Electronic Supplementary Material (ESI) for Journal of Materials Chemistry A.
This journal is © The Royal Society of Chemistry 2021

Supplementary Information for

Data-driven and DFT Assisted Theoretic Guide for Membrane Design in Flow Batteries

Tianyu Li[1], *Wenjing Lu*[1], *Zhizhang Yuan*, *Huamin Zhang*, *Xianfeng Li*[*]

Division of energy storage, Dalian National Laboratory for Clean Energy (DNL), Dalian Institute of Chemical Physics, Chinese Academy of Sciences, Zhongshan Road 457, Dalian 116023, China

[*]Corresponding author, E-mail address: lixianfeng@dicp.ac.cn

[1] These authors contribute equally to this work

Methods

Membrane fabrication. PBI was fabricated based on our previous study[1]. A certain amount (17 wt%) of polybenizimidazole (PBI) was firstly dissolved in dimethylacetamid (DMAc) to form a homogeneous polymer solution. Then the solution was cast on a dry and clean glass plate using a doctor blade (Elcometer 3545 adjustable Bird Coater, Scraper, Elcometer 3545/8) at room temperature with humidity less than 60% to avoid the penetration of water vapor into the polymer solution. The glass plate was immersed into water immediately to form the membrane with a porous structure. Then the membrane was peeled off and soaked in water before solvent treatment.

Subsequentially, membranes with fixed size were immersed into the treating solvent for 30 minutes. The membranes were then put at room temperature for at least 24 h to evaporate the solvent completely. Finally, the membranes were stored in water for use.

Establishment and analysis of dataset. The total dataset includes 98 samples which are from the experiments[2] and the details are list in supplementary Dataset 1. The data is spilt randomly with 75 % of samples in training set and 25 % of samples in testing set during modeling. The Pearson correlation coefficients between each feature and efficiencies (VE and EE) were calculated for correlation analysis. The results are shown in Figure S1. The current density has strong linear correlation with VE and EE, while other features have weak linear correlation with VE and EE so linear regression was chosen. ANN with no hidden layer is capable of representing linear separable functions or decisions. ANN with 1 hidden layer can approximate any function that contains a continuous mapping from one finite space to another. ANN with 2 hidden layers can represent an arbitrary decision boundary to arbitrary accuracy with rational activation functions and can approximate any smooth mapping to any accuracy. ANN with more than 2 hidden layers shows that additional layers can learn complex representations (sort of automatic feature engineering) for layer layers. Since our dataset is simple and a single sufficiently large hidden layer is adequate for approximation of most functions, so we also choose ANN to build the model.

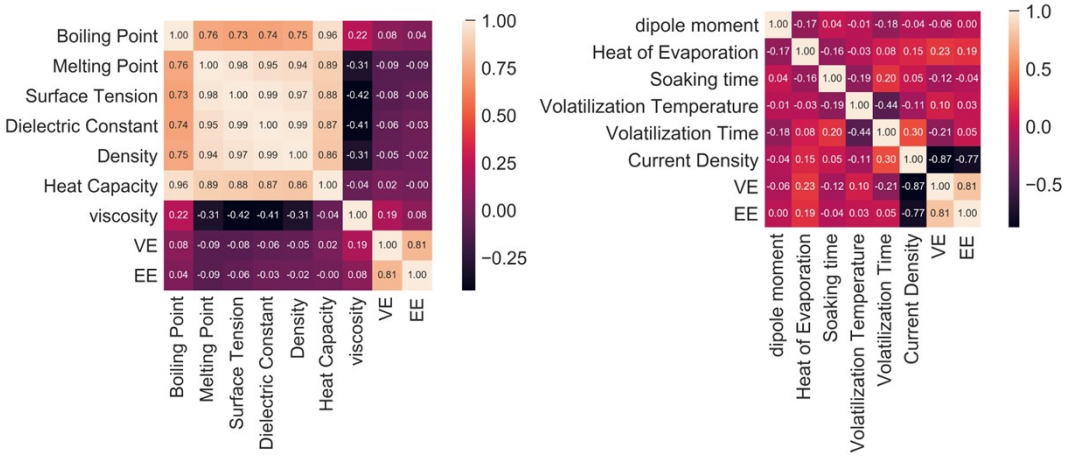

Figure S1. Pearson correlation coefficient between each feature and efficiencies (VE and EE)

Performance prediction model and evaluation criterion. The data pre-processing and ML modeling are carried out using the Python 3.7 package including the NumPy, pandas, sklearn and Tensorflow 2.3.0 packages. The whole dataset is spilt into Linear regression and a fully connected neural network were applied to handle the multidimensional data and to predict the performance of PBI membrane treated with different solvents, respectively. The linear regression model was as follows,

$$f(\mathbf{X},\mathbf{w},b)=\mathbf{w}^{\mathrm{T}}\mathbf{X}+b \tag{1}$$

Where $\mathbf{X}=<x_1, x_2,\ldots, x_n>$ was the n-dimensional feature vector including the experiment parameters and solvent properties. $\mathbf{w}=<w_1, w_2,\ldots, w_n>$ was corresponding coefficient vector of $\mathbf{X}$, and b was the bias (or intercept). Here, $f(\mathbf{X},\mathbf{w},b)$ was the predicted performance (VE or EE) of the PBI membrane. The generalized least squares method was employed to minimize the loss function $L(\mathbf{w},b)$, and the objective formula is as follows,

$$\underset{\mathbf{w},b}{\operatorname{argmin}} L(\mathbf{w},b)=\underset{\mathbf{w},b}{\operatorname{argmin}} \sum_{\mathrm{m}}^{\mathrm{i}=1}\left(f^{\mathrm{i}}(\mathbf{X},\mathbf{w},b)-y^{\mathrm{i}}\right)^2 \tag{2}$$

Where $\underset{\mathbf{w},b}{\operatorname{argmin}} L(\mathbf{w},b)$ means to optimize **w**, b to minimize the cost function $L(\mathbf{w},b)$. **y**=<y^1, y^2,…, y^m> was the experiment value, and y^i was one component in **y**. $f^i(\mathbf{X},\mathbf{w},b)$ was the predicted value of x^i. m was the number of the data in the dataset. ANN was built with one hidden layer including 32 units. The activation function was 'relu', the training epochs was 200. The optimizer was 'adam' with default parameters. The loss of the training set and testing set for VE and EE was shown in Figure S2(a) and S2(b), respectively.

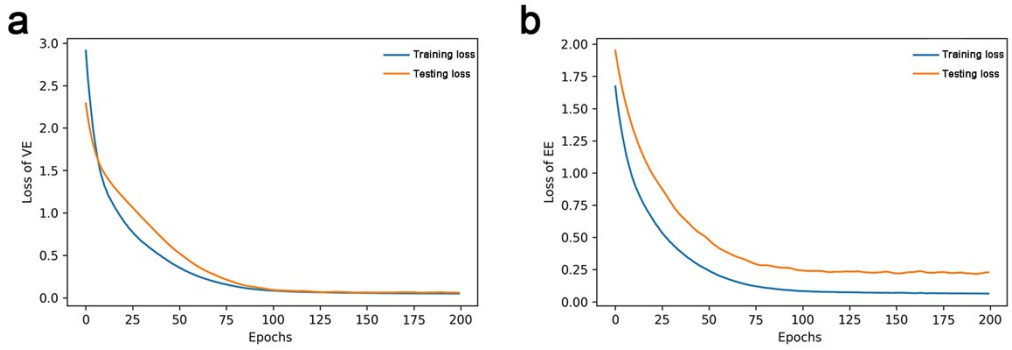

Figure S2. The loss for (a) VE and (b) EE

To evaluate the accuracy of performance prediction model, three standards were applied and can be calculated as follows,

1) Coefficient of determination (**R^2**)

$$R^2 = 1 - \frac{RSS}{TSS} = 1 - \frac{\sum_i (y^i - f^i)^2}{\sum_i (y^i - \hat{y})^2} \quad (3)$$

Where RSS is the residual sum of squares and TSS is the total sum of squares. f^i is the prediction value corresponding to x^i and $\hat{y}$ is the average value experiment data.

2) Mean square error (**MSE**)

$$MSE = \frac{1}{m}\sum_i (f_i - \hat{f})^2 \quad (4)$$

Where $\hat{f}$ is the average value of the predicted data.

3) Mean absolute prediction error (**MAPE**)

$$\mathrm{MAPE} = \frac{1}{m}\sum_{i=1}^{m}\left|\frac{f_i - y^i}{y^i}\right| \times 100\% \qquad (5)$$

Computational details. Density functional theory (DFT) calculations were performed to further reveal the interaction between porous membrane and various solvents. The structure of PBI segment, PBI-10 and solvents were optimized with M06-2X[3] hybrid functional with def2-svp[4] basis set using Gaussian 16[5] software. In order to investigate the weak interaction between molecules, the atom-pairwise dispersion correction (DFT-D3)[6] was also included. The structures of PBI segment combined with each solvent were optimized at the same functional and basis set. The binding energy was calculate with the same functional at triple zeta basis set def2-tzvp[4] and the basis set superposition error (BSSE) correction[7] was taken into account. Molecular electrostatic potential (ESP) on vdW surface was analyzed by wave function analyzer Multiwfn and drawn by VMD software.

The R^2 and MSE of 1, 2 and 3 hidden layers ANN is listed in Table S1.

Table S1. Coefficient of determination (R^2) and mean square error (MSE) of VE and EE for 1, 2 and 3 hidden layers ANN, respectively.

optimizer="adam"	Activation function="relu"	R2		MSE	
		Training	Test	Training	Test
ANN-1L	VE %	0.9465	0.9224	0.2636	0.5013
	EE %	0.9343	0.8244	0.2239	0.8690
ANN-2L	VE %	0.9501	0.9055	0.2501	0.6107
	EE %	0.9477	0.8311	0.1891	0.8358
ANN-3L	VE %	0.9516	0.9260	0.2390	0.4782
	EE %	0.9487	0.8463	0.1816	0.7607

The R^2 and MSE of “adam”, “SGD”, “Adagrad” optimizer in 1 hidden layer ANN is listed in Table S2.

Table S2. Coefficient of determination (R^2) and mean square error (MSE) of VE and EE for optimizer “adam”, “SGD” and “Adagrad”, respectively.

ANN-1L	Activation function="relu"	R^2		MSE	
		Training	Test	Training	Test
adam	VE %	0.9465	0.9224	0.2636	0.5013
	EE %	0.9343	0.8244	0.2239	0.8690
SGD	VE %	0.9374	0.8878	0.3037	0.7248
	EE %	0.9193	0.7682	0.2776	1.1475
Adagrad	VE %	-3.1096	-0.0248	4.2382	6.6205
	EE %	-12.2486	0.2181	2.7798	3.8700

The R^2 and MSE of “relu”, “exponential” and “linear” activation function in 1 hidden layer ANN is listed in Table S3.

Table S3. Coefficient of determination (R^2) and mean square error (MSE) of VE and EE for activation function “relu”, “exponential” and “linear”, respectively.

ANN-1L	optimizer ="adam"	R^2		MSE	
		Training	Test	Training	Test
relu	VE %	0.9465	0.9224	0.2636	0.5013
	EE %	0.9343	0.8244	0.2239	0.8690
exponential	VE %	0.9365	0.9461	0.3106	0.4381
	EE %	0.9037	0.8211	0.3281	0.8857
linear	VE %	0.9322	0.9413	0.3329	0.3793
	EE %	0.8967	0.8069	0.3544	0.9557

The R^2 and MSE of Linear regression, Dummy regression (mean) and Dummy regression (median) is listed in Table S4.

Table S4. Coefficient of determination (R^2) and mean square error (MSE) of VE and EE for Linear regression, Dummy regression (mean) and Dummy regression (median).

		R^2		MSE	
		Training	Test	Training	Test
Linear regression	VE %	0.9363	0.9384	0.3317	0.3977
	EE %	0.9062	0.8141	0.3522	0.9201
Dummy regression (mean)	VE %	0.0000	-0.0364	2.5578	2.5578
	EE %	0.0000	-0.1358	3.6728	2.0707
Dummy regression (median)	VE %	-0.0417	-0.0111	5.4004	2.4954
	EE %	-0.0222	-0.0247	3.7543	1.8681

Supplementary references

1. Z. Yuan, Y. Duan, H. Zhang, X. Li, H. Zhang and I. Vankelecom, *Energy Environ. Sci.*, 2016, **9**, 441-447.
2. W. Lu, Z. Yuan, Y. Zhao, L. Qiao, H. Zhang and X. Li, *Energy Storage Mater.*, 2018, **10**, 40-47.
3. Y. Zhao and D. G. Truhlar, *Acc. Chem. Res.*, 2008, **41**, 157-167.
4. F. Weigend and R. Ahlrichs, *Phys. Chem. Chem. Phys.*, 2005, **7**, 3297-3305.
5. M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A. V. Marenich, J. Bloino, B. G. Janesko, R. Gomperts, B. Mennucci, H. P. Hratchian, J. V. Ortiz, A. F. Izmaylov, J. L. Sonnenberg, Williams, F. Ding, F. Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V. G. Zakrzewski, J. Gao, N. Rega, G. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. Throssell, J. A. Montgomery Jr., J. E. Peralta, F. Ogliaro, M. J. Bearpark, J. J. Heyd, E. N. Brothers, K. N. Kudin, V. N. Staroverov, T. A. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. P. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, J. M. Millam, M. Klene, C. Adamo, R. Cammi, J. W. Ochterski, R. L. Martin, K. Morokuma, O. Farkas, J. B. Foresman and D. J. Fox, *Journal*, 2016.
6. S. Grimme, *J. Comput. Chem.*, 2004, **25**, 1463-1473.
7. S. F. Boys and F. Bernardi, *Mol. Phys.*, 1970, **19**, 553-566.