Supporting Information

Predicting the performance of polyvinylidene fluoride, polyethersulfone and polysulfone filtration membranes using machine learning

Tingli Liu,[ab] Lunyang Liu,[*ab] Fengchao Cui,[ab] Fang Ding,[ab] Qifeng Zhang[ab] and Yunqi Li[*ab]

[a] Key Laboratory of High-Performance Synthetic Rubber and its Composite Materials & Key Laboratory of Polymer Ecomaterials, Changchun Institute of Applied Chemistry (CIAC), Chinese Academy of Sciences, Changchun 130022, China;
[b] University of Science and Technology of China, Hefei 230026, China.

* To whom correspondence should be addressed. E-mail: lyliu@ciac.ac.cn; yunqi@ciac.ac.cn. Phone: +86 (0)431 85262535

Figures and Tables in supplementary

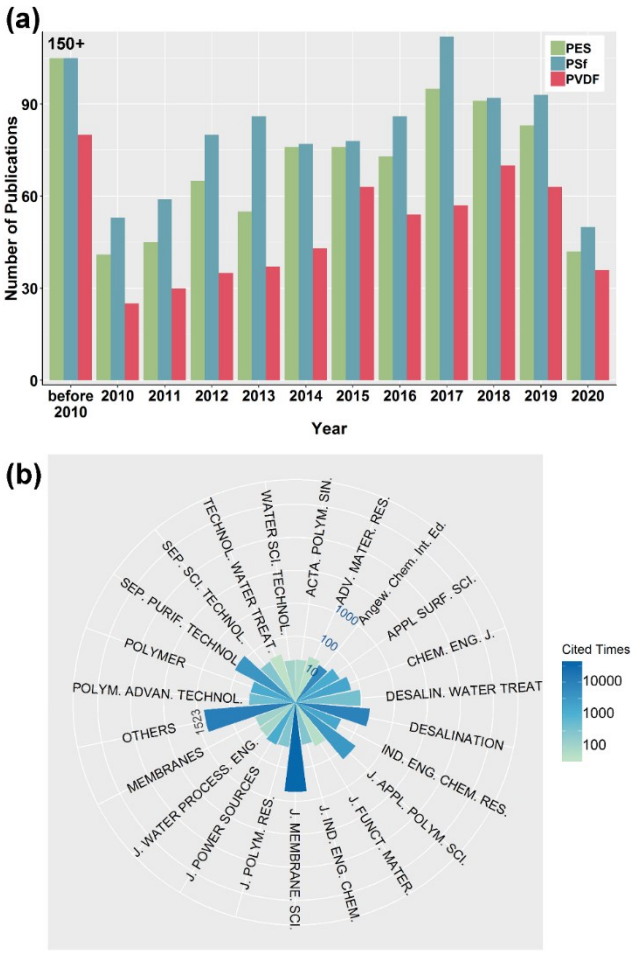

Fig. S1 (a) The histogram of annual publications for polyethersulfone, polysulfone and polyvinylidene fluoride filtration membranes, and (b) the distribution of these publications and their citations in various journals, the height and the color of bars correspond to the number of publications and the cited times accumulated in each journal, respectively.

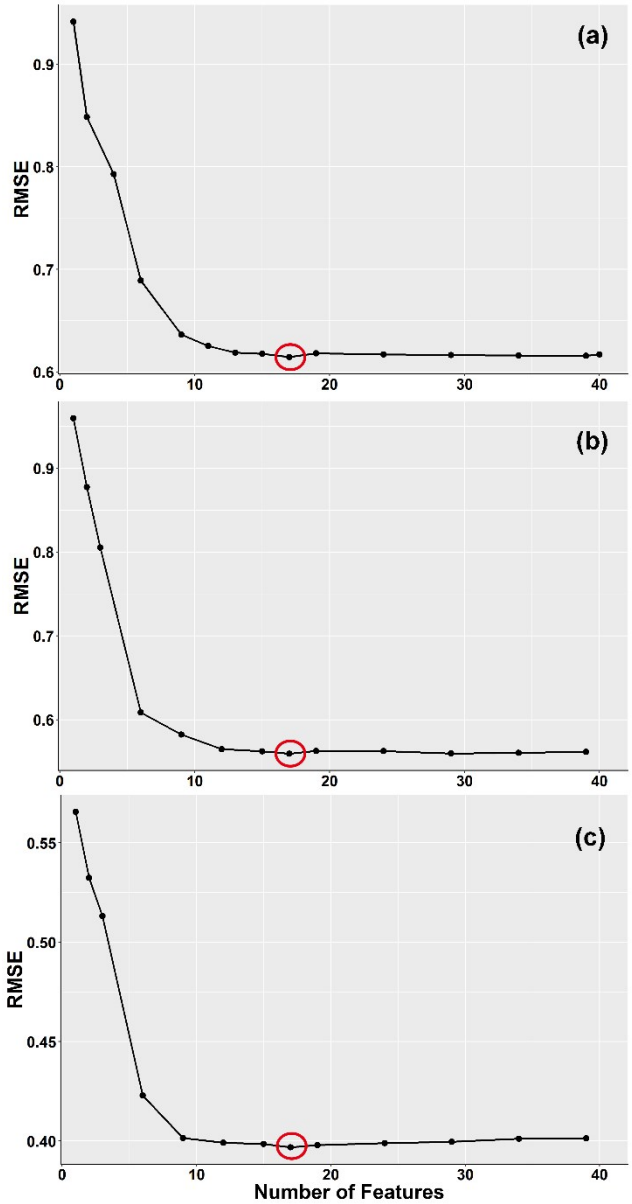

Fig. S2 The Root-Mean Square Error (RMSE) in the numeric regression for Lp (a), $1/S_0$ (b) and Tr (c) using Recursive Feature Elimination against the number of features. The red circle labeled the number of features (17) that finally engaged in predictive model construction.

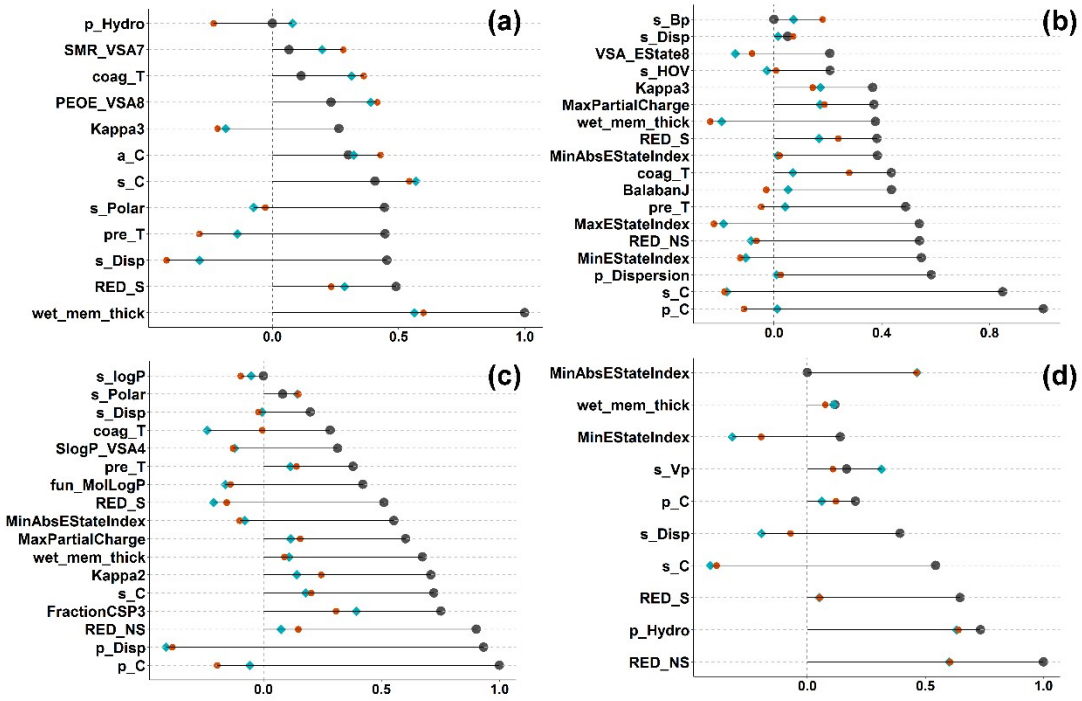

Fig. S3 The ranking of feature importance using RF algorithm (grey circle), Pearson correlation coefficient (Rp, blue diamond) and Spearman correlation coefficient (Rs, red dot) between features and structure parameters including thickness (a), porosity (b), surface water contact angle (CA) (c) and surface roughness (Sa) (d).

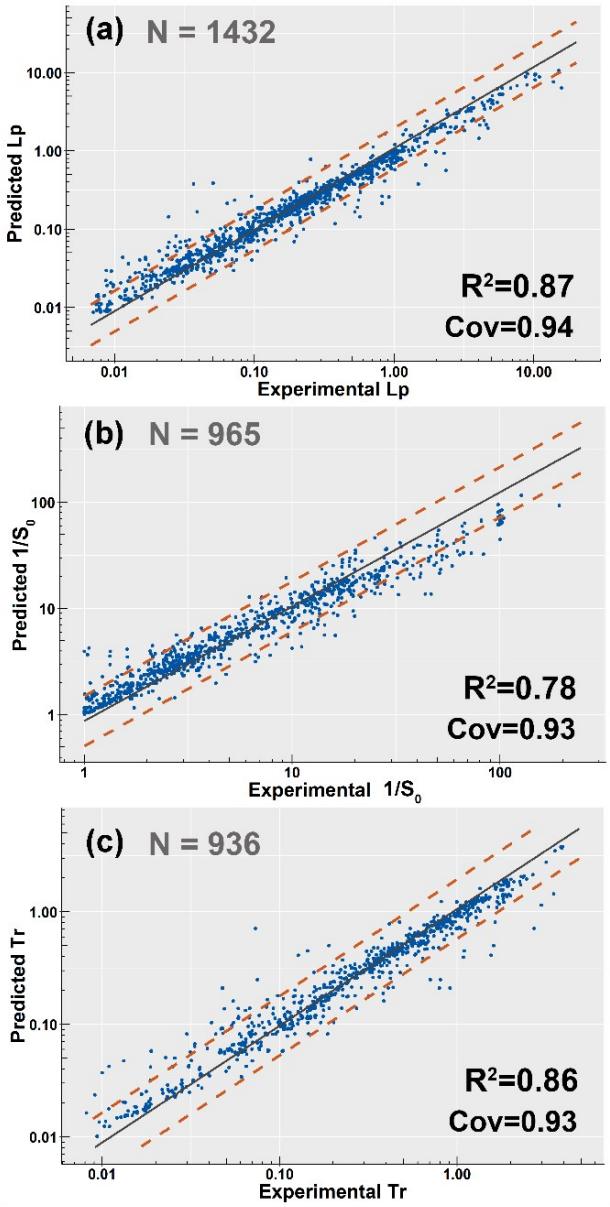

Fig. S4 Plots for the predicted vs. experimental values for models integrated four structural features in the prediction of Lp (a), $1/S_0$ (b) and Tr (c). The orange dashed lines indicate the upper and lower cutoff at 95% confidential coverage.

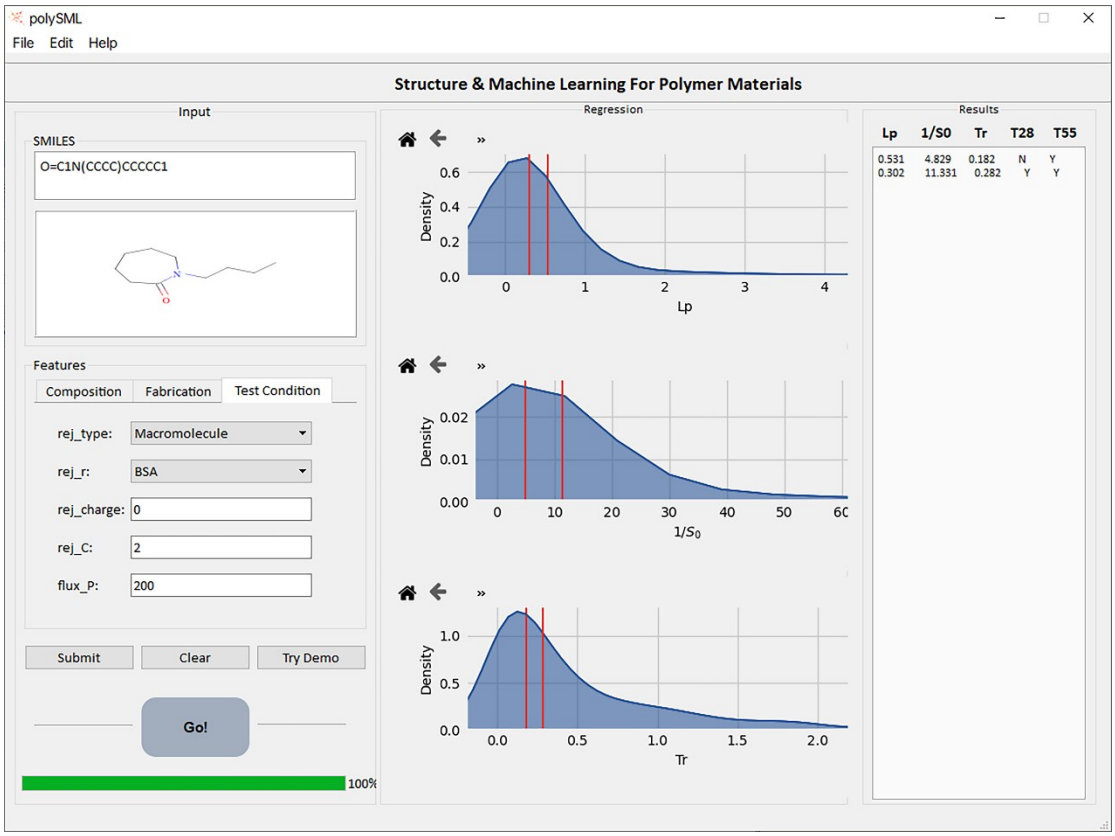

Fig. S5 The user interface of polySML. Parameters in composition, fabrication and test condition, as well as the chemical structure of additives can be readily adjusted in virtual experiments. Predicted performance indexes in the reference to the whole dataset are illustrated and exact predicted values and classification are presented.

Table S1 Description of Features in the construction of predictive models for PES/PSF/PVDF micro-/ultra-/nano-filtration membranes.

ID	Category	Feature	Unit	Description
1	Composition of casting solution	p_C		The weight fraction of base polymer
2		a_C	wt%	The weight fraction of addition agent in casting solution
3		s_C		The weight fraction of solvent in casting solution
4		p_Disp		Hansen solubility parameter (Dispersion force for monomer in polymer)
5		s_Disp		Hansen solubility parameter (Dispersion force for solvent)
6		RED_S	-	The relative energy difference between base polymer and solvent
7		RED_NS		The relative energy difference between base polymer and non-solvent
8		s_Bp	^{o}C	Boiling point for solvent and non-solvent
9		s_Vp	mmHg	The saturated vapor pressure at 25^{o}C for solvent and non-solvent
10		HDT	^{o}C	Heat Deflection Temperature with loading of 1.8MPa
11	Fabrication and test conditions	coag_T	^{o}C	The temperature of coagulation bath
12		pre_T	^{o}C	The temperature during membrane formation
13		exposed.time	s	The exposed time before immersing the casting solution into the non-solvent
14		wet_mem_thick	μm	The thickness of solution on the substrate controlled by the scraper
15		flux_P	kPa	Transmembrane pressure in performance measurement
16		rej_C	wt%	The concentration of substance (protein, salt etc,) in feeding flux

17		rej_type	-	The type of separation substance
18		rej_charge	-	The charge of separation substance
19		rej_r	nm	The radius of rejection substance
20	Structure parameters	thickness	μm	Thickness of dry membrane
21		porosity	%	Volume fraction of water accessible voids in membrane
22		CA	°	Water static contact angle on membrane surface
23		Sa	nm	Mean roughness of membrane surface
24	Computational features based on SMILES	fun_MolLogP	-	Wildman-Crippen LogP value of the functional group
25		MolLogP		Wildman-Crippen LogP value
26		Kappa2		Hall-Kier Kappa values, Rev. Comput. Chem. 2:367-422 (1991)
27		Kappa3		Hall-Kier Kappa values, Rev. Comput. Chem. 2:367-422 (1991)
28		MinAbsEStateIndex		Minimum absolute value of Estate index
29		MinEStateIndex		Minimum value of Estate index
30		MinPartialCharge		Minimum partial charge of molecular
31		MaxPartialCharge		Maximum partial charge of molecular
32		FractionCSP3		Fraction of C atoms that are SP3 hybridized.

Table S2 Prediction evaluation for performance indexes with structure features using Random Forest. Structure information has the best positive impact on the predictive models were shown in bold.

Performance index	**Without structure feature**		**With structure feature**			
	R^2	Cov	Feature added	Number of Vectors	R^2	Cov
Lp	0.85	0.83	porosity	845	0.89	0.94
			CA	**765**	**0.91**	**0.94**
			thickness	298	0.90	0.95
			Sa	196	0.87	0.93
$1/S_0$	**0.79**	**0.94**	**porosity**	**610**	**0.83**	**0.93**
			CA	551	0.80	0.94
			thickness	208	0.76	0.92
			Sa	141	0.76	0.92
Tr	0.83	0.92	porosity	591	0.81	0.94
			CA	**541**	**0.89**	**0.94**
			thickness	**206**	**0.89**	**0.95**
			Sa	**133**	**0.89**	**0.93**