

## Gradient boosting decision tree algorithms for accelerating nanofiltration membrane design and discovery

Weijia Gong<sup>a,\*</sup>, Hangbin Xu<sup>b</sup>, Jinyan Lu<sup>a</sup>, Jungbin Kim<sup>c</sup>, Yan Zhao<sup>d</sup>, Ni Li<sup>e</sup>, Yixuan Zhang<sup>f</sup>, Jiaxuan Yang<sup>b</sup>, Daliang Xu<sup>b</sup>, Heng Liang<sup>b</sup>

<sup>a</sup> School of Engineering, Northeast Agricultural University, 600 Changjiang Street, Xiangfang District, Harbin 150030, China

<sup>b</sup> State Key Laboratory of Urban Water Resource and Environment, School of Environment, Harbin Institute of Technology, Harbin 150090, China

<sup>c</sup> Department of Environmental Science, College of Science, Mathematics and Technology, Wenzhou-Kean University, Wenzhou 325060, China

<sup>d</sup> Department of Chemical Engineering, KU Leuven, Celestijnlaan 200F, B-3001 Leuven, Belgium

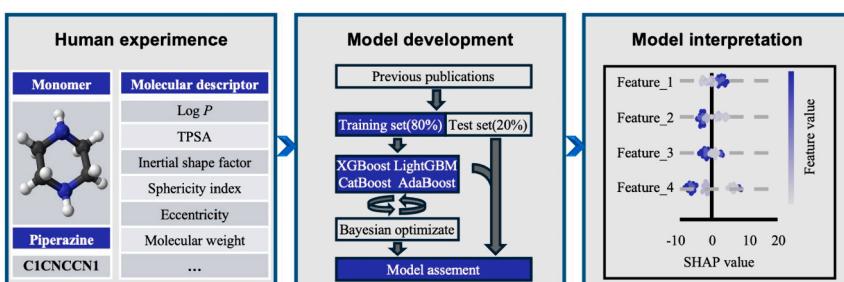
<sup>e</sup> Department of Water and Climate, Vrije Universiteit Brussel, Brussels, Belgium

<sup>f</sup> State Key Laboratory of Loess and Quaternary Geology, Institute of Earth Environment, Chinese Academy of Sciences, Xi'an 710061, China

### HIGHLIGHTS

- Gradient boosting decision tree algorithms accurately predict membrane performance.
- Molecular descriptors are used to represent the candidate monomers.
- Aqueous monomer concentration is the most important feature.
- Integration of monomers and fabrication conditions for the membrane design is highlighted.

### GRAPHICAL ABSTRACT



### ARTICLE INFO

#### Keywords:

Nanofiltration membrane  
Interfacial polymerization  
Gradient boosting decision tree algorithms  
Permeance  
Salt rejection

### ABSTRACT

Interfacial polymerization is the most widely used strategy for nanofiltration membrane fabrication. Despite extensive research on this technology, further improvement in permeance and salt rejection is still essential due to its multidimensional characteristics, including the types of membrane material and the conditions of membrane optimizing fabrication. Herein, we applied four gradient boosting decision tree algorithms to precisely identify the candidate monomers (represented by the molecular descriptors) and their fabrication conditions. The result of the model evaluation indicated the Extreme Gradient Boosting (XGBoost) algorithm had the best predictive performance in accuracy and generalization in predicting membrane permeance and salt rejection, with the corresponding determination coefficients on the test set being 0.76 and 0.88. Shapley additive explanation analysis showed that the aqueous monomer concentration was the most influential fabrication condition in membrane performance. Besides, the partition coefficient (Log P) and topological polar surface area were the most important molecular descriptors in water permeance and salt rejection, respectively. Overall, this study proposed innovative machine learning algorithms to disentangle the multidimensional interactions of various influencing factors on membrane performance, thus initiating a paradigm shift in the development of high-performance nanofiltration membranes.

\* Corresponding author.

E-mail address: [gongweijia@126.com](mailto:gongweijia@126.com) (W. Gong).

## 1. Introduction

Nanofiltration has been increasingly utilized in centralized drinking water purification, wastewater reuse, brine softening and emerging organics removal [1–3]. Polymeric membranes, fabricated through successive soaking in aqueous and organic monomers, are considered as the standard nanofiltration membranes due to their cost-effectiveness and relatively high permeance [4–6]. However, commercially available nanofiltration membranes exhibit low permeance because of a critical trade-off between water permeance and salt rejection, necessitating a larger membrane area and higher operational pressure in practical applications [7–10]. Since the selective layer formed through interfacial polymerization plays a dominant role in nanofiltration, extensive efforts have been made to screen potential new monomers and optimize fabrication conditions for nanofiltration membranes, continually breaking through trade-off limitations [11–14].

Apart from the widely used monomer piperazine, researchers have explored the substitution of the selected amine in interfacial polymerization. To date, the reported selective layers based on monomer substitution primarily consist of polyester, polyamine and polyurea [15–17]. For example, Zheng et al. used glucose, maltose and raffinose as aqueous monomers to react with trimesoyl chloride, respectively [18]. The optimal polyester membrane exhibited a high water permeance of  $33.7 \text{ L}\cdot\text{m}^{-2}\cdot\text{h}^{-1}\cdot\text{bar}^{-1}$  and a  $\text{Na}_2\text{SO}_4$  rejection rate of approximately 95 %. Meanwhile, several researchers have focused on optimizing the fabrication conditions of interfacial polymerization using diamine as primary monomers. Various rational strategies have been implemented to improve the interfacial polymerization reaction, including nanomaterials incorporation [19], substrate modification [20], diffusion regulation [21], and surfactant mediated [22]. Although some of these efforts broke the trade-off limitations, the performance of nanofiltration membranes remained below the theoretical upper bound for water permeance and salt rejection due to the non-optimal selection of aqueous monomers and fabrication condition combinations [23]. Meanwhile, this selection challenge, stemming from the curse of dimensionality in the vast chemical space, renders comprehensive experimental evaluation nearly impossible. Therefore, it is necessary to precisely identify the interfacial polymerization and the corresponding membrane performances, especially in permeance and salt rejection. However, the highly non-linear relationships and intricate interplay of parameters in the interfacial polymerization process pose a significant challenge to identifying membrane performance [24,25].

Recently, machine learning algorithms have emerged as a new paradigm in material science development [26]. Diverging from conventional statistical analysis or mathematical models, machine learning algorithms enable the robust capability to unravel complex non-linear correlations through functions such as classification, regression estimation, clustering, and fitting. Numerous machine learning algorithms have been formulated and effectively implemented in membrane applications, including the random forest [27], artificial neural network [28], support vector machine [29], k-nearest neighbor [30], and gradient-boosted machines [31]. For example, Shetty et al. employed artificial neural networks to precisely predict the nanofiltration permeate concentrations of organic and inorganic contaminants [32]. Besides, Ma et al. utilized random forest to reveal the key structural and operating features that influence the water/salt selectivity of nanofiltration membranes [33]. Among these algorithms, gradient boosting decision tree algorithms build trees on bootstrap datasets, anticipating that each subsequent tree leverages information from preceding trees to enhance its growth and predictive performance [34]. Presently, a range of boosting algorithms, including Light Gradient Boosting Machine (LightGBM) [35], Extreme Gradient Boosting (XGBoost) [36], Adaptive Boosting (AdaBoost) [37], and Categorical Boosting (CatBoost) [38], are commonly employed in conjunction with decision tree models. However, few studies combined the candidate monomers and machine learning algorithms to study the relationship between the interfacial

polymerization process and experimental performance [39]. The development of such algorithms is essential to facilitate the paradigm shift from a trial-and-error mentality to the inverse design strategy of nanofiltration membranes [40].

In this work, four gradient boosting decision tree algorithms were formulated to precisely predict the permeance and salt rejection of nanofiltration membranes fabricated through an interfacial polymerization process. The predictive capabilities of these algorithms were evaluated using the root mean square error ( $\text{RMSE}$ ), the mean absolute error ( $\text{MAE}$ ) and the coefficient of determination ( $R^2$ ). The important features of candidate monomers, represented by 14 molecular descriptors, and their fabrication conditions were identified using Shapley additive explanation (SHAP) coefficients. Additionally, a functional relationship between important features and membrane performance was proposed through univariate and bivariate partial dependence plot (PDP) analyses. This work introduces an inverse design strategy for developing high-performance nanofiltration membranes.

## 2. Materials and methods

### 2.1. Data collection and analysis

**Fig. 1** illustrates the detailed nanofiltration membrane performance prediction process using gradient boosting decision tree algorithms. The dataset about nanofiltration membrane fabrication was sourced from an analysis of previous publications (Text S1), predominantly concentrating on the interfacial polymerization method. This method involves an interfacial reaction between a water monomer and an organic monomer on a porous substrate. Data extraction from Figures within these publications was facilitated using a Plot Digitizer software (version 2.6.11). The fabrication parameters were carefully selected, such as water and organic monomer concentrations, polymerization time, heating time and temperature. Molecular descriptors for the aqueous monomers were computed utilizing RDKit (version 2022.03.5) and Mordred packages (version 1.2.0). To mitigate redundant information, constant, near constant, and missing molecular descriptors were screened based on the Pearson correlation coefficient analysis (see Eq. (1)). The Pearson correlation matrix is calculated via the Pandas package (version 2.1.1).

$$p_{ij} = \frac{\sum_{i=1}^n (a_i - a_{ave}) - \sum_{i=1}^n (b_i - b_{ave})}{\sqrt{\sum_{i=1}^n (a_i - a_{ave})^2 - \sum_{i=1}^n (b_i - b_{ave})^2}} \quad (1)$$

where  $a_i$  and  $b_i$  denote the individual input values,  $a_{ave}$  and  $b_{ave}$  signify the respectively means of variables  $a$  and  $b$ . The symbol  $n$  represents the total number of data points. After the above screening process, the 14 most critical molecular descriptors were left in the dataset (Table S1). Features not explicitly stated in the publications were marked as missing values within the dataset. Finally, two datasets were constructed based on the output variable of permeance or salt rejection. Both datasets had identical fabrication conditions, whereas the only difference was five essential properties of salt ions introduced into the salt rejection dataset, including valence, ionic radius, Stokes radius, hydrated radius, and hydration-free energy.

### 2.2. Artificial intelligence algorithms and statistical accuracy measurement

The gradient boosting algorithm is an ensemble learning method that builds a robust predictor by weighting the results of multiple weak predictors. It can handle various types of features, including numeric features, categorical features, and missing values. Four gradient boosting decision tree algorithms, i.e., XGBoost, LightGBM, AdaBoost and CatBoost, were employed to develop models for predicting water permeance and salt rejection. Generally, XGBoost, LightGBM, and CatBoost were all developed based on the gradient boosting decision tree

algorithm, which learned the first-order negative gradient of the loss function of each weak classifier from the previous weak classifier. For the XGBoost algorithm, the regularization term and the second derivative information of the loss function in the training process are introduced, yielding fast convergence speed and avoiding overfitting problem. The LightGBM algorithm incorporates several innovative techniques, including the histogram algorithm, Gradient-based One-Side Sampling and leaf-wise tree growth strategy. These advancements collectively contribute to substantial enhancements in both the training and running speeds of the machine learning model. The CatBoost algorithm solves the problems of gradient bias and prediction shift, which improves the accuracy and generalization ability of the algorithm. In the AdaBoost algorithm, the weights of samples correctly predicted by the previous weak classifier are decreased in the loss function, while the weights of incorrectly predicted samples increase. This mechanism enhances the learning process for the incorrectly predicted samples, thereby improving the overall performance of the model.

To construct and evaluate machine learning models, 80 % of the dataset was randomly divided into the training set, and the other 20 % of the data was used as the test set. Then, the hyperparameters of each model were optimized using the training set by the Bayesian optimization algorithm, with the root mean square error of the five-fold cross-validation as the objective function (Table S2–5). Finally, the prediction performance was evaluated by  $R^2$ , MAE and RMSE. Generally, smaller MAE and RMSE values and a higher  $R^2$  value represent better prediction performance.  $R^2$ , MAE and RMSE values are calculated using Eqs. (2)–(4), respectively.

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - y_{ave})^2} \quad (2)$$

$$MAE = \frac{\sum_{i=1}^n |\hat{y}_i - y_i|}{n} \quad (3)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}} \quad (4)$$

where  $y_i$  is the actual value of sample  $i$ ,  $\hat{y}_i$  denotes the predicted value of the machine learning model and  $y_{ave}$  signifies the average of all samples in the test set.

### 2.3. Methodology of interpretation analysis

For in-depth interpretation, SHAP analysis, based on cooperative game theory, was used to assess the differences in predictions with and without the feature. In the SHAP analysis, the weighted average of the marginal effects of each feature on the machine learning model is calculated within each feature combination. This process yields a consistent and unbiased estimate of the importance value of each feature. Features with higher absolute SHAP values are regarded as greater contributions, with more information for predicting the target variable. The SHAP value is calculated according to the following formula (Eq. (5)):

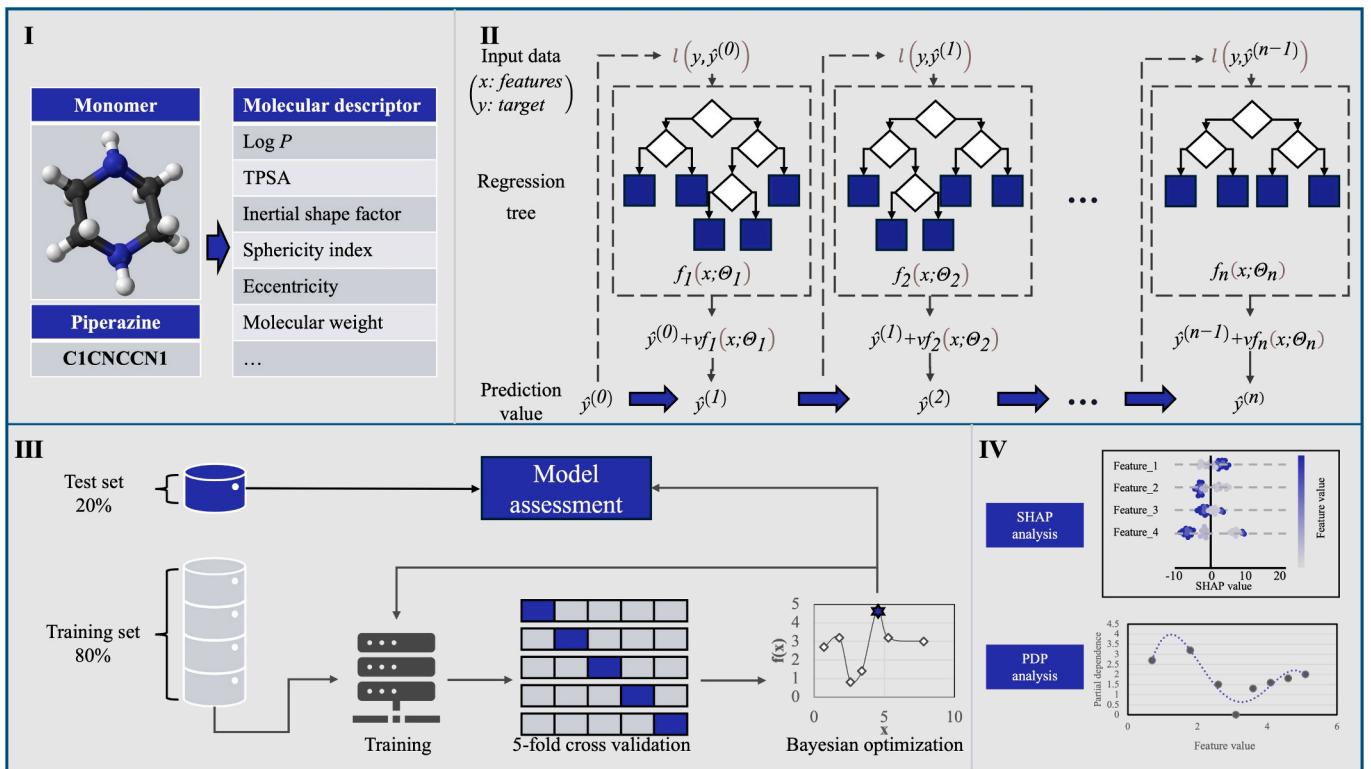
$$\Phi(f, x) = \sum_{S \subseteq N} \frac{(|S|-1)!(n-|S|-1)!}{n!} (f(S) - f(S \setminus x)) \quad (5)$$

where  $S$  is the subset of all features with feature  $x$ ,  $f(S)$  denotes the prediction by the machine learning model considering feature  $x$ , whereas  $f(S \setminus x)$  is the prediction without considering feature  $x$ .

The PDP analysis was employed to analyze the dependence relationships of different features on the prediction results of the machine learning model. The corresponding partial dependence function is defined as follows (Eq. (6)):

$$f_S = \int f(x_S, x_C) dP(x_C) \quad (6)$$

where  $f_S$  is the learned machine learning model,  $x_S$  and  $x_C$  represent the



**Fig. 1.** Schematic of gradient boosting decision tree algorithms framework for nanofiltration membrane performance prediction. (I) data collection and pre-processing; (II) model development; (III) model assessment; (IV) interpretability analysis.

feature to plot the partial dependence relationship and the rest of the feature in the model, respectively. In practice, the Monte Carlo method is used to calculate the average value of the training set to obtain the partial dependency function. The specific formula is shown below (Eq. (7)):

$$f_s = \frac{1}{n} \sum_{i=1}^n f(x_s, x_c^i) \quad (7)$$

where  $x_i c$  represents the feature value of  $x_c$  of sample  $i$ .

### 3. Results and discussion

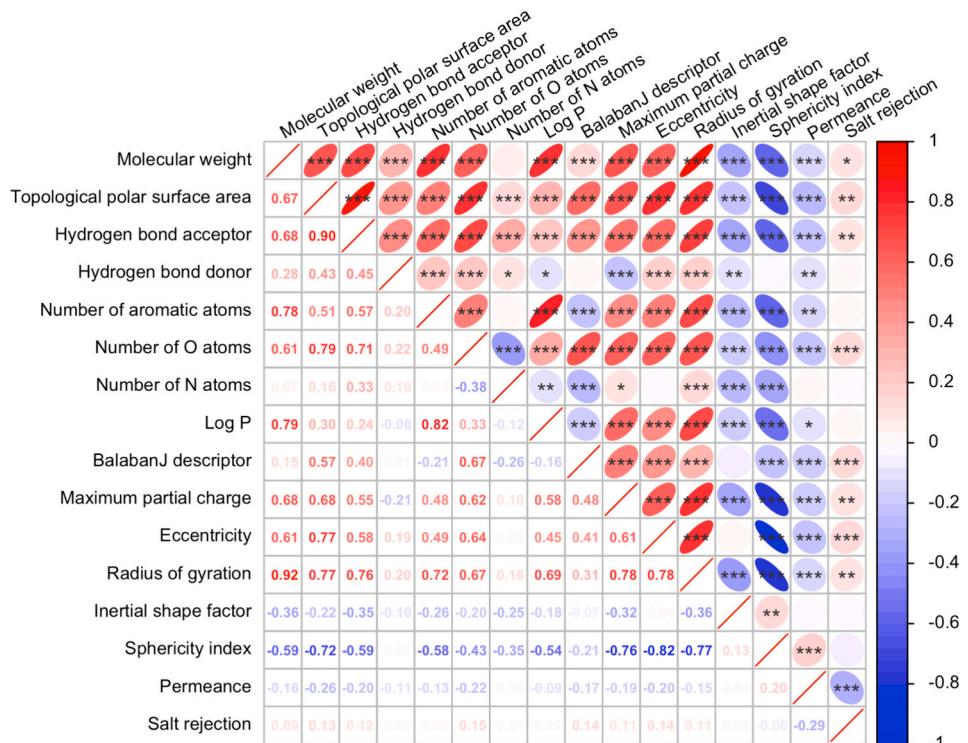
#### 3.1. Pearson correlation coefficient

The linear relationships between molecular descriptors of aqueous monomers and their correlation with nanofiltration membrane performance were investigated by using Pearson correlation analysis (Fig. 2). The Pearson coefficient takes values ranging from  $-1$  to  $+1$  [41]. A value of 0 indicates no linear correlation between two variables, while values of  $+1$  and  $-1$  indicate positive and negative linear correlations, respectively. Considering only molecular descriptors, the molecular weight showed high positive correlations with number of aromatic atoms, Log P and radius of gyration with Pearson coefficient values over 0.7. Generally, the high molecular weight means the complex structure of monomers, which may possess more aromatic atoms and a high radius of gyration. In contrast, the sphericity index had negative correlations with topological polar surface area (TPSA), maximum partial charge, eccentricity and radius of gyration. Several molecular descriptors exhibited poor correlations with others, such as hydrogen bond donor, number of N atoms, BalabanJ descriptor and inertial shape factor. Regarding the target output of nanofiltration membrane permeance and salt rejection, it was observed that all molecular descriptors had low absolute correlation values below 0.3. This result was attributed to the limitations of the Pearson correlation analysis, which only captures linear correlations between the two variables and thus fails to reveal

their non-linear relationships [42]. To further assess the extent of the relationships between molecular descriptors and nanofiltration membrane performance, statistical significance was conducted with a significance level set at  $p < 0.05$ . It was found that the molecular weight, TPSA, hydrogen bond acceptor, number of O atoms, BalabanJ descriptor, maximum partial charge, eccentricity and radius of gyration showed strong correlations ( $p < 0.001$ ) with permeance, indicating that tuning the molecular polarity and weight, hydrogen bond acceptors and topological structure contributed to fabricating nanofiltration membranes with higher permeance. Additionally, the permeance had a negative correlation with sphericity index ( $p < 0.001$ ), confirming that sphericity index is an important parameter for nanofiltration membrane fabrication. In terms of salt rejection, significantly positive correlations were found for the number of O atoms, BalabanJ descriptor and eccentricity. However, relying solely on Pearson correlation analysis and statistical significance indexes made it difficult to discern the individual contributions of each molecular descriptor to the membrane performance, as well as to identify any potential synergistic effects among them. Therefore, further analysis based on machine learning models was deemed necessary to comprehend the non-linear relationships in the nanofiltration fabrication process.

#### 3.2. Predicting performance of gradient boosting decision tree algorithms

Based on the dataset constructed according to the method in Section 2.1, machine learning models for revolutionizing nanofiltration membrane design were trained by XGBoost, LightGBM, CatBoost and AdaBoost algorithms, respectively. In the training process, given that the XGBoost algorithm could not inherently support the input of categorical variables, the categorical features were converted into numeric variables using backward difference encoding. To improve the hyperparameter optimization effect of the Bayesian optimization algorithm, the numeric variables were standardized by the Power transformer. After obtaining the trained model, the water permeance and salt rejection were predicted for all data points in the test set. For a more



**Fig. 2.** Pearson correlation matrix between molecular descriptors and membrane performance. The blue and red ellipse denote positive and negative correlations, respectively. Significant level:  $p < 0.05$  (\*),  $p < 0.01$  (\*\*), and  $p < 0.001$  (\*\*\*)

comprehensive assessment of the prediction performance of gradient boosting decision tree algorithms, joint scatter plots were used to depict the experimental versus predicted values of water permeance and salt rejection. The actual values are on the abscissa, while the predicted values are on the ordinate for each data point. As shown in Fig. 3, four gradient boosting decision tree algorithms showed well prediction ability to predict water permeance of nanofiltration fabricated with the selected two-phase monomers and process. The predictive performance of four algorithms was further evaluated based on the test set using three indices (Table 1). The prediction accuracy of four algorithms is in descending order: XGBoost ( $R^2 = 0.76$ ,  $MAE = 3.15$ ,  $RMSE = 5.08$ )  $\approx$  AdaBoost ( $R^2 = 0.77$ ,  $MAE = 2.87$ ,  $RMSE = 4.02$ )  $>$  LightGBM ( $R^2 = 0.73$ ,  $MAE = 3.45$ ,  $RMSE = 5.52$ )  $>$  CatBoost ( $R^2 = 0.60$ ,  $MAE = 3.78$ ,  $RMSE = 6.44$ ). Given the rapid training and tuning process along with high predictive accuracy, the XGBoost algorithm is recommended as a superior and more robust choice for predicting membrane permeance.

Another critical performance index of nanofiltration membranes, i.e., salt rejection, was predicted using four gradient boosting decision tree algorithms. To broaden the salt rejection prediction encompassing a wider range of salt types, the dataset included five typical salt rejection values:  $Na_2SO_4$ ,  $MgSO_4$ ,  $CaCl_2$ ,  $MgCl_2$  and  $NaCl$ . As shown in Fig. 4, the actual and predicted values of the training and test sets were presented in a scatter plot, distinguished by different colors and shapes. The black dashed line represents the position where the predicted salt rejection matched its actual value as tested by experiments. Most of the experimental results for salt rejection demonstrated a relatively strong agreement with the predicted values, affirming the reliability of four

**Table 1**

Predicting capability evaluation of gradient boosting decision tree algorithms for permeance.

Gradient boosting algorithm	Training set			Test set		
	$R^2$	MAE	RMSE	$R^2$	MAE	RMSE
XGBoost	0.98	0.71	1.30	0.76	3.15	5.08
LightGBM	0.91	2.07	3.22	0.73	3.45	5.52
AdaBoost	0.95	1.95	2.40	0.77	2.87	4.02
CatBoost	0.88	0.75	3.62	0.60	3.78	6.44

algorithms (Fig. 4a-d). Table 2 summarizes the generalization ability of four algorithms for salt rejection. In terms of the test set unused in model development, the XGBoost algorithm exhibited superior accuracy, with the highest  $R^2$  score of 0.88, as well as the lowest  $MAE$  and  $RMSE$  values of 6.42 and 10.44 among all algorithms. Therefore, based on its predicted accuracy and computational efficiency, the XGBoost algorithm demonstrated superior adaptability in predicting the salt rejection of nanofiltration membranes.

### 3.3. Feature importance assessment

Following the development of gradient boosting decision tree algorithms for permeance and salt rejection prediction, the XGBoost algorithm was interpreted using SHAP analysis due to its highest accuracy among the four gradient boosting algorithms. Generally, a negative SHAP value indicates that the input parameter has a negative impact on the output and vice versa [43]. By leveraging the absolute mean SHAP

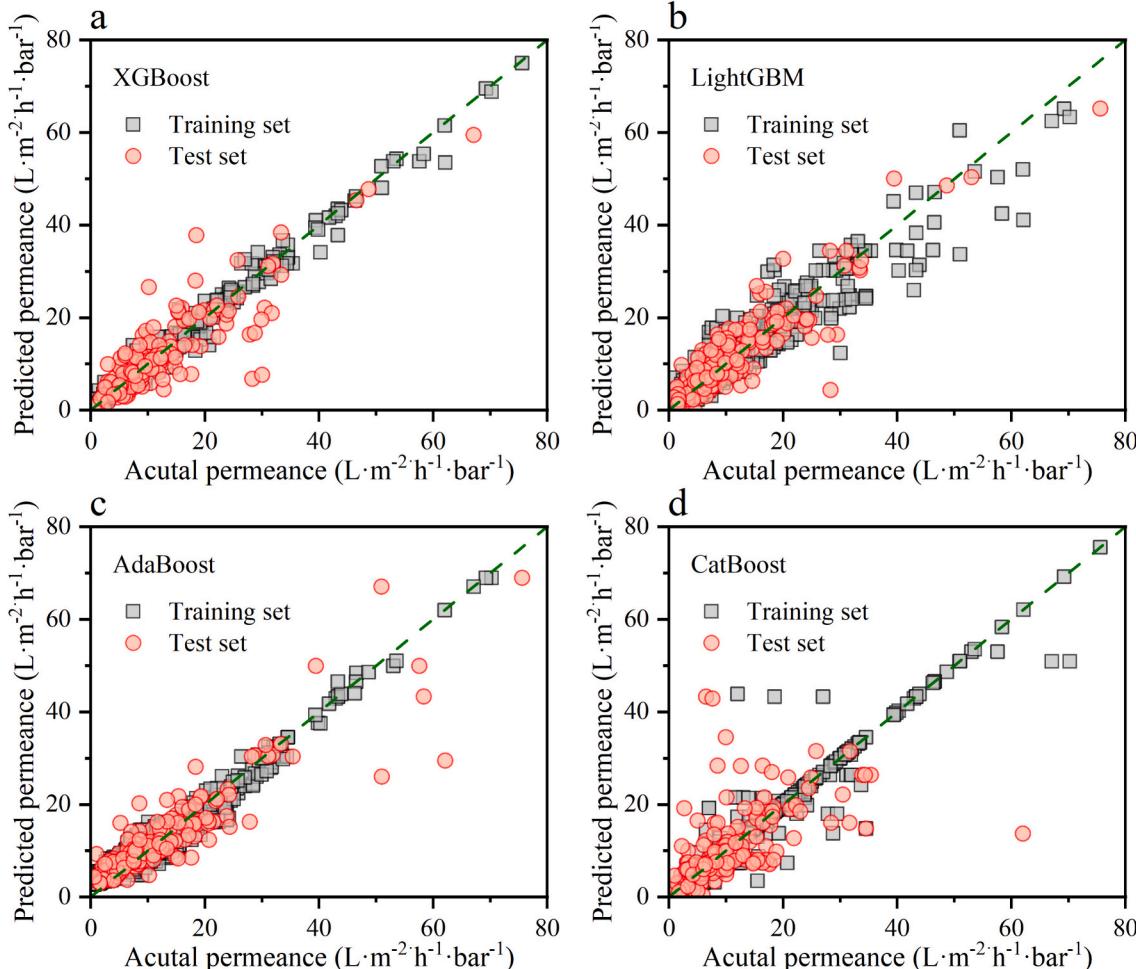
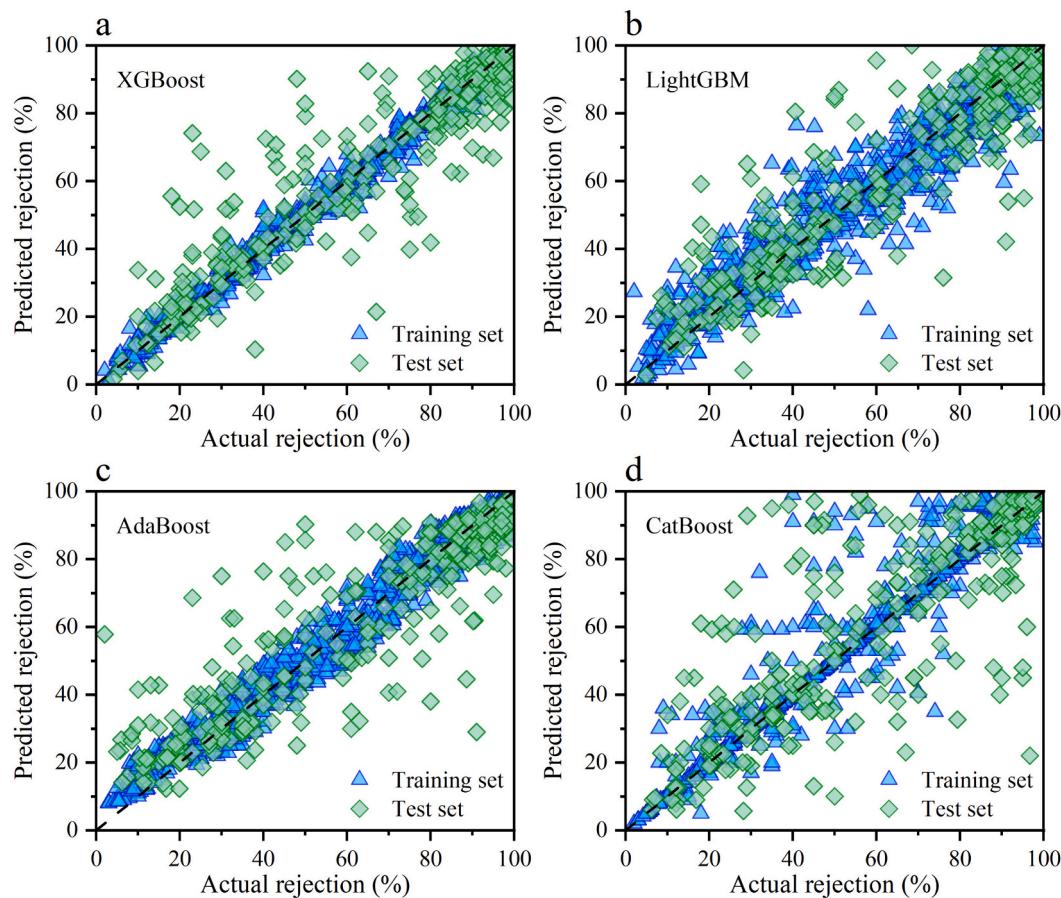


Fig. 3. Correlation of actual and prediction permeance using (a) XGBoost, (b) LightGBM, (c) AdaBoost, and (d) CatBoost algorithms.



**Fig. 4.** Correlation of actual and prediction salt rejection using (a) XGBoost, (b) LightGBM, (c) AdaBoost, and (d) CatBoost algorithms.

**Table 2**  
Predicting capability evaluation of gradient boosting decision tree algorithms for salt rejection.

Gradient boosting algorithm	Training set			Test set		
	R <sup>2</sup>	MAE	RMSE	R <sup>2</sup>	MAE	RMSE
XGBoost	0.99	1.60	2.37	0.88	6.42	10.44
LightGBM	0.95	4.76	6.63	0.84	8.18	11.83
AdaBoost	0.97	4.51	5.41	0.85	7.92	11.36
CatBoost	0.94	2.28	6.99	0.77	8.08	14.08

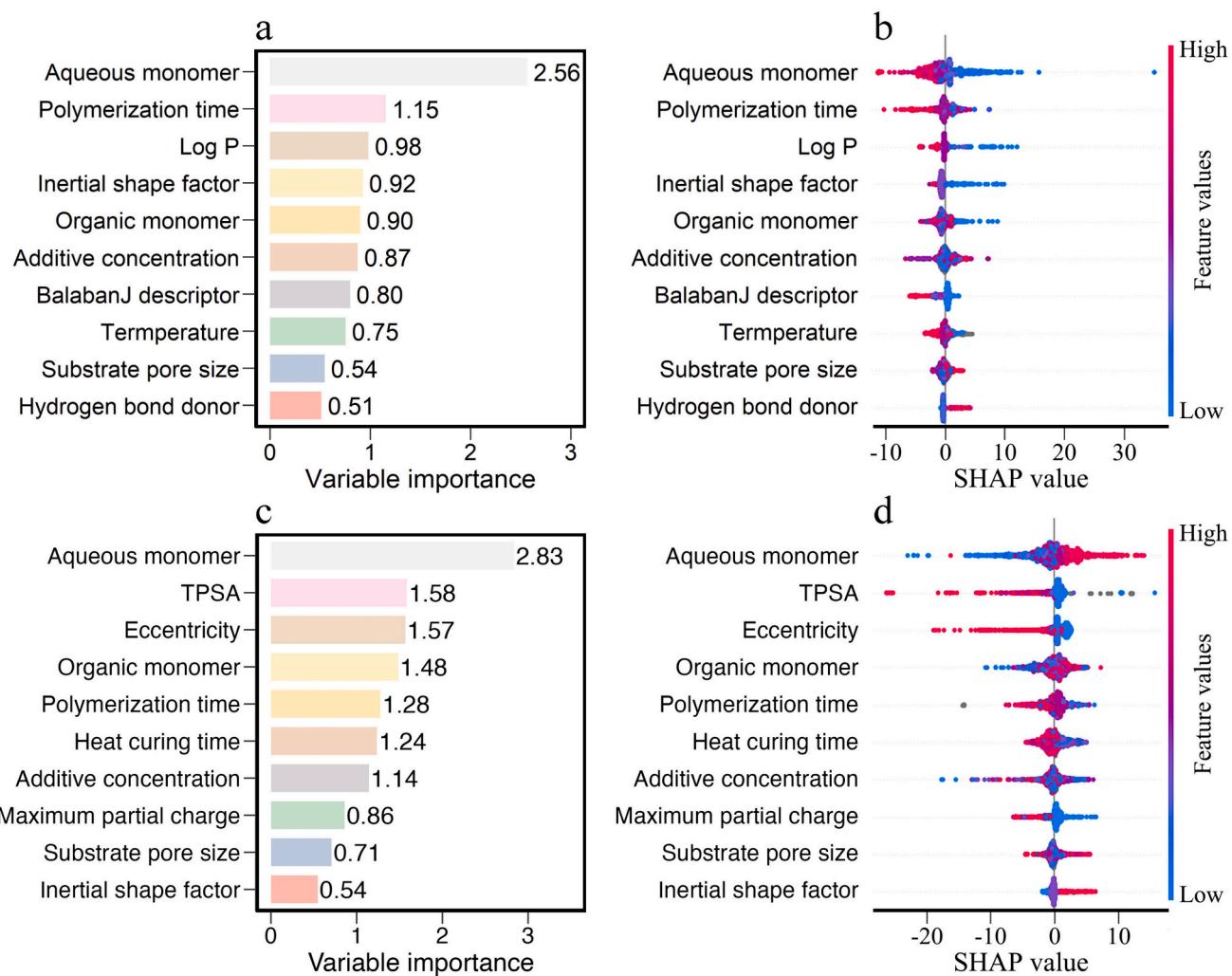
values of the training set, a consistent and impartial assessment of each feature importance was provided. For water permeance, the ten features are listed below in descending order of importance (Fig. 5a): aqueous monomer concentration, polymerization time, Log P, inertial shape factor, organic monomer concentration, additive concentration, BalanbanJ descriptor, temperature, substrate pore size and hydrogen bond donor. The feature of aqueous monomer concentration shows the highest importance than other features. The feature importance values of polymerization time, Log P, inertial shape factor, organic monomer, additive concentration, BalanbanJ descriptor and temperature were above 0.75, which illustrated the importance for membrane permeance by those seven features. The feature importance values of substrate pore size and hydrogen bond donor were 0.54 and 0.51, respectively, consistent with the synergistic optimization method between the substrate and aqueous monomers reported in recent high-performance membrane fabrication [44,45].

The important features affecting permeance were decoupled based on the SHAP value of each sample in the training set. The horizontal location of each point represents its SHAP value, while the color

indicates the level of the corresponding normalized input value. As shown in Fig. 5b, high aqueous monomer concentration exhibited a negative effect on the water permeance. Generally, increasing aqueous monomer concentration would increase the crosslinking degree of nanofiltration membranes, leading to a decrease in membrane permeance [46]. Similarly, the polymerization time and organic monomer concentration were negatively correlated with membrane permeance, which agreed with the domain knowledge [47]. When considering molecular descriptors, Log P and BalanbanJ descriptor were negatively correlated with water permeance, while the hydrogen bond donor exhibited the opposite pattern. Log P represents the partition coefficient of chemicals in water and the organic phase at dynamic equilibrium. A low Log P value indicated high hydrophilicity of the aqueous monomer in the interfacial polymerization process, consistent with results that aqueous monomers with more hydrophilic functional groups could easily fabricate high-flux nanofiltration membranes [48,49].

The feature importance of salt rejection was also evaluated. As illustrated in Fig. 5c, the important membrane fabrication features in the salt rejection were similar to those in the water permeance. With the absolute mean SHAP value of 2.83, the aqueous monomer concentration was the most significant feature. In contrast, the important molecular descriptors were quite different. TPSA and eccentricity showed high importance in predicting salt rejection rate, meaning molecular polarity of aqueous monomer had an important impact on the rejection performance. Some other features such as organic monomer concentration, polymerization time, heat time, additive concentration and maximum partial charge showed significant effects and should be considered when fabricating nanofiltration membranes.

Fig. 5d summarizes the SHAP values of each sample feature on the salt rejection. Given that the monomer concentration played a beneficial



**Fig. 5.** The interpretable analysis for features: (a) feature importance of permeance, (b) SHAP analysis of permeance, (c) feature importance of salt rejection, and (d) SHAP analysis of salt rejection.

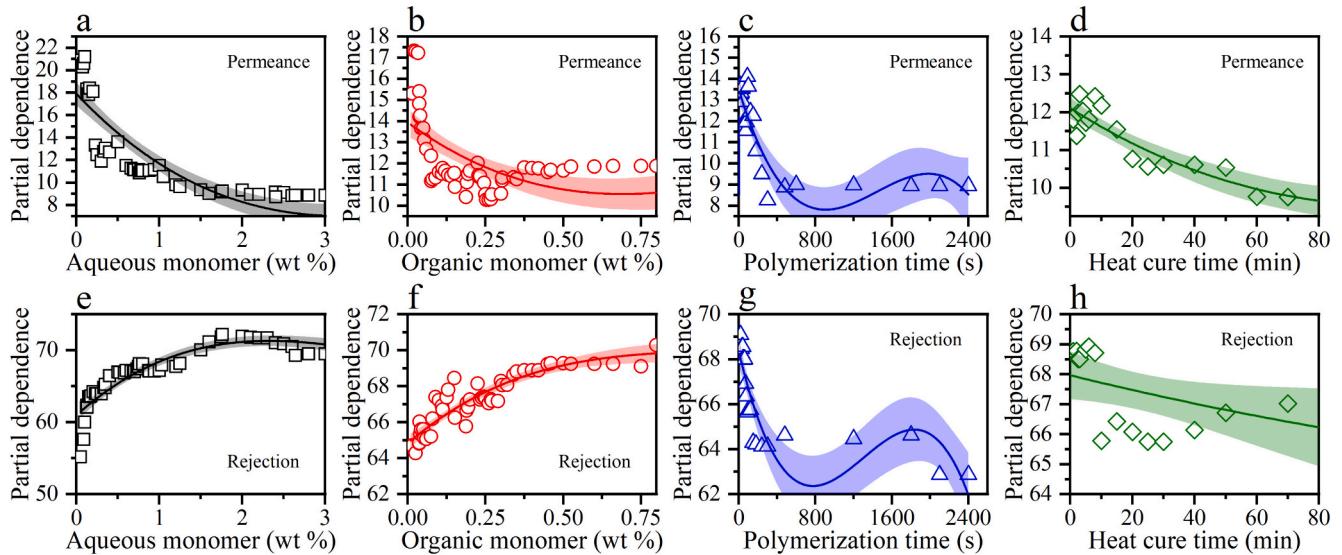
role in the reaction process, both aqueous and organic monomer concentrations exhibited a positive impact on desalination. Notably, the SHAP analysis revealed that a long polymerization time exhibited a negative effect on salt rejection for certain data points, which was contrary to the actual reaction. Combined with the analysis of aqueous monomer concentration, it was found that samples with high polymerization time generally had low reaction activity, yielding nanofiltration membranes with loose structure and low salt rejection. Similarly, the heat time feature exhibited an anomalous effect, likely attributable to the variations in the reaction monomer property employed. In terms of important molecular descriptors, TPSA and eccentricity are related to molecular polarity. Based on their distribution of SHAP values, it could be concluded that aqueous monomers with low molecular polarity contributed to the salt rejection.

### 3.4. Partial dependence analysis for membrane performance

Based on the Pearson correlation and SHAP analysis results, four important operational features, including aqueous monomer concentration, organic monomer concentration, polymerization time and heat cure time, were selected for PDP analysis to reveal the quantitative relationship between these features and nanofiltration membrane performance. Regarding membrane permeance, PDP analysis displayed a strictly negative correlation between aqueous monomer concentration and membrane permeance (Fig. 6a). Specifically, the membrane

permeance exhibited a nearly linear decrease when the aqueous monomer concentration was below 1.8 wt%. However, as the aqueous monomer concentration further increased, the rate of decline in membrane permeance decelerated. Similarly, the organic monomer concentration showed a negative correlation with membrane permeance (Fig. 6b). In general, the effect of organic monomer concentration on membrane permeance was less than that of aqueous monomer concentration [50]. Thus, the membrane permeance tended to be stable when the organic monomer concentration was above 0.37 wt%. The effect of polymerization time on membrane permeance showed remarkable variation (Fig. 6c). When the polymerization time was below 840 s, the membrane permeance decreased rapidly with increasing polymerization time, indicating the formation of a dense selective layer. Subsequently, the water permeance remained stable within the range of 840 to 2400 s of polymerization time due to the usage of low reactivity monomers. The heat cure time showed an approximately linear negative correlation with membrane permeance (Fig. 6d), which was attributed to the further polymerization reaction between residual monomers under the high temperature.

For salt rejection, both the aqueous and organic monomer concentrations showed a significant positive relationship with salt rejection (Fig. 6e, f). Additionally, it was observed that the increase in salt rejection slowed down as the monomer concentration increased, which was attributed to the full crosslinking between aqueous and organic monomers. As a result, the interface diffusion and reaction of monomers

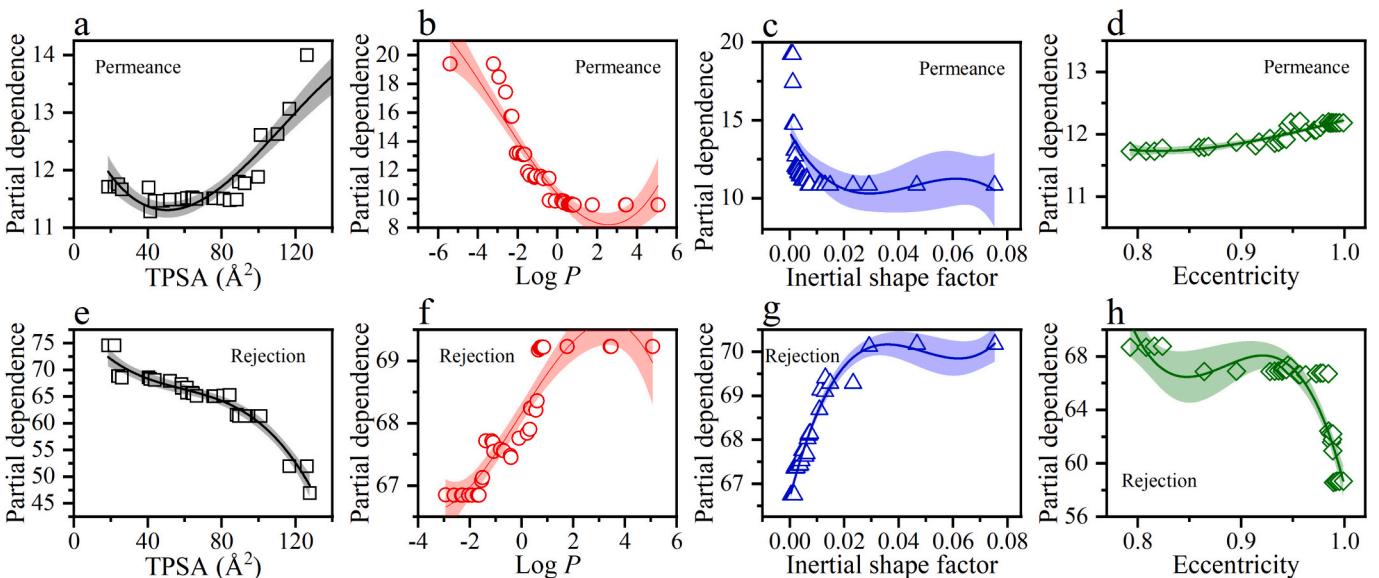


**Fig. 6.** Univariable PDP assessment of critical membrane fabrication conditions on membrane performance. (a, e) Aqueous monomer concentration, (b, f) organic monomer concentration, (c, g) polymerization time, and (d, h) heat cure time. Shaded areas represent the 95 % confidence intervals for the regression lines.

were self-limited due to the high resistance by the formed selective layer [51]. In contrast, the polymerization time had an adverse effect on salt rejection (Fig. 6g). This phenomenon was ascribed to the low reactivity monomers needing more time to form a relatively dense selective layer [52]. However, compared with the traditional polyamide layer fabricated via piperazine and trimethylol chloride, the pore size of such a selective layer was still large, resulting in a poor size sieving effect on salt [53]. Heat cure time also negatively affected salt rejection according to the regressing curve (Fig. 6h), yet this tendency was deemed insignificant due to the wide confidence intervals.

Partial dependence analysis was also adopted to establish the correlation between membrane performance and four important molecular descriptors, including TPSA, Log *P*, inertial shape factor and eccentricity. As illustrated in Fig. 7a, the permeance increased with increasing TPSA when its value was over 80 Å<sup>2</sup>. With the increase in TPSA, aqueous monomers showed a complex three-dimensional topology and several reactive sites, resulting in loose structure or defects in the selective layer.

Thus, aqueous monomers commonly used to prepare dense selective layers generally had low TPSA and molecular weight, such as piperazine and *m*-phenylenediamine. The Log *P* value of aqueous exhibited a remarkable effect on membrane permeance (Fig. 7b). In the range of -6 to 2.5, the Log *P* value showed a significant negative relation to membrane permeance. This result was attributed to the hydrophobicity property of aqueous monomers with a high Log *P* value [54]. Meanwhile, the hydrophobic aqueous monomer tended to form a selective layer with low hydrophilicity and slowed the transport of water molecules due to the hydrophobic interaction [55]. As the Log *P* value continued to increase, the membrane permeance progressively stabilized at a threshold level. Fig. 7c demonstrated that the predicted membrane permeance value rapidly declined before the inertial shape factor value reached 0.005. Then, the contribution of the inertial shape factor almost disappears with its further increase. The sphericity index displayed a poor correlation with membrane permeance (Fig. 7d), consistent with the SHAP analysis result.



**Fig. 7.** Univariable PDP assessment of critical molecular descriptors on membrane performance. (a, e) TPSA, (b, f) Log *P*, (c, g) inertial shape factor, and (d, h) eccentricity. Shaded areas represent the 95 % confidence intervals for the regression lines.

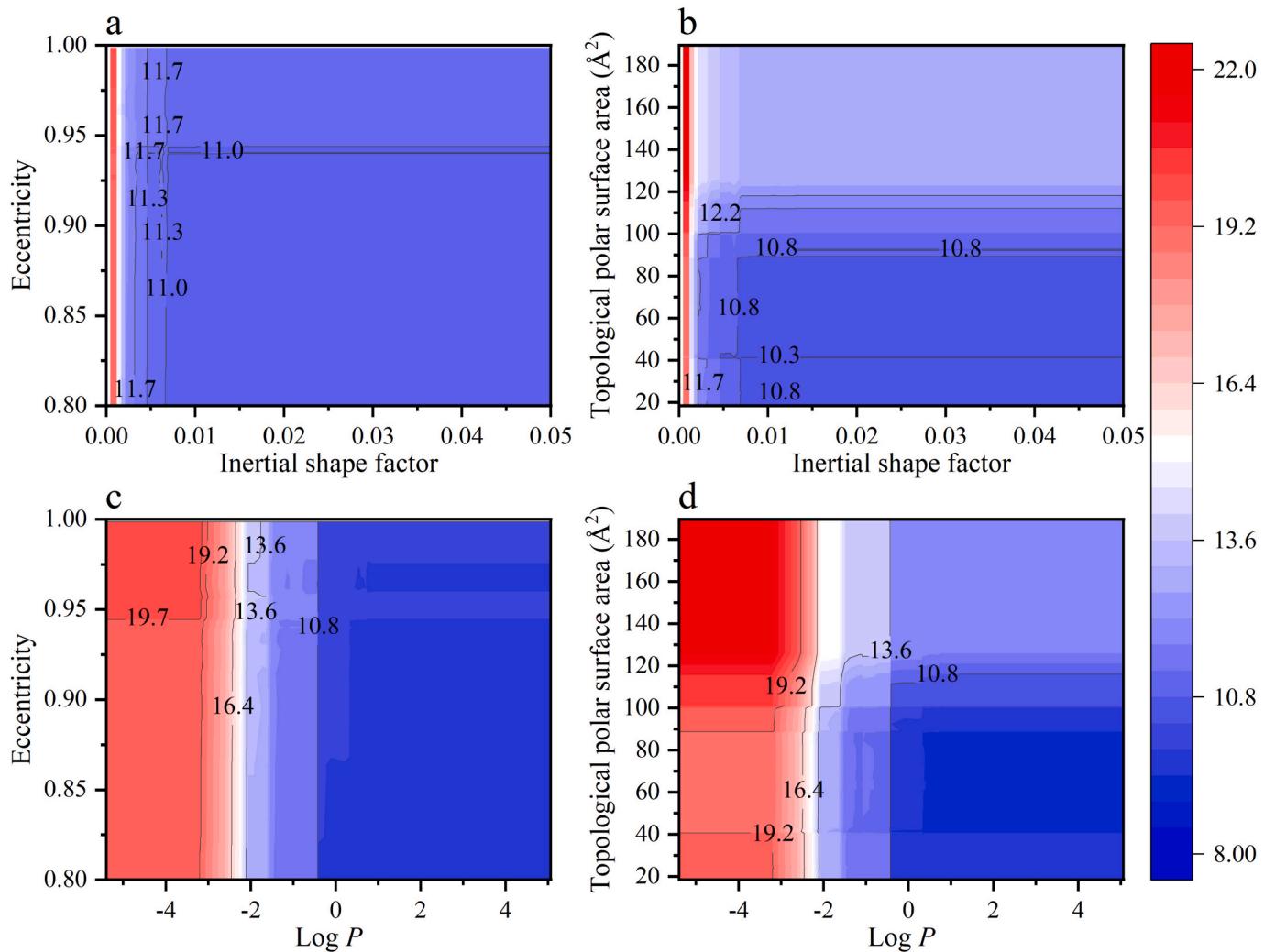
A significant trade-off effect between permeance and salt rejection occurred with the variation of molecule descriptor. For example, TPSA had a distinct negative relation to the salt rejection (Fig. 7e), as opposed to the trend of permeance. This result further indicated that the complex topology of aqueous monomers easily fabricated a nanofiltration membrane with defects [56]. Fig. 7f showed that the Log P had a positive contribution to salt rejection according to the regression curve. Given the relatively small variation range of salt rejection partial dependence, the effect of Log P on salt rejection was less than that on permeance. The predicted salt rejection rate increased with an inertial shape factor value ranging from 0.00 to 0.03 (Fig. 7g). However, further increases in the inertial shape factor had a negligible effect on the predicted salt rejection rate. Eccentricity exhibited a complex relationship with the salt rejection rate. The regression curve indicated that eccentricity had a negative relationship with salt rejection when it was <0.84. Conversely, eccentricity showed a minimal impact on salt rejection when its value was between 0.84 and 0.92. Notably, a significant negative correlation between eccentricity and salt rejection was observed with a continued increase in eccentricity value.

The performance of nanofiltration membranes exhibited a significant correlation with various molecular descriptors of aqueous monomers [57]. To delve deeper into these intricate interactions, a two-dimensional PDP analysis was employed to elucidate the influence of critical molecular descriptors on membrane permeance. Fig. 8a

illustrates the predicted permeance of the XGBoost algorithm across varying inertial shape factors and eccentricity values of aqueous monomers, while the remaining input features were averaged in the dataset. The region of high water permeance appeared compressed to the narrow-left side of the coordinate system. This occurrence stemmed from the discrepancy in the significance of the two descriptors and the pivotal influence of the inertial shape factor primarily manifests within the range of 0 to 0.05.

The combination of low inertia shape factor and high TPSA was found to enhance membrane permeance (Fig. 8b). When the inertial shape factor is lower than 0.002, the change in TPSA value had almost no effect on water permeance. As the inertial shape factor continued to increase, the water permeance rapidly decreased to an equilibrium value, at which point the advantageous effect of a high TPSA on water permeance became evident. In the area where TPSA was above  $120 \text{ \AA}^2$ , the equilibrium value was about 12.67, significantly higher than the equilibrium value where TPSA was below  $120 \text{ \AA}^2$ . Due to the narrow range of inertial shape factor values in the highly permeable region, using aqueous monomers with large TPSA was more feasible for fabricating high water permeance nanofiltration membranes.

Fig. 8c illustrates the effects of Log P and eccentricity on water permeance. Notably, when the Log P value was less than -3.2 and the eccentricity exceeded 0.95, the water permeance was predicted to reach its peak. When the Log P value was greater than -3.2, Log P began to



**Fig. 8.** Bivariate PDP assessment of (a) inertial shape factor and eccentricity, (b) inertial shape factor and TPSA, (c) Log P and eccentricity, and (d) Log P and TPSA on the membrane permeance.

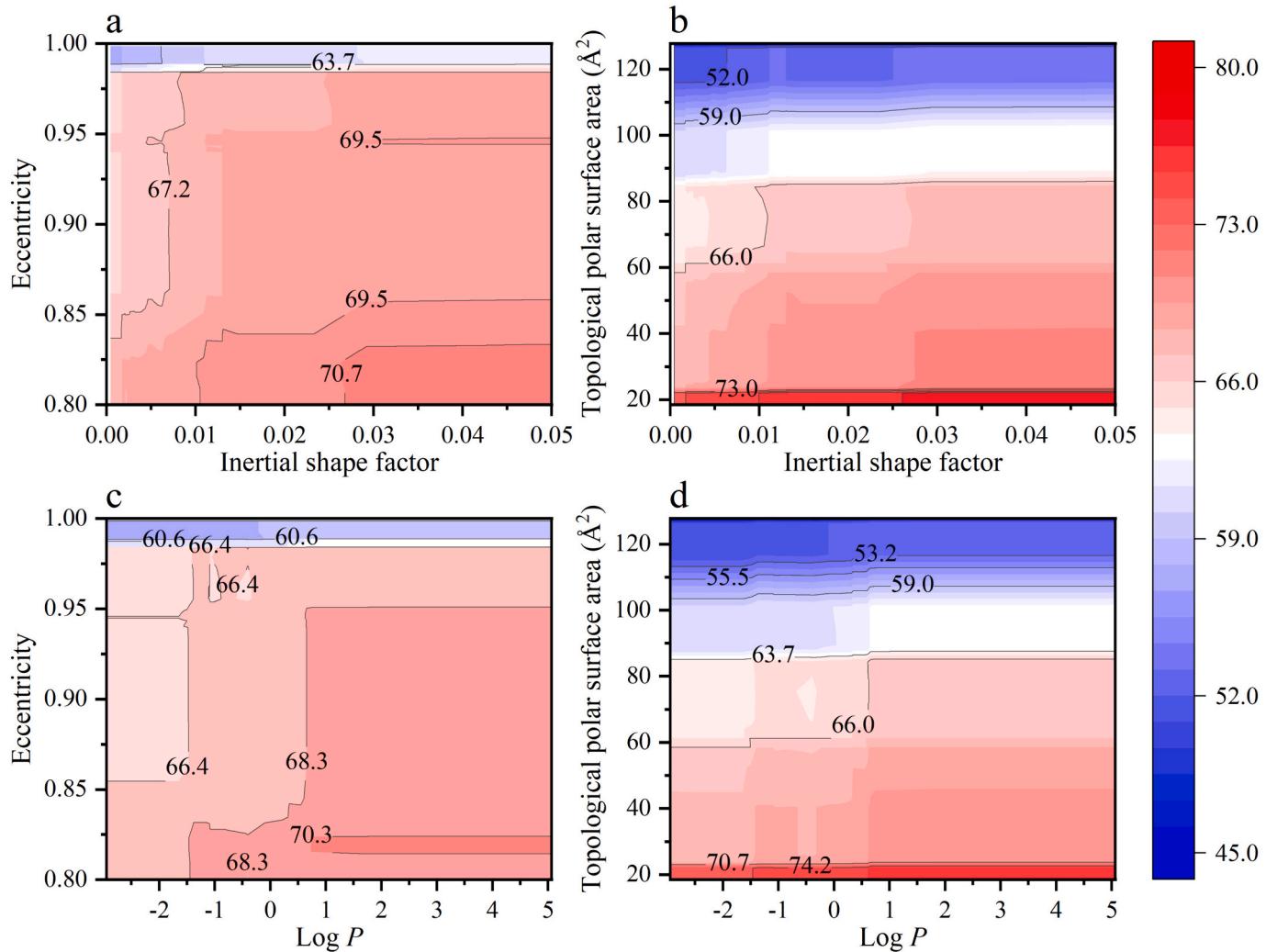
play a dominant role in the membrane permeance with a negative relation. The influence of Log P and TPSA on water permeance was also depicted in Fig. 8d. Interestingly, two regions within the coordinate system exhibited high permeance. Specifically, when the Log P value was lower than  $-2.4$ , TPSA values  $<40 \text{ \AA}^2$  or  $>100 \text{ \AA}^2$  both positively contributed to membrane permeance.

Bivariate PDP analysis was also employed to evaluate the effect of molecular descriptors of aqueous monomers on salt rejection. As shown in Fig. 9a, the predicted salt rejection was highest in the lower right corner of the coordinate system, whereas it was lowest in the upper left corner. This result indicated that the high inertial shape factor endowed the nanofiltration membrane with high predicted salt rejection, while the high eccentricity showed the opposite effect. In addition, the predicted salt rejection is lower at high inertial shape factor and high eccentricity values compared to low inertial shape factor and low eccentricity values, indicating that the inertial shape factor played a more important role in the salt rejection [58]. Meanwhile, the inertial shape factor and TPSA showed a similar trend in the salt rejection (Fig. 9b), except that the salt rejection rate was nearly unaffected by the inertial shape factor when the TPSA value was below  $22 \text{ \AA}^2$ . With a further increase in the TPSA value, the predicted salt rejection is affected by both the TPSA and the inertial shape factor. A high inertial shape factor and low TPSA facilitated the fabrication of nanofiltration membranes with a high salt rejection rate.

The interaction between Log P and eccentricity is illustrated in Fig. 9c. The Log P showed a positive relation to salt rejection, whereas the eccentricity had a negative effect. When the Log P value exceeded  $3$ , its effect on salt rejection diminished, and the predicted value changed almost exclusively with variations in eccentricity. Fig. 9d shows that the decrease in TPSA value could improve the salt rejection by the enhanced size sieving effect. Combining the DPD and SHAP analyses, we found that the Log P and inertial shape factor were significant features for permeance, whereas TPSA and eccentricity were important for salt rejection. Taken together, selecting a Log P value of about  $-2$ , an eccentricity below  $0.85$ , and a TPSA below  $30 \text{ \AA}^2$  proved effective in fabricating nanofiltration membranes with high salt rejection. Increasing eccentricity to  $0.90$  and TPSA to below  $60 \text{ \AA}^2$  were conducive to fabricating high flux nanofiltration membranes with essential salt rejection. This study demonstrates the feasibility of utilizing the gradient boosting decision tree algorithm to gain an in-depth understanding of membrane fabrication parameters and optimize the nanofiltration membrane performance.

#### 4. Conclusions

In summary, we developed a workflow employing machine learning methods to inversely supervise nanofiltration membrane fabrication, eliminating the need for the current trial-and-error approach. Four



**Fig. 9.** Bivariate PDP assessment of (a) inertial shape factor and eccentricity, (b) inertial shape factor and TPSA, (c) Log P and eccentricity, and (d) Log P and TPSA on the salt rejection.

gradient boosting decision tree algorithms were successfully applied to identify non-linear interactions hidden within the datasets and to precisely forecast membrane performance. XGBoost algorithm showed the best predictive performance in accuracy and generalization with  $R^2 = 0.76$ ,  $MAE = 3.15$  and  $RMSE = 5.08$  for permeance. Feature importance and SHAP analysis demonstrated that the aqueous monomer concentration was the most important feature for permeance and salt rejection, with absolute mean SHAP values of 2.56 and 2.83, respectively. Univariate and bivariate PDP analyses quantified the relationship between important features and membrane performance, providing intensive insights regarding the influence of fabrication conditions on permeance and salt rejection. The foundational understanding of these features is instrumental in guiding the modulation of permeance and salt rejection, as well as in directing the design of high-performance nanofiltration membranes.

### Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper. Jungbin Kim and Yan Zhao serve as (Guest) Editors for the journal Desalination, while the editorial handling and review of this manuscript were overseen by a different Editor.

### Acknowledgements

The work was jointly supported by the National Natural Science Foundation of China (No. 52300083), Heilongjiang Province Postdoctoral Fund (No. LBH-Z23181), China Postdoctoral Science Foundation (No. 2023M740918), Postdoctoral Fellowship Program of CPSF (No. GZB20230965), the Natural Science Foundation of the Heilongjiang Province of China (Grant No. LH2021E007). Yan Zhao would like to acknowledge the support provided by the Fonds Wetenschappelijk Onderzoek - Vlaanderen (FWO) (No. 12A6823N).

### Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.desal.2024.118072>.

### References

- [1] W. Liu, B. Liu, X. Zhu, J. Zhu, Novel insight into prior induced crystallization on brackish water nanofiltration, Desalination 568 (2023) 117009, <https://doi.org/10.1016/j.desal.2023.117009>.
- [2] Y. Bai, B. Liu, J. Li, M. Li, Z. Yao, L. Dong, D. Rao, P. Zhang, X. Cao, L.F. Villalobos, C. Zhang, Q.-F. An, M. Elimelech, Microstructure optimization of bioderived polyester nanofibers for antibiotic desalination via nanofiltration, Sci. Adv. 9 (2023) eadg6134, <https://doi.org/10.1126/sciadv.adg6134>.
- [3] S. Han, J. Zhu, A.A. Uliana, D. Li, Y. Zhang, L. Zhang, Y. Wang, T. He, M. Elimelech, Microporous organic nanotube assisted design of high performance nanofiltration membranes, Nat. Commun. 13 (2022) 7954, <https://doi.org/10.1038/s41467-022-35681-9>.
- [4] D. Xu, X. Zhu, X. Luo, Y. Guo, Y. Liu, L. Yang, X. Tang, G. Li, H. Liang, MXene nanosheet templated nanofiltration membranes toward ultrahigh water transport, Environ. Sci. Technol. 55 (2021) 1270–1278, <https://doi.org/10.1021/acs.est.0c06835>.
- [5] D. Xu, J. Zheng, X. Zhang, D. Lin, Q. Gao, X. Luo, X. Zhu, G. Li, H. Liang, B. Van der Bruggen, Mechanistic insights of a thermoresponsive interface for fouling control of thin-film composite nanofiltration membranes, Environ. Sci. Technol. 56 (2022) 1927–1937, <https://doi.org/10.1021/acs.est.1c06156>.
- [6] T.K. Arsene, T. Chen, J. Zhu, Y. Zhang, Elevated water transport of polyamide nanofiltration membranes via aqueous organophosphorus mediated interfacial polymerization, Desalination 584 (2024) 117754, <https://doi.org/10.1016/j.desal.2024.117754>.
- [7] H. Yu, H. Huang, L. Zhong, S. Wu, H. Yang, H. Rong, H. Liang, F. Qu, J. Ma, Evaluation of front-face fluorescence for assessing Cyanobacteria fouling in ultrafiltration, Environ. Sci. Technol. 57 (2023) 17649–17658, <https://doi.org/10.1021/acs.est.3c07397>.
- [8] D. Xu, H. Liang, X. Zhu, L. Yang, X. Luo, Y. Guo, Y. Liu, L. Bai, G. Li, X. Tang, Metal-polyphenol dual crosslinked graphene oxide membrane for desalination of textile wastewater, Desalination 487 (2020) 114503, <https://doi.org/10.1016/j.desal.2020.114503>.
- [9] D. Xu, X. Luo, P. Jin, J. Zhu, X. Zhang, J. Zheng, L. Yang, X. Zhu, H. Liang, B. Van der Bruggen, A novel ceramic-based thin-film composite nanofiltration membrane with enhanced performance and regeneration potential, Water Res. 215 (2022) 118264, <https://doi.org/10.1016/j.watres.2022.118264>.
- [10] D. Janowitz, N. Becker, A. Sweity, A. Margane, N. Al Katheb, S. Groche, S. Yüce, T. Wintgens, Seawater softening by nanofiltration enables ecofriendly Dead Sea level stabilisation while creating the basis for cost-effective inland desalination, Desalination 574 (2024) 117245, <https://doi.org/10.1016/j.desal.2023.117245>.
- [11] K. Wang, X. Wang, B. Januszewski, Y. Liu, D. Li, R. Fu, M. Elimelech, X. Huang, Tailored design of nanofiltration membranes for water treatment based on synthesis-property-performance relationships, Chem. Soc. Rev. 51 (2022) 672–719, <https://doi.org/10.1039/D0CS01599G>.
- [12] X. Zhu, Z. Sun, F. Tan, F. Chen, X. Luo, F. Wang, D. Wu, H. Liang, D. Xu, X. Cheng, Tailoring high-performance polyester loose nanofiltration membrane for selective separation of salt/dyes: the equilibrium of condensation and hydrolysis, Sep. Purif. Technol. 333 (2024) 125848, <https://doi.org/10.1016/j.seppur.2023.125848>.
- [13] Z. Yang, L. Long, C. Wu, C.Y. Tang, High Permeance or high selectivity? Optimization of system-scale nanofiltration performance constrained by the upper bound, ACS EST Eng. 2 (2022) 377–390, <https://doi.org/10.1021/acsestengg.1c00237>.
- [14] S. Park, R. Patel, Y.C. Woo, Polyester-based thin-film composite membranes for nano-filtration of saline water: a review, Desalination 572 (2024) 117138, <https://doi.org/10.1016/j.desal.2023.117138>.
- [15] R. Zhang, Q. Sun, J. Tian, B. Van der Bruggen, J. Zhu, Non-polyamide nanofiltration (NPA-NF) membrane: a non-mainstream but indispensable member of the “membrane family”, Desalination 564 (2023) 116772, <https://doi.org/10.1016/j.desal.2023.116772>.
- [16] M. Wanjiya, J.-C. Zhang, B. Wu, M.-J. Yin, Q.-F. An, Nanofiltration membranes for sustainable removal of heavy metal ions from polluted water: a review and future perspective, Desalination 578 (2024) 117441, <https://doi.org/10.1016/j.desal.2024.117441>.
- [17] A. Alkhouzaam, M. Khraisheh, Development of eco-friendly coating for the fabrication of high performing loose nanofiltration membranes for dye-contaminated wastewater treatment, Desalination 581 (2024) 117609, <https://doi.org/10.1016/j.desal.2024.117609>.
- [18] J. Zheng, Y. Liu, J. Zhu, P. Jin, T. Croes, A. Volodine, S. Yuan, B. Van der Bruggen, Sugar-based membranes for nanofiltration, J. Membr. Sci. 619 (2021) 118786, <https://doi.org/10.1016/j.memsci.2020.118786>.
- [19] B.-H. Jeong, E.M.V. Hoek, Y. Yan, A. Subramani, X. Huang, G. Hurwitz, A. K. Ghosh, A. Jawor, Interfacial polymerization of thin film nanocomposites: a new concept for reverse osmosis membranes, J. Membr. Sci. 294 (2007) 1–7, <https://doi.org/10.1016/j.memsci.2007.02.025>.
- [20] S. Karan, Z. Jiang, A.G. Livingston, Sub-10 nm polyamide nanofilms with ultrafast solvent transport for molecular separation, Science 348 (2015) 1347–1351, <https://doi.org/10.1126/science.aaa5058>.
- [21] Z. Tan, S. Chen, X. Peng, L. Zhang, C. Gao, Polyamide membranes with nanoscale Turing structures for water purification, Science 360 (2018) 518–521, <https://doi.org/10.1126/science.aar6308>.
- [22] Y. Liang, Y. Zhu, C. Liu, K.-R. Lee, W.-S. Hung, Z. Wang, Y. Li, M. Elimelech, J. Jin, S. Lin, Polyamide nanofiltration membrane with highly uniform sub-nanometre pores for sub-1 Å precision separation, Nat. Commun. 11 (2020) 2015, <https://doi.org/10.1038/s41467-020-15771-2>.
- [23] H. Gao, S. Zhong, W. Zhang, T. Igou, E. Berger, E. Reid, Y. Zhao, D. Lambeth, L. Gan, M.A. Afolabi, Z. Tong, G. Lan, Y. Chen, Revolutionizing membrane design using machine learning-Bayesian optimization, Environ. Sci. Technol. 56 (2022) 2572–2581, <https://doi.org/10.1021/acs.est.1c04373>.
- [24] R. Dai, Z. Yang, Z. Qiu, L. Long, C.Y. Tang, Z. Wang, Distinct impact of substrate hydrophilicity on performance and structure of TFC NF and RO polyamide membranes, J. Membr. Sci. 662 (2022) 120966, <https://doi.org/10.1016/j.memsci.2022.120966>.
- [25] L.E. Peng, Z. Yao, Z. Yang, H. Guo, C.Y. Tang, Dissecting the role of substrate on the morphology and separation properties of thin film composite polyamide membranes: seeing is believing, Environ. Sci. Technol. 54 (2020) 6978–6986, <https://doi.org/10.1021/acs.est.0c01427>.
- [26] L. Tao, J. Byrnes, V. Varshney, Y. Li, Machine learning strategies for the structure-property relationship of copolymers, iScience 25 (2022) 104585, <https://doi.org/10.1016/j.isci.2022.104585>.
- [27] B. Deng, Y. Deng, M. Liu, Y. Chen, Q. Wu, H. Guo, Integrated models for prediction and global factors sensitivity analysis of ultrafiltration (UF) membrane fouling: statistics and machine learning approach, Sep. Purif. Technol. 313 (2023) 123326, <https://doi.org/10.1016/j.seppur.2023.123326>.
- [28] S. Mu, Y. Liu, H. Zhang, J. Wang, Characterization of coagulant-induced ultrafiltration membrane fouling by multi-spectral fusion: DOM properties and model prediction based on machine learning, Desalination 531 (2022) 115711, <https://doi.org/10.1016/j.desal.2022.115711>.
- [29] P. Cabrera, J.A. Carta, J. González, G. Melián, Wind-driven SWRO desalination prototype with and without batteries: a performance simulation using machine learning models, Desalination 435 (2018) 77–96, <https://doi.org/10.1016/j.desal.2017.11.044>.
- [30] T. Zhu, Y. Zhang, C. Tao, W. Chen, H. Cheng, Prediction of organic contaminant rejection by nanofiltration and reverse osmosis membranes using interpretable machine learning models, Sci. Total Environ. 857 (2023) 159348, <https://doi.org/10.1016/j.scitotenv.2022.159348>.
- [31] H. Gao, S. Zhong, R. Dangayach, Y. Chen, Understanding and designing a high-performance ultrafiltration membrane using machine learning, Environ. Sci. Technol. 57 (2023) 17831–17840, <https://doi.org/10.1021/acs.est.2c05404>.

- [32] Predicting contaminant removal during municipal drinking water nanofiltration using artificial neural networks, *J. Membr. Sci.* 212 (2003) 99–112, [https://doi.org/10.1016/S0376-7388\(02\)00473-8](https://doi.org/10.1016/S0376-7388(02)00473-8).
- [33] X. Ma, D. Lu, J. Lu, Y. Qian, S. Zhang, Z. Yao, L. Liang, Z. Sun, L. Zhang, Revealing key structural and operating features on water/salts selectivity of polyamide nanofiltration membranes by ensemble machine learning, *Desalination* 548 (2023) 116293, <https://doi.org/10.1016/j.desal.2022.116293>.
- [34] Y. Sun, Z. Zhao, H. Tong, B. Sun, Y. Liu, N. Ren, S. You, Machine learning models for inverse design of the electrochemical oxidation process for water purification, *Environ. Sci. Technol.* 57 (2023) 17900–18000, <https://doi.org/10.1021/acs.est.2c08771>.
- [35] G. Ke, Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, Q. Ye, T.-Y. Liu, LightGBM: A Highly Efficient Gradient Boosting Decision Tree, *Proceedings of the 31st International Conference on Neural Information Processing Systems, 2017*, pp. 3149–3157.
- [36] T. Chen, C. Guestrin, XGBoost: A Scalable Tree Boosting System, *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2016*, pp. 785–794.
- [37] T. Hastie, S. Rosset, J. Zhu, H. Zou, Multi-class AdaBoost, *Stat. Interface* 2 (2009) 349–360, <https://doi.org/10.4310/SII.2009.v2.n3.a8>.
- [38] L. Prokhorenko, G. Gusev, A. Vorobev, A.V. Dorogush, A. Gulin, CatBoost: unbiased boosting with categorical features, in: *Proceedings of the 32nd International Conference on Neural Information Processing Systems, 2018*, pp. 6639–6649.
- [39] H. Jalaei Salmani, R. Hardian, H. Kalani, M.R. Moradi, H. Karkhanechi, G. Szekely, H. Matsuyama, Predicting the performance of organic solvent reverse osmosis membranes using artificial neural network and principal component analysis by considering solvent-solvent and solvent-membrane affinities, *J. Membr. Sci.* 687 (2023) 122025, <https://doi.org/10.1016/j.memsci.2023.122025>.
- [40] G. Ignacz, A.K. Beke, G. Szekely, Data-driven future for nanofiltration: escaping linearity, *J. Membr. Sci. Lett.* 3 (2023) 100040, <https://doi.org/10.1016/j.memlett.2023.100040>.
- [41] Spatiotemporal distribution of polycyclic aromatic hydrocarbons in sediments of a typical river located in the Loess Plateau, China: influence of human activities and land-use changes, *J. Hazard. Mater.* 424 (2022) 127744, <https://doi.org/10.1016/j.jhazmat.2021.127744>.
- [42] Y. Tang, M. Kong, X. Tian, J. Wang, Q. Xie, A. Wang, Q. Zhang, H. Zhou, J. Wu, Y. Tian, A series of terpyridine-based zinc(II) complexes assembled for third-order nonlinear optical responses in the near-infrared region and recognizing lipid membranes, *J. Mater. Chem. B* 5 (2017) 6348–6355, <https://doi.org/10.1039/C7TB01063J>.
- [43] A. Mohammed, H. Alshraideh, F. Alsuwaidi, A holistic framework for improving the prediction of reverse osmosis membrane performance using machine learning, *Desalination* 574 (2024) 117253, <https://doi.org/10.1016/j.desal.2023.117253>.
- [44] Z. Liu, T. Wang, D. Wang, Z. Mi, Regulating the morphology of nanofiltration membrane by thermally induced inorganic salt crystals for efficient water purification, *J. Membr. Sci.* 617 (2021) 118645, <https://doi.org/10.1016/j.memsci.2020.118645>.
- [45] X. Zhu, X. Cheng, X. Luo, Y. Liu, D. Xu, X. Tang, Z. Gan, L. Yang, G. Li, H. Liang, Ultrathin thin-film composite polyamide membranes constructed on hydrophilic poly(vinyl alcohol) decorated support toward enhanced nanofiltration performance, *Environ. Sci. Technol.* 54 (2020) 6365–6374, <https://doi.org/10.1021/acs.est.9b06779>.
- [46] L. Yao, Y. Li, Q. Cheng, Z. Chen, J. Song, Modeling and optimization of metal-organic frameworks membranes for reverse osmosis with artificial neural networks, *Desalination* 532 (2022) 115729, <https://doi.org/10.1016/j.desal.2022.115729>.
- [47] Y. Liu, J.M. Sipe, W. Xu, X. Zhu, L. Bai, D. Xu, G. Li, H. Liang, M.R. Wiesner, Study on the mechanisms for the influence of nanomaterials on the separation performance of nanocomposite membrane from a modeling perspective, *Desalination* 532 (2022) 115740, <https://doi.org/10.1016/j.desal.2022.115740>.
- [48] Y. Liu, G. Van Egen, S. Yuan, H. Wang, M. Chi, D. Xu, Q. Gao, G. Li, J. Zheng, B. Van der Bruggen, Facile and novel fabrication of high-performance loose nanofiltration membranes for textile wastewater recovery, *Sep. Purif. Technol.* 308 (2023) 122867, <https://doi.org/10.1016/j.seppur.2022.122867>.
- [49] J. Li, L. Fang, D. Xu, X. Zhang, L. Jiang, Q. Zhu, Q. Chen, P. Jin, A. Volodine, R. Dewil, X. Gui, Q. Gao, B. Van der Bruggen, Intercalation of small molecules in the selective layer of polyamide nanofiltration membranes facilitates the separation of  $Mg^{2+}/Li^+$ , *Chem. Eng. J.* 487 (2024) 150659 <https://doi.org/10.1016/j.cej.2024.150659>.
- [50] S. Zhou, L. Long, Z. Yang, S.L. So, B. Gan, H. Guo, S.-P. Feng, C.Y. Tang, Unveiling the growth of polyamide nanofilms at water/organic free interfaces: toward enhanced water/salt selectivity, *Environ. Sci. Technol.* 56 (2022) 10279–10288, <https://doi.org/10.1021/acs.est.1c08691>.
- [51] Y. Zhang, H. Wang, J. Guo, X. Cheng, G. Han, C.H. Lau, H. Lin, S. Liu, J. Ma, L. Shao, Ice-confined synthesis of highly ionized 3D-quasilayered polyamide nanofiltration membranes, *Science* 382 (2023) 202–206, <https://doi.org/10.1126/science.adf9531>.
- [52] X. Luo, L. Jiang, R. Zhao, Y. Wang, X. Xiao, S. Ghazouani, L. Yu, Z. Mai, H. Matsuyama, P. Jin, Energy-efficient trehalose-based polyester nanofiltration membranes for zero-discharge textile wastewater treatment, *J. Hazard. Mater.* 465 (2024) 133059, <https://doi.org/10.1016/j.jhazmat.2023.133059>.
- [53] X. Luo, Y. Wang, Y. Guo, Q. Gao, R. Zhao, Y. Liu, R. Liu, L. Jiang, Z. Mai, H. Matsuyama, P. Jin, Cleanable natural-sugar-alcohol-based polyester membrane for highly efficient molecular separation, *Resour. Conserv. Recycl.* 198 (2023) 107149, <https://doi.org/10.1016/j.resconrec.2023.107149>.
- [54] D. Xu, Y. Xie, X. Jin, J. Zheng, Q. Gao, P. Jin, X. Zhu, Z. Zhang, X. Li, G. Li, H. Liang, B. Van der Bruggen, Polyphenol-mediated defect patching of graphene oxide membranes for sulfonamide contaminants removal and fouling control, *J. Hazard. Mater.* 469 (2024) 133890, <https://doi.org/10.1016/j.jhazmat.2024.133890>.
- [55] J. Yin, Z. Yang, C.Y. Tang, B. Deng, Probing the contributions of interior and exterior channels of nanofillers toward the enhanced separation performance of a thin-film nanocomposite reverse osmosis membrane, *Environ. Sci. Technol. Lett.* 7 (2020) 766–772, <https://doi.org/10.1021/acs.estlett.0c00507>.
- [56] Q. Gao, D. Bouwen, S. Yuan, X. Gui, Y. Xing, J. Zheng, H. Ling, Q. Zhu, Y. Wang, S. Depuydt, J. Li, A. Volodine, P. Jin, B. Van der Bruggen, Robust loose nanofiltration membrane with fast solute transfer for dye/salt separation, *J. Membr. Sci.* 674 (2023) 121518, <https://doi.org/10.1016/j.memsci.2023.121518>.
- [57] G. Ignacz, C. Yang, G. Szekely, Diversity matters: widening the chemical space in organic solvent nanofiltration, *J. Membr. Sci.* 641 (2022) 119929, <https://doi.org/10.1016/j.memsci.2021.119929>.
- [58] G. Ignacz, G. Szekely, Deep learning meets quantitative structure–activity relationship (QSAR) for leveraging structure-based prediction of solute rejection in organic solvent nanofiltration, *J. Membr. Sci.* 646 (2022) 120268, <https://doi.org/10.1016/j.memsci.2022.120268>.