



Cite this: *Energy Environ. Sci.*,
2021, **14**, 3965

Machine learning analysis and prediction models of alkaline anion exchange membranes for fuel cells†

Xiuyang Zou, Ji Pan, Zhe Sun, Bowen Wang, Zhiyu Jin, Guodong Xu and Feng Yan *
*

The degradation of anion exchange membranes (AEMs) hindered the practical applications of alkaline membrane fuel cells. This issue has inspired a large number of both experimental and theoretical studies. However, it is highly difficult to draw universal laws from the resulting data. Here, for the first time, artificial intelligence (AI) technology was presented to forecast the chemical stability of AEMs for fuel cells. The chemical stability of AEMs was quantified by Hammett substituent constants based on a materials genomics strategy, and then classified by a decision tree. Among five machine learning algorithms applied, the artificial neural network (ANN) showed the highest accuracy in predicting the chemical stability of AEMs ($R^2 = 0.9978$). Combined with the computational works, long-term chemical stability experiments were conducted to demonstrate the robustness and prediction accuracy of the proposed approach. This study highlights the potential of data-driven modelling for predicting the alkaline stability of AEMs, and thus unnecessary experiments can be avoided for the development of alkaline membrane fuel cells.

Received 19th April 2021,
Accepted 1st June 2021

DOI: 10.1039/d1ee01170g

rsc.li/ees

Broader context

The development of alkaline membrane fuel cells (AMFCs) has attracted great attention since the maximum power density of some AMFCs is comparable to that of proton exchange membrane fuel cells. However, the degradation of anion exchange membranes (AEMs) hindered the practical applications of AMFCs. To help improve the device stability, a larger number of experiments have been conducted to investigate the degradation mechanisms. Since the experimental works are usually high cost, time-consuming and complex, it is extremely difficult to draw universal laws from the resulting data. Combining artificial intelligence (AI) technology to carry out research has now become an emerging trend in the field of materials synthesis, because machine learning can make accurate decisions and predict the performance of hypothetical materials based on big data reported, thereby significantly shortening the material development cycle. For the first time, in this work, AI technology was presented to predict the chemical stability of AEMs for fuel cells. The artificial neural network shows a very high accuracy in predicting the chemical stability of AEMs ($R^2 = 0.9978$). The results illustrate the potential of the proposed research paradigm for elucidating and designing AEMs.

Introduction

Fuel cells are electrochemical devices that could convert the chemical energies (such as hydrogen or methanol) into electrical energy. They have emerged as an efficient source of clean and sustainable energy and attracted a great deal of academic and industrial attention.^{1–4} However, their practical applications have been restricted by the short lifetime caused by the

degradation. To help improve device stability, a larger number of experiments have been conducted to investigate the degradation mechanisms.^{5–8} Since the experimental works are usually high cost, time-consuming, and complex,^{9,10} it is extremely difficult to systematically understand the degradation mechanism of the fuel cells because of the complex interrelationship between the larger number of concerned factors.¹¹

Due to the rapidly developing of artificial intelligence (AI) technology, the way to solve this problem has been broadened.^{12–15} As one of the important branches of AI, machine learning, mainly refers to the improvement of the learning ability of machines through the recognition of systems or knowledge,¹⁶ so that they can acquire new skills or new knowledge.^{17–19} Recent studies have demonstrated that

College of Chemistry, Chemical Engineering and Materials Science, Soochow University, Suzhou, 215123, China. E-mail: fyan@suda.edu.cn

† Electronic supplementary information (ESI) available: Additional analysis details, Fig. S1–S7, Tables S1–S7 as described in the main text. See DOI: 10.1039/d1ee01170g

machine learning could help predict the maximize power output and (or) degradation of proton exchange membrane fuel cells (PEMFCs) by computer simulations and numerical modeling.²⁰ For example, Djerdir and co-workers established a degradation prediction model of PEMFCs by a grey neural network. The proposed method can forecast the degradation for PEMFCs under different operating conditions.²¹ More recently, Liu *et al.* adopted XGBoost regression (XGR) and artificial neural network (ANN) to analyse the impact of a nonprecious metal electro-catalyst on the maximum power density of PEMFCs.²² With only simple experimental parameters as the input, ANN predicted the maximum power density of the PEMFCs with high accuracy ($R^2 = 0.9621$).

Alkaline anion exchange membrane fuel cells (AEMFCs), which could alleviate the shortcomings of PEMFCs and provide a cost-effective approach by using nonprecious metal catalysts,^{23–26} have attracted great attention in the past decade. Anion exchange membranes (AEMs) that comprise polymer backbones and hydrophilic cationic groups are one of the most crucial components in AEMFCs.²⁷ AEMs govern the ion conductivity and long-term chemical stability (mainly alkaline stability) during the operation.²⁸ Therefore, development of AEMs with high hydroxide conductivity, chemical and mechanical stability in high pH solutions at elevated temperatures has been one of the most important and challenging topics, before the extensive application and commercialization of AEMFCs.²⁹ The task of predicting chemical stability of AEMs is critically important given its broad utility but challenging due to nonlinear degradation with polymer backbone and (or) grafted side chain. For example, a polymer backbone containing heteroatoms (such as aryl-ether or sulfonyl group) is less stable as compared with an all-alkane polymer (such as polyethylene) backbone.^{11,30} Besides the polymer backbone, the grafted side chain, such as the type of cationic group, the spacer between the cationic group and the polymer backbone highly affect the alkaline stability of AEMs.^{31–33}

In this study, we design data-driven models that accurately analyse the features and predict the alkaline stability of quaternary ammonium (QA) functionalized AEMs, with no prior knowledge of chemical degradation mechanisms. Inspired by a materials genomics strategy, the polymer was divided into a set of building blocks (substituent groups). Then Hammett constants (ionization constants) of the substituents are used to predict the rate or equilibrium constants for the building blocks in solution, which enable the machine learning algorithms to identify the polymers.³⁴ A decision tree is then used to classify AEMs on the basis of their conductivity retention rate after 168 h and provide suggestions for improving their alkaline stability. Furthermore, five machine learning methods, including support vector regression (SVR), Gaussian process regression (GPR), random forest regression (RFR), XGR, and ANN, were used to predict the conductivity retention of AEMs. Combined with the computational works, long-term stability experiments for four types of AEMs are performed to demonstrate the robustness and prediction accuracy of the proposed approach. These results illustrate the potential of the proposed

research paradigm for elucidating and designing the AEMs for fuel cells.

Results and discussion

Computational works

Data curation. A database containing input and output data is required, regardless of the machine learning algorithm used to construct the AEMs degradation model. The database used in this paper was built based on literature published, which contained the literature between 2010 and 2020. Herein, 176 experimentally measured conductivity retention entries on QA functionalized AEMs (each contains 15 conductivity retention points) are extracted from the 86 high quality literature. The database contains polymer structure information, physical and chemical properties, and membrane degradation operating conditions. The recorded AEMs properties and the operating conditions of long-term stability experiments were used as input features. The missing values of input features are replaced by the average value. Polymers are usually composed of simple, identical structural units repeatedly connected by covalent bonds.³⁵ Accordingly, a gene partition method was proposed, *i.e.*, a set of structural units was divided by mimicking the process of polymer production (Fig. 1a). To ensure that the machine learning algorithm could recognize the chemical structure of the polymer, Hammett substituent constants were used to quantify the main chain structure of the polymer and branched substituent groups. The Hammett substituent constants are the ionization constants of the substituent, considering the steric, inductive, and resonance effects.³⁶ The Hammett equation (and its extended forms) is one of the most widely used means for the study and interpretation of organic reactions and their mechanisms. The Hammett parameters (σ_p and σ_m) were developed to describe the linear free energy relationships between the reaction rate and equilibrium constant. Table S1 (ESI†) shows the Hammett substituent constants of substituent reported.^{36,37} The corresponding value ranges of the input features are listed in Table S2 (ESI†). Finally, the conductivity retention rate of the membrane was used as the output to evaluate the alkaline stability of the AEMs.

Here, ANN was used to predict the alkaline stability of the AEMs because it can learn from existing experimental data to establish numerical models or computer models, control the number of output variables and shows the best regression performance. The structure of ANN was divided into the input layer, hidden layer, and output layer. Each layer contains a certain number of neural units, as shown in Fig. 1b.

Feature importance ranking

For the first time, the Hammett substituent constants were used to describe the polymer information of the AEMs. Therefore, it is necessary to perform characteristic analysis of the collected samples to test the accuracy of the samples before using machine learning algorithms to predict the chemical stability of the AEMs. The decision tree has achieved

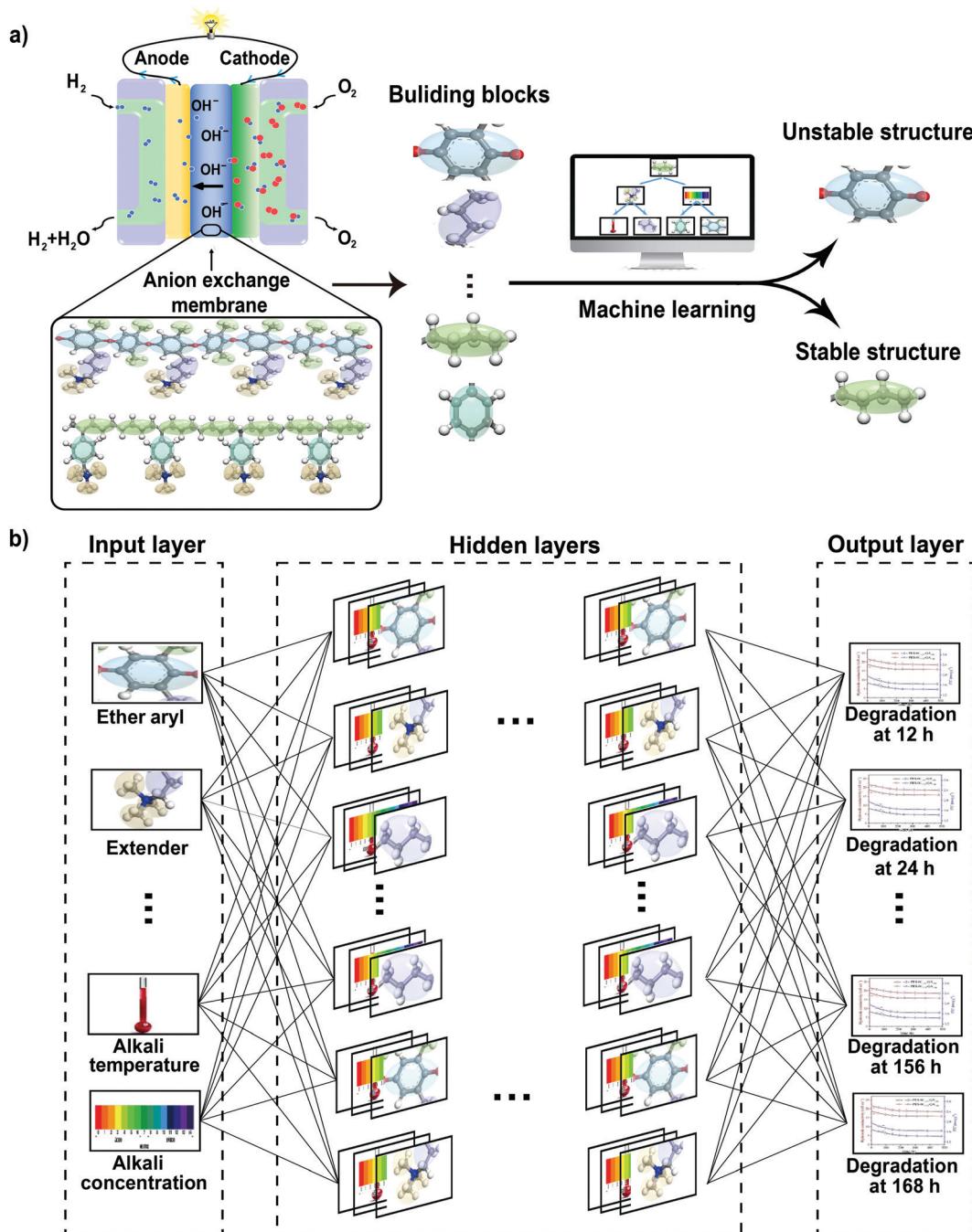


Fig. 1 (a) Overview of polymer structure division and process of machine learning to identify polymer structures. A set of structural units was constructed by separating the simple, identical structural units in the polymer. Machine learning algorithms were used to analyse the structure and predict the stability of the polymer. (b) Schematic diagram of ANN model. The structure of ANN was divided into the input layer, hidden layer, and output layer. Each layer contains a certain number of neural units.

considerable success in analysing the importance of factors because of its speed and accuracy.³⁸ It selects factors with high information gain as the final split attribute and split point of the tree node. The possible over-fitting problems were avoided by adjusting the depth of the tree, the number of samples at the node, the proportion of the largest class at the node, etc. The entire dataset was randomly divided into two parts. The first part contained 85% of the data points and was used to

train the decision tree, while the remaining 15% was used to verify the accuracy of the decision tree and evaluate the chemical stability of the AEMs. By detecting the difference between the input (independent variable) and output (dependent variable) sets, the input data were analysed for its characteristics. First, the conductivity retention of the AEMs after 168 hour-test was ranked. The conductivity retention of higher than 88% after 168 hour-test was considered as highly alkaline stable AEMs

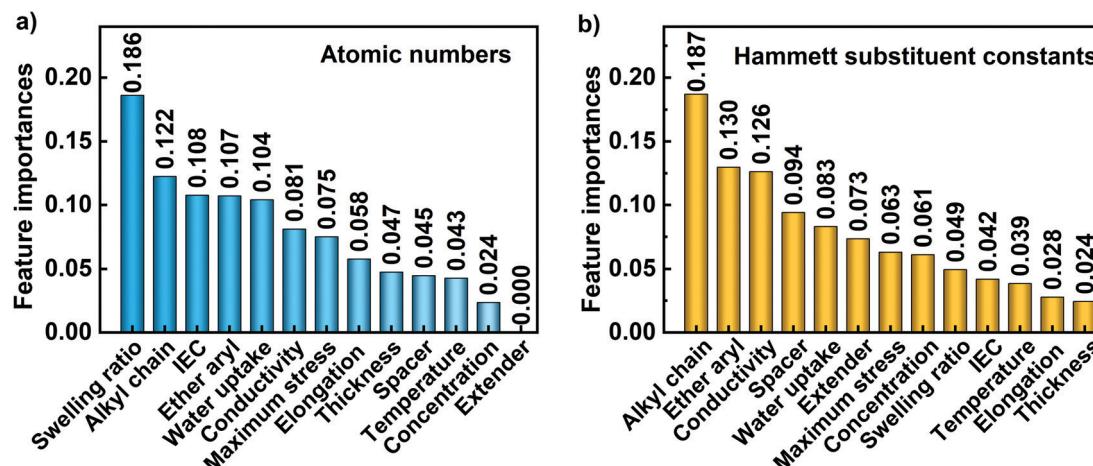


Fig. 2 Feature importance based on (a) atomic numbers and (b) Hammett substituent constants. The larger the value of the feature importance is, the stronger the discrimination ability of the corresponding factor is. The atomic number describes only the size and electronegativity of a single atom, while the Hammett substituent constants describe the steric, inductive, and resonance effects of the substituent group.

(top 10% in database). Next, the decision tree learned the complex relationship between the input factors from the training set and used the criterion of 88% to classify AEMs and analyse the importance of the input factors.

To enable the machine learning algorithms to recognize polymer information, Hammett substituent constants were used to describe the structure of the polymer, including the content of the backbone groups and the structure of the cationic branches. For comparison, the atomic numbers were also used to quantify the structural information of the polymer. The atomic number represents the nuclear charge of each element, which is related to the electronegativity and radius of the atom.³⁹ The feature importance based on the atomic numbers and Hammett substituent constants are shown in Fig. 2a and b, respectively. The larger the value of the feature importance is, the stronger the discrimination ability of the corresponding factor is. As shown in Fig. 2a, the decision tree based on atomic numbers shows that the basic properties of the AEMs has stronger discrimination ability than the structure of the polymer. It contradicts the actual situation that the basic properties of the AEMs, including conductivity, water uptake, and swelling ratio, are determined by the polymer structure in the AEMs. The reason for this problem is that the machine learning algorithm fails to recognize the structure of the polymer correctly because the atomic number describes only the size and electronegativity of a single atom. The confusion matrix of the decision tree, based on the atomic numbers and Hammett substituent constants, is shown in Tables S3 and S4 (ESI†). It is worth noting that the weighted precision of the decision tree based on the Hammett substituent constants (85.3%) was larger than that based on the atomic numbers (71.6%), indicating that the dataset based on the Hammett substituent constants was more suitable for the machine learning algorithms. The difference between the two datasets was analysed by calculating the characteristics of all the input factors. As shown in Fig. 2b, the polymer information, which was considered comparatively less important in the decision

tree based on the atomic numbers, has now ascended. This can be attributed to the Hammett substituent constants, which describe the steric, inductive, and resonance effects of the substituent group. Although the decision tree accurately determines the priority of factors affecting the alkaline stability of the membrane, it may incorrectly estimate the importance of the factors owing to database limitations. For example, the temperature of the alkali lye has been incorrectly evaluated. Most of the data (about 98%) on the alkaline stability of the AEMs were investigated at 60 °C or 80 °C. Due to the high alkali stability of these reported AEMs, the conductivity retention difference between these reported AEMs at 60 °C and 80 °C is relatively small (the average change is 4.5%). Therefore, the influence of temperature on the alkaline stability of the AEMs has not been correctly evaluated till now.

Although the factors with strong discriminative ability have been obtained based on feature importance, the visualized decision tree determines the necessary conditions (specific values of each factor) for the high chemical stability of the AEMs and further guides the synthesis of chemically stable AEMs. The partial decision structure based on the Hammett substituent constants is illustrated in Fig. 3. From the root node at the bottom to the leaf node at the top, the classification samples in the leaf nodes meet the needs of each link node. Therefore, factors with strong discriminating ability appear in tree nodes earlier and larger numbers. To examine whether the decision tree achieved accurate results in improving the chemical stability of the AEMs, the nodes of the decision tree were analysed one by one.

The initial root node suggests that a high alkyl chain contribution in the polymer backbone of the AEMs should be considered first. This is a reasonable conclusion, as the structure of the polymer backbone in the AEMs is very important for the chemical stability of the AEMs.⁴⁰ Several studies have shown that the backbone with heteroatoms is unstable in alkali lye.^{7,11,41} Ramani *et al.* showed that the benzylic side chain is sensitive to nucleophilic attack, regardless of the cationic group chemistry

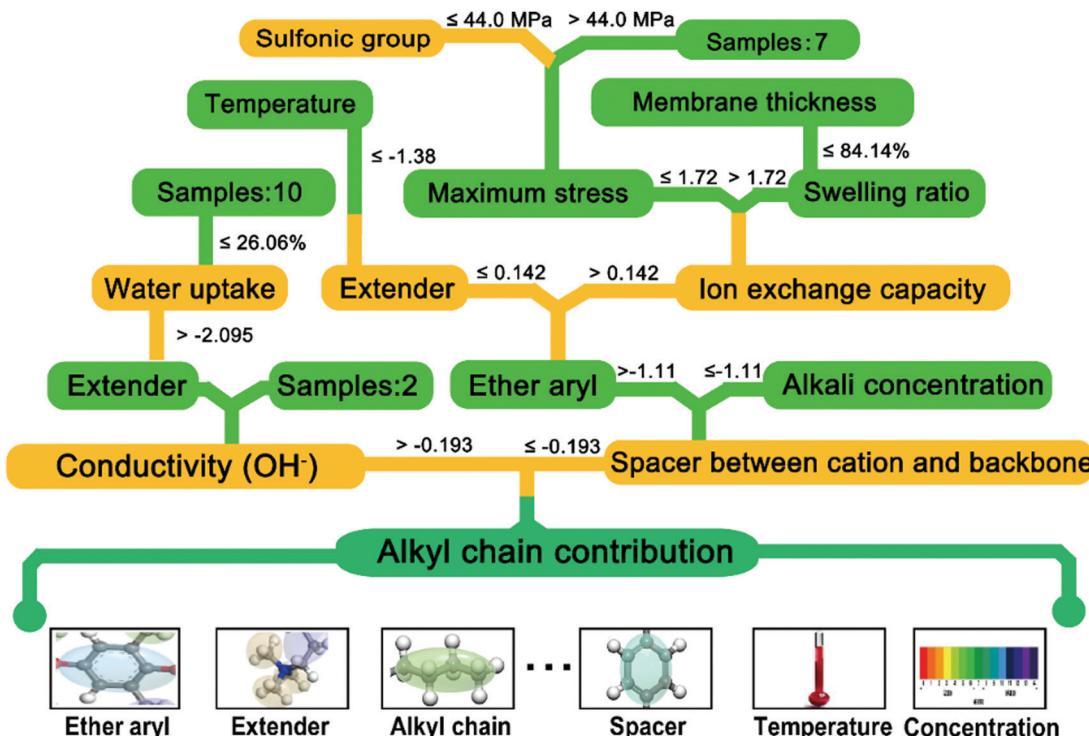


Fig. 3 Partial decision structure based on the Hammett substituent constants used to determine the requirements for high chemical stability of AEMs. To observe the relationship between the various variables clearly, the complex decision tree is simplified.

attached, resulting in the degradation of the polymer backbones at both the quaternary carbon and the ether positions in the aromatic polymer backbone.⁴² Therefore, the decision tree suggests that appropriately increasing the content of alkane chains in the polymer backbone can enhance the alkaline stability of the AEMs. Conductivity is regarded as a crucial factor by the decision tree, as two samples reach the criterion of 88% using the shortest path. The shorter the path of the decision tree is, the stronger the discrimination ability of the corresponding factor is. The conductivity of the AEMs is affected by the grafting level and hydration degree of the cationic groups.⁴³ The decision tree suggests that when the polymer backbone is stable, an AEM with lower conductivity should have higher chemical stability. Therefore, one of the simple ways to reduce the conductivity of the AEM is to reduce the number of cations in the AEM.^{44,45} For the AEMs with hydroxide conductivity greater than 15.23 mS cm^{-1} , the decision tree suggests increasing the protective group contribution of the QA group. The larger substituent groups around the cationic centre (for example, the replacement of two methyl groups by two larger ethyl groups)⁴⁶ cause steric shielding of the cationic domain; hence, the substituent groups hinder Hoffman elimination and protect the cationic group from OH^- attack, thereby slowing down the chemical degradation of the AEMs.⁴⁷ A reasonable request for more ($> 26.06\%$) water uptake is proposed. The details of the experiment and simulation showed that as more water molecules dissolve the hydroxide, the nucleophilicity and alkalinity of the hydroxide are hindered, and QA degradation is slowed down significantly.^{5,33}

In the case of a relatively unstable polymer backbone (alkyl chain contribution < -0.193), such as poly(2,6-dimethyl-1,4-phenylene oxide) (PPO), the spacer between the cationic group and the polymer backbone should be investigated with priority. In general, the introduction of a long alkyl chain spacer between the polymer backbone and the cationic functional group can improve the stability of the polymer backbone.⁴⁸ This is because the appropriate spacer can counter anion condensation and avoid Hofmann eliminations.⁴⁹ A relatively low swelling ratio ($< 84.14\%$) and high maximum stress ($> 44.0 \text{ MPa}$) are recommended. The AEMs with a cross-linked structure tend to have a lower swelling ratio and higher tensile strength. The cross-linked structure protects the cations from OH^- attack.^{50,51} Although the IEC value highly affect the conductivity of AEMs, the decision tree believes that the IEC value has little effects on the chemical stability of the AEMs. For the bottom nodes, the membrane thickness and maximum elongation are considered, implying that the decision tree recommendations are consistent with those reported in the literature. Compared with the factor at the initial root node, the membrane thickness and maximum elongation have a weaker effect on the alkaline stability of the AEMs.⁵²

Optimization of ANN algorithm

One advantage of the ANN is that it can be used to map any type of output results because the number of neurons in the output layer of the ANN can be adjusted according to the actual results obtained. Considering the fitting accuracy of the algorithm and the type of predicted output, ANN was employed to predict the

alkaline stability of AEMs. To further improve the fitting accuracy of ANN, the hyperparameters of ANN need to be optimized.⁵³ The number of hidden layers in the ANN and the number of neurons in a hidden layer were optimized first because they determine the frame structure of the ANN.⁵⁴ In general, as the number of hidden layers in the ANN and the number of neurons in a hidden layer increase, the ANN model becomes complex, and the fitting effect of the ANN is enhanced.^{55,56} However, too many hidden layers and neurons will lead to generalization errors and over-fitting.⁵⁷ Here, by changing the number of neurons and hidden layers, ANN models containing a combination of different hidden layers and neurons were used to fit the database with the same structure. The regression correlation coefficient (R^2) and root mean square error (Rmse) of these models are shown in Fig. S1 (ESI†). The closer the R^2 value of the model is to 1 and the lower the Rmse of the model is, the higher the fitting effect of the model is. It can be observed that the R^2 value of the 4-layer-40-neuron model reached 0.9975 while its Rmse reduced to 1.32, indicating that the model achieved the best fitting effect. The over-fitting issue of ANN is another important factor that reduces the model accuracy. By monitoring the change in loss during the training and cross-validation process, the occurrence of over-fitting was determined (Fig. S2, ESI†). As shown in Fig. S2 (ESI†), the loss of the training and cross-validation process decreased rapidly as the epoch value increased until it reached a trough close to 1.22, and it remained relatively flat thereafter. However, when the epoch value reached 2405, the loss of the verification set started to increase, indicating that over-fitting occurred, which hindered the generalization ability of ANN. Therefore, the best validation performance of ANN was obtained at 2405 epochs, at which time the R^2 value was 0.9978. The above data indicates that the ANN model learned the complex hidden correlations among 15 features and could predict uncharted real experimental data; thus, it could assist researchers in the development of alkali-resistant AEMs.

Complexity and training time cost are important criteria for evaluating machine learning algorithms. Fig. S3 (ESI†) shows the summary of the Rmse and R^2 values for ANN model with different number of input features. As the number of input parameters increases, the Rmse value of the model decreases. For comparison, the top 5 features and the top 10 features are used as inputs to build regression models, respectively. As shown in Fig. S4 and S5 (ESI†), the R^2 value of the model increases with the number of input features, while the difference in training time cost of these models is very small (in the same order of magnitude). In addition, since the decision tree did not correctly identify the effects of temperature feature on the chemical stability of the AEMs, an input set without temperature feature was also used to train the new model (Fig. S6, ESI†). However, it could be noted that the accuracy of new model without temperature feature ($R^2 = 0.9933$ and Rmse = 2.14) is lower than those of model with the temperature feature ($R^2 = 0.9978$ and Rmse = 1.22). According to the accuracy of the model, 15 features were selected as the input of the model.

Comparison of multiple algorithm models

To this end, various machine learning algorithms were extensively screened to find the best-performing machine learning algorithm for the regression task. It is noted that the division of the training and test sets used by the five algorithms was consistent. The predictions on the training set of these algorithms are shown in Fig. S7 (ESI†). The R^2 -value of both XGR and ANN is close to 1.0, which justifies the reliability of the fit. Fig. 4a–e shows scatter diagrams of the predictions on the test set of these algorithms. The fitting accuracy of the algorithm was evaluated by the R^2 . The x-axis of the scatter plot represents the conductivity retention of the AEMs actually reported in the literature, and the y-axis represents the conductivity retention of the AEMs predicted by the machine learning algorithm. The closer the predicted value is to the real value, the closer is the R^2 value to 1. As shown in Fig. 4a–e, the R^2 values of GPR, SVR, RFR, XGR, and ANN are 0.5106, 0.9213, 0.9862, 0.9971, and 0.9978, respectively. Except for the GPR algorithm, all the R^2 values are above 0.92. Among them, the R^2 values of XGB and ANN are higher, exceeding 0.99. This is because the database samples come from high-quality studies and the integrated algorithms have strong anti-noise ability.

In order to verify the robustness of the algorithm, the database was divided into different training and test sets by modifying the random seed number. Then multiple simulations were performed on the database containing different training and test sets.⁵⁸ Fig. 4f shows the R^2 and Rmse values of the five algorithms. It can be observed that the ANN model has smaller errors, although the R^2 values of ANN and XGB are both above 0.997, indicating that the predicted values are very close to the actual values. Based on the above data, it can be concluded that the ANN model is more suitable for analysing the alkaline stability of AEMs.

Conductivity retention rate predicted by ANN

Compared with the single-output model, the multi-output model, *i.e.*, the conductivity of AEMs as a function of time, can describe the alkaline stability of the membrane more intuitively. Therefore, here, the time variable was deleted from the input set, and the number of outputs of the ANN model was changed from one output to more outputs. The hyperparameters of this new ANN model were adjusted by grid search, as in the previous case. Four new records that have never been used in the ANN model (training or testing) were selected from the database to test the learning effect of the ANN model. Fig. 5 shows the predicted and experimental conductivity retention rate curves of these four records. It can be observed that the predicted conductivity degradation curves of the four records (blue solid line) basically coincide with the conductivity degradation curves reported in the literature (red scattered dots). The R^2 values of all four records are greater than 0.95 (0.9692, 0.9966, 0.9944, and 0.9931, respectively), and the Rmse values of all four records are less than 4 (0.9746, 3.9987, 2.2265, and 2.6653, respectively). It is noted that the chemical degradation degree of these four AEMs has a large

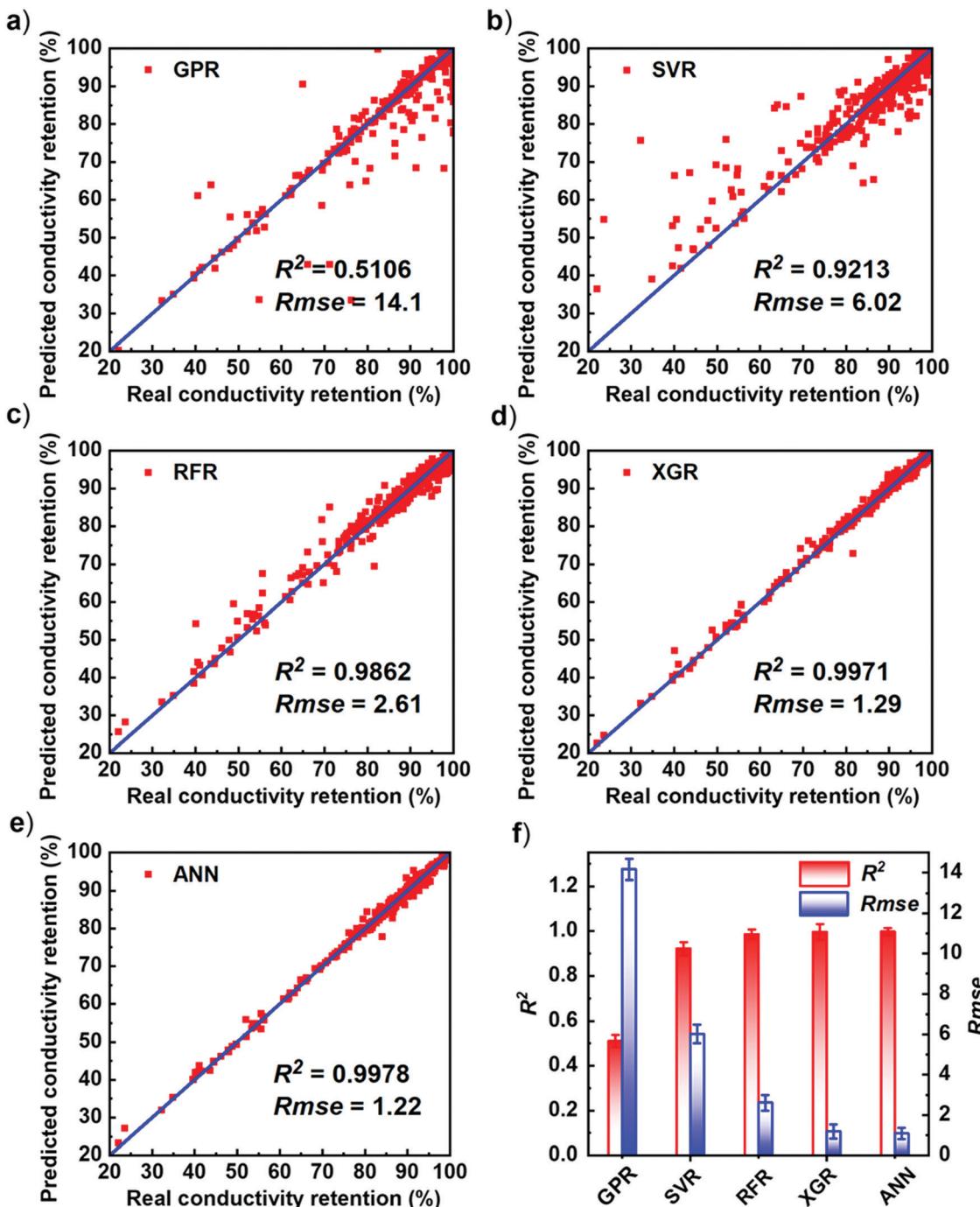


Fig. 4 Regression correlation coefficient (R^2) of (a) GPR, (b) SVR, (c) RFR, (d) XGR, and (e) ANN algorithms for test set under the same structure of the database. (f) Summary of the $Rmse$ and R^2 values for each machine learning algorithm. By modifying the random seed number, different training and test sets were generated to verify the robustness of the algorithms.

span (the conductivity of the AEMs decreased by 30%, 50%, 70%, and 90%, respectively, after 168 h) because four different curves were derived from totally different feature inputs. In this case, the ANN model still achieves accurate predictions very well, indicating that the ANN model can describe the complex correlation between the input variables, which is difficult to directly express by mathematical relationships. Therefore, it can be concluded that the ANN model

correctly predicts the conductivity of the AEM as a function of time.

Computational works

Since AI has the ability to predict the results of experiments, the best possible performance can be obtained through ANN model to guide the experiments (Table S5 and Fig. S8, ESI†). In this work, however, there are 15 complex input features used for the

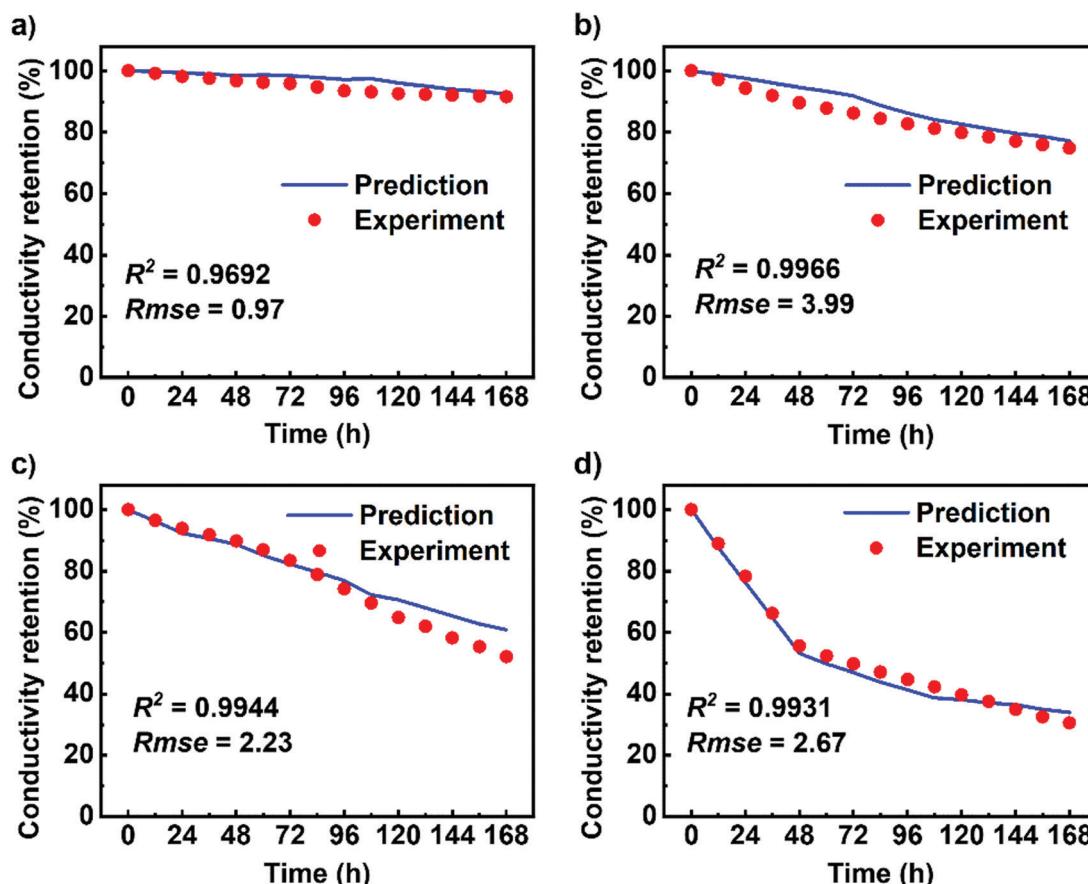


Fig. 5 Predicted and experimental conductivity retention curves of reported the AEMs: (a) poly(p-phenylene-co-aryl ether ketone) AEMs,⁵⁹ (b) poly(ether sulfone)-based AEMs,⁶⁰ (c) PPO-based AEMs,⁶¹ and (d) polystyrene-N,N-diethyl-3-(4-vinylphenyl)propan-1-amine AEMs.⁵²

ANN model. These input features may be contradictory to each other when AI designs a new AEM (polymers 1–8, chemical structures see Fig. S8, ESI†). For example, the swelling ratio and water uptake of polymer 1 are set to be 1.92% and 55.64%, respectively. However, such a design is difficult to be realized experimentally. Due to the inherent shortcomings of the current ANN model, the optimal chemical structures of AEMs given by the AI were not adopted in this work.

Although ANN could not provide the AEMs with best performances, it can still guide our experimental design, due to its high accuracy ($R^2 = 0.9978$). In the case of a relatively unstable polymer backbone, such as PPO, the spacer between the cationic group and the polymer backbone or the extender of the cationic group should be adjusted to improve the alkaline stability of the polymer as much as possible. To verify whether the ANN model can predict the alkaline stability of PPO-based AEMs with different spacers or extenders, the long-term stability experiments were performed using four types of simple AEMs.

Four AEMs based on trimethyl ammonium cation (PPO-MTMA), triethyl ammonium cation (PPO-ETMA), *N,N*-dimethylbutyl ammonium cation (PPO-DMBA), and *N,N*-dimethyl-hydroxy ammonium cation (PPO-DMHA) were synthesized. The chemical structures and NMR spectra of four AEMs are shown in Fig. 6a

and Fig. S9 (ESI†), respectively. The physical and chemical properties of the AEMs, and the detailed operating conditions of the long-term stability experiments are summarized in Table S6 (ESI†). Fig. 6b–e show the predicted and real conductivity retention rates of four types of AEMs at 60 °C. It can be seen that the R^2 values of the four records are 0.9811, 0.9572, 0.9989, and 0.9354, respectively. The Rmse values of four records are 4.03, 4.69, 1.72, and 3.70, respectively. We can conclude that ANN successfully predicted the conductivity of these AEMs with different extenders at 60 °C. Although the ANN model still has some limitations owing to the limited existing database, they do not affect the prediction of the alkaline stability of AEMs via the ANN model. With the increase in experimental data, the ANN model will show better performance.

Methods

To build a successful AEMs degradation model, different kinds of machine learning algorithms were applied to analyze the database. These machine learning algorithms were performed using Python 3.6 package. The ANN was constructed by Keras 2.2.4 with Tensorflow-GPU 1.12.0 as a backend. All the other machine learning methods were accomplished by Scikit-learn

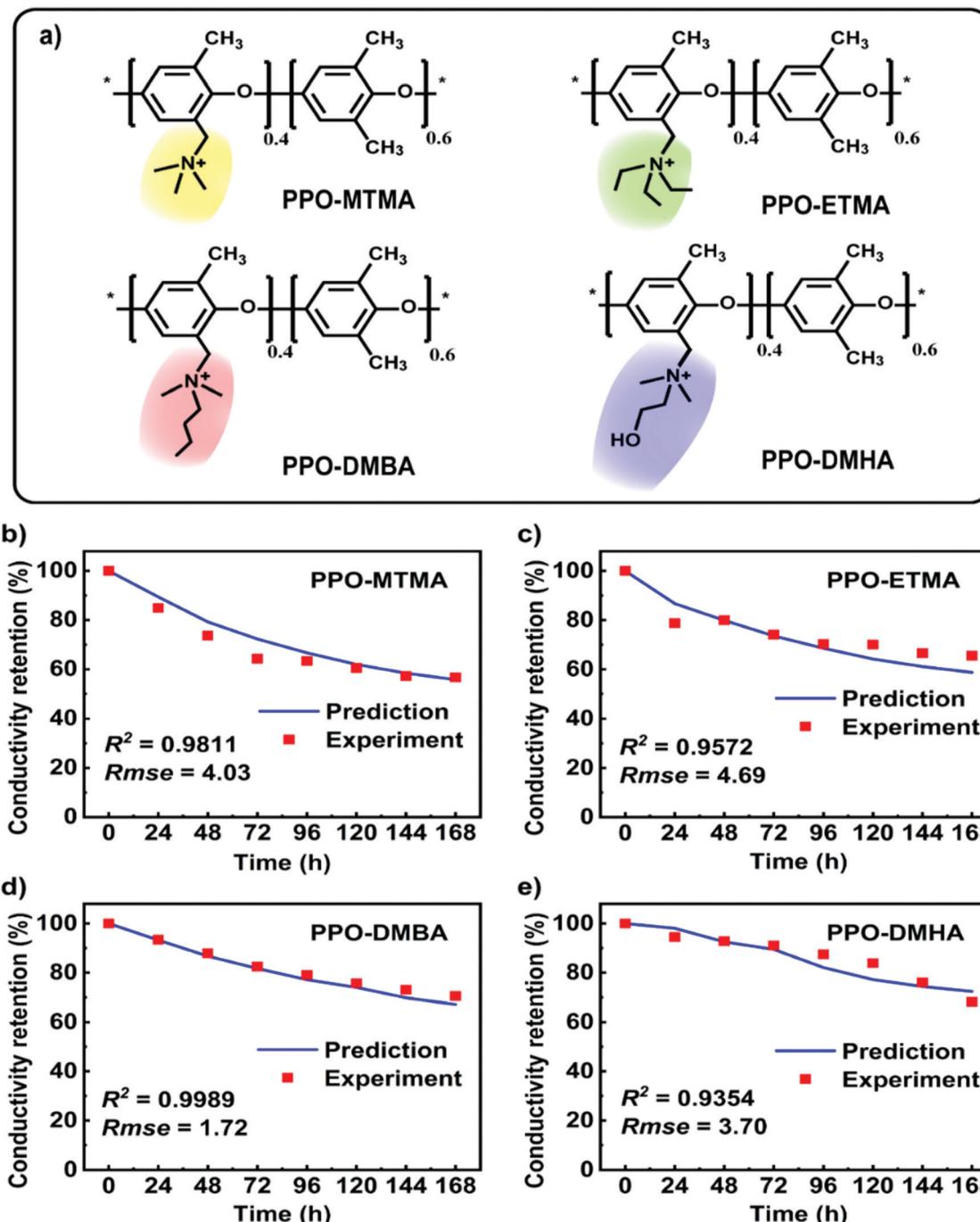


Fig. 6 Predicted and experimental conductivity retention rate curves of (a) PPO-MTMA (b) PPO-ETMA, (c) PPO-DMBA, and (d) PPO-DMHA at 60 °C. The illustrations show the structure of these four AEMs.

0.19.0. For different tasks, the grid search method was used to optimize the hyperparameters of the corresponding algorithms to obtain a satisfactory model. The optimized hyperparameters of different algorithms were listed in Table S7 (ESI†).

XGBoost algorithm is one of the most popular gradient enhancement machine decision tree algorithms. It is an optimized distributed gradient boosting algorithm and uses the residuals of the first n trees to build the $(n + 1)^{\text{st}}$ tree. According to the importance of the features in the database, the XGBoost

algorithm evaluates a score for each leaf node, and adds up the scores corresponding to each tree to get the predicted value of the features. Therefore, compared with deep learning algorithms, XGBoost is easier to use to process smaller data sets.

The neurons in each hidden layer contain a bias value and are connected to the neurons in the previous layer through interconnections with certain weights. A brief introduction of the ANN and concise mathematical derivation in this article were introduced in ESI.† One advantage of ANN is that ANN can

be optimized by itself, through learning based on database information. That is, the error between the real value and the predicted value could be traced backward to retrain the previous neurons and update their bias values and weights. Another advantage of ANN is that it can be used to map any type of input and output. This is because that the number of neurons in the input layer and output layer of the ANN can be adjusted according to the actual input and output. The input layer of ANN in this paper had fifteen neurons (conductivity, swelling ratio, water uptake, the thickness and mechanical properties of AEMs, temperature, and concentration of alkali lye and polymer structure information) and the output layer had fifteen neurons (the conductivity of AEMs as a function of time). The “learning” process of ANN is to divide the entire input and output database into three subsets: training, testing, and verification, which account for 70%, 15%, and 15% of the database, respectively. In this paper, a typical trial-and-error method was used to select the number of the hidden layers and the neurons in each hidden layer of ANN. The performance of AEMs degradation model was evaluated statistically using the R^2 and Rmse. The best ANN model (maximum R^2 and minimum Rmse) was obtained by grid search of combinations of different numbers of hidden layers and neurons.

Conclusions

This study proved that data-driven modelling is an excellent tool for the analysis and prediction of the alkaline stability of AEMs. First, the chemical structure of polymers was quantified by Hammett substituent constants, and the digital polymer structures were used to classify AEMs on the basis of their conductivity retention rates after 168 h. Then, the decision tree ranked the importance of 15 complex functions and prioritized chemically stable polymer backbones to determine AEMs with high alkaline stability. However, when a relatively unstable polymer backbone was used, such as PPO, the decision tree suggested that the spacer between the cationic group and the polymer backbone should be prioritized. Furthermore, five machine learning algorithms were used to predict the alkaline stability of AEMs. The R^2 values of SVR, GPR, RFR, XGR, and ANN were 0.5106, 0.9213, 0.9862, 0.9971, and 0.9978, respectively. Among five machine learning algorithms, ANN rapidly and accurately predicts the conductivity of AEMs as a function of time when the user enters their physical and chemical properties; thus, a large number of repeated experiments can be avoided. Finally, the long-term chemical stability experiments were performed to demonstrate the robustness and prediction accuracy of the proposed approach. The proposed approach improves the quality and efficiency of experimental studies for understanding and designing AEMs.

Author contributions

X. Z., J. P., and F. Y. designed machine learning algorithms and analysed the key parameters of anion exchange membranes

with high alkaline stability. X. Z., Z. S., and B. W. prepared anion exchange membranes and performed alkali resistance stability experiments. Z. J. and G. X. collected data for machine learning from the literature. All authors reviewed and commented on the paper.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

This work was supported by the National Natural Science Foundation of China (21835005, U1862109, 21905193), the National Natural Science Foundation of the Jiangsu Higher Education Institutions of China (19KJB150016), the Collaborative Innovation Center of Suzhou Nano Science and Technology, and the Priority Academic Program Development of Jiangsu Higher Education Institutions.

References

- 1 J. H. Wang, Y. Zhao, B. P. Setzler, S. Rojas-Carbonell, C. Ben Yehuda, A. Amel, M. Page, L. Wang, K. Hu, L. Shi, S. Gottesfeld, B. J. Xu and Y. S. Yan, *Nat. Energy*, 2019, **4**, 392–398.
- 2 L. Wang, X. Peng, W. E. Mustain and J. R. Varcoe, *Energy Environ. Sci.*, 2019, **12**, 1575–1579.
- 3 D. R. Dekel, *J. Power Sources*, 2018, **375**, 158–169.
- 4 J. R. Varcoe, P. Atanassov, D. R. Dekel, A. M. Herring, M. A. Hickner, P. A. Kohl, A. R. Kucernak, W. E. Mustain, K. Nijmeijer, K. Scott, T. W. Xu and L. Zhuang, *Energy Environ. Sci.*, 2014, **7**, 3135–3191.
- 5 D. R. Dekel, M. Amar, S. Willdorf, M. Kosa, S. Dhara and C. E. Diesendruck, *Chem. Mater.*, 2017, **29**, 4425–4431.
- 6 W. E. Mustain, M. Chatenet, M. Page and Y. S. Kim, *Energy Environ. Sci.*, 2020, **13**, 2805–2838.
- 7 J. Fan, S. Willdorf-Cohen, E. M. Schibli, Z. Paula, W. Li, T. J. G. Skalski, A. T. Sergeenko, A. Hohenadel, B. J. Frisken, E. Magliocca, W. E. Mustain, C. E. Diesendruck, D. R. Dekel and S. Holdcroft, *Nat. Commun.*, 2019, **10**, 2306.
- 8 J. Ponce-González, D. K. Whelligan, L. Wang, R. Bance-Soualhi, Y. Wang, Y. Peng, H. Peng, D. C. Apperley, H. N. Sarode, T. P. Pandey, A. G. Divekar, S. Seifert, A. M. Herring, L. Zhuang and J. R. Varcoe, *Energy Environ. Sci.*, 2016, **9**, 3724–3735.
- 9 Z. Sun, B. Lin and F. Yan, *ChemSusChem*, 2018, **11**, 58–70.
- 10 Z. Sun, J. Pan, J. Guo and F. Yan, *Adv. Sci.*, 2018, **5**, 1800065.
- 11 D. Li, I. Matanovic, A. S. Lee, E. J. Park, C. Fujimoto, H. T. Chung and Y. S. Kim, *ACS Appl. Mater. Interfaces*, 2019, **11**, 9696–9701.
- 12 K. A. Severson, P. M. Attia, N. Jin, N. Perkins, B. Jiang, Z. Yang, M. H. Chen, M. Aykol, P. K. Herring, D. Fragedakis, M. Z. Bazan, S. J. Harris, W. C. Chueh and R. D. Braatz, *Nat. Energy*, 2019, **4**, 383–391.

- 13 I. G. Clayson, D. Hewitt, M. Hutereau, T. Pope and B. Slater, *Adv. Mater.*, 2020, **32**, e2002780.
- 14 Q. Yang, Y. Li, J. D. Yang, Y. Liu, L. Zhang, S. Luo and J. P. Cheng, *Angew. Chem., Int. Ed.*, 2020, **59**, 19282–19291.
- 15 Baekjun Kim, Sangwon Lee and J. Kim, *Sci. Adv.*, 2020, **6**, eaax9324.
- 16 Z. Zhang, A. Mansouri Tehrani, A. O. Oliynyk, B. Day and J. Brgoch, *Adv. Mater.*, 2021, **33**, e2005112.
- 17 A. Mansouri Tehrani, A. O. Oliynyk, M. Parry, Z. Rizvi, S. Couper, F. Lin, L. Miyagi, T. D. Sparks and J. Brgoch, *J. Am. Chem. Soc.*, 2018, **140**, 9844–9853.
- 18 L. Mennel, J. Symonowicz, S. Wachter, D. K. Polyushkin, A. J. Molina-Mendoza and T. Mueller, *Nature*, 2020, **579**, 62–66.
- 19 K. H. Tu, H. Huang, S. Lee, W. Lee, Z. Sun, A. Alexander-Katz and C. A. Ross, *Adv. Mater.*, 2020, **32**, e2005713.
- 20 B. Wang, B. Xie, J. Xuan and K. Jiao, *Energy Convers. Manage.*, 2020, **205**, 112460.
- 21 K. Chen, S. Lagrouche and A. Djerdir, *Energy Convers. Manage.*, 2019, **195**, 810–818.
- 22 R. Ding, R. Wang, Y. Ding, W. Yin, Y. Liu, J. Li and J. Liu, *Angew. Chem., Int. Ed.*, 2020, **59**, 19175–19183.
- 23 X. Peng, D. Kulkarni, Y. Huang, T. J. Omasta, B. Ng, Y. W. Zheng, L. Q. Wang, J. M. LaManna, D. S. Hussey, J. R. Varcoe, I. V. Zenyuk and W. E. Mustain, *Nat. Commun.*, 2020, **11**, 3561.
- 24 A. Smith, A. Keane, J. A. Dumesic, G. W. Huber and V. M. Zavala, *Appl. Catal., B*, 2020, **263**, 118257.
- 25 N. Ramaswamy and S. Mukerjee, *Chem. Rev.*, 2019, **119**, 11945–11979.
- 26 K. Tanaka, K. Hachiya, W. Zhang, K. Matsuda and Y. Miyauchi, *ACS Nano*, 2019, **13**, 12687–12693.
- 27 N. Chen, C. Hu, H. H. Wang, S. P. Kim, H. M. Kim, W. H. Lee, J. Y. Bae, J. H. Park and Y. M. Lee, *Angew. Chem., Int. Ed.*, 2021, **60**, 7710–7718.
- 28 L. Wang, M. Bellini, H. A. Miller and J. R. Varcoe, *J. Mater. Chem. A*, 2018, **6**, 15404–15412.
- 29 S. Zaman, L. Huang, A. I. Douka, H. Yang, B. You and B. Y. Xia, *Angew. Chem., Int. Ed.*, 2021, **60**, 2–23.
- 30 W.-H. Lee, E. J. Park, J. Han, D. W. Shin, Y. S. Kim and C. Bae, *ACS Macro Lett.*, 2017, **6**, 566–570.
- 31 Y. Zhang, W. Chen, X. Yan, F. Zhang, X. Wang, X. Wu, B. Pang, J. Wang and G. He, *J. Membr. Sci.*, 2020, **598**, 117650.
- 32 L. Wang, J. J. Brink, Y. Liu, A. M. Herring, J. Ponce-González, D. K. Whelligan and J. R. Varcoe, *Energy Environ. Sci.*, 2017, **10**, 2154–2167.
- 33 D. R. Dekel, S. Willdorf, U. Ash, M. Amar, S. Pusara, S. Dhara, S. Srebnik and C. E. Diesendruck, *J. Power Sources*, 2018, **375**, 351–360.
- 34 M. Bragato, G. F. von Rudorff and O. A. von Lilienfeld, *Chem. Sci.*, 2020, **11**, 11859–11868.
- 35 R. L. Greenaway and K. E. Jelfs, *Adv. Mater.*, 2021, **33**, e2004831.
- 36 D. H. McDANIEL and H. C. Brown, *J. Org. Chem.*, 1958, **23**, 420–427.
- 37 C. Hansch, A. Leo and R. W. Taft, *Chem. Rev.*, 1991, **91**, 165–195.
- 38 S. Kim, N. Kim, J. Seo, J. E. Park, E. H. Song, S. Y. Choi, J. E. Kim, S. Cha, H. H. Park and J. M. Nam, *Sci. Adv.*, 2020, **6**, eabb3348.
- 39 M. Dou and M. Fyta, *J. Mater. Chem. A*, 2020, **8**, 23511–23518.
- 40 J. Y. Chu, K. H. Lee, A. R. Kim and D. J. Yoo, *ACS Sustainable Chem. Eng.*, 2019, **7**, 20077–20087.
- 41 S. Maurya, S. Noh, I. Matanovic, E. J. Park, C. Narvaez Villarrubia, U. Martinez, J. Han, C. Bae and Y. S. Kim, *Energy Environ. Sci.*, 2018, **11**, 3283–3291.
- 42 C. G. Arges and V. Ramani, *Proc. Natl. Acad. Sci. U. S. A.*, 2013, **110**, 2490–2495.
- 43 C. M. Simon, J. Kim, D. A. Gomez-Gualdrón, J. S. Camp, Y. G. Chung, R. L. Martin, R. Mercado, M. W. Deem, D. Gunter, M. Haranczyk, D. S. Sholl, R. Q. Snurr and B. Smit, *Energy Environ. Sci.*, 2015, **8**, 1190–1199.
- 44 B. Lin, F. Xu, F. Chu, Y. Ren, J. Ding and F. Yan, *J. Mater. Chem. A*, 2019, **7**, 13275–13283.
- 45 X. Tan, Z. Sun, J. Pan, J. Zhao, H. Cao, H. Zhu and F. Yan, *J. Membr. Sci.*, 2021, **618**, 118689.
- 46 M. Zhu, M. Zhang, Q. Chen, Y. Su, Z. Zhang, L. Liu, Y. Wang, L. An and N. Li, *Polym. Chem.*, 2017, **8**, 2074–2086.
- 47 N. Li, Y. Leng, M. A. Hickner and C. Y. Wang, *J. Am. Chem. Soc.*, 2013, **135**, 10124–10133.
- 48 C. M. Tuan and D. Kim, *J. Membr. Sci.*, 2016, **511**, 143–150.
- 49 H. S. Dang and P. Jannasch, *J. Mater. Chem. A*, 2016, **4**, 17138–17153.
- 50 X. J. Zhang, Y. J. Cao, M. Zhang, Y. D. Huang, Y. G. Wang, L. Liu and N. W. Li, *J. Membr. Sci.*, 2020, **596**, 117700.
- 51 J. K. Hao, Y. Y. Jiang, X. Q. Gao, W. T. Lu, Y. Xiao, Z. G. Shao and B. L. Yi, *J. Membr. Sci.*, 2018, **548**, 1–10.
- 52 J. Pan, H. Zhu, H. Cao, B. Wang, J. Zhao, Z. Sun and F. Yan, *J. Membr. Sci.*, 2021, **620**, 118794.
- 53 C. Empel and R. M. Koenigs, *Angew. Chem., Int. Ed.*, 2019, **58**, 17114–17116.
- 54 V. L. Deringer, M. A. Caro and G. Csanyi, *Adv. Mater.*, 2019, **31**, e1902765.
- 55 A. F. Belhaj, K. A. Elraies, M. S. Alnarabiji, F. A. Abdul Kareem, J. A. Shuhli, S. M. Mahmood and H. Belhaj, *Chem. Eng. J.*, 2021, **406**, 127081.
- 56 Y. Xie, J. Zou, Z. Li, F. Gao and C. Peng, *IEEE Access*, 2020, **8**, 176661–176675.
- 57 S. Abdullah, R. C. Pradhan, D. Pradhan and S. Mishra, *Food Chem.*, 2021, **339**, 127862.
- 58 W. Beker, E. P. Gajewska, T. Badowski and B. A. Grzybowski, *Angew. Chem., Int. Ed.*, 2019, **58**, 4515–4519.
- 59 X. Dong, B. X. Xue, H. D. Qian, J. F. Zheng, S. H. Li and S. B. Zhang, *J. Power Sources*, 2017, **342**, 605–615.
- 60 F. H. Liu, C. X. Lin, E. N. Hu, Q. Yang, Q. G. Zhang, A. M. Zhu and Q. L. Liu, *J. Membr. Sci.*, 2018, **564**, 298–307.
- 61 Y. L. Jiang, J. B. Liao, S. S. Yang, J. Li, Y. Q. Xu, H. M. Ruan, A. Sotto, B. Van der Bruggen and J. N. Shen, *React. Funct. Polym.*, 2018, **130**, 61–69.
- 62 H. Lim and T. H. Kim, *Macromol. Res.*, 2017, **25**, 1220–1229.