Article

Understanding and Designing a High-Performance Ultrafiltration Membrane Using Machine Learning

Haiping Gao,[§] Shifa Zhong,[§] Raghav Dangayach, and Yongsheng Chen*

Cite This: *Environ. Sci. Technol.* 2023, 57, 17831−17840

Read Online

ACCESS | Metrics & More | Article Recommendations | SI Supporting Information

ABSTRACT: Ultrafiltration (UF) as one of the mainstream membrane-based technologies has been widely used in water and wastewater treatment. Increasing demand for clean and safe water requires the rational design of UF membranes with antifouling potential, while maintaining high water permeability and removal efficiency. This work employed a machine learning (ML) method to establish and understand the correlation of five membrane performance indices as well as three major performance-determining membrane properties with membrane fabrication conditions. The loading of additives, specifically nanomaterials ($A\_wt\%$), at loading amounts of >1.0 wt % was found to be the most significant feature affecting all of the membrane performance indices. The polymer content ($P\_wt\%$), molecular weight of the pore maker (M_Da), and pore maker content ($M\_wt\%$) also made considerable contributions to predicting membrane performance. Notably, M_Da was more important than $M\_wt\%$ for predicting membrane performance. The feature analysis of ML models in terms of membrane properties (i.e., mean pore size, overall porosity, and contact angle) provided an unequivocal explanation of the effects of fabrication conditions on membrane performance. Our approach can provide practical aid in guiding the design of fit-for-purpose separation membranes through data-driven virtual experiments.

KEYWORDS: *ultrafiltration membrane, machine learning, antifouling potential, water permeability, membrane properties*
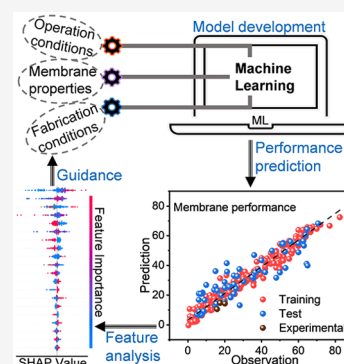
INTRODUCTION

One of the most prominent issues of our time is the increasing supply−demand gap of clean and safe water due to industrialization, population growth, and global climate change. There is a critical need to develop sustainable water treatment technologies. In this context, membrane technologies such as reverse osmosis (RO), nanofiltration (NF), and ultrafiltration (UF) have emerged as promising alternatives to traditional water treatment practices owing to their compelling advantages such as high energy efficiency and fewer chemical additives.[1,2] Of these, UF has been recognized as one of the mainstream membrane-based separation technologies in water and wastewater treatment applications, including the pretreatment stage for RO processes, membrane bioreactors, and water reclamation of effluent from wastewater treatment plants (WWTPs).[3−5]

At the core of the UF separation process is the UF membrane capable of rejecting large molecules [e.g., natural organic matter (NOM)] and bacteria from impaired water sources to produce a clean permeate. UF membranes can also be tuned to target the removal of specific compounds such as heavy metals and dyes, which makes it an attractive candidate for the treatment of industrial effluents.[6−11] In the past several decades, there has been a rapid growth in UF membrane fabrication, including the exploration of membrane materials and the development of fabrication approaches (e.g., phase inversion, electrospinning, etc.).[12,13] Notable examples include the emergence of adsorptive membranes[14,15] and mixed matrix membranes (MMM) involving the incorporation of additives (e.g., nanomaterials) into the membrane matrix.[16−19] However, due to the existence of various organic compounds (e.g., NOM and polysaccharides) and pathogenic microorganisms in water streams, membrane fouling has always been a major obstacle hampering the more widespread application of UF membranes.

Generally, membrane fouling can be classified as reversible fouling and irreversible fouling.[20−22] The impaired membrane performance caused by reversible fouling can be recovered through hydraulic backwashing. However, irreversible fouling induced by the foulants bonded to membrane surface and trapped in pores can be eliminated only by chemical cleaning.[23] Membrane fouling not only causes significant loss of water permeation and a shorter membrane lifetime but also increases operational cost and complexity.[24,25] Therefore, fabricating UF membranes with desirable water permeance and removal efficiency accompanied by high antifouling potential is crucial. A polymeric membrane fabricated through

a phase inversion method has been recognized as the state-of-the-art UF membrane. However, most polymeric membranes were fabricated using hydrophobic polymers such as polysulfone (PSf), poly(ether sulfone) (PES), and polyvinylidene fluoride (PVDF), suffering from poor water flux and membrane fouling. Many efforts, including developing block polymers, surface grafting, blending with hydrophilic polymers, and incorporation of inorganic fillers, have been devoted to developing high-performance UF membranes.[26−32] One popular approach is to fabricate nanocomposite UF membranes by embedding nanomaterials into a polymer matrix. A wide variety of nanomaterials with different physical and chemical properties, ranging from nonporous nanoparticles (e.g., TiO_2) and porous nanomaterials (e.g., carbon nanotubes) to two-dimensional (2D) materials (e.g., graphene oxide), have been investigated.[33−35] With so much interest in nanocomposite UF membranes, identifying the desirable loading and properties of the nanomaterials as well as the suitable membrane fabrication conditions to optimize membrane performance is essential. As membrane performance is mostly determined by membrane properties such as surface hydrophilicity, surface roughness, effective pore size, and porosity, the correlation of membrane properties with fabrication conditions is also worth examining.

Due to the variety of membrane backbone materials and available additives as well as the complexity of the fabrication process, rationally designing UF membranes by developing efficient methods to alleviate the time and resource constraints posed by the iterative trial-and-error approach is pivotal. Because of its powerful ability in processing and learning from large, complex, and multidimensional data sets to develop predictive models, machine learning (ML) as a data-driven method has become increasingly important in chemistry and material science communities for accelerating the discovery of new materials and chemical synthesis.[36−40] In most recent years, ML has been employed to guide gas separation membrane design.[41−43] In addition, ML models such as tree-based models (e.g., random forest, XGBoost, etc.) and artificial neural networks (ANNs) have been developed to predict permeance and rejection for RO and NF membranes in water treatment and resource recovery, including but not limited to solvent recovery.[44−48] Several studies have been acknowledged for the application of ML models in ML-assisted UF membrane fabrication.[49,50] However, the performance of these reported models was often limited by the incomplete input variables and unclear classification of the input features. Furthermore, previous work has mainly focused on prediction of water flux and removal efficiency. Limited attention has been paid to predicting membrane fouling-related performance comprising the flux decline ratio, flux recovery ratio, reversible fouling ratio, and irreversible fouling ratio. Additionally, the quantitative relationship between fabrication conditions and membrane performance-determining properties has not been well established.

In this work, we developed tree-based ML models using extreme gradient boosting (XGBoost) and categorical boosting (CatBoost) as potential candidates to analyze a data set containing input features associated with both fabrication and operational conditions and to target membrane performance. As for membrane performance, we considered water permeability, removal efficiency, and indices associated with membrane antifouling performance, such as the flux decline ratio, flux recovery ratio, and reversible fouling ratio. The relative importance and impact of each feature on the target were evaluated using the Shapley additive explanations (SHAP) method to provide guidance on fabricating UF membranes with desirable water permeability, removal efficiency, and antifouling potential. Moreover, we also developed predictive ML models by correlating fabrication conditions with membrane properties and carried out model interpretations with the SHAP method. This interpretation facilitated a better understanding of the underneath mechanics in terms of the influence and contribution of each fabrication parameter to membrane performance. This work demonstrated the potential of ML methods in providing guidance to the fit-for-purpose membrane development to meet challenges in water and wastewater treatment and resource recovery.

■ DATA SET AND METHODS

Data Collection and Data Set Construction. The quantity and quality of collected data used to develop ML models are crucial to the model prediction performance. To develop ML models with accurate prediction, we mined data from research articles associated with flat-sheet polymeric UF membranes fabricated by the most used non-solvent-induced phase separation (NIPS) method at room temperature to construct the data sets with the total size of 320. The full data sets can be found in the Supporting Information. Parameters and descriptors used as input features were exhaustively collected from tables, text, and graphical data. As summarized in Table S1, on the basis of the empirical domain knowledge, these input features were assembled into three categories consisting of fabrication conditions (11 variables), operational conditions (six variables), and membrane properties (four variables). The ratio between variables and total data size was calculated to be 6.6%. Features describing the fabrication and operational conditions involved both numerical parameters such as polymer concentration ($P\_$wt %), pore maker content ($M\_$wt %), and the loading of the additives ($A\_$wt %) and categorical ones such as the types of pore makers and organic solvents (as summarized in Table S1). The distribution of the data of the numerical features is illustrated in Figure S1. In the constructed data sets, the additives were specifically referred to various types of nanomaterials (as categorized in Figure S2). The categorical features were first converted into numeric ones by the encoding methods. We screened eight encoding methods (as listed in Table S2) and selected the optimum one to convert these categorical features into numeric features. Notably, features with insufficient information such as the absence of molecular weights for polymers and pore makers were designated as missing values rather than being excluded. The base polymers were represented by molecular fingerprints encoding the repeating unit as a binary vector (0, 1) by converting its SMILES in Python's RDKit package. The obtained molecular fingerprints of the polymers combined with other numerical features (including those from converted categorical features) were used as the final input features to develop ML models. Generally, three data sets were compiled, one for the prediction of membrane performance from fabrication and operational conditions, one for the prediction of membrane performance with the combination of fabrication/operational conditions and membrane properties as input, and the third for the prediction of membrane properties from fabrication conditions.

Machine Learning Model Development and Evaluation. As all of the features have their specific physicochemical

meanings, the missing values present in the input features were not imputed. The percentages of missing values for each feature are listed in Table S3. Hence, it is necessary to employ ML algorithms with the capability of processing missing values. Tree-based algorithms have been reported to exibit satisfactory performance in handling missing values containing data sets. Here, we utilized two tree-based ML algorithms, i.e., XGBoost and CatBoost, as candidates. The tree-based ML model for each data set was developed in a similar manner with some modifications depending on the specific data set. Typically, the data set was randomly split into two parts: 80% of the whole data set as the training set and the remaining 20% of the data set as the test set for model evaluation. Five-fold cross-validation was employed on the training set to screen the ML algorithms, encoding methods, and hyper-parameter tuning. After the optimum configuration of ML algorithms and/or encoding methods had been screened (Table S4), the hyperparameters of the ML algorithm were then tuned by using the Bayesian optimization algorithm (Table S5). The length and radius of the molecular fingerprint were considered as hyperparameters, which were tuned together with other hyperparameters of ML algorithms. After obtaining the optimum hyperparameters, we retrained the ML algorithm with the optimal hyperparameters on the whole training set (without using 5-fold cross-validation) to deliver the final ML models. The predictive performance (i.e., generalization ability) of ML models was evaluated on the test set. The coefficient of determination (R^2) and root-mean-square error (RMSE) were utilized to evaluate the prediction accuracy as defined by eqs 1 and 2. The lower RMSE and higher R^2 indicate the better predictive performance of ML models.

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (x_p^i - x_t^i)^2}{\sum_{i=1}^{n} (x_t^i - x_m)^2} \tag{1}$$

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{n} (x_t^i - x_p^i)^2}{n}} \tag{2}$$

where x_p^i is the predicted value of the output, x_t^i is the actual value of the output reported in the literature, x_m is the mean value of all of the output, and n is the number of data points in the training or test set.

Feature Analysis. To understand the built models and thus provide insightful guidance for future membrane fabrication, identifying potential controlling fabrication parameters for membrane properties and performance is essential. In previous work, feature importance analysis has been performed to explain different models (such as random forest and neural network), providing insight into the roles of input features.[49,51] Here, the importance and impact of each feature on the targets were analyzed using the SHAP method.[48] The Shapley value for input feature x (of n total input features) given the prediction p by the built ML model was calculated by[52]

$$\varnothing_x(p) = \sum_{S \subseteq N/x} \frac{|S|!(n - |S| - 1)!}{n!} [p(S \cup x) - p(S)] \tag{3}$$

where S is the subset for each feature without feature x, $p(S \cup x)$ represents the predictions by the built ML model considering feature x, and $p(S)$ represents the predictions without considering feature x. The differences among all possible subsets of $S \subseteq n$ are calculated due to the dependency

of the effect of withholding a feature from other features in the ML model.

Characterization of Membrane Properties. To further validate the model prediction accuracy, three UF membranes were fabricated using NIPS methods as described in Texts S1 and S2 and Table S6. The water contact angle (CA) indicating membrane surface hydrophilicity was measured by a Ramé-Hart model 250 goniometer (Ramé-Hart Instrument Co.). The average membrane pore radius (micrometers) was determined on the basis of the Guerout−Elford−Ferry equation[53] as follows:

$$r_m = \sqrt{\frac{(2.9 - 1.75\varepsilon) \times 8\eta l Q}{\varepsilon A \Delta P}} \tag{4}$$

where η is the water viscosity at 23 °C (9.3×10^{-4} Pa s), Q is the permeate flow rate (cubic meters per second), and ΔP is the operational pressure (pascals).

The overall porosity (percent) of a membrane was measured by the dry−wet weight gravimetric method as expressed by the following equation:

$$\varepsilon = \frac{w_w - w_d}{A l \rho} \tag{5}$$

where w_w is the weight of the hydrated membrane (grams), w_d is the weight of the dried membrane (grams), A is the surface area of the membrane (square centimeters), l is the membrane thickness (centimeters) determined by the cross-section SEM image, and ρ is the water density at 23 °C (0.998 g cm^{-3}).

Evaluation of Membrane Performance. The pure water permeability of the fabricated membranes was measured using a dead-end ultrafiltration cell (Amicon stirred cell, Millipore Sigma) with an effective membrane area of 13.4 cm^2. The membranes were precompacted under 4 bar for 2 h before switching to the operating pressure of 1−1.5 bar. A bovine serum albumin (BSA) solution was used as the feed solution to test the rejection performance of the membranes. The operational conditions for each of the fabricated membranes are listed in Table S6. The concentration changes of BSA were determined using a Shimadzu TOC-L analyzer (Shimadzu Scientific Instruments). The water permeability (A) and rejection efficiency (R) were calculated using eqs 6 and 7.

$$A = \frac{J_w}{\Delta P} \tag{6}$$

$$R = \frac{C_0 - C_t}{C_0} \times 100\% \tag{7}$$

where J_w is the water flux, ΔP is the applied hydraulic pressure, and C_0 and C_t refer to the concentrations of BSA in the feed solution and permeate, respectively.

Membrane fouling performance was evaluated using humic acid (HA) as the model foulant. The membranes under investigation were first compacted for 2 h under applied hydraulic pressure to reach a steady water flux. The water flux was recorded for an additional 1 h to obtain the pure water flux (J_w). Subsequently, the pure water was converted into the HA solution and allowed to run for 6 h at the same applied hydraulic pressure. The steady flux with a 20 mg L^{-1} HA solution as the feed was recorded as J_t. Then, the fouled membrane was physically cleaned with deionized (DI) water. After physical cleaning, the recovered pure water flux (J_r) was recorded for an additional 1 h with DI water as the feed

solution. The flux decline ratio (FDR) coupled with the water flux recovery ratio (FRR) was used to evaluate the membrane antifouling potential based on eqs 8 and 9.[54] Generally, the lower FDR and the higher FRR suggest a better antifouling performance of the membranes.

$$FDR = \left(1 - \frac{J_t}{J_w}\right) \times 100\% \tag{8}$$

$$FRR = \frac{J_r}{J_w} \times 100\% \tag{9}$$

Membrane fouling can be generally classified as reversible and irreversible fouling. The flux decline ratio caused by reversible and irreversible fouling during the filtration process was calculated by

$$R_r = \frac{J_r - J_t}{J_w} \times 100\% \tag{10}$$

$$R_{ir} = \frac{J_w - J_r}{J_w} \times 100\% \tag{11}$$

■ RESULTS AND DISCUSSION

Predictive ML Models. To simplify the ML model and enhance its performance, the Pearson correlation coefficient (PCC) was determined to identify the correlations between features. Figure S3 shows the PCC results of all of the features. With respect to membrane performance, the irreversible fouling ratio showed a completely negative correlation with the flux recovery ratio (i.e., PCC = −1), which is in line with the calculation result according to eqs 9 and 11. Therefore, the irreversible fouling ratio was sorted out from the output target and the flux recovery ratio was chosen as the representative one.

Overfitting is controlled by tunning the hyperparameters of the machine learning algorithm to control the complexity of the model. We used the Bayesian optimization algorithm to tune the hyperparameters of the machine learning algorithm, in which we observed the training and validation performance together. We took the hyperparameters that can afford the best validation performance as the optimum hyperparameters. Figure S4 demonstrates the plot of training RMSE (RMSEcv-training) versus validation RMSE (RMSEcv-validation), in which the black dotted line indicated the lowest RMSEcv-validation. With a decrease in RMSEcv-training, RMSEcv-validation first decreased and then increased, indicating the underfitting and overfitting process. The location of the lowest RMSEcv-validation indicates the optimum complexity of the model.

The predictive performance of the built ML model for each target is listed in Table 1 and plotted in Figure 1. The prediction of ML models in the test data sets exhibited an R^2 value of ≥0.78 for water permeability, removal efficiency, and flux decline ratio. These results indicated the strongly quantitative correlation of membrane performance with the fabrication conditions and operational conditions. In comparison with the prediction performance of these three targets, the models exhibited relatively lower testing R^2 values of 0.62 and 0.73 for the flux recovery ratio and reversible fouling ratio, respectively. As expressed in eqs 9 and 10, these two targets were highly relevant to the recovered water flux after physical

Table 1. Prediction Performance of the ML Models for Membrane Performance

performance index	total data size	training set R^2	training set RMSE	test set R^2	test set RMSE
water permeability (LMH bar^{-1})	320	0.97	29.61	0.83	68.24
removal efficiency (%)	320	0.99	1.83	0.84	6.60
flux decline ratio (%)	320	0.96	3.82	0.78	9.50
flux recovery ratio (%)	320	0.87	4.73	0.62	7.89
reversible fouling ratio (%)	320	0.96	3.91	0.73	10.33

cleaning. Therefore, the relatively limited prediction performance for the flux recovery ratio and reversible fouling ratio may be attributable to the case-by-case variations in the physical cleaning procedures. Improvements in the prediction performance could be achieved by expanding the current data sets to include more reasonable variables as input features such as the cleaning time of the physical cleaning process.

As membrane performance mainly depends on membrane properties, we also developed ML models for predicting membrane performance by including membrane properties (e.g., mean pore radius, overall porosity, contact angle, and surface roughness) together with fabrication/operational conditions as input features. As listed in Table S7, the prediction of the ML models with membrane properties exhibited an R^2 value comparable to those of the ML models without membrane properties. These results revealed the quantitative relationship among fabrication, property, and performance.

To further validate the generalization ability of the developed ML models, we fabricated three UF membranes (as listed in Table S6) and tested their performance (Table S8). Our experimental data points were independent of the 320 data points collected from the literature. We found that the experimental results for all of the targets showed relatively good agreement with the predicted values (Figure 1). This finding indicated that all of the built ML models were reliable enough to provide satisfactory predictions of membrane performance based on the selected input features. Notably, the polymers used to fabricate these three membranes have molecular weights different from those of the polymers enclosed in the collected data sets. The application of the molecular fingerprint method made it possible to deliver acceptable performance predictions for membranes fabricated with new polymers. However, the additive as a categorical feature was encoded to the numerical feature in this work, which limited the prediction capability of the built model for cases with new additives (i.e., those not present in the training data set).

To understand the correlation between membrane properties and fabrication conditions, we also developed ML models to predict three major membrane performance-determining properties, i.e., mean pore radius, overall porosity, and contact angle (revealing the hydrophilicity of the membrane surface), using fabrication conditions as input features. As shown in Table 2, for the test data sets, the R^2 values were 0.87 and 0.76 on the prediction of mean pore radius and contact angle, respectively, and their corresponding RMSEs were 7.59 and 5.87, respectively, suggesting satisfactory prediction performance of the built ML models with fabrication conditions as
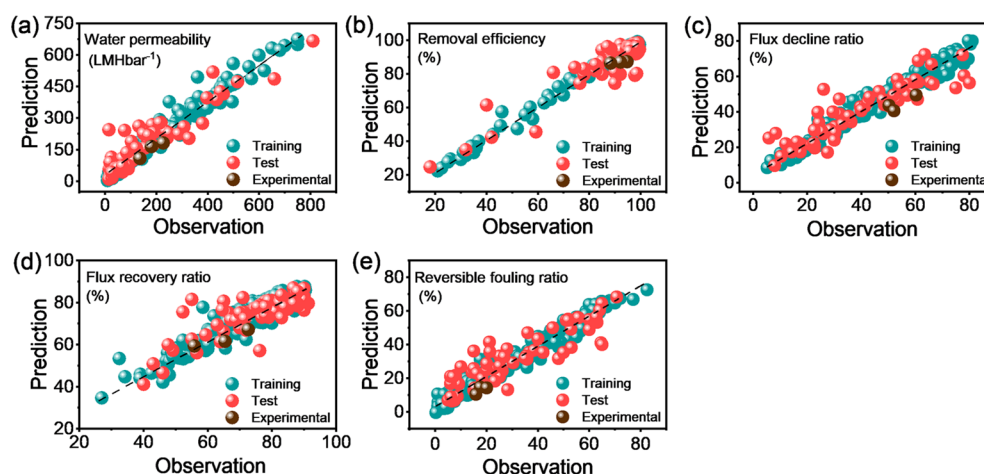
Figure 1. Prediction results of the ML models for each target: (a) water permeability, (b) removal efficiency, (c) flux decline ratio, (d) flux recovery ratio, and (e) reversible fouling ratio.

Table 2. Prediction Performance of the ML Models for Membrane Properties

membrane property	total data size	training set R^2	training set RMSE	test set R^2	test set RMSE
overall porosity (%)	320	0.96	2.51	0.66	7.61
mean pore radius (μm)	320	0.98	3.47	0.87	7.59
contact angle (deg)	320	0.98	1.87	0.76	5.87

input features. In comparison with the prediction performance on mean pore radius and cotact angle, the R^2 value of 0.66 for overall porosity was relatively lower, which could be partially explained by the variations in data quality due to the dry−wet gravimetric measurement method as calculated in eq 5. Notably, the observed membrane properties of our fabricated membranes (as listed in Table S9) were in line with the predicted values as indicated by the blue dots in Figure 2. This experimental validation further confirmed the prediction accuracy of the three predictive ML models on membrane properties.

Feature Analysis using the SHAP Method. According to the built ML models, the contributions of each input feature to the targets (i.e., membrane performance) were evaluated using the SHAP method. A feature's Shapley value quantifies its contribution, whether negative or positive. A feature with a higher absolute Shapley value implies a greater contribution to membrane performance. Figure 3a and Figure S5 illustrate the

importance and impact of each feature on membrane performance. In general, the loading of additives (A_wt %), the polymer content of the total casting solution (P_wt %), the molecular weight of the pore maker (M_Da), and the pore maker content (M_wt %) were found to be the four most influential fabrication parameters for predicting membrane performance. Notably, A_wt % was found to be the most important fabrication parameter for predicting membrane performance indices except for the flux decline ratio, for which A_wt % ranked second in importance. With respect to water permeability as shown in Figure 3a, it is noteworthy that the feature A_5 also played an important role in improving water permeability. As described in the data collection and data set construction section, the category features were first converted into numerical features. Here, A_number stands for the encoder of the category feature (i.e., the type of additive). The SHAP plot demonstrated that A_5 positively contributed to water permeability, which means that the additive having an A_5 value of 1 (such as UiO-66) might be a desirable additive for enhancing water permeability. Because the additive was identified as a significant contributor to membrane performance, to optimize the robustness of the ML models and gain more insights into the effects of the additive, future work may focus on compiling data sets with structural parameters of the additives, especially nanomaterials (such as size, length, shape, and diameter) as well as its chemical properties (e.g., ζ potential, hydrophilicity, and surface functional groups).
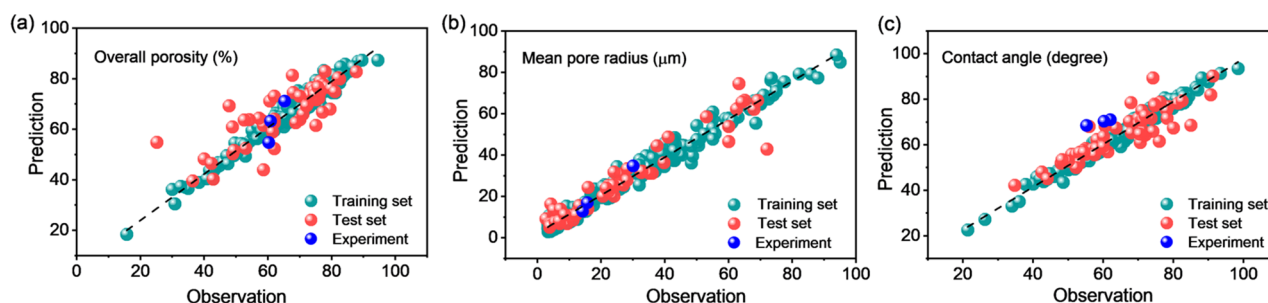

Figure 2. Prediction performance of the ML model for each membrane property: (a) prediction performance with experimental validation of the overall porosity, (b) prediction performance with experimental validation of the mean pore radius, and (c) prediction performance with experimental validation of the contact angle.
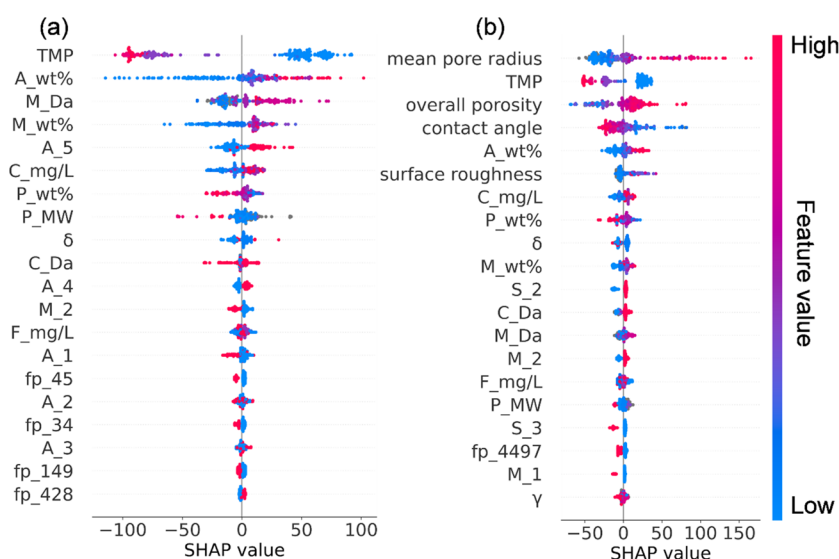
Figure 3. SHAP plot for water permeability based on ML models (a) without membrane properties and (b) with membrane properties. For panel a, the model was developed by using fabrication conditions as input features and water permeability as the target, while for panel b, the model was developed by using fabrication conditions combined with membrane properties as input features and water permeability as the target. A_number (e.g., A_5) denotes the encoder for the category feature (i.e., the type of additive). The feature number (e.g., fp_45) stands for the feature position in the Morgan fingerprint vector.
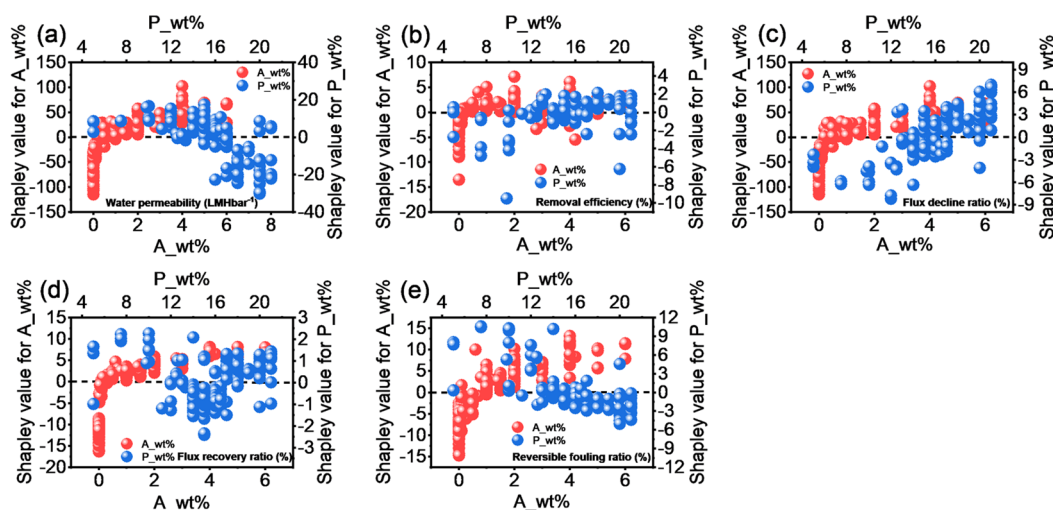

Figure 4. Shapley values of additive loading (A_wt %) and polymer content (P_wt %) for each of the membrane performance indices: (a) water permeability, (b) removal efficiency, (c) flux decline ratio, (d) flux recovery ratio, and (e) reversible fouling ratio.

Additionally, the operational conditions played considerably important roles in membrane performance. In particular, the transmembrane pressure (TMP) was integral to water permeation, while the molecular weight of contaminants (C_Da), concentration of contaminants (C_mg/L), and concentration of foulants (F_mg/L) played crucial roles in removal efficiency- and membrane fouling-related performance. As demonstrated in Figure 3a, TMP was negatively correlated with water permeation. Such a negative effect of applied hydraulic pressure on water flux was mainly ascribed to the compaction of the polymeric membrane.[55,56] The compaction under different applied pressures resulted in a reduction in the membrane pore size and porosity, which agreed well with the contribution of pore size and porosity to water permeability (Figure 3b). C_Da was positively correlated with removal efficiency (Figure S5a). The separation of UF membranes was dictated by the molecular-sieving mechanism

indicated by the molecular weight cutoff (MWCO), where the large solutes were retained by the smaller pores to achieve high removal efficiency.[57] The negative correlation of the mean pore radius with removal efficiency in Figure S6a is consistent with this domain knowledge. These findings verified the necessity of developing an ML model with operational conditions included (in addition to the fabrication parameters) as input features.

As shown in Figure 4, A_wt % was positively correlated with water permeability, removal efficiency, and flux recovery ratio with a loading of >0.5 wt % (defined by the weight percentage of an additive to the base polymer), while the beneficial loading for the flux decline ratio and reversible fouling ratio was found to be >1.0 wt % (Figure 4c,e). These findings revealed that incorporating 1.0 wt % additives (specifically nanomaterials) into a polymer matrix could potentially afford a membrane with high water permeability and removal efficiency, as well as antifouling performance. As the backbone
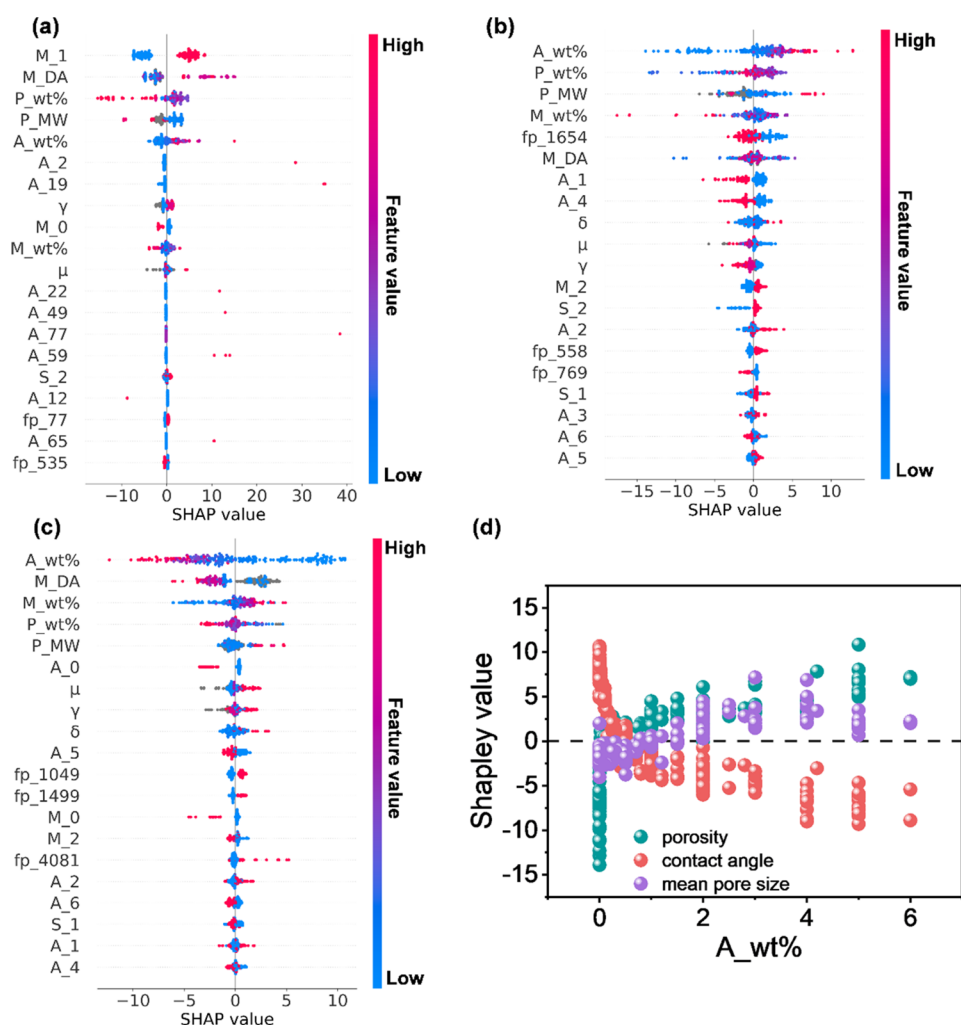
Figure 5. SHAP plot for ML models on membrane properties: (a) mean pore size, (b) overall porosity, (c) contact angle, and (d) Shapley values of $A\_$wt % for each of the membrane properties. $A\_$number (e.g., A_2) denotes the encoder for the category feature (i.e., the type of additive). The feature number (e.g., fp_1654) denotes the feature position in the Morgan fingerprint vector. The chemical structure of fp_1654 is shown in Figure S7.

of a membrane, the polymer used to fabricate the UF membrane was expected to be a crucial feature. As shown in Figure 4, $P\_$wt % had a significant impact on membrane performance, especially on water permeability. With $P\_$wt % ranging from 10 to 16 wt %, it was positively correlated with water permeability. Beyond 16 wt %, a strongly negative correlation with water permeability was observed. It might be attributed to the delayed phase inversion due to the higher polymer content, which normally resulted in less porosity and a small pore sizes in the membrane.[58] This trend was in accordance with the results shown in Figure 3b, wherein the pore size and porosity exhibited a positive correlation with water permeability. Notably, as for removal efficiency, flux recovery ratio, and reversible fouling ratio, the Shapley value of $P\_$wt % was comparable to that of $A\_$wt %, while in terms of water permeability and flux decline ratio, the Shapley value of $A\_$wt % was much larger than that of $P\_$wt %, suggesting that in comparison with tuning the polymer content, tailoring the addition loading of nanomaterials into the polymer matrix might be a more effective method for achieving a UF membrane with desirable membrane performance, especially enhanced water permeability and a decreased flux decline ratio. In practical water and wastewater treatment applications, UF

membranes with a lower flux decline ratio are desired as it indicates that such a membrane holds better antifouling potential.

Figure 5 displays the feature analysis of the ML models for the prediction of membrane properties. The ranking of the features' importance for predicting membrane properties was in accordance with that for membrane performance prediction. In the analogue to the prediction of membrane performance, $A\_$wt %, $P\_$wt %, $M\_$Da, and $M\_$wt % were also found to be the four most significant fabrication factors dominating membrane property prediction. Therefore, the influence of these factors on membrane performance prediction can be explained by their contributions to each of the membrane properties. $A\_$wt % was positively correlated with the mean pore size and overall porosity and negatively correlated with the contact angle at a loading of >1.0 wt % (Figure 5d), which was highly consistent with the beneficial range of $A\_$wt % for membrane performance. This revealed that an increased addition of additives (referring to nanomaterials in the collected data sets) contributed to the formation of a larger pore size, a higher porosity, and a smaller contact angle indicating higher surface hydrophilicity,[59,60] typically leading to

desirable water permeability,[61,62] which agreed well with the results shown in Figure 3b.

It has been reported that irreversible fouling rapidly occurred as a result of internal pore blockage, followed by the formation of a cake layer on the membrane surface.[63] The pore constriction-induced irreversible fouling resulted in the progressive decline of the membrane water flux. Therefore, a smaller pore might make it difficult for foulants to enter and constrict the pores, which in turn could afford a lower flux decline ratio. As shown in Figure S6, the mean pore radius positively correlated with the flux decline ratio, revealing that a smaller mean pore radius contributed to a lower flux decline ratio, which was a desirable performance for separation membranes. $P\_wt \%$, $A\_wt \%$, P_MW, and M_Da were found to be the four most influential factors in predicting the flux decline ratio (Figure S5b). A possible reason could be these four factors showed significant effects on the mean pore size of the membrane as illustrated in Figure 5a. The flux recovery ratio and reversible fouling ratio, which could be promoted through hydraulic cleaning, largely depended on the membrane surface hydrophilicity inferred from the contact angle. The ranking of features for the flux recovery ratio and the reversible fouling ratio indicated that the contact angle was the most important property for both performance indices. The four most important factors governing the prediction of the flux recovery ratio and the reversible fouling ratio were identified as $A\_wt \%$, $P\_wt \%$, M_Da, and $M\_wt \%$, which was consistent with the top four factors contributing to the contact angle.

■ ASSOCIATED CONTENT

Ⓢ Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.est.2c05404.

> Reagents and materials (Text S1), ultrafiltration membrane fabrication (Text S2), input features and target used in the data sets (Table S1), screening candidates for encoder methods (Table S2), percentage of the missing value for each feature (Table S3), predictive performance of different configurations of machine learning algorithms and encoder methods for the water permeability data set (Table S4), range of candidate hyperparameters for each ML algorithm (Table S5), detailed conditions for fabricating and testing UF membranes (Table S6), predictive performance of the ML models with membrane properties included as input features (Table S7), predicted and experimental results of the fabricated membranes (Table S8), predicted and characterized membrane properties of the fabricated membranes (Table S9), data distribution of the numeric input features (Figure S1), types of additives involved in the UF membrane as collected from the literature (Figure S2), Pearson correlation coefficients (PCCs) (Figure S3), plot of training RMSE (RMSEcv-training, Loss) versus validation RMSE (RMSEcv-validation, Train_loss) (Figure S4), SHAP plot for ML models (Figure S5), SHAP plot for ML models accounting for membrane properties (Figure S6), and chemical structure of feature fp_1654 (Figure S7) (PDF)
>
> Supplementary data sets (XLSX)

Accession Codes

Codes related to this paper can be found in Github (https://github.com/Shifa-Zhong/UF).

■ AUTHOR INFORMATION

Corresponding Author

Yongsheng Chen − *School of Civil and Environmental Engineering, Georgia Institute of Technology, Atlanta, Georgia 30332, United States;* ⓞ orcid.org/0000-0002-9519-2302; Phone: +14048943089; Email: yongsheng.chen@ce.gatech.edu

Authors

Haiping Gao − *School of Civil and Environmental Engineering, Georgia Institute of Technology, Atlanta, Georgia 30332, United States; Shandong Provincial Key Laboratory of Water Pollution Control and Resource Reuse, School of Environmental Science and Engineering, Shandong University, Qingdao, Shandong 266237, China;* ⓞ orcid.org/0000-0001-9980-8146

Shifa Zhong − *School of Civil and Environmental Engineering, Georgia Institute of Technology, Atlanta, Georgia 30332, United States; School of Ecological and Environmental Sciences, East China Normal University, Shanghai 200241, China;* ⓞ orcid.org/0000-0002-5822-0837

Raghav Dangayach − *School of Civil and Environmental Engineering, Georgia Institute of Technology, Atlanta, Georgia 30332, United States*

Complete contact information is available at:
https://pubs.acs.org/10.1021/acs.est.2c05404

Author Contributions

§H.G. and S.Z. contributed equally to this work.

Notes

The authors declare no competing financial interest.

■ REFERENCES

(1) Ahmed, F. E.; Khalil, A.; Hilal, N. Emerging desalination technologies: Current status, challenges and future trends. *Desalination* **2021**, *517*, 115183.

(2) Elimelech, M.; Phillip, W. A. The future of seawater desalination: energy, technology, and the environment. *science* **2011**, *333* (6043), 712−717.

(3) Fane, A. G.; Wang, R.; Hu, M. X. Synthetic membranes for water purification: status and future. *Angew. Chem., Int. Ed.* **2015**, *54* (11), 3368−3386.

(4) Norouzbahari, S.; Roostaazad, R.; Hesampour, M. Crude oil desalter effluent treatment by a hybrid UF/RO membrane separation process. *Desalination* **2009**, *238* (1−3), 174−182.

(5) Huang, H.; Young, T. A.; Schwab, K. J.; Jacangelo, J. G. Mechanisms of virus removal from secondary wastewater effluent by low pressure membrane filtration. *J. Membr. Sci.* **2012**, *409*, 1−8.

(6) Lam, B.; Déon, S.; Morin-Crini, N.; Crini, G.; Fievet, P. Polymer-enhanced ultrafiltration for heavy metal removal: Influence

of chitosan and carboxymethyl cellulose on filtration performances. *Journal of cleaner production* **2018**, *171*, 927−933.

(7) Trivunac, K.; Stevanovic, S. Removal of heavy metal ions from water by complexation-assisted ultrafiltration. *Chemosphere* **2006**, *64* (3), 486−491.

(8) Katsou, E.; Malamis, S.; Haralambous, K. J. Industrial wastewater pre-treatment for heavy metal reduction by employing a sorbent-assisted ultrafiltration system. *Chemosphere* **2011**, *82* (4), 557−564.

(9) Qiu, Y.-R.; Mao, L.-J. Removal of heavy metal ions from aqueous solution by ultrafiltration assisted with copolymer of maleic acid and acrylic acid. *Desalination* **2013**, *329*, 78−85.

(10) Ahmad, A.; Puasa, S.; Zulkali, M. Micellar-enhanced ultra-filtration for removal of reactive dyes from an aqueous solution. *Desalination* **2006**, *191* (1−3), 153−161.

(11) Al-Ani, D. M.; Al-Ani, F. H.; Alsalhy, Q. F.; Ibrahim, S. S. Preparation and characterization of ultrafiltration membranes from PPSU-PES polymer blend for dye removal. *Chem. Eng. Commun.* **2021**, *208* (1), 41−59.

(12) Omi, F. R.; Choudhury, M. R.; Anwar, N.; Bakr, A. R.; Rahaman, M. S. Highly conductive ultrafiltration membrane via vacuum filtration assisted layer-by-layer deposition of functionalized carbon nanotubes. *Ind. Eng. Chem. Res.* **2017**, *56* (30), 8474−8484.

(13) Kim, J. F.; Kim, J. H.; Lee, Y. M.; Drioli, E. Thermally induced phase separation and electrospinning methods for emerging membrane applications: A review. *AIChE J.* **2016**, *62* (2), 461−490.

(14) Mondal, S.; Kumar Majumder, S. Fabrication of the polysulfone-based composite ultrafiltration membranes for the adsorptive removal of heavy metal ions from their contaminated aqueous solutions. *Chemical Engineering Journal* **2020**, *401*, 126036.

(15) Yu, T.; Zhou, J.; Liu, F.; Xu, B.-M.; Pan, Y. Recent Progress of Adsorptive Ultrafiltration Membranes in Water Treatment—A Mini Review. *Membranes* **2022**, *12* (5), 519.

(16) Khanzada, N. K.; Farid, M. U.; Kharraz, J. A.; Choi, J.; Tang, C. Y.; Nghiem, L. D.; Jang, A.; An, A. K. Removal of organic micropollutants using advanced membrane-based water and waste-water treatment: A review. *Journal of membrane science* **2020**, *598*, 117672.

(17) Zhu, J.; Zhou, S.; Li, M.; Xue, A.; Zhao, Y.; Peng, W.; Xing, W. PVDF mixed matrix ultrafiltration membrane incorporated with deformed rebar-like Fe3O4-palygorskite nanocomposites to enhance strength and antifouling properties. *J. Membr. Sci.* **2020**, *612*, 118467.

(18) Qadir, D.; Mukhtar, H.; Keong, L. K. Mixed Matrix Membranes for Water Purification Applications. *Separation & Purification Reviews* **2017**, *46* (1), 62−80.

(19) Salehi, E.; Heidary, F.; Daraei, P.; Keyhani, M.; Behjomanesh, M. Carbon nanostructures for advanced nanocomposite mixed matrix membranes: a comprehensive overview. *Reviews in Chemical Engineering* **2020**, *36* (6), 723−748.

(20) Tsuyuhara, T.; Hanamoto, Y.; Miyoshi, T.; Kimura, K.; Watanabe, Y. Influence of membrane properties on physically reversible and irreversible fouling in membrane bioreactors. *Water Sci. Technol.* **2010**, *61* (9), 2235−2240.

(21) Kimura, K.; Hane, Y.; Watanabe, Y.; Amy, G.; Ohkuma, N. Irreversible membrane fouling during ultrafiltration of surface water. *Water research* **2004**, *38* (14−15), 3431−3441.

(22) Zheng, X.; Khan, M. T.; Cao, X.; Croue, J.-P. Importance of origin and characteristics of biopolymers in reversible and irreversible fouling of ultrafiltration membranes. *Science of The Total Environment* **2021**, *784*, 147157.

(23) Shi, X.; Tal, G.; Hankins, N. P.; Gitis, V. Fouling and cleaning of ultrafiltration membranes: A review. *Journal of Water Process Engineering* **2014**, *1*, 121−138.

(24) Huang, H.; Lee, N.; Young, T.; Gary, A.; Lozier, J. C.; Jacangelo, J. G. Natural organic matter fouling of low-pressure, hollow-fiber membranes: Effects of NOM source and hydrodynamic conditions. *Water research* **2007**, *41* (17), 3823−3832.

(25) Peldszus, S.; Hallé, C.; Peiris, R. H.; Hamouda, M.; Jin, X.; Legge, R. L.; Budman, H.; Moresoli, C.; Huck, P. M. Reversible and irreversible low-pressure membrane foulants in drinking water

treatment: Identification by principal component analysis of fluorescence EEM and mitigation by biofiltration pretreatment. *Water research* **2011**, *45* (16), 5161−5170.

(26) Hampu, N.; Werber, J. R.; Chan, W. Y.; Feinberg, E. C.; Hillmyer, M. A. Next-Generation Ultrafiltration Membranes Enabled by Block Polymers. *ACS Nano* **2020**, *14* (12), 16446−16471.

(27) Hoek, E. M.; Ghosh, A. K.; Huang, X.; Liong, M.; Zink, J. I. Physical-chemical properties, separation performance, and fouling resistance of mixed-matrix ultrafiltration membranes. *Desalination* **2011**, *283*, 89−99.

(28) Fane, A.; Fell, C.; Waters, A. The relationship between membrane surface pore characteristics and flux for ultrafiltration membranes. *J. Membr. Sci.* **1981**, *9* (3), 245−262.

(29) Chang, H.; Qu, F.; Liu, B.; Yu, H.; Li, K.; Shao, S.; Li, G.; Liang, H. Hydraulic irreversibility of ultrafiltration membrane fouling by humic acid: Effects of membrane properties and backwash water composition. *J. Membr. Sci.* **2015**, *493*, 723−733.

(30) Tran, T. T. V.; Kumar, S. R.; Lue, S. J. Separation mechanisms of binary dye mixtures using a PVDF ultrafiltration membrane: Donnan effect and intermolecular interaction. *J. Membr. Sci.* **2019**, *575*, 38−49.

(31) Lee, J.; Walker, H. W. Mechanisms and factors influencing the removal of microcystin-LR by ultrafiltration membranes. *J. Membr. Sci.* **2008**, *320* (1−2), 240−247.

(32) Zhang, P.; Rajabzadeh, S.; Istirokhatun, T.; Shen, Q.; Jia, Y.; Yao, X.; Venault, A.; Chang, Y.; Matsuyama, H. A novel method to immobilize zwitterionic copolymers onto PVDF hollow fiber membrane surface to obtain antifouling membranes. *J. Membr. Sci.* **2022**, *656*, 120592.

(33) Rabiee, H.; Farahani, M. H. D. A.; Vatanpour, V. Preparation and characterization of emulsion poly (vinyl chloride)(EPVC)/TiO2 nanocomposite ultrafiltration membrane. *J. Membr. Sci.* **2014**, *472*, 185−193.

(34) De Lannoy, C.; Jassby, D.; Davis, D.; Wiesner, M. A highly electrically conductive polymer-multiwalled carbon nanotube nano-composite membrane. *Journal of membrane science* **2012**, *415*, 718−724.

(35) Shukla, A. K.; Alam, J.; Alhoshan, M.; Dass, L. A.; Muthumareeswaran, M. Development of a nanocomposite ultra-filtration membrane based on polyphenylsulfone blended with graphene oxide. *Sci. Rep.* **2017**, *7* (1), 41976.

(36) Butler, K. T.; Davies, D. W.; Cartwright, H.; Isayev, O.; Walsh, A. Machine learning for molecular and materials science. *Nature* **2018**, *559* (7715), 547−555.

(37) Wei, J.; Chu, X.; Sun, X. Y.; Xu, K.; Deng, H. X.; Chen, J.; Wei, Z.; Lei, M. Machine learning in materials science. *InfoMat* **2019**, *1* (3), 338−358.

(38) Morgan, D.; Jacobs, R. Opportunities and challenges for machine learning in materials science. *Annu. Rev. Mater. Res.* **2020**, *50*, 71−103.

(39) Townsend, J.; Micucci, C. P.; Hymel, J. H.; Maroulas, V.; Vogiatzis, K. D. Representation of molecular structures with persistent homology for machine learning applications in chemistry. *Nat. Commun.* **2020**, *11* (1), 3230.

(40) Shields, B. J.; Stevens, J.; Li, J.; Parasram, M.; Damani, F.; Alvarado, J. I. M.; Janey, J. M.; Adams, R. P.; Doyle, A. G. Bayesian reaction optimization as a tool for chemical synthesis. *Nature* **2021**, *590* (7844), 89−96.

(41) Barnett, J. W.; Bilchak, C. R.; Wang, Y.; Benicewicz, B. C.; Murdock, L. A.; Bereau, T.; Kumar, S. K. Designing exceptional gas-separation polymer membranes using machine learning. *Sci. Adv.* **2020**, *6* (20), No. eaaz4301.

(42) Hasnaoui, H.; Krea, M.; Roizard, D. Neural networks for the prediction of polymer permeability to gases. *J. Membr. Sci.* **2017**, *541*, 541−549.

(43) Bai, X.; Shi, Z.; Xia, H.; Li, S.; Liu, Z.; Liang, H.; Liu, Z.; Wang, B.; Qiao, Z. Machine-Learning-Assisted High-Throughput computa-tional screening of Metal-Organic framework membranes for hydrogen separation. *Chemical Engineering Journal* **2022**, *446*, 136783.

(44) Yeo, C. S. H.; Xie, Q.; Wang, X.; Zhang, S. Understanding and optimization of thin film nanocomposite membranes for reverse osmosis with machine learning. *J. Membr. Sci.* **2020**, *606*, 118135.

(45) Fetanat, M.; Keshtiara, M.; Keyikoglu, R.; Khataee, A.; Daiyan, R.; Razmjou, A. Machine learning for design of thin-film nanocomposite membranes. *Sep. Purif. Technol.* **2021**, *270*, 118383.

(46) Gao, H.; Zhong, S.; Zhang, W.; Igou, T.; Berger, E.; Reid, E.; Zhao, Y.; Lambeth, D.; Gan, L.; Afolabi, M. A.; Tong, Z.; Lan, G.; Chen, Y. Revolutionizing Membrane Design Using Machine Learning-Bayesian Optimization. *Environ. Sci. Technol.* **2022**, *56* (4), 2572−2581.

(47) Hu, J.; Kim, C.; Halasz, P.; Kim, J. F.; Kim, J.; Szekely, G. Artificial intelligence for performance prediction of organic solvent nanofiltration membranes. *J. Membr. Sci.* **2021**, *619*, 118513.

(48) Jeong, N.; Chung, T.-h.; Tong, T. Predicting Micropollutant Removal by Reverse Osmosis and Nanofiltration Membranes: Is Machine Learning Viable? *Environ. Sci. Technol.* **2021**, *55* (16), 11348−11359.

(49) Liu, T.; Liu, L.; Cui, F.; Ding, F.; Zhang, Q.; Li, Y. Predicting the performance of polyvinylidene fluoride, polyethersulfone and polysulfone filtration membranes using machine learning. *Journal of Materials Chemistry A* **2020**, *8* (41), 21862−21871.

(50) Fetanat, M.; Keshtiara, M.; Low, Z.-X.; Keyikoglu, R.; Khataee, A.; Orooji, Y.; Chen, V.; Leslie, G.; Razmjou, A. Machine learning for advanced design of nanocomposite ultrafiltration membranes. *Ind. Eng. Chem. Res.* **2021**, *60* (14), 5236−5250.

(51) Yang, S. D.; Ali, Z. A.; Kwon, H.; Wong, B. M. Predicting Complex Erosion Profiles in Steam Distribution Headers with Convolutional and Recurrent Neural Networks. *Ind. Eng. Chem. Res.* **2022**, *61* (24), 8520−8529.

(52) Zhang, K.; Zhong, S.; Zhang, H. Predicting aqueous adsorption of organic compounds onto biochars, carbon nanotubes, granular activated carbons, and resins with machine learning. *Environ. Sci. Technol.* **2020**, *54* (11), 7008−7018.

(53) Wu, G.; Gan, S.; Cui, L.; Xu, Y. Preparation and characterization of PES/TiO2 composite membranes. *Appl. Surf. Sci.* **2008**, *254* (21), 7080−7086.

(54) Bai, L.; Liu, Y.; Bossa, N.; Ding, A.; Ren, N.; Li, G.; Liang, H.; Wiesner, M. R. Incorporation of Cellulose Nanocrystals (CNCs) into the Polyamide Layer of Thin-Film Composite (TFC) Nanofiltration Membranes for Enhanced Separation Performance and Antifouling Properties. *Environ. Sci. Technol.* **2018**, *52* (19), 11178−11187.

(55) Persson, K. M.; Gekas, V.; Trägårdh, G. Study of membrane compaction and its influence on ultrafiltration water permeability. *J. Membr. Sci.* **1995**, *100* (2), 155−162.

(56) Demirel, E.; Zhang, B.; Papakyriakou, M.; Xia, S.; Chen, Y. Fe2O3 nanocomposite PVC membrane with enhanced properties and separation performance. *J. Membr. Sci.* **2017**, *529*, 170−184.

(57) Mehta, A.; Zydney, A. L. Permeability and selectivity analysis for ultrafiltration membranes. *J. Membr. Sci.* **2005**, *249* (1), 245−249.

(58) Hołda, A. K.; Aernouts, B.; Saeys, W.; Vankelecom, I. F. J. Study of polymer concentration and evaporation time as phase inversion parameters for polysulfone-based SRNF membranes. *J. Membr. Sci.* **2013**, *442*, 196−205.

(59) Shen, J.-n.; Ruan, H.-m.; Wu, L.-g.; Gao, C.-j. Preparation and characterization of PES-SiO2 organic-inorganic composite ultrafiltration membrane for raw water pretreatment. *Chemical engineering journal* **2011**, *168* (3), 1272−1278.

(60) Yan, L.; Li, Y. S.; Xiang, C. B.; Xianda, S. Effect of nano-sized Al2O3-particle addition on PVDF ultrafiltration membrane performance. *J. Membr. Sci.* **2006**, *276* (1−2), 162−167.

(61) Guillen, G. R.; Farrell, T. P.; Kaner, R. B.; Hoek, E. M. V. Pore-structure, hydrophilicity, and particle filtration characteristics of polyaniline-polysulfone ultrafiltration membranes. *J. Mater. Chem.* **2010**, *20* (22), 4621−4628.

(62) Zheng, Q.-Z.; Wang, P.; Yang, Y.-N.; Cui, D.-J. The relationship between porosity and kinetics parameter of membrane formation in PSF ultrafiltration membrane. *J. Membr. Sci.* **2006**, *286* (1−2), 7−11.

(63) Katsoufidou, K.; Yiantsios, S. G.; Karabelas, A. J. Experimental study of ultrafiltration membrane fouling by sodium alginate and flux recovery by backwashing. *J. Membr. Sci.* **2007**, *300* (1), 137−146.