Environmental Science & Technology

pubs.acs.org/est

Article

Revolutionizing Membrane Design Using Machine Learning-Bayesian Optimization

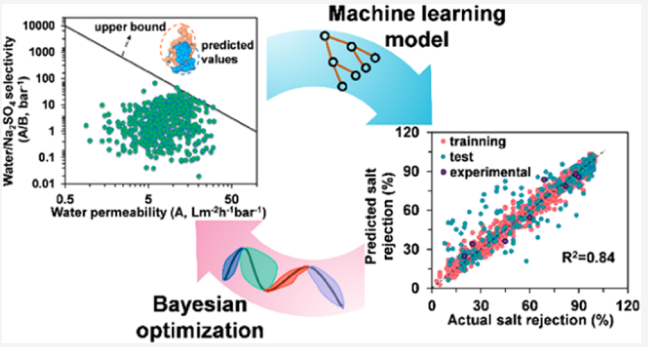
Haiping Gao,[‖] Shifa Zhong,[‖] Wenlong Zhang, Thomas Igou, Eli Berger, Elliot Reid, Yangying Zhao, Dylan Lambeth, Lan Gan, Moyosore A. Afolabi, Zhaohui Tong, Guanghui Lan, and Yongsheng Chen*

Cite This: https://doi.org/10.1021/acs.est.1c04373

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Polymeric membrane design is a multidimensional process involving selection of membrane materials and optimization of fabrication conditions from an infinite candidate space. It is impossible to explore the entire space by trial-and-error experimentation. Here, we present a membrane design strategy utilizing machine learning-based Bayesian optimization to precisely identify the optimal combinations of unexplored monomers and their fabrication conditions from an infinite space. We developed ML models to accurately predict water permeability and salt rejection from membrane monomer types (represented by the Morgan fingerprint) and fabrication conditions. We applied Bayesian optimization on the built ML model to inversely identify sets of monomer/fabrication condition combinations with the potential to break the upper bound for water/salt selectivity and permeability. We fabricated eight membranes under the identified combinations and found that they exceeded the present upper bound. Our findings demonstrate that ML-based Bayesian optimization represents a paradigm shift for next-generation separation membrane design.



KEYWORDS: *membrane design, machine learning, Bayesian optimization, Morgan fingerprint, water/salt selectivity*

1. INTRODUCTION

Polyamide (PA)-based polymeric membranes (e.g., nanofiltration (NF) and reverse osmosis (RO)) are widely used in separation processes including water desalination and purification,[1,2] wastewater reclamation,[3,4] resource recovery,[5,6] and industrial chemical separation.[7,8] PA-based membrane separation performance is typically evaluated by water permeability and water/salt selectivity. However, a critical trade-off exists between water permeability and water/salt selectivity. Namely, increased water permeability produces reduced salt rejection.[9] This trade-off can be defined by the present upper bound correlation,[10] which empirically illustrates the state of the art in separation performance (Figure 1a).

Water permeability and water/salt selectivity are determined by the properties of the PA layer. PA layer formation is tedious and requires a high degree of experimental control (e.g., monomer selection, monomer concentration, polymerization time, heat curing time, incorporation of nanomaterials, addition of additives, and selection of substrate types) (Figure 1b). To disrupt the long-standing trade-off in current PA-based membranes, emphasis is often placed on the incorporation of nanomaterials into the membrane.[11,12] However, most of these efforts have been unsuccessful in breaking the upper bound limitation (Figure 1a) due to the nonoptimal selection of monomer/fabrication combinations. This selection challenge is inherited from the theoretically limitless availability of combinations, which forms a highly dimensional, infinite space, making comprehensive experimental evaluation nearly impossible. Therefore, precise identification of optimal combinations from this infinite space is essential to achieve superior membrane performance. However, achieving an efficient and cost-effective strategy for optimal performance remains an open challenge.

Machine learning algorithms can manage complex, multidimensional datasets with the powerful fitting ability and have garnered increasing attention in the membrane community. Several ML models have been developed and successfully deployed to predict water permeability, salt rejection, and organic solvent separation from fabrication conditions.[13−19] However, previous models were limited to established membrane materials as reported in the literature and were unable to directly predict membrane performance from candidate materials without experimental data. A functional

Received: July 1, 2021
Revised: December 11, 2021
Accepted: December 20, 2021

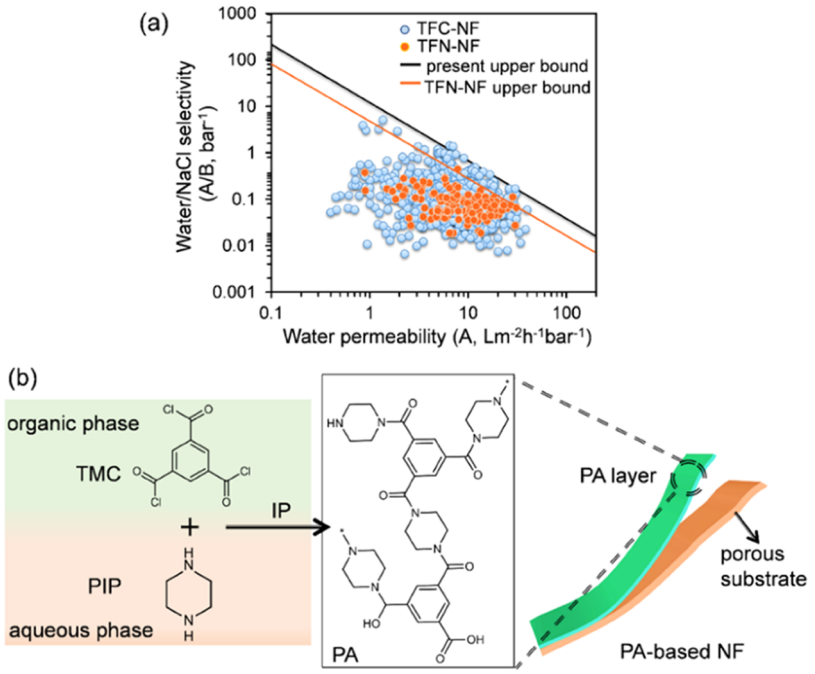


Figure 1. Polyamide (PA)-based thin-film composite (TFC) NF membrane. (a) Upper-bound correlation of selectivity versus permeability for water/NaCl separation. (b) An example of the formation of PA layer with piperazine (PIP) serving as the amine monomer in the aqueous phase, and trimesoyl chloride (TMC) as acyl chloride in the organic phase on a porous substrate. Thin-film nanocomposite (TFN) nanofiltration (NF) is one type of TFC NF fabricated by interfacial polymerization (IP) with the incorporation of nanomaterials into the aqueous or organic phase.

description of materials is critical for the development of ML models with application to unexplored candidates. Molecular fingerprint has demonstrated its value in mapping candidate polymers for the synthesis of gas separation membranes.[20] However, its application to liquid-phase polymeric membrane ML models remains nascent. Further, ML models developed up to date are unable to identify optimal combinations from an infinite space, and their sole function is the prediction of membrane performance from fabrication conditions. Alternatively, Bayesian optimization algorithms minimize the target loss of any objective functions whether discrete or continuous.[21,22] Bayesian optimization has been widely applied in hyperparameter tuning for ML algorithms;[23,24] however, its ability to search an infinite space for high-quality candidates through its optimization procedure has the potential for novel applications. Recently, Bayesian optimization was successfully applied in chemical reaction and synthesis optimization to augment efficiency and chemical yields.[25−27] However, Bayesian optimization has not yet been applied to membrane fabrication optimization.

Herein, we proposed an inverse membrane design strategy that applies Bayesian optimization on a constructed ML model, which can support: (1) the discovery of unexplored monomers and (2) the precise identification of optimal monomer/fabrication condition combinations across an infinite space by understanding relationships between monomer structures, fabrication conditions, and membrane performance. We first developed ML models from literature-based datasets. We then interpreted our model using the Shapley Additive exPlanation (SHAP) method to select monomer atomic groups with positive contributions toward membrane performance. Further, we used SHAP-identified beneficial atomic groups to screen new monomers. Next, we applied Bayesian optimization on the well-developed ML model to inversely identify optimal combinations of monomers and fabrication conditions with the potential to deliver membranes that could break the upper bound of water/salt selectivity and permeability. Finally, we experimentally validated the performance of fabricated membranes subject to the identified combinations.

2. METHODS

2.1. Dataset Construction. To construct the datasets, we mined all of the data for any fabrication conditions that may affect PA-based flat-sheet NF membrane performance from 218 reports published in the last 20 years. The full datasets were provided in the Supporting Data. Fabrication conditions included not only numeric features such as monomer concentration, polymerization time, and heat curing time but also categorical features such as additive type, organic solvent, and substrate membrane as listed in Table S1. However, not all listed features (i.e., fabrication conditions) were reported in each publication, and we left unreported features as missing values. In this way, we can collect as many data points as possible to develop a highly accurate, predictive ML model. Finally, we constructed two datasets containing: (1) water permeability (A) and (2) salt rejection (R). The total number of data points for these two datasets varied due to the data availability reported in the literature, with 567 data points for the A dataset and 1524 data points for the R dataset. These two datasets have the same fabrication conditions—the sole difference was in the R dataset, wherein five key properties of salt ions were added: (1) valence, (2) ionic radius, (3) Stokes radius, (4) hydrated radius, and (5) hydration free energy, which were used to differentiate salts (NaCl, Na_2SO_4, $MgSO_4$, $MgCl_2$, and $CaCl_2$) as listed in Table S2. This treatment enabled a broader collection of data and widened application of the R dataset for a variety of other salt predictions. These two datasets are summarized in Table S3, including the number of data points and the statistical feature information.

https://doi.org/10.1021/acs.est.1c04373
Environ. Sci. Technol. XXXX, XXX, XXX−XXX

2.2. Machine Learning Model Development. One of the key factors determining the successful development of accurate and reliable ML models using our constructed two datasets is how to handle missing values in the input features. Missing values referred to the unreported features (i.e., fabrication conditions) in the literature. While missing values were commonly imputed by the mean or median values in previous works,[19,28] we chose to keep them in their raw format (i.e., true missing values) because all of these features have specific physicochemical meanings and affect the membrane performance. The presence of missing values limits the application of certain ML algorithms such as deep neural network (DNN), making it necessary to employ other ML algorithms, which can process their raw format of missing values. For this work, we utilized two tree-based ML algorithms as candidates: (1) XGBoost (XGBoost Python package) and (2) CatBoost (CatBoost Python package), which are both capable of handling missing values.[29,30] CatBoost treats missing values for a feature as the minimum or maximum values of that feature, while XGBoost allocates the missing values to the side that reduces the loss in each split. Besides, input features also consist of categorical features (e.g., organic solvent type), which are necessary to be encoded as numeric values prior to developing ML. Eight encoding methods were selected as screening candidates rather than relying on arbitrary selection (as described in Table S4).

One challenge for developing ML models lies in the functional description of membrane materials (e.g., aqueous phase amine monomers in this work). Common descriptions include molecular fingerprint,[20] molecular descriptor,[31,32] molecular image,[33] and molecular graph[34] (as briefly described in Text S1 and Figure S1). Molecular images and graphs were not used due to the presence of missing values incompatible with candidate ML algorithms such as the convolutional neural network for molecular images and graph neural networks for molecular graphs. Here, we chose the Morgan fingerprint to represent monomers. Compared with other molecular fingerprints (e.g., atom-pair fingerprint), the Morgan fingerprint is more flexible and capable of accurately representing chemical species due to the tunability of atomic group size.[35]

The Morgan fingerprint decomposes the chemical structure into several atomic groups to produce a binary vector containing 0's and 1's. The position of the 1's in the vector defines specific atomic groups, which exist in the chemicals. An example of how the Morgan fingerprint is used to describe the chemical structure can be found in Text S2 and Figure S2. Atomic group size (i.e., the radius of the Morgan fingerprint) and number of bits in the vector (i.e., the length of the Morgan fingerprint) are freely tunable, guaranteeing flexibility. With increasing radius, more atomic groups (i.e., those containing 1's in vectors) are included, which increases the possibility of different atomic groups overlapping in the vector. Since the real number of candidate monomers is much larger than our dataset space, we set the minimum radius of the Morgan fingerprint to 0 to avoid overlapping cases (as demonstrated in Text S3 and Figure S3). Meanwhile, the length of the Morgan fingerprint was tuned along with hyperparameters of the ML algorithms. In cases where polymers (such as poly(vinyl amine) and poly(amidoamine))[36,37] were used for membrane fabrication, we applied the Morgan fingerprint of the repeating unit for this polymer and the polymer's molecular weight (MW) to represent this polymer. When processing polymers using the Morgan fingerprint method, the number of each atom type in a repeating unit and the chemical connectivity between different units of each polymer were read and then decomposed into a binary fingerprint (i.e., vector) as mentioned above. In cases with two monomers (e.g., monomer A1 and A2 as listed in Table S1), we combined their Morgan fingerprints.

2.3. Model Interpretation. After model development, we apply the SHAP method to calculate the Shapley value for each feature. The SHAP method works by checking the differences in prediction before and after the feature is removed. Feature-to-feature interaction information is also considered by including all possible ways the feature can be removed. The Shapley value for feature x (out of n total features) given the prediction p by the built ML model was calculated as follows[38]

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

where S is the subsets of all features with feature x; $p(S \cup x)$ denotes the prediction by the built ML model considering feature x, and $p(S)$ is the prediction 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 on other features in the ML model. The SHAP method was chosen for ML model interpretations and represents a thorough theoretical demonstration of consistent and unbiased interpretation methods for any ML algorithm.[39,40] 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.

2.4. Virtual Reference Morgan Fingerprint Construction and Monomer Screening. The SHAP interpretation provided important information regarding atomic groups and their effects on membrane performance. For example, an amine group had a positive Shapley value for water permeability, suggesting that its presence improves water permeability. Based on this information, monomers containing the carbonyl group are preferred when a high permeability is desired. With this knowledge, we constructed a reference Morgan fingerprint to record all atomic groups with positive Shapley values (i.e., positive contributions). This reference Morgan fingerprint was used to screen potential monomers. The screening process is to compare similarities between the Morgan fingerprint of each candidate monomer and our constructed reference Morgan fingerprint. Morgan fingerprints of candidate monomers closer to the reference are more likely to contain positive Shapley values, so we encourage their selection. A Morgan fingerprint similarity between candidate monomer and the reference is determined by the Tanimoto coefficient ($S_{A,B}$), which is computed as the number of bits in common divided by the number of total bits as follows[41]

$$S_{A,B} = c/(a + b - c) \tag{2}$$

where α is the number of bits in molecule A, b is the number of bits in molecule B, and c denotes the number of bits that are in both molecules. The Tanimoto coefficient is an intuitive measure of the number of common substructures shared by two molecules. A Tanimoto coefficient of 1 means a completely identical molecule, whereas a value of 0 suggests no similarity between two Morgan fingerprints.

https://doi.org/10.1021/acs.est.1c04373
Environ. Sci. Technol. XXXX, XXX, XXX−XXX

2.5. Membrane Fabrication and Performance Evaluation. Two new amine materials (not included in training datasets) (i.e., polyethylenimine and 1,2-diaminopropane) and one commonly used monomer (i.e., PIP) were used to fabricate PA-based NF membranes by IP to verify ML model predictions on water permeability and salt rejection toward NaCl, Na_2SO_4, $MgSO_4$, $MgCl_2$, $CaCl_2$, LiCl, and $NaNO_3$ (as summarized in Table S5). Commercial polyethersulfone microfiltration membranes (PES, MF, 0.2 μm) were chosen as the porous substrates and were immersed in deionized water before use. In brief, a PES substrate was first secured on a glass plate with a funnel and then impregnated with an aqueous monomer solution at a certain concentration for several minutes. The solution was drained, and excess solution was removed from the substrate surface using a rubber roller. Subsequently, TMC dissolved in anhydrous *n*-hexane was poured onto the impregnated PES membrane surface for 30 s or 60 s, resulting in the formation of a PA active layer on the substrate. The resultant PA-based NF membrane was rinsed with n-hexane to remove unreacted TMC, then cured at an elevated temperature (60 °C) in an oven to enhance the cross-linking degree of the PA layer. The heat-cured PA-based NF membranes were then stored in water at 4 °C before testing. Using the same procedure, eight PA-based NF membranes were fabricated according to the optimized fabrication combinations identified by Bayesian optimization (as listed in Tables S6 and S7).

Water permeability and salt rejection toward different salts for the as-prepared membranes were measured using a crossflow testing cell with an effective testing area of 4.1 cm^2 under a crossflow velocity of 0.5 m s^{-1}. The salt concentration was determined by a conductivity meter (Thermo Scientific). Water permeability (A), salt rejection (R), and water/salt selectivity (A/B) were calculated from the following equations

$$A = \frac{J_w}{\Delta P} \tag{3}$$

$$R = \frac{C_f - C_p}{C_f} \tag{4}$$

$$\frac{A}{B} = \frac{R}{(1 - R) \times (\Delta P - \Delta\pi)} \tag{5}$$

where J_w is the water flux, ΔP is the applied hydraulic pressure, C_f and C_p are the solute concentrations of the feed and permeate solutions, respectively, B denotes the salt permeability coefficient, and $\Delta\pi$ is the osmotic pressure difference of a specific salt. In this work, the equations for the calculation of A and A/B were simplified by assuming that the concentration polarization coefficient (f_{cp}) is equivalent to 1.[10]

3. RESULTS AND DISCUSSION

3.1. Model Performance. For the development of the ML model, we randomly divided each dataset into a training set (80% of the data points) and a test set (20% of the data points). Numeric features were converted into the same range or distribution through feature scaling. Candidate scaling methods are available in Table S4. This conversion may not be necessary for the tree-based ML algorithms used here (XGBoost and CatBoost), but it may modify predictions of the test set. To screen optimal configuration of ML algorithms, encoding methods and feature scaling methods, we applied five cross-validations on the training datasets to evaluate each configuration. Table S8 lists the average predictive performance of the top five configurations on the A and R datasets. To prevent data leakage, the scaling and encoding methods were only trained on the subtraining dataset rather than the entire training dataset and were then applied on the subvalidation dataset during the cross-validation. The configuration with the best predictive performance was chosen as the final optimal configuration. Under optimal configuration, predictive performance on the training dataset was far superior to that of the validation dataset for both A and R datasets (i.e., overfitting problem). Overfitting was alleviated by tuning hyperparameters of the corresponding ML algorithms through Bayesian optimization. We first defined a space containing any possible hyperparameter values. Then, we set the average root-mean-square error (RMSE) on the validation datasets as the target loss. The Bayesian optimization algorithms gradually chose a set of optimal hyperparameters, which minimized the loss. The ML algorithms were then retrained on the entire training dataset using optimum hyperparameters to obtain the final models.

The generalization ability of these ML models was evaluated by unseen test datasets, which were not used in model development. As described in Table 1, our trained ML models achieved predictions on the test datasets with a coefficient of determination (R^2) value of 0.78 for water permeability. The R^2 value was improved to 0.84 for salt rejection as the larger size of the R dataset than the A dataset.

Table 1. Evaluation of Model Performance

objective	training size	training R^2	training RMSE[a]	test size	test R^2	test RMSE
water permeability	567	0.96	1.17	141	0.78	3.01
salt rejection	1524	0.98	4.17	381	0.84	11.74

[a]RMSE has the same unit as its corresponding target, that is, water permeability (LMH bar^{-1}) and salt rejection (%) in this work.

To further evaluate the predictive accuracy of the built models using the test datasets, we also performed experimental validation by testing 10 fabricated NF membranes (named from 1 to 10). In addition to the commonly used monomer, PIP, other two new amine materials, polyethylenimine (PEI, MW1300) and 1,2-diaminopropane (MW 74.125), which were not present in the training datasets, were also used to fabricate PA-based NF membranes. We extended the salt rejection tests in terms of more diverse salt types (i.e., $MgSO_4$, $MgCl_2$, $CaCl_2$, $NaNO_3$, and LiCl). Membrane materials used for the fabrication and detailed fabrication conditions are summarized in Table S5. We found that experimental results for both permeability and salt rejection showed relatively good agreement with the predicted values, supporting the reliability of our built models (Figure 2). Our results suggested that the application of the molecular fingerprint method allowed us to explore new monomers not previously studied in the literature. On the other hand, the good agreement on the wide range of salt types tested here proved that it is reasonable to describe salts using the key characteristics of salt ions.

3.2. Interpretation of the ML Model. Following validation of the ML model for these two datasets, we then interpreted the model to understand the mechanism of prediction from the monomer structures and fabrication

https://doi.org/10.1021/acs.est.1c04373
Environ. Sci. Technol. XXXX, XXX, XXX−XXX

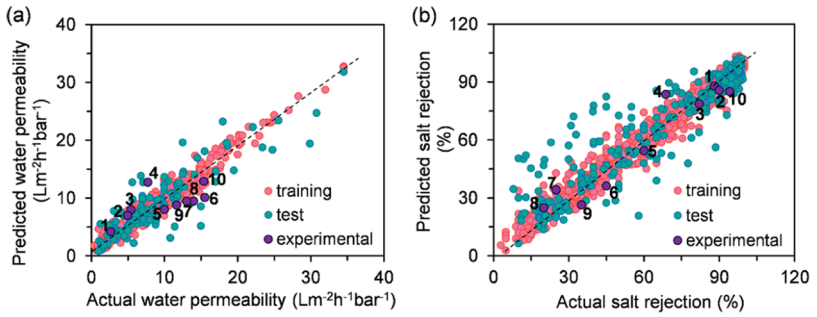


Figure 2. Correlation of experimental results with predicted values. (a) Water permeability dataset and (b) salt rejection dataset. Na_2SO_4, $MgSO_4$, $MgCl_2$, $CaCl_2$, LiCl, and $NaNO_3$ were used for salt rejection tests. The fabricated membranes were named from 1 to 10. Materials used and the corresponding fabrication conditions are summarized in Table S5.

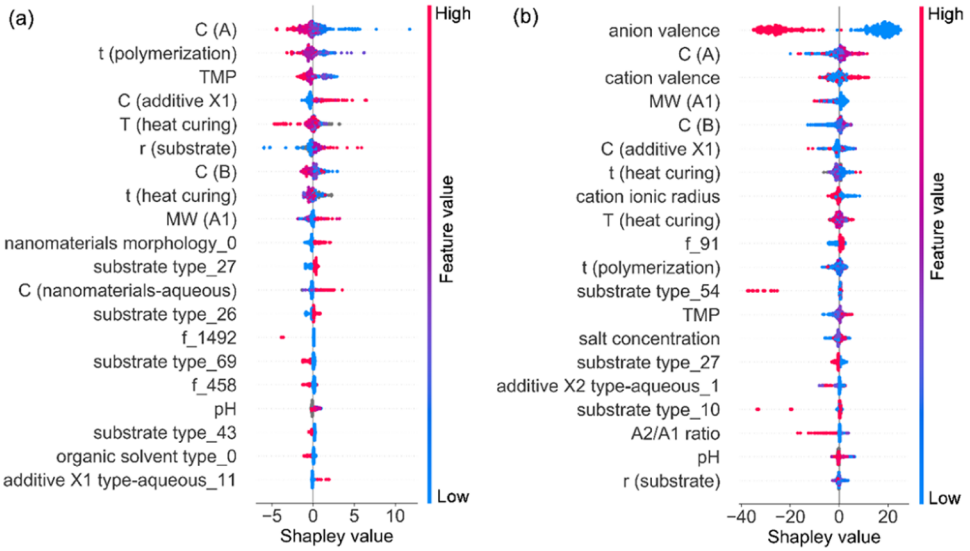


Figure 3. SHAP plot used to interpret the models. (a) Contribution of fabrication conditions to water permeability in the training dataset. (b) Contribution of fabrication conditions to salt rejection in the training dataset. The *X*-axis has the Shapley values, where positive values indicate that the water permeability and the salt rejection can be increased by the specific features listed in both (a) and (b), while negative values indicate a resultant reduction in water permeability and salt rejection. The size of each feature's value is colored from blue to red, corresponding to the smallest and largest values. The pattern for each feature is composed of small dots, and each dot represents one sample containing this feature.

conditions. This understanding is essential in the evaluation of ML model predictions while ensuring consistency with fundamental domain knowledge and experimental experience. It also offers insights into desirable atomic groups for improved salt rejection and water permeability.

Figure 3 summarizes Shapley values in terms of water permeability and salt rejection. The most critical contributors for both predicted water permeability and salt rejection were the aqueous-phase monomer concentration (*C* (A)), aqueous-phase additive concentration (*C* (additive X1)), heat curing temperature (*T* (heat curing)), organic phase monomer concentration (*C* (B)), heat curing time (*t* (heat curing)), and polymerization time (*t* (polymerization)). These results agree well with the widely recognized knowledge of PA-based NF membranes fabricated by IP. The performance of PA-based NF membranes depends significantly on membrane properties (e.g., effective membrane thickness, membrane surface charge density, and membrane pore size) as explained by the Donnan-steric pore model with dielectric exclusion (DSPM-DE).[42] Typically, these properties are tuned by altering fabrication conditions consistent with those identified by Shapley values. From the perspective of positive or negative contributions, a high aqueous-phase monomer concentration forms a thicker PA layer with a greater extent of cross-linking, resulting in reduced water permeability and increased salt rejection. In terms of salt rejection, anion/cation valence plays a crucial role, which is reasonable when considering the mechanisms of ion transport. The transport of charged solutes through NF membranes is largely governed by electrostatic effects as most PA-NF membranes carry a surface charge.[43] A fixed-charge membrane surface repels co-ions while attracting counterions. The association of Shapley values with underlying NF separation mechanisms enhances the reliability of our built models.

3.3. Virtual Reference Morgan Fingerprint and Monomer Selection. Based on the Shapley value of each Morgan fingerprint (as shown in Figure S4), all of the atomic groups with positive contributions to the desirable water permeability and salt rejection are demonstrated in Figure 4. We constructed two reference Morgan fingerprints by integrating all of them to screen unexplored monomers with the potential to achieve target membrane performance. We obtained 310 candidates of new amine monomers from the National Institute for Materials Science (NIMS) materials database[44] to screen; these monomers have been synthesized to date, but they have not been experimentally tested in the

https://doi.org/10.1021/acs.est.1c04373
Environ. Sci. Technol. XXXX, XXX, XXX−XXX

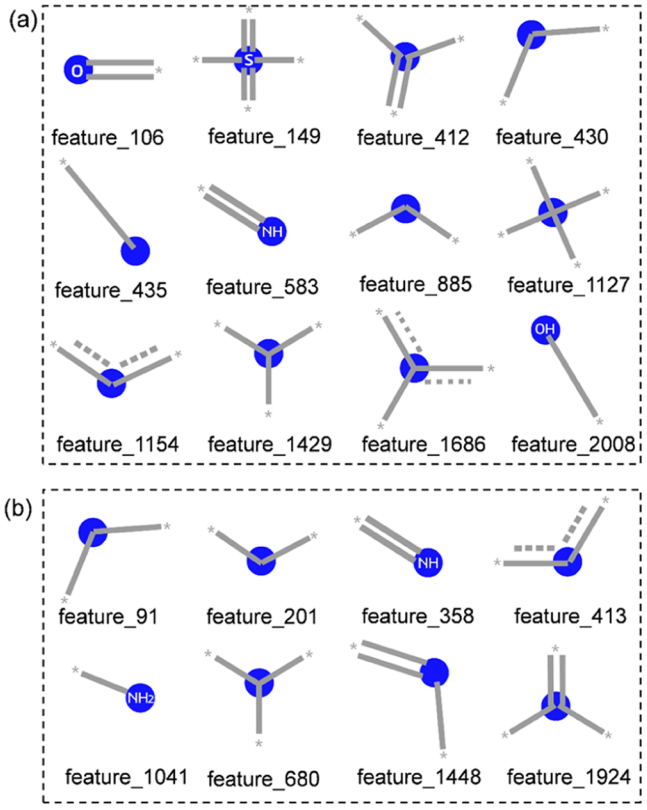


Figure 4. Atomic groups serving as positive contributors. (a) Water permeability and (b) salt rejection. The unlabeled blue dots represent the carbon atoms. Feature number denotes the feature position in the Morgan fingerprint vector. Atoms are colored by blue dots. The gray lines represent the bonds that are not included in the features.

context of membrane separation. The similarity between the Morgan fingerprint of each new monomer with the two constructed references was then calculated based on eq 2. Accordingly, we identified 20 new monomer candidates with Morgan fingerprint similarities to both reference Morgan fingerprints to execute Bayesian optimization. The 20 monomers were classified into two groups. Group 1 consists of 10 monomers with high Morgan fingerprint similarity. Group 2 is composed of 10 commercially available monomers (Figure S5) with lower Morgan fingerprint similarity than that of monomer in Group 1. These monomers contain chemical structures with more positive Shapley value attributes, so it is intuitive that they will exhibit strong performance.

As shown in Figure 4a, we found that the atomic groups, feature_106, feature_149, feature_2008, and feature_583, were linked to membrane water permeability. This implies that the presence of hydrophilic functional groups such as carboxyl groups (referring to feature_106), sulfonate groups (referring to feature_149), and hydroxyl groups (referring to feature_2008) largely contributed to water permeability. And the amine groups (referring to feature_583) offered the possibility to get involved in an interfacial polymerization process. Thus, it was observed that a monomer (i.e., 2,5-diaminopentanoic acid shown in Figure S5) containing a carboxyl group and an amine group was chosen due to its relatively high similarity to the reference Morgan fingerprint. Similarly, as illustrated in Figure 4b, amine groups (referring to feature_358 and feature_1041) were found to be beneficial for salt rejection. Monomers consisting of several branched amine groups such as tris(2-aminoethyl)amine and poly(ethylenimine) were selected and held the potential to form a highly cross-linked PA network, which is desirable for enhanced salt rejection.

3.4. Bayesian Optimization for the Identification of Optimal Combinations and Experimental Validation.

Bayesian optimization offers the opportunity to identify a set of optimal combinations of monomers and fabrication conditions that enable the fabrication of membranes with upper bound-breaking performance. To execute Bayesian optimization, we began by defining the combination space and initializing reasonable ranges for different fabrication conditions. It is worth noting that our custom objective function (eq 6) is calculated from our ML model and is not necessarily continuous over the feasible range of fabrication conditions. This is the reason for adopting Bayesian optimization, which can process discrete objective functions, rather than other large-scale continuous optimization techniques, tackling problems of higher dimensionality but requiring continuity of the objective function.[45]

$$\text{loss} = |A/B - y_i| + |A - x_i| \tag{6}$$

where A is the water permeability and A/B is the water/salt selectivity. By optimizing eq 6, we can obtain multiple combinations of monomers and fabrication conditions that can deliver membranes with A and A/B close to (x_i, y_i), an arbitrarily selected point with a water permeability larger than 10 above the upper bound. By changing this point to any other upper bound-breaking point, the optimizer can provide a series of fabrication conditions delivering membranes with corresponding performance.

All 20 selected monomers lie significantly above the upper bound of water/Na_2SO_4 selectivity (Figure 5a,b), although they remain just above the upper bound of water/NaCl selectivity (Figure 5c,d). As listed in Tables S6 and S7, most of the fabrication conditions provided by Bayesian optimization used a mixture of two amine monomers. It has been reported that the cross-linking reaction between amine and acryl chloride might be retarded because of the competing effect between two amine monomers, which could be controlled by the concentration and ratio of the two amine monomers.[46] Under optimal conditions, the relatively less dense PA active layer induced by the retarded cross-linking might offer increased water permeability while maintaining high salt rejection.[46−48] Moreover, for membranes in terms of Na_2SO_4 rejection, sodium dodecyl sulfate (SDS) was used as an additive (as listed in Table S6). The addition of SDS in aqueous phase has been proved to form uniform pore size distribution with high salt rejection, especially toward multivalent ions.[49] These beneficial conditions synergistically contributed to the upper-bound-breaking performance shown in Figure 5. However, it is still noteworthy that the target membrane in this work is a polyamide-based NF membrane fabricated through interfacial polymerization. As is well known, the IP process involving the reaction between aromatic amine monomer (such as MPD) and acryl chloride (such as TMC) forms a highly cross-linked polyamide active layer giving NaCl rejection of higher than 90%.[50] The aqueous phase monomers used in this work were semiaromatic or aliphatic amine monomers less reactive than the aromatic amine monomers resulted in the formation of a less dense PA layer.[51] Although the combinations of amine monomer and fabrication conditions could be significantly optimized using Bayesian

https://doi.org/10.1021/acs.est.1c04373
Environ. Sci. Technol. XXXX, XXX, XXX−XXX

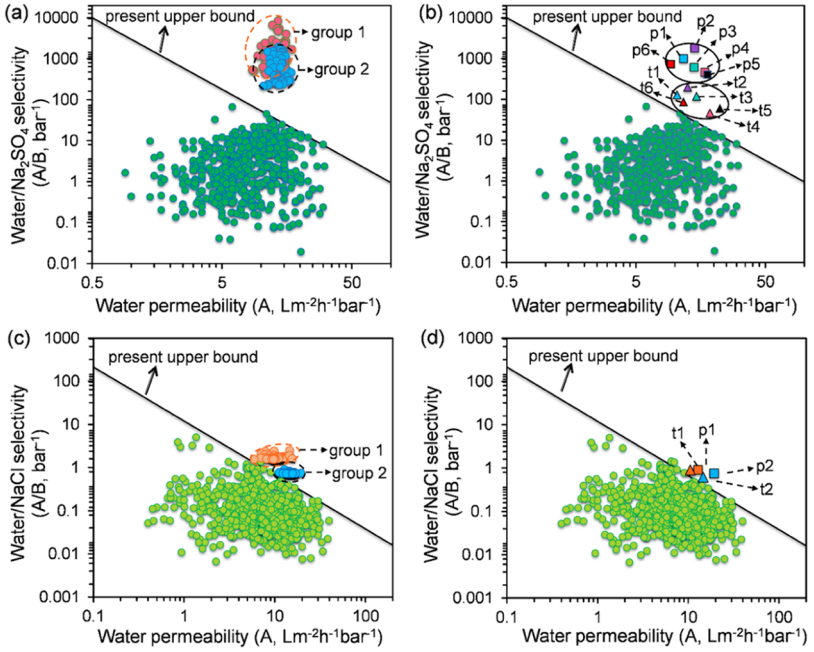


Figure 5. Identification of optimal combinations from Bayesian optimization. (a) Predicted results with Group 1 and Group 2 monomers for water/Na_2SO_4 selectivity versus water permeability. (b) Predicted values from identified combinations and the corresponding experimental performance. (c) Predicted results with Group 1 and Group 2 monomers for water/Na_2SO_4 selectivity versus water permeability. (d) Predicted values from identified combinations and corresponding experimental performance. The axes are logarithmic base10. Predicted values are marked with squares and experimental results are denoted with triangles. The sets of fabrication conditions were differentiated by color. Predicted values and their corresponding experimental results were illustrated in the same color. For Na_2SO_4 rejection, tests are numbered from 1 to 6, and for NaCl rejection, tests are numbered between 1 and 2. Likewise, each prediction # corresponds to the same test number # when comparing membrane performance; p stands for prediction, and t stands for test.

optimization, the obtained membranes were still in the NF membrane region with NaCl rejection <90%. In this work, when executing Bayesian optimization, the water permeability was targeted at 10 LMH bar^{-1} or higher, which is higher than that could be achieved by membranes in the RO region.[10] Under such conditions, the overall permeability-NaCl selectivity of the membranes provided by the Bayesian optimization was limited to reside just above the upper bound.

To validate the identified combinations, we fabricated eight PA-based NF membranes with new monomers from group 2 according to the top −10 identified combinations. The monomers used and their detailed fabrication conditions are summarized in Tables S6 and S7. We tested the water permeability of these membranes and their salt rejection in terms of NaCl and Na_2SO_4 to verify the optimization results. Experimental results were transformed to the corresponding water/salt selectivity and plotted with their predicted values (as shown in Figure 5). For both NaCl and Na_2SO_4, we observed discrepancies between experimental data and predicted values. Bayesian optimization was performed based on the built ML models, which were developed using literature-based data for polyamide NF membranes fabricated interfacial polymerization. For training ML models, a wide range of input features (i.e., fabrication conditions) such as monomer concentration, polymerization time, organic type, additives, nanomaterials, etc. were included. However, the candidate fabrication conditions used for Bayesian optimization were restricted to a smaller space than those in the training dataset. We mainly focused our attention on the most influential conditions to both water permeability and salt rejection as highlighted by the Shapley values (as shown in Figure 3), referring to the concentration of aqueous phase monomer (C(A)), the concentration of additive concentration (C(additive X1)), the concentration of organic phase monomer (C(B)), polymerization time (t), heat curing time, and temperature. Therefore, the model may show a weaker predictive performance in this smaller space. In addition, when developing ML models, concentration polarization was not included as an input feature. However, the influence of concentration polarization on salt rejection might be another reason for the discrepancy.

As revealed by the Shapley values in Figure 3, the incorporated nanomaterials also made positive contributions to improve water permeability. Thus, the candidate input features used in the Bayesian optimization were extended to include the nanomaterial type in aqueous phase, the loading of nanomaterial used, and nanomaterial morphology in addition to the most influential fabrication conditions mentioned above. Cellulose nanocrystal (CNC) was selected as the nanomaterial used to fabricate NF membranes. Based on the monomers and fabrication conditions provided by Bayesian optimization (as listed in Table S6), two TFN membranes were fabricated. As shown in Figure 5b, overall, the TFN membranes referring to membranes 5 and 6 did not show excellent performance compared to the TFC membranes referring to membranes 1−4. According to the Shapley values, although the addition of nanomaterials played roles in enhancing water permeability, the contributions of nanomaterials to salt rejection were not as important as other fabrication conditions (e.g., monomer concentrations, polymerization time, etc.), which might lead to the fair permeability-selectivity performance of TFN membranes. On the other hand, differing from the monomers which could be new ones selected from a large material database, the selection of nanomaterials for the Bayesian optimization was

https://doi.org/10.1021/acs.est.1c04373
Environ. Sci. Technol. XXXX, XXX, XXX−XXX

limited to those collected in the constructed datasets (as found in the Supporting Data). The addition of nanomaterials selected from the datasets such as the CNC used here did not always guarantee better performance than membranes without nanomaterials, which aligned with the statistical trend shown in Figure 1a.

3.5. Implications of ML-Based Bayesian Optimization Strategy. The ML model-based Bayesian optimization demonstrated here is an effective and efficient strategy to inversely supervise membrane design, which is free of the current trial-and-error approach. We can construct a reference Morgan fingerprint based on the chosen atomic groups derived from ML model interpretation, allowing rapid screening of unexplored materials. Hereafter, Bayesian optimization on the well-developed ML model is a useful and innovative tool to explore numerous space possibilities for more efficient membrane design. By identifying the optimal combinations of membrane materials and fabrication conditions, Bayesian optimization allows us to fabricate PA-based NF membranes with upper-bound-breaking performance. These NF membranes can be utilized for efficient water purification and water softening especially for multivalent ions removal. These membranes are supposed to have the potential to achieve high emerging contaminants removal since the size of some contaminants (such as antibiotic, sulfamethoxazole) is larger than SO_4^{2-}. Research interest is increasing in emerging applications of polymeric separation membranes (e.g., solute−solute selectivity from multicomponent systems, emerging contaminants removal, recovering nutrients and valuable metals from wastewater waters). Membranes fabricated based on the current optimized combinations might not be able to achieve satisfactory solute−solute selectivity since we only focused on single-salt systems in this work due to data sparsity in mixed-salt or multicomponent (e.g., the presence of salts and emerging contaminants) systems. Additionally, in this work, we mainly paid attention to the flat-sheet membranes. Therefore, the models developed here could not be applied to hollow fiber membranes directly. However, since the prediction performance of the ML models strongly depends on the availability, accuracy, and size of a dataset, with more studies related to these applications being published, the strategy demonstrated in this contribution can be easily extended to develop appropriate models and guide in designing different types of membranes for those emerging applications.

Our ML algorithm and the subsequent Bayesian optimization mainly focus on testing and optimization of commercially available monomers. Most monomers are derived from petroleum-based raw materials, which, upon disposal, contribute to the omnipresent issue of inert plastic waste. Our developed method will enable researchers to map the chemistries and structures of unexplored membrane materials, enabling the identification and selection of high-performance environmentally friendly, sustainable membrane materials derived from bio-based raw materials. More importantly, our work provided a data-driven computational framework for the development of membranes and would be useful across the membrane field, which could potentially make a massive difference in how fast membranes are tailored for one or another application. By extension, this strategy could also serve as a roadmap for the development of materials for environmental remediation in other technologies (e.g., adsorbers, catalysts, etc.).

Our built framework was designed to extract new, high-performance membrane materials from a large set of unexplored membrane materials and to facilitate the efficient design of separation membranes. We did not emphasize the underlying mechanisms of water and solute transport through membranes relating to membrane properties (e.g., membrane pore size, surface charge density, and hydrophilicity) as reported data do not meet the availability threshold to support ML algorithms. Molecular dynamics (MD) has seen growing research interest in the exploration of solute transport within membrane structures at the atomic level.[49,52,53] Constructing membrane configurations and MD-based simulations could augment currently limited membrane property data availability. Elucidation of the underlying molecular mechanisms for water and solute transport through separation membranes by a synergistic MD−ML approach remains an open challenge for future research.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.est.1c04373.

Four descriptors to represent chemicals (Text S1); brief introduction of Morgan fingerprint (Text S2); brief introduction of molecular fingerprint overlapping (Text S3); four representations for piperazine (PIP) (Figure S1); decomposition of the Morgan fingerprint of PIP to their representative atoms or groups (Figure S2); example of fingerprint overlapping (Figure S3); SHAP plot of features in the Morgan fingerprint (Figure S4); Group 2 monomers used to perform Bayesian fabrication optimization (Figure S5); description of input features for model development (Table S1); description of key characteristics for salt ions (Table S2); statistical information of the numerical features as model inputs (Table S3); screening candidates for encoder and scaler methods (Table S4); materials and fabrication conditions used for membrane fabrication to validate prediction accuracy of the built ML models (Table S5); monomers and fabrication conditions used to fabricate membranes for validation of Bayesian optimization on water permeability and water/Na_2SO_4 selectivity (Table S6); monomers and fabrication conditions used to fabricate membranes for validation of Bayesian optimization on water permeability and water/NaCl selectivity (Table S7); and predictive performance of different configurations of machine learning algorithms, encoder methods, and scaler methods (Table S8) (PDF)

Datasets (XLSX)

■ 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: 4048943089; Email: yongsheng.chen@ce.gatech.edu

Authors

Haiping Gao − *School of Civil and Environmental Engineering, Georgia Institute of Technology, Atlanta,*

https://doi.org/10.1021/acs.est.1c04373
Environ. Sci. Technol. XXXX, XXX, XXX−XXX

Georgia 30332, United States; orcid.org/0000-0001-9980-8146
Shifa Zhong − *School of Civil and Environmental Engineering, Georgia Institute of Technology, Atlanta, Georgia 30332, United States;* orcid.org/0000-0002-5822-0837
Wenlong Zhang − *School of Civil and Environmental Engineering, Georgia Institute of Technology, Atlanta, Georgia 30332, United States*
Thomas Igou − *School of Civil and Environmental Engineering, Georgia Institute of Technology, Atlanta, Georgia 30332, United States*
Eli Berger − *School of Civil and Environmental Engineering, Georgia Institute of Technology, Atlanta, Georgia 30332, United States*
Elliot Reid − *School of Civil and Environmental Engineering, Georgia Institute of Technology, Atlanta, Georgia 30332, United States*
Yangying Zhao − *School of Civil and Environmental Engineering, Georgia Institute of Technology, Atlanta, Georgia 30332, United States;* orcid.org/0000-0002-2486-4845
Dylan Lambeth − *School of Civil and Environmental Engineering, Georgia Institute of Technology, Atlanta, Georgia 30332, United States*
Lan Gan − *School of Civil and Environmental Engineering, Georgia Institute of Technology, Atlanta, Georgia 30332, United States*
Moyosore A. Afolabi − *School of Civil and Environmental Engineering, Georgia Institute of Technology, Atlanta, Georgia 30332, United States*
Zhaohui Tong − *School of Chemical and Biomolecular Engineering, Georgia Institute of Technology, Atlanta, Georgia 30332, United States;* orcid.org/0000-0002-0847-5612
Guanghui Lan − *H. Milton Stewart School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, Georgia 30332, United States*

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.est.1c04373

Author Contributions

[‖]H.G. and S.Z. contributed equally to this work.

Notes

The authors declare no competing financial interest.
The full datasets collected from the literature can be found in the Supporting Data. All of the codes written for this work will be available at GitHub (https://github.com/nogoodnameye/ML-guided-polymer-synthesis).

■ ACKNOWLEDGMENTS

This work was performed in part at the Georgia Tech Institute for Electronics and Nanotechnology, a member of the National Nanotechnology Coordinated Infrastructure (NNCI), which is supported by the National Science Foundation (Award No. ECCS-2025462); U.S. Department of Agriculture (Award No. 2018-68011-28371); National Science Foundation (Award No. 1936928); National Science Foundation-U.S. Department of Agriculture (Award No. 2020-67021-31526); National Science Foundation (Award No. 2112533); and U.S. Environmental Protection Agency (Award No. 840080010).

■ REFERENCES

(1) Elimelech, M.; Phillip, W. A. The Future of Seawater Desalination: Energy, Technology, and the Environment. *Science* **2011**, *333*, 712−717.
(2) Fane, A. G.; Wang, R.; Hu, M. X. Synthetic membranes for water purification: status and future. *Angew. Chem., Int. Ed.* **2015**, *54*, 3368−3386.
(3) Wilf, M.; Alt, S. Application of low fouling RO membrane elements for reclamation of municipal wastewater. *Desalination* **2000**, *132*, 11−19.
(4) Bes-Piá, A.; Cuartas-Uribe, B.; Mendoza-Roca, J.; Galiana-Aleixandre, M.; Iborra-Clar, M.; Alcaina-Miranda, M. Pickling wastewater reclamation by means of nanofiltration. *Desalination* **2008**, *221*, 225−233.
(5) Zhao, Y.; Tong, T.; Wang, X.; Lin, S.; Reid, E. M.; Chen, Y. Differentiating Solutes with Precise Nanofiltration for Next Generation Environmental Separations: A Review. *Environ. Sci. Technol.* **2021**, *55*, 1359−1376.
(6) Blöcher, C.; Niewersch, C.; Melin, T. Phosphorus recovery from sewage sludge with a hybrid process of low pressure wet oxidation and nanofiltration. *Water Res.* **2012**, *46*, 2009−2019.
(7) Gin, D. L.; Noble, R. D. Designing the next generation of chemical separation membranes. *Science* **2011**, *332*, 674−676.
(8) Marchetti, P.; Jimenez Solomon, M. F.; Szekely, G.; Livingston, A. G. Molecular separation with organic solvent nanofiltration: a critical review. *Chem. Rev.* **2014**, *114*, 10735−10806.
(9) Geise, G. M.; Park, H. B.; Sagle, A. C.; Freeman, B. D.; McGrath, J. E. Water permeability and water/salt selectivity tradeoff in polymers for desalination. *J. Membr. Sci.* **2011**, *369*, 130−138.
(10) Yang, Z.; Guo, H.; Tang, C. Y. The upper bound of thin-film composite (TFC) polyamide membranes for desalination. *J. Membr. Sci.* **2019**, *590*, No. 117297.
(11) Lau, W.; Gray, S.; Matsuura, T.; Emadzadeh, D.; Chen, J. P.; Ismail, A. A review on polyamide thin film nanocomposite (TFN) membranes: History, applications, challenges and approaches. *Water Res.* **2015**, *80*, 306−324.
(12) Yang, Z.; Sun, P.-F.; Li, X.; Gan, B.; Wang, L.; Song, X.; Park, H.-D.; Tang, C. Y. A Critical Review on Thin-Film Nanocomposite Membranes with Interlayered Structure: Mechanisms, Recent Developments, and Environmental Applications. *Environ. Sci. Technol.* **2020**, *54*, 15563−15583.
(13) Lee, S.; Kim, J. Prediction of Nanofiltration and Reverse-Osmosis-Membrane Rejection of Organic Compounds Using Random Forest Model. *J. Environ. Eng.* **2020**, *146*, No. 04020127.
(14) Goebel, R.; Skiborowski, M. Machine-based learning of predictive models in organic solvent nanofiltration: Pure and mixed solvent flux. *Sep. Purif. Technol.* **2020**, *237*, No. 116363.
(15) Goebel, R.; Glaser, T.; Skiborowski, M. Machine-based learning of predictive models in organic solvent nanofiltration: Solute rejection in pure and mixed solvents. *Sep. Purif. Technol.* **2020**, *248*, No. 117046.
(16) Fetanat, M.; Keshtiara, M.; Keyikoglu, R.; Khataee, A.; Daiyan, R.; Razmjou, A. Machine learning for designing of thin-film nanocomposite membrane. *Sep. Purif. Technol.* **2021**, *270*, No. 118383.
(17) 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*, No. 118513.
(18) 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*, No. 118135.
(19) 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*, 5236−5250.
(20) 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*, No. eaaz4301.

(21) Pelikan, M. Bayesian Optimization Algorithm. In *Hierarchical Bayesian Optimization Algorithm*; Springer, 2005; pp 31−48.
(22) Snoek, J.; Larochelle, H.; Adams, R. P. Practical bayesian optimization of machine learning algorithms. 2012, arXiv:1206.2944. arXiv.org e-Printarchive.https://arxiv.org/abs/1206.2944.
(23) Wu, J.; Chen, X.-Y.; Zhang, H.; Xiong, L.-D.; Lei, H.; Deng, S.-H. Hyperparameter optimization for machine learning models based on Bayesian optimization. *J. Electron. Sci. Technol.* **2019**, *17*, 26−40.
(24) Victoria, A. H.; Maragatham, G. Automatic tuning of hyperparameters using Bayesian optimization. *Evol. Syst.* **2020**, *12*, 217−223.
(25) Griffiths, R.-R.; Hernández-Lobato, J. M. Constrained Bayesian optimization for automatic chemical design using variational autoencoders. *Chem. Sci.* **2020**, *11*, 577−586.
(26) 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*, 89−96.
(27) Häse, F.; Roch, L. M.; Kreisbeck, C.; Aspuru-Guzik, A. Phoenics: a Bayesian optimizer for chemistry. *ACS Cent. Sci.* **2018**, *4*, 1134−1145.
(28) 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. *J. Mater. Chem. A* **2020**, *8*, 21862−21871.
(29) Chen, T. In *Guestrin, C. Xgboost: A Scalable Tree Boosting System*, Proceedings of the 22Nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (New York, NY, USA, 2016), KDD '16, ACM, 2016; 2016; pp 785−794.
(30) Dorogush, A. V.; Ershov, V.; Gulin, A. CatBoost: gradient boosting with categorical features support. 2018, arXiv:1810.11363. arXiv.org e-Printarchive. https://arxiv.org/abs/1810.11363.
(31) Yap, C. W. PaDEL-descriptor: An open source software to calculate molecular descriptors and fingerprints. *J. Comput. Chem.* **2011**, *32*, 1466−1474.
(32) Moriwaki, H.; Tian, Y.-S.; Kawashita, N.; Takagi, T. Mordred: A Molecular Descriptor C alculator. *J. Cheminf.* **2018**, *10*, No. 4.
(33) Zhong, S.; Hu, J.; Yu, X.; Zhang, H. Molecular image-convolutional neural network (CNN) assisted QSAR models for predicting contaminant reactivity toward OH radicals: Transfer learning, data augmentation and model interpretation. *Chem. Eng. J.* **2021**, *408*, No. 127998.
(34) Kearnes, S.; McCloskey, K.; Berndl, M.; Pande, V.; Riley, P. Molecular graph convolutions: moving beyond fingerprints. *J. Comput.-Aided Mol. Des.* **2016**, *30*, 595−608.
(35) Rogers, D.; Hahn, M. Extended-connectivity fingerprints. *J. Chem. Inf. Model.* **2010**, *50*, 742−754.
(36) Yu, S.; Ma, M.; Liu, J.; Tao, J.; Liu, M.; Gao, C. Study on polyamide thin-film composite nanofiltration membrane by interfacial polymerization of polyvinylamine (PVAm) and isophthaloyl chloride (IPC). *J. Membr. Sci.* **2011**, *379*, 164−173.
(37) Tang, Y.-J.; Xu, Z.-L.; Huang, B.-Q.; Wei, Y.-M.; Yang, H. Novel polyamide thin-film composite nanofiltration membrane modified with poly(amidoamine) and SiO2 gel. *RSC Adv.* **2016**, *6*, 45585−45594.
(38) 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*, 7008−7018.
(39) Sundararajan, M.; Najmi, A. In *The Many Shapley Values for Model Explanation*, International Conference on Machine Learning, 2020; PMLR: 2020; pp 9269−9278.
(40) Merrick, L.; Taly, A. *The Explanation Game: Explaining Machine Learning Models Using Shapley Values*, International Cross-Domain Conference for Machine Learning and Knowledge Extraction, 2020; Springer, 2020; pp 17−38.
(41) Butina, D. Unsupervised Data Base Clustering Based on Daylight's Fingerprint and Tanimoto Similarity: A Fast and Automated Way To Cluster Small and Large Data Sets. *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 747−750.
(42) Wang, R.; Lin, S. Pore model for nanofiltration: History, theoretical framework, key predictions, limitations, and prospects. *J. Membr. Sci.* **2021**, *620*, No. 118809.
(43) Mohammad, A. W.; Teow, Y.; Ang, W.; Chung, Y.; Oatley-Radcliffe, D.; Hilal, N. Nanofiltration membranes review: Recent advances and future prospects. *Desalination* **2015**, *356*, 226−254.
(44) PolyInfo. https://polymer.nims.go.jp/en/.
(45) Lan, G. *First-order and Stochastic Optimization Methods for Machine Learning*; Springer, 2020.
(46) Fang, W.; Shi, L.; Wang, R. Mixed polyamide-based composite nanofiltration hollow fiber membranes with improved low-pressure water softening capability. *J. Membr. Sci.* **2014**, *468*, 52−61.
(47) Ahmad, A. L.; Ooi, B. S.; Mohammad, A. W.; Choudhury, J. P. Composite Nanofiltration Polyamide Membrane: A Study on the Diamine Ratio and Its Performance Evaluation. *Ind. Eng. Chem. Res.* **2004**, *43*, 8074−8082.
(48) Zhao, Y.; Tong, X.; Chen, Y. Fit-for-Purpose Design of Nanofiltration Membranes for Simultaneous Nutrient Recovery and Micropollutant Removal. *Environ. Sci. Technol.* **2021**, *55*, 3352−3361.
(49) Liang, Y.; Zhu, Y.; Liu, C.; Lee, K.-R.; Hung, W.-S.; Wang, Z.; Li, Y.; Elimelech, M.; Jin, J.; Lin, S. Polyamide nanofiltration membrane with highly uniform sub-nanometre pores for sub-1 Å precision separation. *Nat. Commun.* **2020**, *11*, No. 2015.
(50) Tang, C. Y.; Kwon, Y.-N.; Leckie, J. O. Effect of membrane chemistry and coating layer on physiochemical properties of thin film composite polyamide RO and NF membranes: II. Membrane physiochemical properties and their dependence on polyamide and coating layers. *Desalination* **2009**, *242*, 168−182.
(51) Ma, X.-H.; Yao, Z.-K.; Yang, Z.; Guo, H.; Xu, Z.-L.; Tang, C. Y.; Elimelech, M. Nanofoaming of Polyamide Desalination Membranes To Tune Permeability and Selectivity. *Environ. Sci. Technol. Lett.* **2018**, *5*, 123−130.
(52) Willcox, J. A.; Kim, H. J. Molecular dynamics study of water flow across multiple layers of pristine, oxidized, and mixed regions of graphene oxide. *ACS Nano* **2017**, *11*, 2187−2193.
(53) Dai, H.; Xu, Z.; Yang, X. Water permeation and ion rejection in layer-by-layer stacked graphene oxide nanochannels: a molecular dynamics simulation. *J. Phys. Chem. C* **2016**, *120*, 22585−22596.

https://doi.org/10.1021/acs.est.1c04373
Environ. Sci. Technol. XXXX, XXX, XXX−XXX