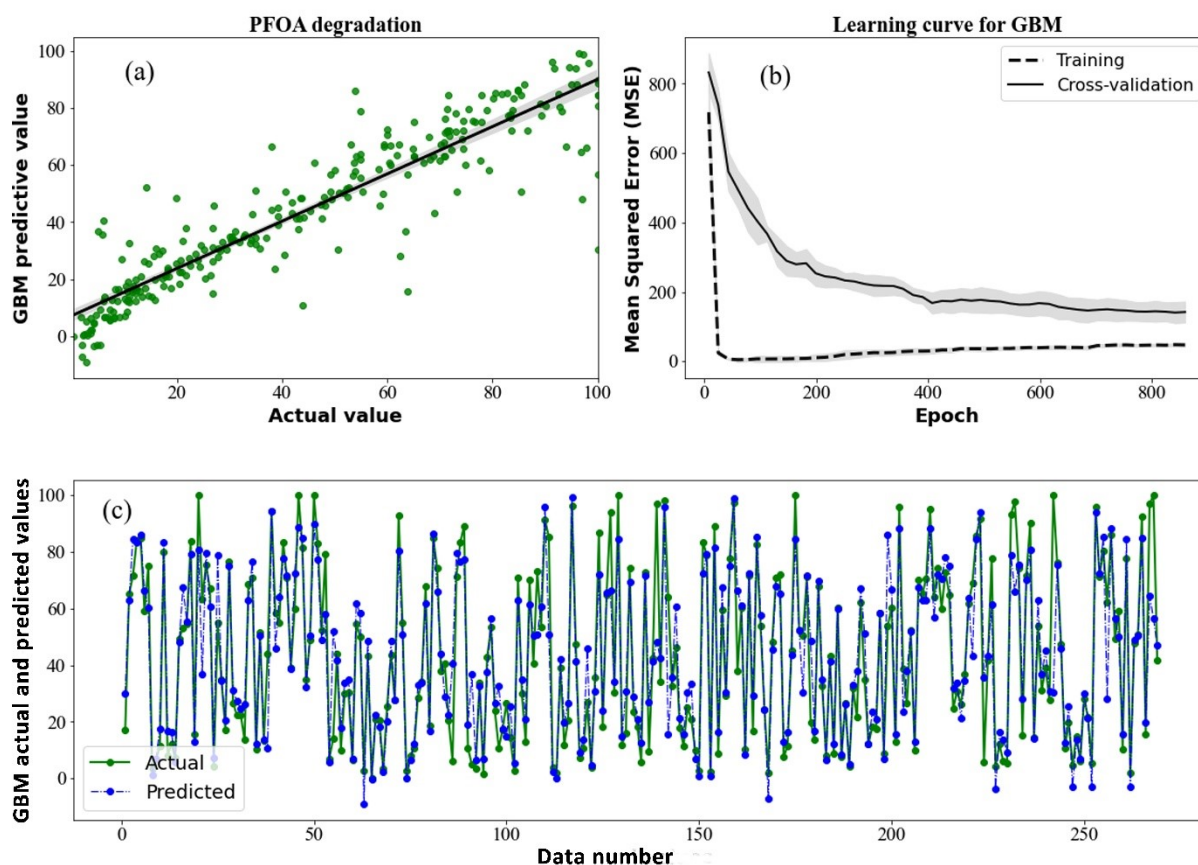


Figure 2. (a) Coefficient of correlation (in test phase), (b) learning curve, and (c) strength of prediction (in test phase) of the model developed by GBM.

Overfitting potential and goodness of fitting are among major criteria used for evaluation of the model performance. Figure 2(b) depicts the learning curve for cross-validation and training conditions of the GBM model developed. As evident, the MSEs obtained were sufficiently low and became stable at epochs higher than 600. Moreover, their insignificant difference at high epochs represents lack of overfitting and good strength of prediction in the GBM model.

3.4. Random forest

A grid search was employed to adapt the hyperparameters in development of the RF model. The optimized conditions of the hyperparameters including the minimum number of samples in each split (*min_samples_split*), the maximum number of features for the best split (*max_features*), the minimum number of samples in each leaf (*min_samples_leaf*), and the number of boosted trees (*n_estimators*) were 2, 2, 1 and 300, consecutively. The corresponding MSE values were 169.457 and 57.950; and R^2 values were 0.802 and 0.932 for validation and training phases, respectively. Consequently, the MSE values were 107.500 and 58.426; and R^2 values were 0.867 and 0.932 for test and total train phases, respectively, representing a remarkable strength of prediction equals to 86.7% in estimation of photocatalytic degradation of PFOA over different photocatalysts.

Considering the test phase, the fitting situations of the test dataset are illustrated in Fig. 3 where the promising strength of prediction obtained by RF model can be evidenced. Figure 3(b) depicts the MSEs of both cross-validation and training phases versus epoch for the model developed by RF. It is noteworthy that similar to the model developed by GBM, a decreasing trend is observed by increasing epochs in both phases. Furthermore, any significant overfitting is not evidenced by the model obtained by RF. Although both phases follow similar patterns, their difference became smaller at sufficiently high epochs (≥ 600) where the MSEs became stable. These findings clearly signify an acceptable outcome for the developed model by RF.

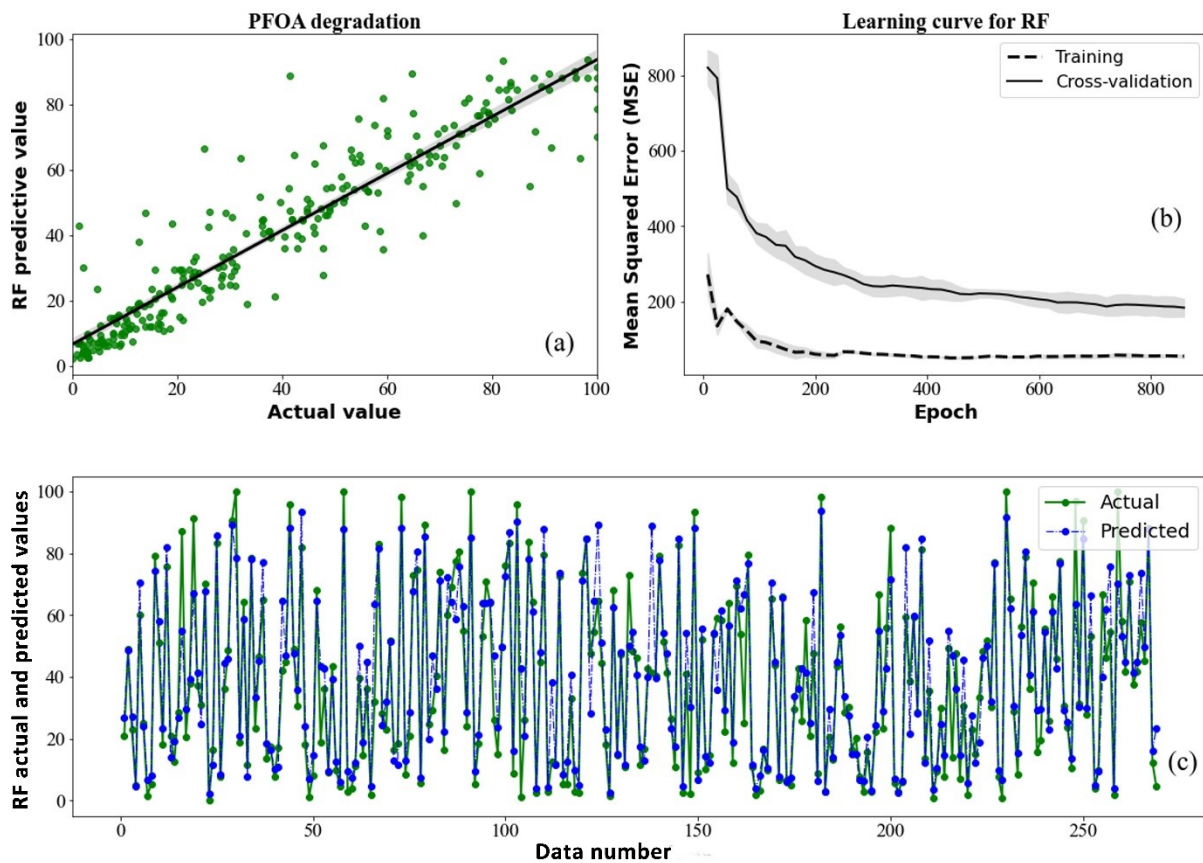


Figure 3. (a) Coefficient of correlation (in test phase), (b) learning curve, and (c) strength of prediction (in test phase) of the model developed by RF.

3.5. Evaluation of variable importance

PVI approach was employed to evaluate the relative importance of all variables in the three models developed by AdaBoost, GBM and RF. As presented in Fig. 4, the relative importance values of variables are highly dependent on the methods of modeling. In addition, variables are of significantly different importance in each model.

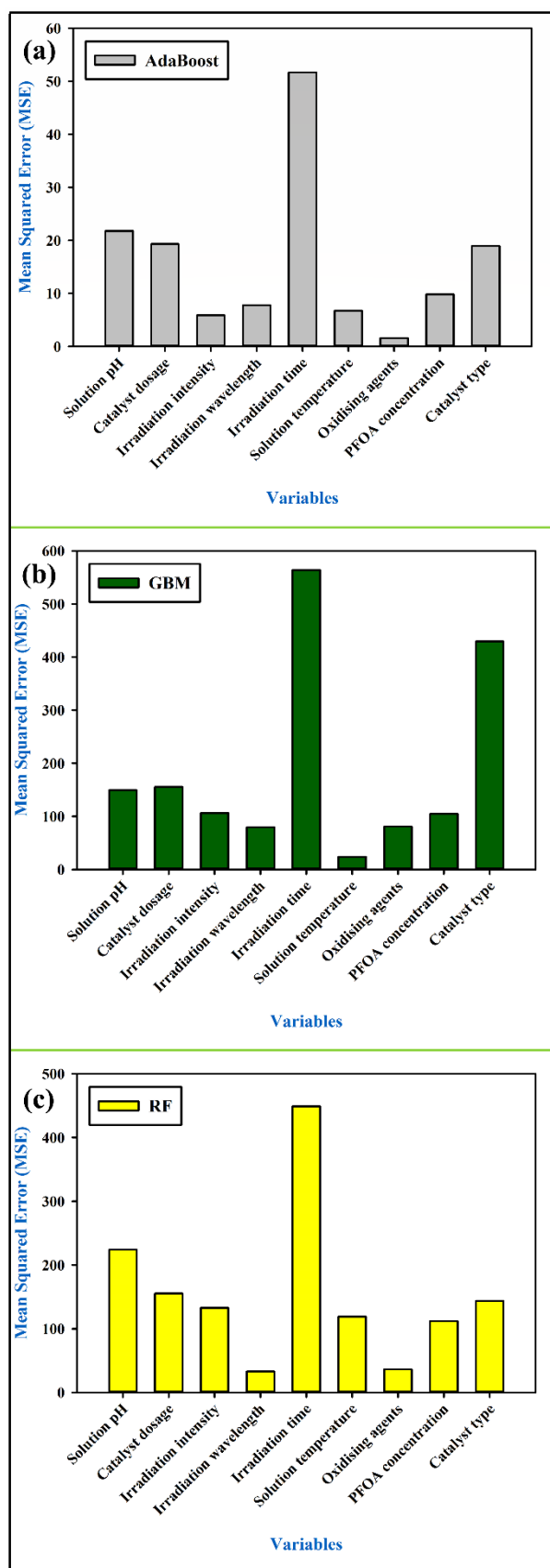


Figure 4. Relative importance of variables evaluated by PVI approach for the models developed by (a) AdaBoost, (b) GBM, and (c) RF.

3.5.1. Light irradiation time

Among all variables, the light irradiation time showed the most significant effect on the photocatalytic degradation of PFOA in the AdaBoost, GBM, and RF models. The results are consistent with the experimental findings. For instance, in studying the photocatalytic degradation of PFOA using TiO₂-MWCNT (mass ratio 2:1) composite catalyst (1.6 g/L), initial pH of 5, and initial PFOA concentration of 30 mg/L^[51], Song et al. (2012) reported approximately 34%, 55%, 67%, 74%, 75%, 80%, 83% and 85% degradation after 1 h, 2 h, 3 h, 4 h, 5 h, 6 h, 7 h and 8 h, respectively. It is obvious that increasing irradiation time enhances the degradation efficiency of PFOA though at varying degrees. Notably, a remarkable degradation efficiency of 34% was observed after just 1 h. When it comes to the last interval, increasing irradiation time from 7 h to 8 h has slightly improved the degradation efficiency from 83% to 85%. Such a remarkable increase of degradation efficiency between the first and last intervals signifies the crucial importance of irradiation time on PFOA photocatalytic degradation. In addition, increasing irradiation time from 1 h to 8 h drastically enhanced the efficiency of PFOA decomposition from 34% to 85%. The photocatalytic process undergoes some sequential steps in the liquid phase as follows:^[70]

- i. transfer of pollutants to the surface of semiconductor material,
- ii. pollutants adsorption on the surface of photo-activated semiconducting material,
- iii. photogeneration of ROSs such as hydroxyl radicals, followed by the photodegradation of organic pollutants,
- iv. products/intermediates desorption from the surface of semiconductor material, and
- v. transfer of the final products/intermediates into the liquid phase.

It is notable that the rates of these steps change significantly at different irradiation times which highly affects the degradation efficiency of organic pollutants. Overall, among all the variables, irradiation time can play the most important role in the photocatalytic degradation of PFOA which is consistent with the results obtained by the AdaBoost, GBM, and RF models.

3.5.2. Type of catalyst

The type of catalyst is another major variable that highly affects PFOA photocatalysis, which could be even more important than that of irradiation time. Although there are various kinds of semiconductors, only a few of which could be highly efficient for PFOA photocatalytic degradation. In other words, the effect of irradiation time on PFOA degradation efficiency is high on the condition of using an appropriate photocatalyst. Otherwise, a negligible effect is expected. For instance, Li et al. (2012) have assessed the photocatalytic degradation of PFOA over both TiO_2 and In_2O_3 under similar experimental conditions.^[12] As shown in Fig. 5(a), unlike In_2O_3 , a negligible efficiency was observed for TiO_2 whereas increasing irradiation time did not affect the degree of degradation significantly. Although TiO_2 is the most commonly studied catalyst for the degradation of organic pollutants, it has not exhibited high photocatalytic activities for PFOA decomposition.^[11,12] On the other hand, other catalysts including In_2O_3 and Ga_2O_3 has shown promising efficiencies for such a purpose.^[11,12,26] Promising activity of Ga_2O_3 -based semiconductors for the photocatalytic remediation of PFOA has been attributed to their high energy sustainability and oxidizing potential.^[71]

Apart from the oxidizing potential, which is related to the position of valence and conduction bands of the semiconducting material, surface properties of the catalyst are among key parameters affecting PFOA degradation efficacy. The schematic illustration of adsorption of PFOA on the surface of TiO_2 and In_2O_3 is shown in Fig. 5(b). In the case of In_2O_3 , PFOA strongly coordinates to its surface, in either bridging or bidentate modes, which leads to the vertical array of PFOA chain along with a good order on In_2O_3 . As per TiO_2 , a tilted array is observed originated from surface binding between TiO_2 and the carboxylate group of PFOA. Thus, unlike TiO_2 , the inner CF_2 groups of PFOA may interact rarely with surface of In_2O_3 . In addition, the bidentate mode between In_2O_3 and PFOA facilitates the transfer of electrons from PFOA to the holes (in In_2O_3) which are responsible for superior efficacy of In_2O_3 than TiO_2 for photocatalytic decomposition of PFOA.^[12]

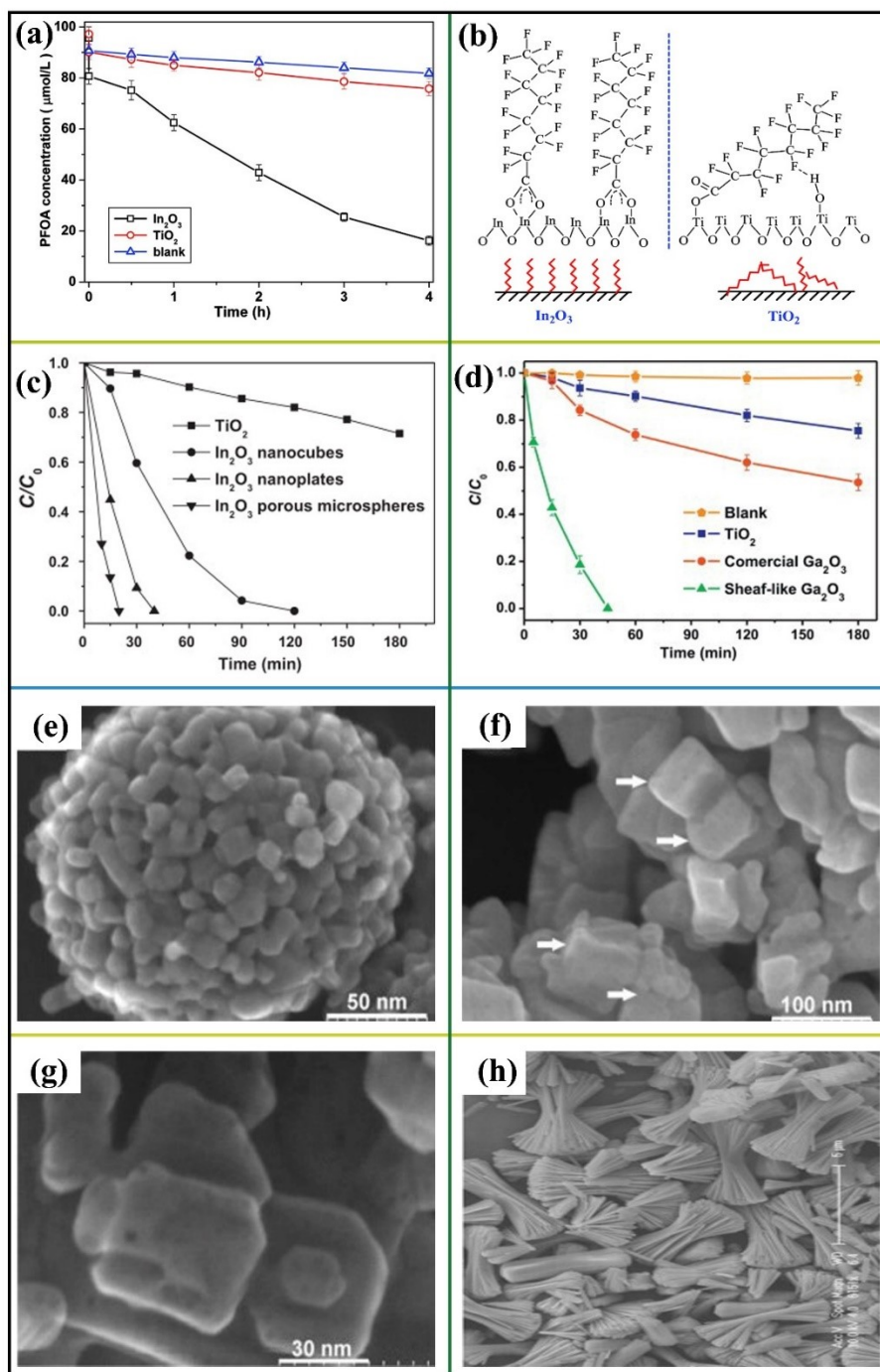


Figure 5. (a) Comparison of the photocatalytic activity of In_2O_3 and TiO_2 for PFOA decomposition. Reproduced with permission from .^[12] Copyright © 2012, Li et al., ACS Publications. (b) Schematic illustration of adsorption of PFOA on the surface of TiO_2 and In_2O_3 . Reproduced with permission from .^[12] Copyright © 2012, Li et al., ACS Publications. (c) Effect of various morphologies of In_2O_3 on the photocatalytic degradation of PFOA. Reproduced with permission from .^[11] Copyright © 2013, Li et al., Elsevier. (d) Effect of various morphologies of Ga_2O_3 on the photocatalytic degradation of PFOA. Reproduced with permission from .^[14] Copyright © 2013, Shao et al., Elsevier. (e) In_2O_3 microspheres. Reproduced with permission from .^[11] Copyright © 2013, Li et al., Elsevier. (f) In_2O_3 nanocubes. Reproduced with

permission from .^[11] Copyright © 2013, Li et al., Elsevier. (g) In₂O₃ nanoplates. Reproduced with permission from .^[11] Copyright © 2013, Li et al., Elsevier. and (h) Sheaf-like Ga₂O₃ nanostructures. Reproduced with permission from .^[14] Copyright © 2013, Shao et al., Elsevier.

Moreover, the morphology of the catalysts could remarkably affect their performance in photocatalysis. Li et al. (2013) used nanostructured In₂O₃ with various morphologies including nanoplates, nanocubes, and porous microsphere for photocatalytic decomposition of PFOA.^[11] As shown in Fig. 5(c), the efficacy of In₂O₃ highly depends on its morphology. Shao et al. (2013) also evaluated the important effects of morphology on the photocatalytic decomposition of PFOA over Ga₂O₃, as shown in Fig. 5(d). The presence of a large number of nanopores along with the higher surface area of sheaf-like Ga₂O₃ than that of commercial Ga₂O₃ resulted in the higher efficacy of sheaf-like Ga₂O₃ than that of commercial Ga₂O₃. Notably, enlarged surface area could increase the reaction centers and provides more adsorption.^[14] Various morphologies of In₂O₃ microspheres, In₂O₃ nanocubes, In₂O₃ nanoplates, and sheaf-like nanostructured Ga₂O₃ are shown in Fig. 5(e-h). Various morphologies not only result in different specific surface areas, but also influence the light adsorption of catalysts which can significantly affect their efficiencies in photocatalysis. For instance, the improved absorption of scattered radiation or the reduced loss of photons in solution led to the higher efficiency of titania nanotubes with 3D structures than those with planar arrays.^[24]

The recombination rate of photo-generated electron-hole pairs is another important factor which markedly influences the efficiency of PFOA photodecomposition. Generally, the lower the recombination rate, the higher the efficiency is expected. The production of composites, loading by metallic particles, and doping are among main methods which can highly reduce the recombination rate of charge carriers. For instance, Cu- and Fe-loaded TiO₂ catalysts significantly increased the decomposition efficiency of PFOA to 91% and 69%, respectively, compared with only 14% for unloaded TiO₂.^[49]

Considering several aspects including nature of the catalyst (position of valence and conduction bands), surface properties, morphology, specific surface area, pore size, recombination rate of charge carriers and photo-adsorption ability, it is obvious that catalyst type is one of the most crucial parameters affecting the photocatalytic decomposition percentage of PFOA. This finding is successfully confirmed by the GBM model.

3.5.3. Other variables

Tang et al. (2021) have evaluated the effects of various parameters including pH, irradiation intensity, reaction temperature, initial concentration of PFOA, and catalyst dosage on the photocatalytic decomposition of PFOA over different catalysts including CeO₂, NiAl-LDHs, and CeO₂@NiAl-LDHs composites. Among all the variable, the greatest effect has been observed for catalyst dosage, pH, and irradiation intensity.^[23] As shown in Fig. 4(b), catalyst dosage, pH, and irradiation intensity are the most effective parameters after irradiation time and catalyst type which clearly confirm the high accuracy of the model developed by GBM in estimating the photocatalytic degradation of PFOA. It is noteworthy that the dosage of catalyst and initial concentration of pollutant can highly affect decomposition of organic pollutants, though at varying degrees, since the rate of the five sequential steps in photocatalysis is strongly related to these variables. The initial solution pH is also of high importance since it strongly affects the adsorption of PFOA on the surface of catalyst. However, its effect is highly related to the type of catalyst. For instance, the optimal pH value was reported to be pH 9 (among pH 3, 5, 7, 9, 11) for CeO₂, NiAl-LDHs and CeO₂@NiAl-LDHs composites,^[23] and pH 3 for Ga₂O₃ (among 3, 5, 7, 10).^[26]

The higher the irradiation intensity, the greater the degradation efficiency is expected. Moreover, the photo-adsorption ability of catalysts is strongly dependent on the wavelength. Therefore, the source of light, in terms of either intensity or wavelength, can remarkably affect the photocatalytic decomposition of PFOA. In addition to the source of light, oxidizing agents

(i.e. PMS, APS and NaPS) significantly increase the photocatalytic degradation of PFOA.^[25-27] Such relative importance of these variables is appropriately demonstrated by the developed GBM model.

Solution temperature is also another major variable affecting the decomposition of PFOA. It should be taken into account that solution temperature has been set at room temperature, in the overwhelming majority of photocatalytic experiments. The main reason might be the challenges associated with increasing or decreasing temperature. On the one hand, increasing temperature might result in evaporation of water. On the other hand, decreasing temperature necessitates using cooling facilities. Overall, photocatalytic experiments have been mainly carried out in a limited range of solution temperature. Furthermore, the maximum degradation efficiency of PFOA has been observed at room temperature in some cases.^[25,46] However, it highly depends on the type of catalyst since Tang et al. (2021) reported higher decomposition efficiencies at higher temperatures.^[23] Considering the limited temperature range used for evaluation of the temperature effect on the decomposition of PFOA, the minimal relative importance of variables could be devoted to solution temperature which is evidenced by the GBM model (Fig. 4(b)).

3.6. Model comparison

The strengths of the models developed by AdaBoost, GBM, and RF in predicting the photocatalytic decomposition of PFOA were evaluated by different statistical indices (Table 3). Compared with the model developed by AdaBoost, those developed by GBM and RF showed higher prediction strengths in term of R^2 . In terms of MAE and MSE, the model developed by AdaBoost showed more errors than those developed by GBM and RF. Overall, considering both error values and R^2 , the AdaBoost model showed a much lower performance than RF and GBM models. Notably, both RF and GBM models showed approximately similar performances in predicting the photocatalytic degradation of PFOA over various photocatalysts. The

application of several ML algorithms in estimating the photocatalytic degradation of different organic pollutants is provided in Table 3. As evident, the number of datapoints in this research is remarkably more than in the other studies demonstrating the diversity of the considered experimental conditions in this study. As a result, the prediction strengths of the GBM and RF models in this study are similar to those by using other models for estimating the photocatalytic degradation of different organic pollutants. The insignificant differences between the prediction strengths of current models, i.e. GBM and RF, and the models used for other applications could be attributed to the nature of data, e.g. diversity of photocatalysts, number and type of inputs, and type of the algorithms. Notably, different algorithms could yield different performances in various applications.^[36,69]

Table 3. Comparing the prediction strengths of models from ML algorithms for photocatalytic degradation of different organic pollutants.

Model	Pollutant	Statistical indices				Number of data	Reference
		R ²	MAE	MSE	RMSE		
CGCNN-MF-ANN	Methylene Blue	-	0.286	-	-	67	[72]
CGCNN-MF-ANN	Rhodamine B	-	0.338	-	-	50	[72]
CGCNN-MF-ANN	Rose Bengal	-	0.095	-	-	33	[72]
CGCNN-MF-ANN	Toluidine Blue	-	0.127	-	-	31	[72]
CGCNN-MF-ANN	Azure B	-	0.142	-	-	31	[72]
CGCNN-MF-ANN	Carmine Indigo	-	0.275	-	-	22	[72]
CGCNN-MF-ANN	Phenoxyacetic acid	-	0.113	-	-	20	[72]
CatBoost	Tetracycline	0.989	-	-	3.164	374	[33]
LightGBM	Tetracycline	0.980	-	-	4.190	374	[33]
XGBoost	Tetracycline	0.984	-	-	3.717	374	[33]
AdaBoost	Tetracycline	0.981	-	-	4.086	374	[33]
AdaBoost	PFOA	0.574	16.480	388.369	19.707	1343	This research
GBDT	Tetracycline	0.981	-	-	4.112	374	[33]

Extra Tree	Tetracycline	0.978	-	-	4.378	374	[33]
DT	Tetracycline	0.981	-	-	4.058	374	[33]
RF	Tetracycline	0.979	-	-	4.341	374	[33]
RF	PFOA	0.867	6.796	107.500	10.368	1343	This research
GBM	PFOA	0.878	6.009	106.660	10.328	1343	This research

CGCNN-MF-ANN: crystal graph convolutional neural network-molecular fingerprint-artificial neural network; RMSE: root mean squared error.

4. Conclusions

The prediction of the photocatalytic degradation of PFOA as an emerging persistent organic pollutant is of great importance. Seven ML processes were pre-screened for such a purpose, among which AdaBoost, RF and GBM showed more promising performances considering statistical criteria including MSE, MAE and R^2 values. Being optimized by grid search, the models developed by RF and GBM showed superior performances than that developed by AdaBoost in predicting the photocatalytic decomposition of PFOA. Considering the relative importance of process variables evaluated by PVI, the GMB model resulted in better outcomes than RF model with the light irradiation time, type of catalyst, dosage of catalyst, solution pH, irradiation intensity, initial PFOA concentration, oxidizing agents (PMS, APS, and NaPS), irradiation wavelength, and solution temperature as the most effective process variables in decreasing order.

References

- [1] Xu, B.; Liu, S.; Zhou, J. L.; Zheng, C.; Jin, W.; Chen, B.; Zhang, T.; Qiu, W. PFAS and Their Substitutes in Groundwater: Occurrence, Transformation and Remediation. *J. Hazard. Mater.* 2021, 412, 125159.
- [2] Kwiatkowski, C. F.; Andrews, D. Q.; Birnbaum, L. S.; Bruton, T. A.; DeWitt, J. C.; Knappe, D. R. U.; Maffini, M. V.; Miller, M. F.; Pelch, K. E.; Reade, A.; et al. A Scientific Basis for Managing PFAS as a Chemical Class. *Environ. Sci. Technol. Lett.* 2020, 7 (8), 532-543.
- [3] Abunada, Z.; Alazaiza, M. Y. D.; Bashir, M. J. K. An Overview of Per- and Polyfluoroalkyl Substances (PFAS) in the Environment: Source, Fate, Risk and Regulations. *Water* 2020, 12 (12), 3590.
- [4] DeWitt, J. C.; Peden-Adams, M. M.; Keller, J. M.; Germolec, D. R. Immunotoxicity of perfluorinated compounds: recent developments. *Toxicol. Pathol.* 2012, 40(2), 300-311.

- [5] Boonya-atichart, A.; Boontanon, S. K.; Boontanon, N. Removal of Perfluorooctanoic Acid (PFOA) in Groundwater by Nanofiltration Membrane. *Water Sci. Technol.* 2016, 74 (11), 2627-2633.
- [6] Gao, P.; Cui, J.; Deng, Y. Direct Regeneration of Ion Exchange Resins with Sulfate Radical-Based Advanced Oxidation for Enabling a Cyclic Adsorption – Regeneration Treatment Approach to Aqueous Perfluorooctanoic Acid (PFOA). *Chem. Eng. J.* 2021, 405, 126698.
- [7] Rattanaoudom, R.; Visvanathan, C. Removal of PFOA by Hybrid Membrane Filtration Using PAC and Hydrotalcite. *Desalin. Water Treat.* 2011, 32 (1-3), 262-270.
- [8] Parsons, J.; Saez, M.; Dolfing, J.; De Voogt, P. Biodegradation of Perfluorinated Compounds. *Rev. Environ. Contam. Toxicol.* 2008, 196, 53-71.
- [9] Son, H.; Kim, T.; Yoom, H.-S.; Zhao, D.; An, B. The Adsorption Selectivity of Short and Long Per- and Polyfluoroalkyl Substances (PFASs) from Surface Water Using Powder-Activated Carbon. *Water* 2020, 12 (11), 3287.
- [10] Wang, S.; Yang, Q.; Chen, F.; Sun, J.; Luo, K.; Yao, F.; Wang, X.; Wang, D.; Li, X.; Zeng, G. Photocatalytic Degradation of Perfluorooctanoic Acid and Perfluorooctane Sulfonate in Water: A Critical Review. *Chem. Eng. J.* 2017, 328, 927-942.
- [11] Li, Z.; Zhang, P.; Shao, T.; Wang, J.; Jin, L.; Li, X. Different Nanostructured In₂O₃ for Photocatalytic Decomposition of Perfluorooctanoic Acid (PFOA). *J. Hazard. Mater.* 2013, 260, 40-46.
- [12] Li, X.; Zhang, P.; Jin, L.; Shao, T.; Li, Z.; Cao, J. Efficient Photocatalytic Decomposition of Perfluorooctanoic Acid by Indium Oxide and Its Mechanism. *Environ. Sci. Technol.* 2012, 46 (10), 5528-5534.
- [13] Mahapatra, C. T.; Damayanti, N. P.; Guffey, S. C.; Serafin, J. S.; Irudayaraj, J.; Sepúlveda, M. S. Comparative in Vitro Toxicity Assessment of Perfluorinated Carboxylic Acids. *J. Appl. Toxicol.* 2017, 37 (6), 699-708.

- [14] Shao, T.; Zhang, P.; Jin, L.; Li, Z. Photocatalytic Decomposition of Perfluorooctanoic Acid in Pure Water and Sewage Water by Nanostructured Gallium Oxide. *Appl. Catal. B: Environ.* 2013, 142-143, 654-661.
- [15] Sariket, D.; Shyamal, S.; Hajra, P.; Mandal, H.; Bera, A.; Maity, A.; Bhattacharya, C. Improvement of Photocatalytic Activity of Surfactant Modified In₂O₃ Towards Environmental Remediation. *New J. Chem.* 2018, 42 (4), 2467-2475.
- [16] Zhao, B.; Zhang, P. Photocatalytic Decomposition of Perfluorooctanoic Acid with β -Ga₂O₃ wide Bandgap Photocatalyst. *Catal. Commun.* 2009, 10 (8), 1184-1187.
- [17] Navidpour, A. H.; Salehi, M.; Salimijazi, H. R.; Kalantari, Y.; Azarpour Siahkali, M. Photocatalytic Activity of Flame-Sprayed Coating of Zinc Ferrite Powder. *J. Therm. Spray Technol.* 2017, 26 (8), 2030-2039.
- [18] Navidpour, A. H.; Fakhrzad, M. Photocatalytic and Magnetic Properties of ZnFe₂O₄ Nanoparticles Synthesised by Mechanical Alloying. *Int. J. Environ. Anal. Chem.* 2022, 102 (3), 690-706.
- [19] Navidpour, A. H.; Kalantari, Y.; Salehi, M.; Salimijazi, H. R.; Amirmasr, M.; Rismanchian, M.; Azarpour Siahkali, M. Plasma-Sprayed Photocatalytic Zinc Oxide Coatings. *J. Therm. Spray Technol.* 2017, 26 (4), 717-727.
- [20] Kamranifar, M.; Al-Musawi, T. J.; Amarzadeh, M.; Hosseinzadeh, A.; Nasseh, N.; Qutob, M.; Arghavan, F. S. Quick Adsorption Followed by Lengthy Photodegradation Using FeNi₃@SiO₂@ZnO: A Promising Method for Complete Removal of Penicillin G from Wastewater. *J. Water Process Eng.* 2021, 40, 101940.
- [21] Djilali, T.; Lebouachera, S. E. I.; Dechir, S.; Chekir, N.; Benhabiles, O.; Bentahar, F. Comparison between TiO₂ and ZnO Photocatalytic Efficiency for the Degradation of Tartrazine Contaminant in Water. *Int. J. Environ. Sci. Technol.* 2016, 1, 357-360.

- [22] Shinde, D.; Tambade, P.; Chaskar, M.; Gadave, K. Photocatalytic Degradation of Dyes in Water by Analytical Reagent Grades ZnO, TiO₂ and SnO₂ : A Comparative Study. *Drink. Water Eng. Sci.* 2017, 10, 109-117.
- [23] Tang, H.; Zhang, W.; Meng, Y.; Xia, S. A Direct Z-Scheme Heterojunction with Boosted Transportation of Photogenerated Charge Carriers for Highly Efficient Photodegradation of PFOA: Reaction Kinetics and Mechanism. *Appl. Catal. B: Environ.* 2021, 285, 119851.
- [24] Navidpour, A. H.; Hosseinzadeh, A.; Zhou, J. L.; Huang, Z. Progress in the Application of Surface Engineering Methods in Immobilizing TiO₂ and ZnO Coatings for Environmental Photocatalysis. *Catal. Rev.* 2022, <https://doi.org/10.1080/01614940.2021.1983066>.
- [25] Wu, D.; Li, X.; Zhang, J.; Chen, W.; Lu, P.; Tang, Y.; Li, L. Efficient PFOA Degradation by Persulfate-Assisted Photocatalytic Ozonation. *Sep. Purif. Technol.* 2018, 207, 255-261.
- [26] Xu, B.; Zhou, J. L.; Altaee, A.; Ahmed, M. B.; Jahir, M. A. H.; Ren, J.; Li, X. Improved Photocatalysis of Perfluorooctanoic Acid in Water and Wastewater by Ga₂O₃/UV System Assisted by Peroxymonosulfate. *Chemosphere* 2020, 239, 124722.
- [27] Xu, B.; Ahmed, M. B.; Zhou, J. L.; Altaee, A. Visible and UV Photocatalysis of Aqueous Perfluorooctanoic Acid by TiO₂ and Peroxymonosulfate: Process Kinetics and Mechanistic Insights. *Chemosphere* 2020, 243, 125366.
- [28] Sanongraj, W.; Chen, Y.; Crittenden, J. C.; Destailats, H.; Hand, D. W.; Perram, D. L.; Taylor, R. Mathematical Model for Photocatalytic Destruction of Organic Contaminants in Air. *J. Air Waste Manag. Assoc.* 2007, 57 (9), 1112-1122.
- [29] Zekri, M. e. M.; Colbeau-Justin, C. A Mathematical Model to Describe the Photocatalytic Reality: What Is the Probability that a Photon Does Its Job? *Chem. Eng. J.* 2013, 225, 547-557.
- [30] Sannino, D.; Vaiano, V.; Sacco, O.; Ciambelli, P. Mathematical Modelling of Photocatalytic Degradation of Methylene Blue under Visible Light Irradiation. *J. Environ. Chem. Eng.* 2013, 1 (1), 56-60.

- [31] Tokode, O.; Prabhu, R.; Lawton, L. A.; Robertson, P. K. J. Mathematical Modelling of Quantum Yield Enhancements of Methyl Orange Photooxidation in Aqueous TiO₂ Suspensions under Controlled Periodic UV LED Illumination. *Appl. Catal. B: Environ.* 2014, 156-157, 398-403.
- [32] Hamed, H.; Mohammadzadeh, O.; Rasouli, S.; Zendejboudi, S. A Critical Review of Biomass Kinetics and Membrane Filtration Models for Membrane Bioreactor Systems. *J. Environ. Chem. Eng.* 2021, 106406.
- [33] Abdi, J.; Hadipour, M.; Hadavimoghaddam, F.; Hemmati-Sarapardeh, A. Estimation of Tetracycline Antibiotic Photodegradation from Wastewater by Heterogeneous Metal-Organic Frameworks Photocatalysts. *Chemosphere* 2022, 287, 132135.
- [34] Heijungs, R. On the Number of Monte Carlo Runs in Comparative Probabilistic LCA. *Int. J. Life Cycle Assess.* 2020, 25 (2), 394-402.
- [35] Afzal, A.; Saleel, C. A.; Bhattacharyya, S.; Satish, N.; Samuel, O. D.; Badruddin, I. A. Merits and Limitations of Mathematical Modeling and Computational Simulations in Mitigation of COVID-19 Pandemic: A Comprehensive Review. *Arch. Comput. Methods Eng.* 2022, 29, 1311-1337.
- [36] Hosseinzadeh, A.; Zhou, J. L.; Altaee, A.; Li, D. Machine Learning Modeling and Analysis of Biohydrogen Production from Wastewater by Dark Fermentation Process. *Bioresour. Technol.* 2022, 343, 126111.
- [37] Wuest, T.; Weimer, D.; Irgens, C.; Thoben, K.-D. Machine Learning in Manufacturing: Advantages, Challenges, and Applications. *Prod. Manuf. Res.* 2016, 4 (1), 23-45.
- [38] Sultana, N.; Hossain, S. M. Z.; Abusaad, M.; Alanbar, N.; Senan, Y.; Razzak, S. A. Prediction of Biodiesel Production from Microalgal Oil Using Bayesian Optimization Algorithm-Based Machine Learning Approaches. *Fuel* 2022, 309, 122184.

- [39] Chen, M.-J.; Lo, S.-L.; Lee, Y.-C.; Kuo, J.; Wu, C.-H. Decomposition of Perfluorooctanoic Acid by Ultraviolet Light Irradiation with Pb-Modified Titanium Dioxide. *J. Hazard. Mater.* 2016, 303, 111-118.
- [40] Duan, L.; Wang, B.; Heck, K.; Guo, S.; Clark, C. A.; Arredondo, J.; Wang, M.; Senftle, T. P.; Westerhoff, P.; Wen, X.; et al. Efficient Photocatalytic PFOA Degradation over Boron Nitride. *Environ. Sci. Technol. Lett.* 2020, 7 (8), 613-619.
- [41] Yao, X.; Zuo, J.; Wang, Y.-J.; Song, N.-N.; Li, H.-H.; Qiu, K. Enhanced Photocatalytic Degradation of Perfluorooctanoic Acid by Mesoporous Sb₂O₃/TiO₂ Heterojunctions. *Front. Chem.* 2021, 9 (352), 690520.
- [42] Xu, T.; Zhu, Y.; Duan, J.; Xia, Y.; Tong, T.; Zhang, L.; Zhao, D. Enhanced Photocatalytic Degradation of Perfluorooctanoic Acid Using Carbon-Modified Bismuth Phosphate Composite: Effectiveness, Material Synergy and Roles of Carbon. *Chem. Eng. J.* 2020, 395, 124991.
- [43] Qanbarzadeh, M.; Wang, D.; Ateia, M.; Sahu, S. P.; Cates, E. L. Impacts of Reactor Configuration, Degradation Mechanisms, and Water Matrices on Perfluorocarboxylic Acid Treatment Efficiency by the UV/Bi₃O(OH)(PO₄)₂ Photocatalytic Process. *ACS ES&T Eng.* 2021, 1 (2), 239-248.
- [44] Song, Z.; Dong, X.; Fang, J.; Xiong, C.; Wang, N.; Tang, X. Improved Photocatalytic Degradation of Perfluorooctanoic Acid on Oxygen Vacancies-Tunable Bismuth Oxychloride Nanosheets Prepared by a Facile Hydrolysis. *J. Hazard. Mater.* 2019, 377, 371-380.
- [45] Li, Z.; Zhang, P.; Shao, T.; Li, X. In₂O₃ Nanoporous Nanosphere: A Highly Efficient Photocatalyst for Decomposition of Perfluorooctanoic Acid. *Appl. Catal. B: Environ.* 2012, 125, 350-357.
- [46] Wu, D.; Li, X.; Tang, Y.; Lu, P.; Chen, W.; Xu, X.; Li, L. Mechanism Insight of PFOA Degradation by ZnO Assisted-Photocatalytic Ozonation: Efficiency and Intermediates. *Chemosphere* 2017, 180, 247-252.

- [47] Wu, Y.; Hu, Y.; Han, M.; Ouyang, Y.; Xia, L.; Huang, X.; Hu, Z.; Li, C. Mechanism Insights into the Facet-Dependent Photocatalytic Degradation of Perfluorooctanoic Acid on BiOCl Nanosheets. *Chem. Eng. J.* 2021, 425, 130672.
- [48] Li, M.; Yu, Z.; Liu, Q.; Sun, L.; Huang, W. Photocatalytic Decomposition of Perfluorooctanoic Acid by Noble Metallic Nanoparticles Modified TiO₂. *Chem. Eng. J.* 2016, 286, 232-238.
- [49] Chen, M.-J.; Lo, S.-L.; Lee, Y.-C.; Huang, C.-C. Photocatalytic Decomposition of Perfluorooctanoic Acid by Transition-Metal Modified Titanium Dioxide. *J. Hazard. Mater.* 2015, 288, 168-175.
- [50] Shang, E.; Li, Y.; Niu, J.; Li, S.; Zhang, G.; Wang, X. Photocatalytic Degradation of Perfluorooctanoic Acid over Pb-BiFeO₃/rGO Catalyst: Kinetics and Mechanism. *Chemosphere* 2018, 211, 34-43.
- [51] Song, C.; Chen, P.; Wang, C.; Zhu, L. Photodegradation of Perfluorooctanoic Acid by Synthesized TiO₂-MWCNT Composites under 365nm UV Irradiation. *Chemosphere* 2012, 86 (8), 853-859.
- [52] Sahu, S. P.; Qanbarzadeh, M.; Ateia, M.; Torkzadeh, H.; Maroli, A. S.; Cates, E. L. Rapid Degradation and Mineralization of Perfluorooctanoic Acid by a New Petitjeanite Bi₃O(OH)(PO₄)₂ Microparticle Ultraviolet Photocatalyst. *Environ. Sci. Technol. Lett.* 2018, 5 (8), 533-538.
- [53] Hosseinzadeh, A.; Baziar, M.; Alidadi, H.; Zhou, J. L.; Altaee, A.; Najafpour, A. A.; Jafarpour, S. Application of Artificial Neural Network and Multiple Linear Regression in Modeling Nutrient Recovery in Vermicompost under Different Conditions. *Bioresour. Technol.* 2020, 303, 122926.
- [54] Hosseinzadeh, A.; Zhou, J. L.; Altaee, A.; Baziar, M.; Li, D. Effective modelling of hydrogen and energy recovery in microbial electrolysis cell by artificial neural network and adaptive network-based fuzzy inference system. *Bioresour. Technol.* 2020, 316, 123967.

- [55] Jerome, H. F. Greedy Function Approximation: A Gradient Boosting Machine. *Ann. Stat.* 2001, 29 (5), 1189-1232.
- [56] Cai, J.; Xu, K.; Zhu, Y.; Hu, F.; Li, L. Prediction and Analysis of Net Ecosystem Carbon Exchange Based on Gradient Boosting Regression and Random Forest. *Appl. Energy* 2020, 262, 114566.
- [57] Sarker, I. H. Machine Learning: Algorithms, Real-World Applications and Research Directions. *SN Comput. Sci.* 2021, 2 (3), 160.
- [58] Natekin, A.; Knoll, A. Gradient Boosting Machines, A Tutorial. *Front. Neurorobotics* 2013, 7, 21.
- [59] Nguyen, H.; Vu, T.; Vo, T. P.; Thai, H.-T. Efficient Machine Learning Models for Prediction of Concrete Strengths. *Constr. Build. Mater.* 2021, 266, 120950.
- [60] Grillone, B.; Danov, S.; Sumper, A.; Cipriano, J.; Mor, G. A Review of Deterministic and Data-Driven Methods to Quantify Energy Efficiency Savings and to Predict Retrofitting Scenarios in Buildings. *Renew. Sustain. Energy Rev.* 2020, 131, 110027.
- [61] Zhou, L.; Fujita, H.; Ding, H.; Ma, R. Credit Risk Modeling on Data with Two Timestamps in Peer-To-Peer Lending by Gradient Boosting. *Appl. Soft Comput.* 2021, 110, 107672.
- [62] Li, Y.; Zou, C.; Berecibar, M.; Nanini-Maury, E.; Chan, J. C. W.; van den Bossche, P.; Van Mierlo, J.; Omar, N. Random Forest Regression for Online Capacity Estimation of Lithium-Ion Batteries. *Appl. Energy* 2018, 232, 197-210.
- [63] Breiman, L. Random Forests. *Mach. Learn.* 2001, 45 (1), 5-32.
- [64] Min, H.; Luo, X. Calibration of Soft Sensor by Using Just-In-Time Modeling and AdaBoost Learning Method. *Chin. J. Chem. Eng.* 2016, 24 (8), 1038-1046.
- [65] Wei, P.; Lu, Z.; Song, J. A Comprehensive Comparison of Two Variable Importance Analysis Techniques in High Dimensions: Application to an Environmental Multi-Indicators System. *Environ. Model. Softw.* 2015, 70, 178-190.

- [66] Altmann, A.; Toloşi, L.; Sander, O.; Lengauer, T. Permutation Importance: A Corrected Feature Importance Measure. *Bioinformatics* 2010, 26 (10), 1340-1347.
- [67] Antoniadis, A.; Lambert-Lacroix, S.; Poggi, J.-M. Random Forests for Global Sensitivity Analysis: A Selective Review. *Reliab. Eng. Syst. Saf.* 2021, 206, 107312.
- [68] Mohammadifar, A.; Gholami, H.; Comino, J. R.; Collins, A. L. Assessment of the Interpretability of Data Mining for the Spatial Modelling of Water Erosion Using Game Theory. *Catena* 2021, 200, 105178.
- [69] Hosseinzadeh, A.; Zhou, J. L.; Altaee, A.; Baziar, M.; Li, X. Modeling Water Flux in Osmotic Membrane Bioreactor by Adaptive Network-Based Fuzzy Inference System and Artificial Neural Network. *Bioresour. Technol.* 2020, 310, 123391.
- [70] Ahmed, S. N.; Haider, W. Heterogeneous Photocatalysis and Its Potential Applications in Water and Wastewater Treatment: A Review. *Nanotechnology* 2018, 29 (34), 342001.
- [71] Tan, X.; Chen, G.; Xing, D.; Ding, W.; Liu, H.; Li, T.; Huang, Y. Indium-Modified Ga₂O₃ Hierarchical Nanosheets as Efficient Photocatalysts for the Degradation of Perfluorooctanoic Acid. *Environ. Sci. Nano* 2020, 7 (8), 2229-2239.
- [72] Jiang, Z.; Hu, J.; Tong, M.; Samia, A. C.; Zhang, H.; Yu, X. A Novel Machine Learning Model to Predict the Photo-Degradation Performance of Different Photocatalysts on a Variety of Water Contaminants. *Catalysts* 2021, 11 (9), 1107.