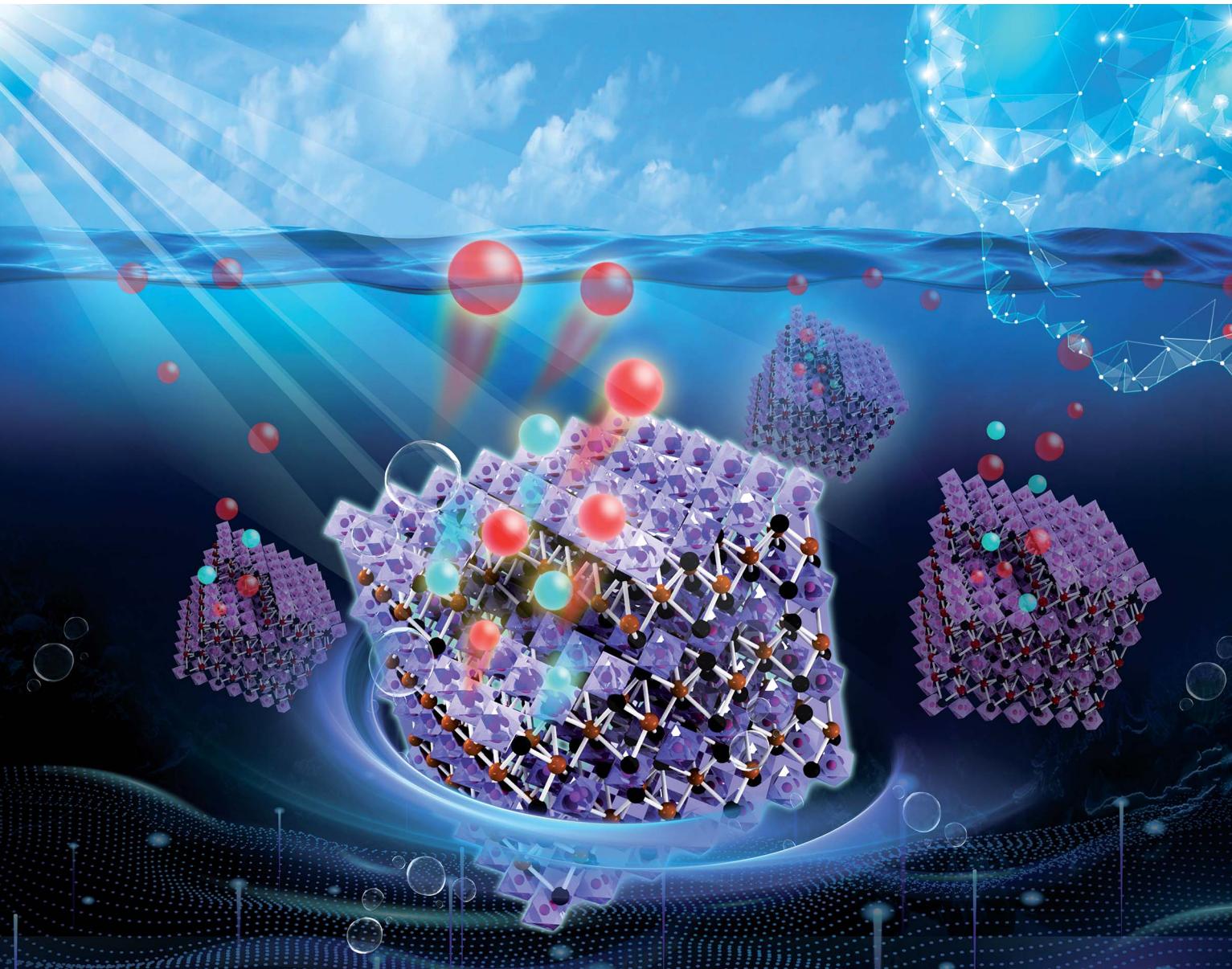


# Nanoscale Advances

[rsc.li/nanoscale-advances](http://rsc.li/nanoscale-advances)



ISSN 2516-0230

**PAPER**

Xiuyun Zhai and Mingtong Chen

A machine learning-based nano-photocatalyst module  
for accelerating the design of  $\text{Bi}_2\text{WO}_6/\text{MIL-53(Al)}$   
nanocomposites with enhanced photocatalytic activity



Cite this: *Nanoscale Adv.*, 2023, 5, 4065

## A machine learning-based nano-photocatalyst module for accelerating the design of $\text{Bi}_2\text{WO}_6/\text{MIL-53}(\text{Al})$ nanocomposites with enhanced photocatalytic activity†

Xiuyun Zhai \*<sup>a</sup> and Mingtong Chen<sup>b</sup>

It is a great challenge to acquire novel  $\text{Bi}_2\text{WO}_6/\text{MIL-53}(\text{Al})$  (BWO/MIL) nanocomposites with excellent catalytic activity by the trial-and-error method in the vast untapped synthetic space. The degradation rate of Rhodamine B dye ( $\text{DR}_{\text{RhB}}$ ) can be used as the main parameter to evaluate the catalytic activity of BWO/MIL nanocomposites. In this work, a machine learning-based nano-photocatalyst module was developed to speed up the design of BWO/MIL with desirable performance. Firstly, the  $\text{DR}_{\text{RhB}}$  dataset was constructed, and four key features related to the synthetic conditions of BWO/MIL were filtered by the forward feature selection method based on support vector regression (SVR). Secondly, the SVR model with radical basis function for predicting the  $\text{DR}_{\text{RhB}}$  of BWO/MIL was established with the key features and optimal hyperparameters. The correlation coefficients ( $R$ ) between predicted and experimental  $\text{DR}_{\text{RhB}}$  were 0.823 and 0.884 for leave-one-out cross-validation (LOOCV) and the external test, respectively. Thirdly, potential BWO/MIL nanocomposites with higher  $\text{DR}_{\text{RhB}}$  were discovered by inverse projection, the prediction model, and virtual screening from the synthesis space. Meanwhile, an online web service ([http://1.14.49.110/online\\_predict/BWO2](http://1.14.49.110/online_predict/BWO2)) was built to share the model for predicting the  $\text{DR}_{\text{RhB}}$  of BWO/MIL. Moreover, sensitivity analysis was brought into boosting the model's explainability and illustrating how the  $\text{DR}_{\text{RhB}}$  of BWO/MIL changes over the four key features, respectively. The method mentioned here can provide valuable clues to develop new nanocomposites with the desired properties and accelerate the design of nano-photocatalysts.

Received 24th February 2023  
Accepted 20th May 2023

DOI: 10.1039/d3na00122a  
[rsc.li/nanoscale-advances](http://rsc.li/nanoscale-advances)

## 1 Introduction

Against the background of the energy crisis and environmental pollution worldwide, photocatalysts have been widely paid attention and developed<sup>1,2</sup> due to their distinct advantages of energy conservation, high efficiency, green environmental protection, etc.<sup>3–5</sup> They can capture solar energy to completely decompose organic pollutants into  $\text{CO}_2$  and  $\text{H}_2\text{O}$ , decompose  $\text{H}_2\text{O}$  into  $\text{O}_2$  and  $\text{H}_2$ , etc. The  $\text{Bi}_2\text{WO}_6/\text{MIL-53}(\text{Al})$  (BWO/MIL) composites that we studied are a type of nanomaterial<sup>6–10</sup> as well as a novel kind of visible-light photocatalyst whose morphology and size affect the specific surface area and the number of surface active sites, subsequently impacting the visible-light catalytic activity. Therefore, the key to the preparation of efficient visible-light photocatalysts is to adjust their

morphologies and sizes by selecting different preparation methods and suitable surfactants.

The photocatalytic principle is shown in Fig. 1. The photocatalyst is excited to produce photogenerated electron–hole pairs under irradiation of light ( $h\nu$ ). The electrons jump from the top of the valence band (VB) to the bottom of the conduction band (CB). During migration to the surface under electric field or through diffusion, some photogenerated carriers recombine and then release absorbed energy in the forms of heat or light.

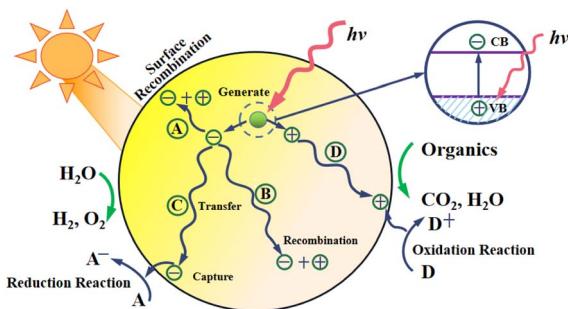


Fig. 1 Diagram of the photocatalytic principle.

<sup>a</sup>College of Intelligent Manufacturing, Hunan University of Science and Engineering, Yongzhou, 425100, Hunan, China. E-mail: [cqfbb2008@shu.edu.cn](mailto:cqfbb2008@shu.edu.cn)

<sup>b</sup>Public Experimental Teaching Center, Panzhihua University, Panzhihua, 617000, Sichuan, China

† Electronic supplementary information (ESI) available. See DOI: <https://doi.org/10.1039/d3na00122a>



The electrons and holes that migrate to the surface of the catalyst can react with the adsorbed substances on the surface by reduction and oxidation, and finally, further thermal and catalytic reactions are carried out.

Photocatalysis technology has been applied in many frontier fields, such as environmental purification,<sup>11,12</sup> self-cleaning materials,<sup>13</sup> advanced new energy,<sup>14,15</sup> cancer treatment,<sup>16</sup> high-efficiency antibacterial agents,<sup>17</sup> etc. Especially, the visible-light photocatalysts have become part of mainstream research in the environmental protection field<sup>18–20</sup> because they can utilize approximately 50% solar energy, well above the nano-semiconductor materials.<sup>21</sup> Among them,  $\text{Bi}_2\text{WO}_6$  is gaining more and more attention for its many important assets, such as low cost, non-toxicity, narrow band gap and high physico-chemical stability.<sup>22–25</sup> In spite of the many advantages, it has intrinsic limitations and disadvantages that affect its photocatalytic activity, such as high combination rate of photo-generated electrons and holes, along with a narrow photoabsorption region.<sup>26,27</sup> Constructing heterojunctions is an effective way to resolve the above problems.<sup>28–30</sup> MIL-53(Al), an Al-based metal-organic framework (MOF), is an excellent catalyst support due to its large specific surface area and high thermal stability.<sup>31,32</sup> Hu *et al.* reported that the formation of the BWO/MIL heterojunction could enhance the degradation of Rhodamine B (RhB) dye in water under visible-light irradiation.<sup>22</sup> This is mainly because the formation of heterojunctions can improve the separation efficiency of photogenerated carriers. In the process of combining MOFs with  $\text{Bi}_2\text{WO}_6$ , MOFs serve as both semiconductors with different CB or VB structures and the carriers for the  $\text{Bi}_2\text{WO}_6$  nanomaterials, thus forming effective composite photocatalysts. However, it is difficult to achieve the controllable synthesis of BWO/MIL because there are many factors impacting the photocatalytic activity of BWO/MIL. Considering the existence of multiple experimental parameters and the wide compositions, the number of potential BWO/MIL nanocomposites is far greater than that of synthesized BWO/MILs. The traditional trial-and-error method which completely depends on experiments can't meet the urgent requirements of the novel material development with rapid growth.

The computational methods that can efficiently accelerate material design have attracted more and more attention, which mainly include first-principles calculations,<sup>33,34</sup> molecular dynamics,<sup>35</sup> finite element method<sup>36</sup> and machine learning (ML).<sup>37,38</sup> Though the first three methods can promote the development of new materials to a certain extent, it proves costly to quickly discover new photocatalysts with ideal performance, considering the highly time-consuming and excessive computation costs caused by too many experimental factors. In recent years, material design with the assistance of ML has become a hot research topic and an alternative approach to the trial-and-error method, and it can greatly shorten the development cycle of materials with high performance.<sup>38–40</sup> ML is an interdisciplinary subject involving probability theory, statistics, convex analysis, algorithm complexity theory, and so on. It specializes in studying how computers simulate or realize human learning behaviors for acquiring relative knowledge or

skills, reorganizing existing knowledge to continuously improve their own performance.<sup>41,42</sup> It is the core of artificial intelligence (AI), and its application scope is very extensive and covers nearly all sectors of AI.<sup>43–45</sup> A ML-based alloy design system to facilitate the rational design of high-entropy alloys with enhanced hardness was developed by Yang *et al.*<sup>46</sup> Lamoureux *et al.* reviewed how cutting-edge data infrastructures and ML methods were being used to address problems in computational heterogeneous catalysis.<sup>47</sup> Raccuglia *et al.* reported that the ML model they developed outperformed traditional human strategies and successfully predicted the formation conditions for new templated inorganic products.<sup>48</sup> Despite all this, most ML models lack interpretation and prevent researchers from acquiring more material design mechanisms. Furthermore, it is still a difficult problem to fulfill the inverse design of ideal materials from performance to synthetic conditions, and most experimental researchers are often not good at using the developed prediction models they are interested in because of their lack of in-depth understanding of ML methods.

In our work, a ML-based nano-photocatalyst module (MLM-BWO/MIL) for accelerating the design of BWO/MIL with enhanced photocatalytic activity was constructed in four steps, namely, constructing a dataset, establishing the model, optimization of synthesis conditions, and sensitivity analysis (SA). The BWO/MIL dataset was formed by assembling the experimental condition parameters and composition ratio from the related reference. Then, the forward feature selection method based on support vector regression (FFS-SVR)<sup>49</sup> was employed to find the key features to establish the prediction model of BWO/MIL performance. Furthermore, the support vector regression (SVR)<sup>50</sup> model with radical basis function (RBF), the inverse projection based on Fisher method, and virtual screening were employed to search for the optimal parameters of BWO/MIL synthesis. A web service to easily utilize the model is provided for readers without requiring understanding of the ML principle. Furthermore, SA was used to explain the changes of the target variable with the key features, help guide the subsequent experiments, and discover the mechanism of BWO/MIL synthesis.

The contributions and novelties of our work are as follows: (1) The BWO/MIL nanocomposites are a novel kind of visible-light nano-photocatalysts attracting the attention of researchers. For the first time, a dataset of BWO/MIL nanocomposites was established by collecting relevant literature, which can be conveniently used for subsequent extensions of BWO/MIL experiments and data mining of nano-photocatalytic materials. (2) To our knowledge, detailed research on the ML-based nano-photocatalyst module for developing BWO/MIL with enhanced photocatalytic activity is new. (3) Just four features ( $C_{\text{HN}}$ ,  $S_{\text{ur}}$ ,  $T_{\text{hyd}}$  and  $R_m$ ) were recognized as the key factors affecting photocatalytic performance and the inputs of the SVR-RBF model for predicting the  $\text{DR}_{\text{RhB}}$  of BWO/MIL, which can greatly reduce the complexity of the model and save computational time. (4) Five promising samples with likely higher performance were discovered *via* our MLM-BWO/MIL. (5) An online web service ([http://1.14.49.110/online\\_predict/BWO2](http://1.14.49.110/online_predict/BWO2)) was established to quickly and effectively predict the



$DR_{RhB}$  of BWO/MIL, which can help to save significant time and cost of synthetic experiments of BWO/MIL. Moreover, all researchers who are interested in BWO/MIL nanocomposites don't need to be familiar with ML technology to access and utilize the service freely.

## 2 Material and methods

### 2.1 The formation of MLM-BWO/MIL

A MLM-BWO/MIL for speeding up the design of BWO/MIL nanocomposites with enhanced photocatalytic activity was constructed by the following steps (shown in Fig. 2): construct dataset, establish prediction model, optimize experimental parameters and conduct SA.

**2.1.1 Construct dataset.** A series of BWO/MIL nanocomposites in the dataset were fabricated by hydrothermal method. Their photocatalytic activity was evaluated by the degradation of RhB in aqueous solution under visible-light irradiation. The typical preparation procedure of BWO/MIL is described in detail as follows. The meaning of each parameter is listed in Table 1.

(a) First, 2 mmol  $Bi(NO_3)_3 \cdot 5H_2O$  in 25 mL  $HNO_3$  was dissolved at the molar concentration of  $C_{HN}$  M, and then stirred continuously at room temperature to fully dissolve  $Bi(NO_3)_3 \cdot 5H_2O$ , yielding the  $Bi(NO_3)_3$  solution.

(b) 1 mmol  $Na_2WO_4 \cdot 2H_2O$  and  $D_{sur}$  g  $S_{ur}$  were dissolved in 20 mL  $S_{ol}$  solution to obtain  $Na_2WO_4$  solution.

(c)  $1/R_m$  mmol of the MIL-53(Al) was uniformly mixed with  $Na_2WO_4$  solution, then ultrasonically dispersed for 30 min, and added with  $Bi(NO_3)_3$  solution drop by drop under stirring.

(d) The pH of the mixed solution was adjusted to  $V_{pH}$  with  $R_{pH}$  solution, then magnetically stirred for  $T_{stir}$  hours.

(e) The obtained mixture was transferred into a 100 mL autoclave with PTFE lining, placed into an oven at  $T_{hyd}$  °C, and reacted for  $T_{hy}$  hours.

(f) After being cooled to room temperature by  $C_M$ , the reaction product was centrifuged, washed twice with secondary distilled water, washed once with absolute ethanol, placed in an oven at  $T_{dry}$  °C and dried for 10 h by  $D_M$ .

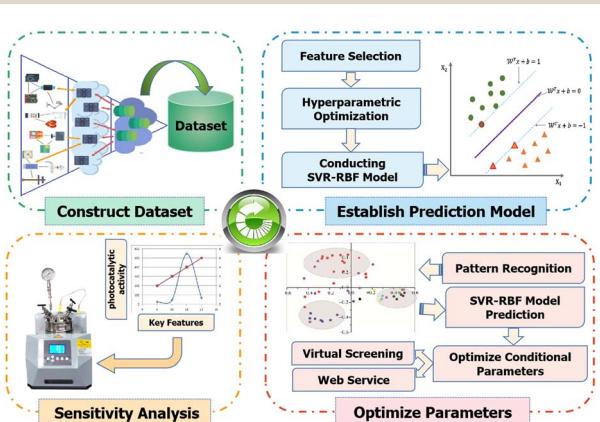


Fig. 2 Diagram of MLM-BWO/MIL constructed by four steps.

Table 1 The meaning of the thirteen features in the dataset

No.	Meanings	Features
1	Reagent for pH adjustment	$R_{pH}$
2	Cooling method	$C_M$
3	Drying method	$D_M$
4	Concentration of $HNO_3$ (mol L <sup>-1</sup> )	$C_{HN}$
5	Surfactant	$S_{ur}$
6	Surfactant dosage (g)	$D_{sur}$
7	pH	$V_{pH}$
8	Stirring time (h)	$T_{stir}$
9	Hydrothermal temperature (°C)	$T_{hyd}$
10	Hydrothermal time (h)	$T_{hy}$
11	Solvent of $Na_2WO_4 \cdot 2H_2O$	$S_{ol}$
12	Drying temperature (°C)	$T_{dry}$
13	Mole ratio of $Bi_2WO_6$ to MIL-53(Al)	$R_m$

The growth of BWO/MIL is a typical Ostwald aging<sup>51</sup> and self-assembly process. Based on the analysis of the preparation process and characterization results of BWO/MIL, its phase formation mechanism (shown in Fig. 3) is described as follows. Two kinds of tiny crystal nuclei ( $WO_6$  and  $Bi_2O_2^{2+}$ ) are formed in the precursor fluid, grown in parallel on the surface of MIL-53(Al), and gradually matured to form two-dimensional  $Bi_2WO_6$  nanosheets with hydrothermal temperature and pressure increasing and with the assistance of the surfactants. The heterojunction generated at the contact interface between  $Bi_2WO_6$  and MIL-53(Al) can effectively inhibit the recombination of photogenerated electron–hole pairs of the BWO/MIL photocatalyst, and then improve the catalytic activity.

The dataset was established by collecting the experimental data of BWO/MIL synthesis from ref. 52, which is a graduate student's dissertation. The mechanism of photocatalytic degradation of BWO/MIL is also provided in ref. 52. The degradation rate of RhB dye ( $DR_{RhB}$ ) (listed in Table S1 of ESI<sup>†</sup>) is as the target variable of the dataset. Thirteen features (shown in Table 1), including the twelve experimental parameters and one composition ratio of BWO/MIL ( $R_m$ ), served as the candidate inputs of the prediction model. The meanings of values for the five features ( $R_{pH}$ ,  $C_M$ ,  $D_M$ ,  $S_{ur}$  and  $S_{ol}$ ) are explained in Table S2.<sup>†</sup> Before modeling, two samples (No. 5 and 27) were basically without photocatalytic activity due to their very low  $DR_{RhB}$  and were deleted from the dataset. Then, the dataset with 53 samples was divided into two parts: the training set, which included 45 samples for modeling, and the testing set,

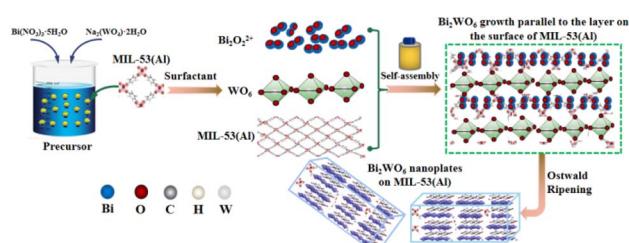


Fig. 3 The formation mechanism of BWO/MIL.



comprised by the remaining samples for validation and marked by asterisks in Table S1.†

**2.1.2 Establish prediction model.** Removing the redundant features from the candidate feature pool is a critical step to constructing a successful model. The optimal feature subset obtained by feature selection involves essential and sufficient information with very little redundancy. Apparently, feature selection can reduce the training time and the risk of overfitting, and further improve the prediction ability and generalization performance of the model. In this work, FFS-SVR was introduced to find the key features related to the target variable.

The RBF parameters of SVR were optimized by cross-validation and grid research techniques before building the model. Cross-validation was conducted to avoid the overfitting problem. The SVR-RBF model<sup>53</sup> was established for predicting the  $DR_{RhB}$  of BWO/MILs with the optimal feature subset. The independent testing set was used to validate the generality of the SVR-RBF model.

**2.1.3 Optimize experimental parameters.** Inverse projection of pattern recognition was exploited to search some BWO/MILs with possibly higher  $DR_{RhB}$ . The SVR-RBF model established was utilized to predict the  $DR_{RhB}$  of the BWO/MILs. Virtual screening and web service based on inverse projection and the SVR model were used to find the synthesis parameters of the BWO/MILs with a higher  $DR_{RhB}$  than the highest value in the existing dataset.

**2.1.4 Sensitivity analysis.** To further explain the prediction model and guide the later synthesis experiment, SA based on the model was conducted. SA refers to an uncertainty analysis technique to study the influence of certain changes of key features on one or a group of indicators from the perspective of quantitative analysis. Its essence is to interpret the rule that the  $DR_{RhB}$  of BWO/MIL is affected by changing the value of one key feature.

## 2.2 Inverse projection of pattern recognition

Inverse projection (IP) of pattern recognition refers to mapping of the designed sample points from the two-dimensional space back to the multi-dimensional space, where the experimental condition for material synthesis can be found. A discriminant function related to the known samples and their categories needs to be set up to complete IP. Fisher discriminant analysis<sup>54,55</sup> was used to determine the weight vector and threshold value in the projection direction by using the given training data, before constructing the discriminant function. Fisher method is a popular method to measure the separation degree between two categories.<sup>56</sup> Although it was proposed earlier, it is still being used widely and plays an important part in pattern recognition.<sup>57,58</sup> Certain constraints must be introduced to make the result of IP unique. The constraint for linear IP is to take the fixed values (such as mean value or optimal value) as the coordinate values of the design points on other projection vectors, while the constraint for nonlinear IP is found by minimizing the error function of the inverse projection.

## 2.3 Software availability

The ML calculations of our work were performed using the HyperMiner software package<sup>59,60</sup> and the Online Computational Platform of Material Data Mining (OCPMDM),<sup>61,62</sup> which we developed. HyperMiner can be freely downloaded on the website: <http://materials-data-mining.com/home>. OCPMDM can be freely used on the website of the Laboratory of Materials Data Mining at Shanghai University: <http://materials-data-mining.com/ocpmdm/>.

## 3 Results and discussion

In order to seek out the most suitable algorithm for modeling, the results of 10-fold cross-validation of several algorithms, including SVR-RBF, SVR based on linear kernel function (SVR-LKF), SVR based on polynomial kernel function (SVR-PKF), and multiple linear regression (MLR), were compared. The results reflected that SVR-RBF algorithm was more suitable for modeling than the others. Therefore, SVR-RBF was identified as the method to establish the model for predicting  $DR_{RhB}$  of BWO/MIL nanocomposites.

### 3.1 Feature selection

Feature selection is the process of selecting an optimal subset from the original features, which falls into three categories according to formation mode: exhaustive method, heuristic method and random method. The forward feature selection method (FFS)<sup>63,64</sup> used in this work is a heuristic method which starts from an empty set and incrementally adds a feature to the target feature subset from the candidate features. The search process of FFS doesn't end until a feature subset close to the optimal solution is obtained. A SVR model is constructed with the features on each pass of the FFS search. Root mean square error (RMSE) of the 10-fold cross-validation is used as the evaluation indicator of the SVR model. The above feature selection approach is termed the FFS-SVR method. Fig. 4 reveals how FFS-SVR and cross-validation were used to search for the optimal feature set. The result illustrated that RMSE showed concave parabola variation with the increase of feature number, and the minimal and maximal RMSEs occur when the feature numbers are 4 and 12, respectively. It was obvious that the

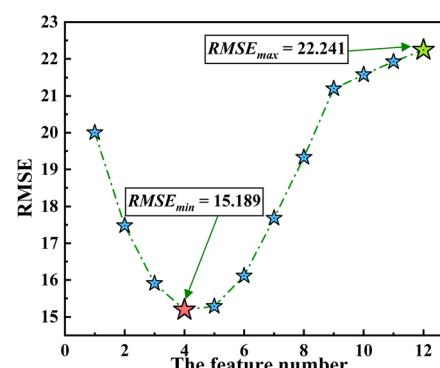


Fig. 4 The process of feature selection in FFS-SVR.



feature set including  $C_{\text{HN}}$ ,  $S_{\text{ur}}$ ,  $T_{\text{hyd}}$  and  $R_m$  (highlighted in bold in Table S1†) is best when the RMSE of the model is minimal.

### 3.2 Model establishment

**3.2.1 Optimizing hyperparameters.** Moderate parameter adjustment for the great majority of modeling techniques is required to achieve optimal performance. Three hyperparameters of SVR-RBF are the regularization parameter  $C$ , insensitive parameter  $\varepsilon$  and gamma value  $\gamma$ . Their optimization ranges are [1, 100], [0.01, 0.09] and [0.5, 1.5], and the step sizes are 2, 0.02 and 0.1, respectively. Parameter  $C$  determines the tradeoff between the complexity and precision of the model. The larger  $C$  is, the worse the generalization ability of the model is. The smaller  $C$  is, the more insensitive  $\varepsilon$  is, and the larger the training error. Parameter  $\varepsilon$  controls the sparsity of support vector in SVR and affects the precision of the regression model. Gamma controls the influence distance of a single training point. A model with very large gamma value can easily cause overfitting.

Grid research method and 10-fold cross-validation were employed to optimize the three hyperparameters, and RMSE was used as the evaluation index of the SVR model. The optimization process is shown in Fig. 5, from which the minimum RMSE is 13.724 when  $C$ ,  $\varepsilon$  and  $\gamma$  are 29, 0.05 and 1.3, respectively.

**3.2.2 SVR-RBF model.** The SVR-RBF algorithm with the optimal hyperparameters and four key features was used to

establish the model for predicting the  $\text{DR}_{\text{RhB}}$  of BWO/MILs. The correlation coefficient ( $R$ ) and RMSE were used as evaluation indices of model performance. The SVR-RBF model is expressed as follows.

$$\text{DR}_{\text{RhB}} = \sum_{i=1}^n \beta_i \exp\left(-0.5 \times \|\mathbf{X} - \mathbf{X}_i\|^2\right) + 0.438 \quad (1)$$

where  $\mathbf{X}$  and  $\mathbf{X}_i$  are the unknown and the support vector, respectively.  $\beta_i$  and  $n$  are the Lagrange multiplier of the support vector and corresponding number, respectively.

Fig. 6(a) and (b) show the experimental *versus* the predictive  $\text{DR}_{\text{RhB}}$  of BWO/MILs for the training and testing samples, and the LOOCV and 10-fold cross-validation of the training set, respectively. The subscripts "TR", "TS", "LCV" and "CRS" in the figure denote "training", "testing", "LOOCV" and "10-fold cross-validation", respectively. It can be seen from Fig. 6(a) that the sample points, whether in the training set or in the testing set, scatter around the diagonal, illustrating that the model has preferable prediction performance and practicability. The results of the model evaluated by LOOCV and 10-fold cross-validation are shown in Fig. 6(b). The  $R_{\text{LCV}}$  and  $R_{\text{CRS}}$  of the training set were 0.823 and 0.826, respectively. There are some deviations between the predictive and the experimental values of some samples. That may be due to some errors in experiments and requiring more data for building a predictive model in the vast synthesis space of BWO/MILs. If more experimental data can be gathered in future work, the deviations will be reduced.

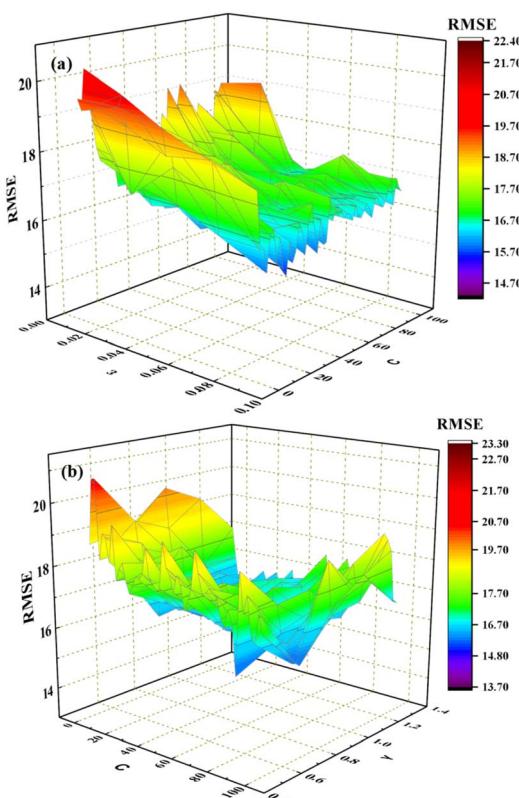


Fig. 5 The optimization process of three hyperparameters of SVR-RBF: (a) RMSE versus  $\varepsilon$  and  $C$ ; (b) RMSE versus  $C$  and  $\gamma$ .

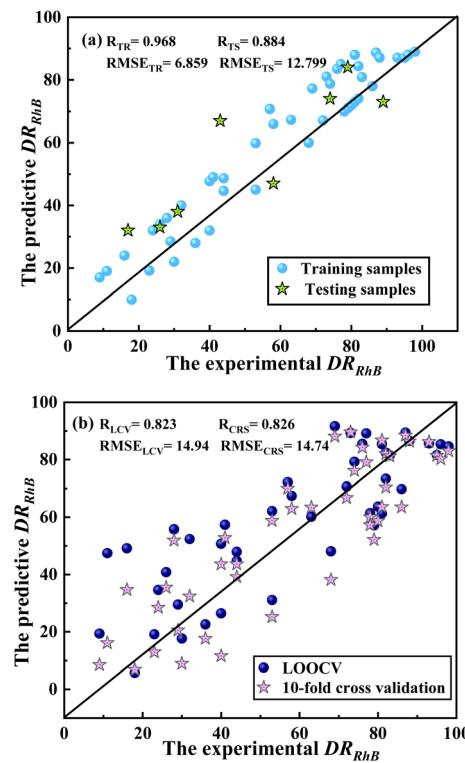


Fig. 6 Experimental *versus* predictive  $\text{DR}_{\text{RhB}}$  of BWO/MILs for (a) the training and testing samples. (b) LOOCV and 10-fold cross-validation of the training set.



**3.2.3 Optimization of experimental parameters.** The basic principle of pattern recognition is that similar samples approach each other in the pattern space and form a “group”, that is, “birds of a feather flock together”. A given pattern is classified into C classes according to the eigenvector measured by the pattern, and then the classification is discriminated according to the distance between the patterns. The goal of IP based on pattern recognition is to find the synthetic conditions of superior samples in the original multidimensional space. In this work, samples are divided into two categories: superior class with better performance and inferior class with worse performance. The distribution region of superior samples is sought for and built in the pattern recognition diagram. Then, unsynthesized sample points with likely higher performance are designed in the optimal area, whose features in the original space can be derived by IP. Finally, the parameters of the samples with high performance are recognized as the inputs of the SVR-RBF model. The points whose prediction values of the model meet the requirements are the samples we are looking for.

The values of  $DR_{RhB}$  in the dataset range from 9 to 98. The maximum of  $DR_{RhB}$  can reach 100. The value (55) is close to the median between 9 and 100. Considering unbiased classification, the  $DR_{RhB}$  value dividing superior and inferior samples is 55. Samples with greater  $DR_{RhB}$  than this value are superior; otherwise, they are inferior. Fig. 7 plots the pattern recognition projection using Fisher method according to the distribution of superior and inferior samples in the training set and the testing set. It can be seen from Fig. 7 that the two classes of samples are clearly distinguished by a purple dotted line apart from four misclassified points. The line is expressed as a linear combination of the key features, which is shown as follows:

$$\text{Fisher}(1) = 0.295C_{HN} - 0.8559S_{ur} - 0.02086T_{hyd} + 0.3427R_m + 2.704 = -0.037 \quad (2)$$

The superior samples are gathered on the right side of the dotted line in Fig. 7, forming the optimal region in which the newly designed samples with high  $DR_{RhB}$  should be distributed. It can be seen from Fig. 7 and eqn (2) that the restraint condition of the discriminant function of new promising samples is  $\text{Fisher}(1) \geq -0.037$ . The features and  $DR_{RhB}$  of the designed

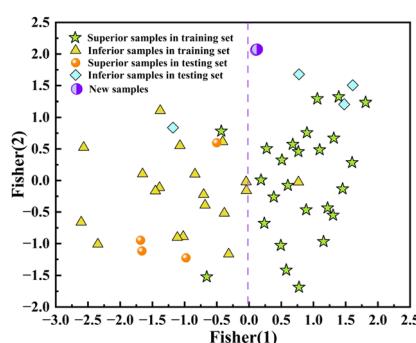


Fig. 7 Pattern recognition projection of Fisher method.

BWO/MILs can be derived by IP and predicted by the constructed SVR-RBF model. Obviously, IP can not only supply a suitable projection for classification to prevent the calculation and experiment with inferior samples, but also help us quickly seek the candidates with high  $DR_{RhB}$  in the optimal region.

Moreover, massive amounts of virtual samples were generated to improve the probability of BWO/MIL discovery and the quality of new BWO/MILs. Based on the bound of each feature in the existing dataset, the following restrictions were obeyed when generating the new samples.

① Concentration of  $HNO_3$  ( $C_{HN}$ ) changes from  $0.5 \text{ mol L}^{-1}$  to  $4.0 \text{ mol L}^{-1}$  in steps of  $0.1 \text{ mol L}^{-1}$ .

② The value of surfactant ( $S_{ur}$ ) is 0, 1 or 2.

③ Hydrothermal temperature ( $T_{hyd}$ ) ranges from  $120 \text{ }^\circ\text{C}$  to  $200 \text{ }^\circ\text{C}$  with steps of  $5 \text{ }^\circ\text{C}$ .

④ Mole ratio of  $Bi_2WO_6$  to MIL-53(Al) ( $R_m$ ) ranges from 0.5 to 2 with steps of 0.01.

The IP of pattern recognition and the SVR-RBF model were employed to predict the class and  $DR_{RhB}$  of each virtual sample by using the four key features. According to the results of virtual screening, the key features and the Fisher (1) values of the five samples with the top performance are shown in Table S3.† The experimental  $DR_{RhB}$  values could likely reach 100 (the maximum value of  $DR_{RhB}$ ) because the predictive  $DR_{RhB}$  values of the samples were already above the highest values (98) in the existing dataset. The two samples with the top  $DR_{RhB}$  values among them are shown in Fig. 7 and marked by a purple half-empty circle. Therefore, the above is helpful for experimental researchers to explore the new BWO/MILs with enhanced photocatalytic activity.

**3.2.4 Web service.** Web services make it very easy to implement the prediction tasks of the clients. A web service based on the SVR-RBF model we constructed was used to predict the  $DR_{RhB}$  of BWO/MILs, and a screenshot of the web page is shown in Fig. 8. The researchers who are interested in BWO/MIL synthesis don't need to master the principle of the ML model and can use this service on the clients to know whether the new sample of BWO/MIL has high  $DR_{RhB}$ . When using the service, the four key features ( $C_{HN}$ ,  $S_{ur}$ ,  $T_{hyd}$  and  $R_m$ ) need to be input to the corresponding boxes, and clicking the “Predict” button generates the predictive value of  $DR_{RhB}$ . This tool can help researchers accelerate the design of BWO/MILs

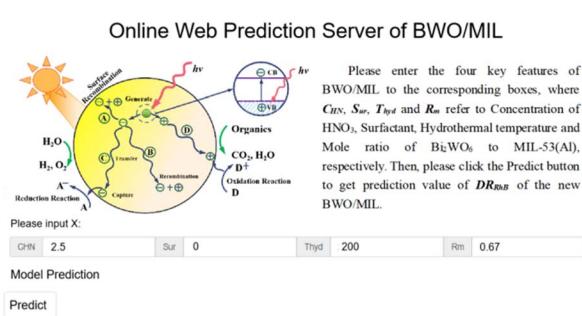


Fig. 8 The web service based on the SVR-RBF model for predicting  $DR_{RhB}$  of BWO/MIL.



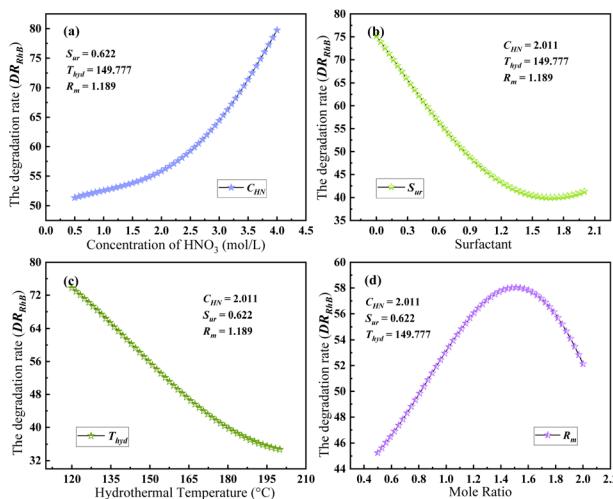


Fig. 9 SA of the four features based on the SVR-RBF model (a)  $C_{HN}$ , (b)  $S_{ur}$ , (c)  $T_{hyd}$  and (d)  $R_m$ .

with enhanced  $DR_{RhB}$ . The web service is available freely on the website: [http://1.14.49.110/online\\_predict/BWO2](http://1.14.49.110/online_predict/BWO2).

## 4 Sensitivity analyses

The essence of SA is to explain how the key indicator is affected by changing the values of the relevant variables one by one. Generally, the main parameters are selected for analyses of the sensitivity factors. In this work, the change of  $DR_{RhB}$  with one of the key features was observed when the other features were set as their mean values. Fig. 9 shows the SA for the four features ( $C_{HN}$ ,  $S_{ur}$ ,  $T_{hyd}$  and  $R_m$ ) based on the SVR-RBF model. Fig. 9(a) illustrates that  $DR_{RhB}$  increases with  $C_{HN}$  when  $S_{ur}$ ,  $T_{hyd}$  and  $R_m$  are 0.622, 149.777 and 1.189, respectively. In Fig. 9(b) and (c),  $DR_{RhB}$  decreased with  $S_{ur}$  and  $T_{hyd}$  increase, respectively. The convex parabolic relationship between  $DR_{RhB}$  and  $R_m$  is shown in Fig. 9(d). As a whole,  $S_{ur}$  and  $T_{hyd}$  are more sensitive factors than  $C_{HN}$  and  $R_m$  because a small change in the two prior parameters could lead to a large change in  $DR_{RhB}$ . From the trends of the four graphs in Fig. 9, the larger  $C_{HN}$ , the smaller  $S_{ur}$  and  $T_{hyd}$ , and the appropriate  $R_m$  could result in a larger target value ( $DR_{RhB}$ ). In other words, it is beneficial to obtain a BWO/MIL with high  $DR_{RhB}$  in the above adjustment direction of the experimental parameters. Apparently, SA can interpret the relationships between  $DR_{RhB}$  and the features, instruct the adjustment direction of experimental parameters, and assist in revealing the mechanism of BWO/MIL synthesis.

## 5 Conclusions

A ML-based photocatalyst module for advancing the design of BWO/MIL nano-photocatalysts with the desired photocatalytic performance was established in four steps, including dataset establishment, constructing the prediction model, optimization of experimental parameters and sensitivity analysis. By using FFS-SVR method, four key features ( $C_{HN}$ ,  $S_{ur}$ ,  $T_{hyd}$  and  $R_m$ ) were selected as the inputs of the model for predicting the  $DR_{RhB}$  of

BWO/MIL. The SVR-RBF model constructed with the key features and the optimal hyperparameters has better predictive performance and robustness, and can meet the requirements of rapid BWO/MIL design. The inverse projection based on Fisher method, SVR-RBF model and virtual screening were used to optimize the synthesis parameters of BWO/MIL and find BWO/MILs with higher  $DR_{RhB}$  than the highest value in the existing dataset. A web service ([http://1.14.49.110/online\\_predict/BWO2](http://1.14.49.110/online_predict/BWO2)) based on the SVR-RBF model was constructed to predict the  $DR_{RhB}$  of BWO/MILs, and it can be freely shared by all researchers interested in BWO/MILs. Furthermore, sensitivity analysis was introduced to analyze the relationship between  $DR_{RhB}$  and the four key features. The MLM-BWO/MIL we constructed could be employed to accelerate the design of BWO/MILs with ideal performance and further promote research on ML-assisted material design. This research paradigm is suitable for accelerating the synthesis and development of new materials, including but not limited to photocatalysts. An online prediction service like those mentioned in the paper can be established when enough historical data for the materials are collected.

## Author contributions

X. Z. contributed to the writing of the original draft, review and editing, and assisted with the methodology. M. C. contributed to the writing of the original draft and assisted with the formal analysis, visualization, project administration, and investigation.

## Conflicts of interest

The authors state that there are no conflicts to declare.

## Acknowledgements

The authors acknowledge the financial supports from Sichuan Science and Technology Program of China (No. 2022YFG0318), Panzhihua Instructional Science and Technology Program of China (No. 2020ZD-G-11) and Panzhihua University Science and Technology Program of China (No. 2021PY009).

## References

- Y. Yang, H. Y. Tan, B. Cheng, *et al.*, Near-Infrared-Responsive Photocatalysts, *Small Methods*, 2021, **5**(4), 2001042.
- L. Q. Yang, D. Q. Fan, Z. L. Li, *et al.*, A Review on the Bioinspired Photocatalysts and Photocatalytic Systems, *Adv. Sustainable Syst.*, 2022, **6**(5), 2100477.
- S. Ullah, E. P. Ferreira-Neto, A. A. Khan, *et al.*, Supported nanostructured photocatalysts: the role of support-photocatalyst interactions, *Photochem. Photobiol. Sci.*, 2023, **22**(1), 219–240.
- G. Lofrano, F. Ubaldi, L. Albarano, *et al.*, Antimicrobial Effectiveness of Innovative Photocatalysts: A Review, *Nanomaterials*, 2022, **12**(16), 2831.

5 I. Barba-Nieto, U. Caudillo-Flores, M. Fernandez-Garcia, *et al.*, Sunlight-Operated  $TiO_2$ -Based Photocatalysts, *Molecules*, 2020, **25**(17), 4008.

6 A. Almasian, N. M. Mahmoodi and M. E. Olya, Tectomer grafted nanofiber: Synthesis, characterization and dye removal ability from multicomponent system, *J. Ind. Eng. Chem.*, 2015, **32**, 85–98.

7 N. M. Mahmoodi, M. Ghezelbash, M. Shabanian, *et al.*, Efficient removal of cationic dyes from colored wastewaters by dithiocarbamate-functionalized graphene oxide nanosheets: From synthesis to detailed kinetics studies, *J. Taiwan Inst. Chem. Eng.*, 2017, **81**, 239–246.

8 F. Hosseini, S. Sadighian, H. Hosseini-Monfared, *et al.*, Dye removal and kinetics of adsorption by magnetic chitosan nanoparticles: Desalination and Water Treatment, *Desalin. Water Treat.*, 2016, **57**(51), 24378–24386.

9 S. A. Hosseini, M. Vossoughi, N. M. Mahmoodi, *et al.*, Clay-based electrospun nanofibrous membranes for colored wastewater treatment, *Appl. Clay Sci.*, 2019, **168**, 77–86.

10 S. R. Mousavi, M. Asghari and N. M. Mahmoodi, Chitosan-wrapped multiwalled carbon nanotube as filler within PEBA thin film nanocomposite (TFN) membrane to improve dye removal, *Carbohydr. Polym.*, 2020, **237**, 116128.

11 C. L. Yu, W. Q. Zhou, H. Liu, *et al.*, Design and fabrication of microsphere photocatalysts for environmental purification and energy conversion, *Chem. Eng. J.*, 2016, **287**, 117–129.

12 S. Yanagida, Nano/microsized  $TiO_2$  composite photocatalysts for environmental purification, *J. Ceram. Soc. Jpn.*, 2018, **126**(8), 625–631.

13 E. Luevano-Hipolito, L. M. Torres-Martinez and L. V. F. Cantu-Castro, Self-cleaning coatings based on fly ash and bismuth-photocatalysts:  $Bi_2O_3$ ,  $Bi_2O_2CO_3$ ,  $BiOI$ ,  $BiVO_4$ ,  $BiPO_4$ , *Constr. Build. Mater.*, 2019, **220**, 206–213.

14 Y. Horiuchi, T. Toyao, M. Takeuchi, *et al.*, Recent advances in visible-light-responsive photocatalysts for hydrogen production and solar energy conversion - from semiconducting  $TiO_2$  to MOF/PCP photocatalysts, *Phys. Chem. Chem. Phys.*, 2013, **15**(32), 13243–13253.

15 T. Takata and K. Domen, Particulate Photocatalysts for Water Splitting: Recent Advances and Future Prospects, *ACS Energy Lett.*, 2019, **4**(2), 542–549.

16 M. Padervand, F. Nasiri, S. Hajiahmadi, *et al.*,  $Ag@Ag_2MoO_4$  decorated polyoxomolybdate/ $C_3N_4$  nanostructures as highly efficient photocatalysts for the wastewater treatment and cancer cells killing under visible light, *Inorg. Chem. Commun.*, 2022, **141**, 109500.

17 B. Leng, X. L. Zhang, S. S. Chen, *et al.*, Highly efficient visible-light photocatalytic degradation and antibacterial activity by  $GaN:ZnO$  solid solution nanoparticles, *J. Mater. Sci. Technol.*, 2021, **94**, 67–76.

18 E. J. Li, L. Chen, Q. A. Zhang, *et al.*, Bismuth-Containing Semiconductor Photocatalysts, *Prog. Chem.*, 2010, **22**(12), 2282–2289.

19 M. Kitano, K. Tsujimaru and M. Anpo, Hydrogen production using highly active titanium oxide-based photocatalysts, *Top. Catal.*, 2008, **49**(1–2), 4–17.

20 X. Zhang and J. Fan, Carbon Materials Modified Bismuth Based Photocatalysts and Their Applications, *Prog. Chem.*, 2016, **28**(4), 438–449.

21 J. D. Li, W. Fang, C. L. Yu, *et al.*, Ag-based semiconductor photocatalysts in environmental purification, *Appl. Surf. Sci.*, 2015, **358**, 46–56.

22 L. Hu, Y. Zhang, W. Lu, *et al.*, Easily recyclable photocatalyst  $Bi_2WO_6$ /MOF/PVDF composite film for efficient degradation of aqueous refractory organic pollutants under visible-light irradiation, *J. Mater. Sci.*, 2019, **54**(8), 6238–6257.

23 D. Q. He, H. C. Liu, Q. Wang, *et al.*, Photo-Induced Charge Transfer on  $Pt/Bi_2WO_6$  Composite Photocatalysts, *J. Nanosci. Nanotechnol.*, 2020, **20**(3), 1838–1844.

24 M. S. Gui, P. F. Wang, D. Yuan, *et al.*, Synthesis and Visible-Light Photocatalytic Activity of  $Bi_2WO_6/g-C_3N_4$  Composite Photocatalysts, *Chin. J. Inorg. Chem.*, 2013, **29**(10), 2057–2064.

25 T. Chen, L. Z. Liu, C. Hu, *et al.*, Recent advances on  $Bi_2WO_6$ -based photocatalysts for environmental and energy applications, *Chin. J. Catal.*, 2021, **42**(9), 1413–1438.

26 Y. K. Li, L. Chen, Y. Wang, *et al.*, Advanced nanostructured photocatalysts based on reduced graphene oxide-flower-like  $Bi_2WO_6$  composites for an augmented simulated solar photoactivity activity, *Mater. Sci. Eng., B*, 2016, **210**, 29–36.

27 M. S. Gui, W. D. Zhang, Q. X. Su, *et al.*, Preparation and visible light photocatalytic activity of  $Bi_2O_3/Bi_2WO_6$  heterojunction photocatalysts, *J. Solid State Chem.*, 2011, **184**(8), 1977–1982.

28 J. X. Low, J. G. Yu, M. Jaroniec, *et al.*, Heterojunction Photocatalysts, *Adv. Mater.*, 2017, **29**(20), 1601694.

29 X. Li, R. C. Shen, S. Ma, *et al.*, Graphene-based heterojunction photocatalysts, *Appl. Surf. Sci.*, 2018, **430**, 53–107.

30 J. W. Fu, J. G. Yu, C. J. Jiang, *et al.*,  $g-C_3N_4$ -Based Heterostructured Photocatalysts, *Adv. Energy Mater.*, 2018, **8**(3), 1701503.

31 A. Karami, R. Shomal, R. Sabouni, *et al.*, Photocatalytic degradation of diclofenac using hybrid  $MIL-53(Al)@TiO_2$  and  $MIL-53(Al)@ZnO$  catalysts, *Can. J. Chem. Eng.*, 2023, **101**(5), 2660–2676.

32 F. Martinez, G. Orcajo, D. Briones, *et al.*, Catalytic advantages of  $NH_2$ -modified  $MIL-53(Al)$  materials for Knoevenagel condensation reaction, *Microporous Mesoporous Mater.*, 2017, **246**, 43–50.

33 B. L. Zhang, Design novel hard materials  $B_3N_4$  via first-principles calculation, *J. Alloys Compd.*, 2016, **663**, 862–866.

34 Z. S. Lin, X. X. Jiang, L. Kang, *et al.*, First-principles materials applications and design of nonlinear optical crystals, *J. Phys. D: Appl. Phys.*, 2014, **47**(25), 253001.

35 F. P. Wang, G. Y. Du, X. C. Liu, *et al.*, Molecular dynamics application of cocrystal energetic materials: A review, *Nanotechnol. Rev.*, 2022, **11**(1), 2141–2153.

36 D. S. Chen, D. I. W. Levin, S. Sueda, *et al.*, Data-Driven Finite Elements for Geometry and Material Design, *ACM Trans. Graph.*, 2015, **34**(4), 74.



37 C. Chen, Y. X. Zuo, W. K. Ye, *et al.*, A Critical Review of Machine Learning of Energy Materials, *Adv. Energy Mater.*, 2020, **10**(8), 1903242.

38 J. Wei, X. Chu, X. Y. Sun, *et al.*, Machine learning in materials science, *InfoMat*, 2019, **1**(3), 338–358.

39 Y. Liu, T. L. Zhao, W. W. Ju, *et al.*, Materials discovery and design using machine learning, *J. Materomics*, 2017, **3**(3), 159–177.

40 Y. Liu, B. R. Guo, X. X. Zou, *et al.*, Machine learning assisted materials design and discovery for rechargeable batteries, *Energy Stor. Mater.*, 2020, **31**, 434–450.

41 J. M. Bone, C. M. Childs, A. Menon, *et al.*, Hierarchical Machine Learning for High-Fidelity 3D Printed Biopolymers, *ACS Biomater. Sci. Eng.*, 2020, **6**(12), 7021–7031.

42 T. G. Dietterich, P. Domingos, L. Getoor, *et al.*, Structured machine learning: the next ten years, *Mach. Learn.*, 2008, **73**(1), 3–23.

43 N. Kuhl, M. Schemmer, M. Goutier, *et al.*, Artificial intelligence and machine learning, *Electron. Mark.*, 2022, **32**(4), 2235–2244.

44 P. Ongsulee, *Ieee, Artificial Intelligence, Machine Learning and Deep Learning, in 2017 15th International Conference on ICT and Knowledge Engineering, ICT&KE*, 2017, pp. 92–97.

45 A. Merkin, R. Krishnamurthi and O. N. Medvedev, Machine learning, artificial intelligence and the prediction of dementia, *Curr. Opin. Psychiatry*, 2022, **35**(2), 123–129.

46 C. Yang, C. Ren, Y. Jia, *et al.*, A machine learning-based alloy design system to facilitate the rational design of high entropy alloys with enhanced hardness, *Acta Mater.*, 2022, **222**, 117431.

47 P. S. Lamoureux, K. T. Winther, J. A. G. Torres, *et al.*, Machine Learning for Computational Heterogeneous Catalysis, *ChemCatChem*, 2019, **11**(16), 3579–3599.

48 P. Raccuglia, K. C. Elbert, P. D. F. Adler, *et al.*, Machine-learning-assisted materials discovery using failed experiments, *Nature*, 2016, **533**(7601), 73–+.

49 J. M. Valente, S. Maldonado and SVR-FFS, A novel forward feature selection approach for high-frequency time series forecasting using support vector regression, *Expert Syst. Appl.*, 2020, **160**, 113729.

50 S. Feng, H. Y. Zhou and H. B. Dong, Using deep neural network with small dataset to predict material defects, *Mater. Des.*, 2019, **162**, 300–310.

51 S. Y. Dong, X. H. Ding, T. Guo, *et al.*, Self-assembled hollow sphere shaped  $\text{Bi}_2\text{WO}_6$ /RGO composites for efficient sunlight-driven photocatalytic degradation of organic pollutants, *Chem. Eng. J.*, 2017, **316**, 778–789.

52 Y. Zhang, *Degradation of aqueous refractory organic pollutants by PMS activated with  $\text{MnFe}_2\text{O}_4/\text{MIL-53(Al)}$  or visible-light photocatalysts composite  $\text{Bi}_2\text{WO}_6/\text{MIL-53(Al)}$  and composite film  $\text{Bi}_2\text{WO}_6/\text{MIL-53(Al)}/\text{PVDF}$* , MA thesis, Shanghai University, 2018.

53 X. Zhai and M. Chen, Comparison of Data-driven Prediction Methods for Comprehensive Coke Ratio of Blast Furnace, *High Temp. Mater. Process.*, 2023, **42**(1), 20220261.

54 Y. Y. Ma, J. W. Wang, X. Y. Luo, *et al.*, Image steganalysis feature selection based on the improved Fisher criterion, *Math. Biosci. Eng.*, 2020, **17**(2), 1355–1371.

55 X. Y. Wu, X. Mao, L. J. Chen, *et al.*, Kernel optimization using nonparametric Fisher criterion in the subspace, *Pattern Recognit. Lett.*, 2015, **54**, 43–49.

56 Y. T. Wang, J. D. Wang, H. Y. Chen, *et al.*, Semi-Supervised Local Fisher Discriminant Analysis Based on Reconstruction Probability Class, *Int. J. Pattern Recognit. Artif. Intell.*, 2015, **29**(2), 1550007.

57 Q. X. Gao, J. J. Liu, H. J. Zhang, *et al.*, Enhanced fisher discriminant criterion for image recognition, *Pattern Recognit.*, 2012, **45**(10), 3717–3724.

58 Y. X. Liang, C. R. Li, W. G. Gong, *et al.*, Uncorrelated linear discriminant analysis based on weighted pairwise Fisher criterion, *Pattern Recognit.*, 2007, **40**(12), 3606–3615.

59 Q. Zhang, X. Y. Zhai, P. Xiong, *et al.*, Prediction and synthesis of novel layered double hydroxide with desired basal spacing based on relevance vector machine, *Mater. Res. Bull.*, 2017, **93**, 123–129.

60 X. Zhai, M. Chen and W. Lu, Accelerated search for perovskite materials with higher Curie temperature based on the machine learning methods, *Comput. Mater. Sci.*, 2018, **151**, 41–48.

61 Q. Tao, T. Lu, Y. Sheng, *et al.*, Machine learning aided design of perovskite oxide materials for photocatalytic water splitting, *J. Energy Chem.*, 2021, **60**, 351–359.

62 Q. Tao, D. Chang, T. Lu, *et al.*, Multiobjective Stepwise Design Strategy-Assisted Design of High-Performance Perovskite Oxide Photocatalysts, *J. Phys. Chem. C*, 2021, **125**(38), 21141–21150.

63 F. Macedo, M. R. Oliveira, A. Pacheco, *et al.*, Theoretical foundations of forward feature selection methods based on mutual information, *Neurocomputing*, 2019, **325**, 67–89.

64 L. K. Luo, L. J. Ye, M. X. Luo, *et al.*, Methods of forward feature selection based on the aggregation of classifiers generated by single attribute, *Comput. Biol. Med.*, 2011, **41**(7), 435–441.

