

Supplementary Information

Machine Learning-Enabled Screening and Experimental Validation of Ionic Liquids for High-efficient Cellulose Dissolution

Bocheng Zhao^{‡ab}, Zixin Chen^{‡ab}, Kunchi Xie^a, Yuxin Qiu^a, Zhiwen Qi^a, Linfeng Lei^{*ab}, Zhen
Song^{*a}

^a *State Key Laboratory of Chemical Engineering, School of Chemical Engineering, East China
University of Science and Technology, 130 Meilong Road, Shanghai 200237, China*

^b *Suzhou Laboratory, Suzhou, 215100, China*

[‡] *These authors contributed equally to this work.*

**Corresponding authors: linfeng.lei@ecust.edu.cn; songz@ecust.edu.cn*

Number of Pages: 35

Number of Figures: 5

Number of Tables: 5

Contents

Figure Captions

Fig. S1. Original data distribution and Log-processed data distribute. -----Page S1

Fig. S2. Fitting results for different cellulose models combined with solubility units.

(a) Cellobiose-Mass fraction. (b) Cellotriose-Mass fraction. (c) Glucose-Molar solubility. (d) Cellobiose-Molar solubility. (e) Cellotriose-Molar solubility. -----

---Page S2

Fig. S3 Linear fitting results of (a) ILs contained in COSMObase and (b) ILs not contained in COSMObase-----Page S3

Fig. S4. Statistical results of the 5 times cross-validation of (a) morgan fingerprints-based models, (b) RDKit descriptors-based models and (c) Transformer-CNN model. -----Page S4

Fig. S5. Plot of experimental versus predicted values for the Transformer-CNN. -----
-----Page

S5

Table Captions

Table S1. The structure of all ILs in database. -----Page S6

Table S2. Search range and optimized hyperparameters of each model. -----Page S31

Table S3. The optimal hyperparameter for each fold of all models. -----Page S32

Table S4. Supplementary Metrics for All Models. -----Page S34

Table S5. Experimental validation results. -----Page

S35

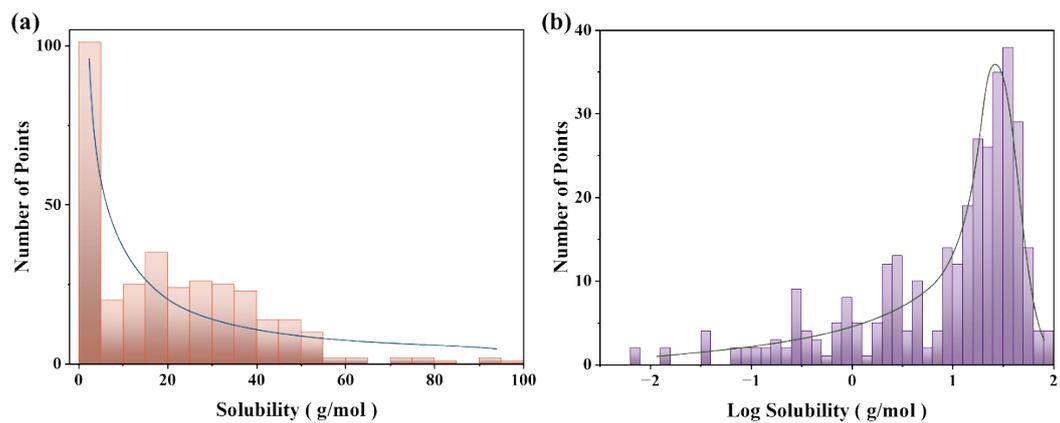


Fig. S1 Distribution of cellulose solubility data. (a)Original data distribution. (b) Log-processed data distribute.

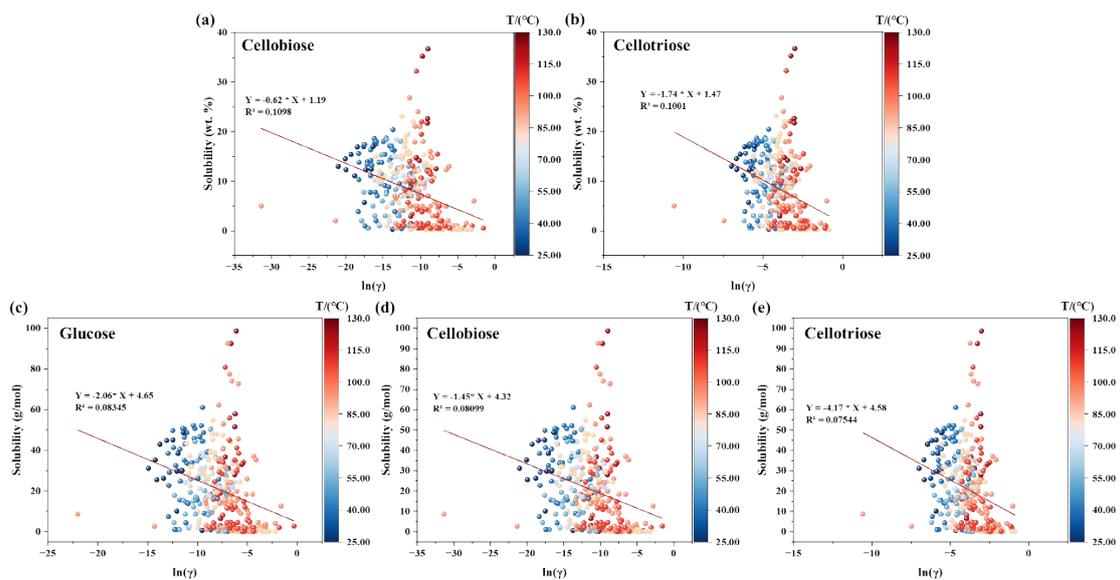


Fig. S2 Fitting results for different cellulose models combined with solubility units. (a) Cellulose-Mass fraction. (b) Cellotriose-Mass fraction. (c) Glucose-Molar solubility. (d) Cellulose-Molar solubility. (e) Cellotriose-Molar solubility.

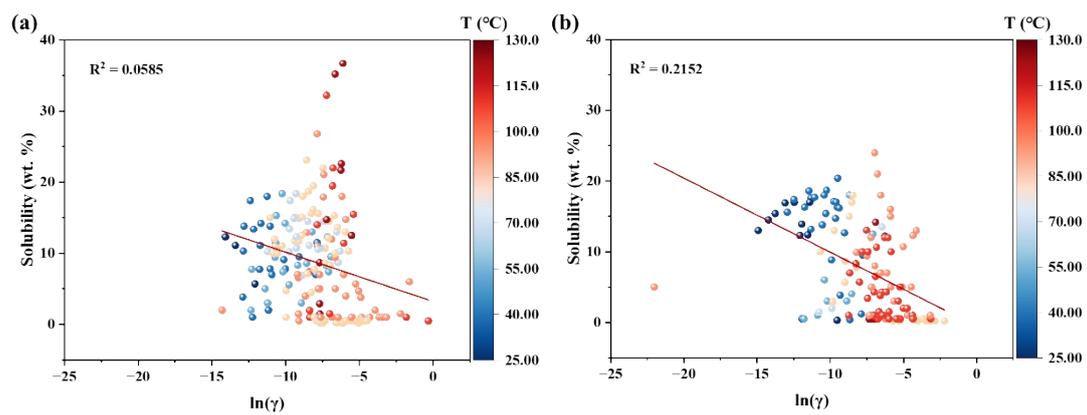


Fig. S3 Linear fitting results of (a) ILs contained in COSMObase and (b) ILs not contained in COSMObase.

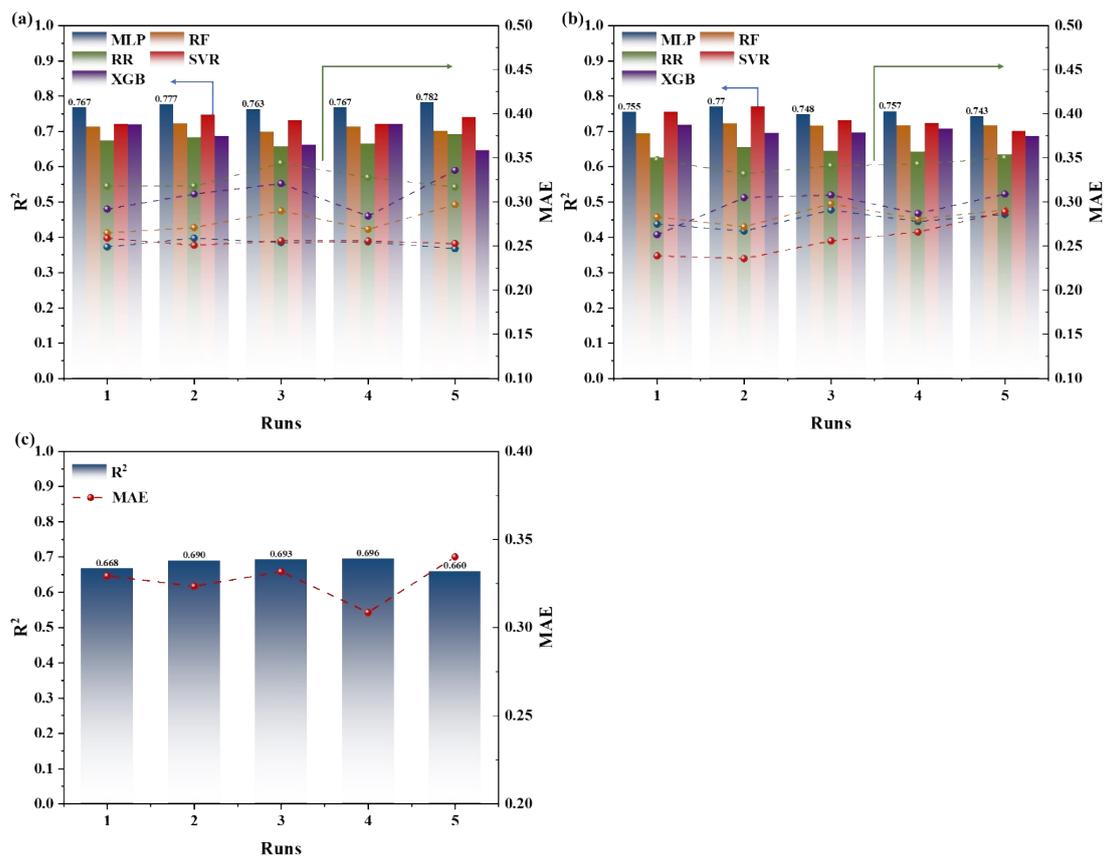


Fig. S4 Statistical results of the 5 times cross-validation of (a) morgan fingerprints-based models, (b) RDKit descriptors-based models and (c) Transformer-CNN model.

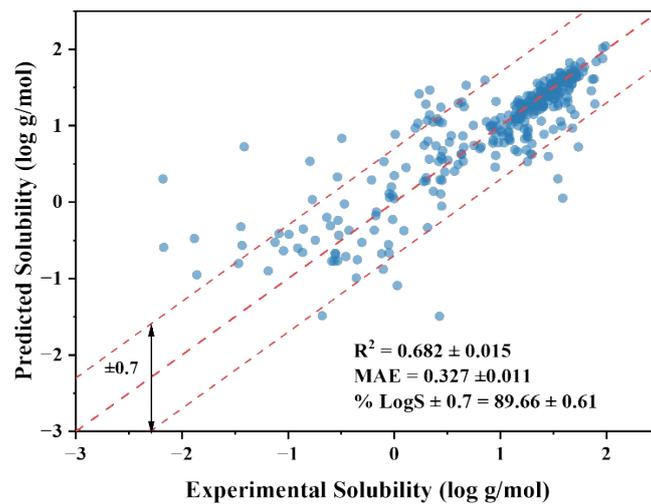
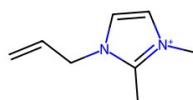


Fig. S5 Plot of experimental versus predicted values for the Transformer-CNN model.

Table S1. The structure of all ILs in database.

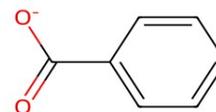
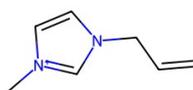
IL	Cation Structure	Anion Structure
$[(\text{CH}_3)_2\text{Aim}][\text{CH}_3\text{OCH}_2\text{COO}]$		
$[\text{A}_2\text{im}][\text{CH}_2=\text{CHCOO}]$		
$[\text{A}_2\text{im}][\text{CH}_3(\text{CH}_2)_2\text{COO}]$		
$[\text{A}_2\text{im}][\text{CH}_3\text{CH}_2\text{COO}]$		
$[\text{A}_2\text{im}][\text{CH}_3\text{CH}_2\text{OCH}_2\text{COO}]$		
$[\text{A}_2\text{im}][\text{CH}_3\text{COO}]$		
$[\text{A}_2\text{im}][\text{CH}_3\text{OCH}_2\text{COO}]$		
$[\text{A}_2\text{im}][\text{HOCH}_2\text{COO}]$		

[Admim][Br]

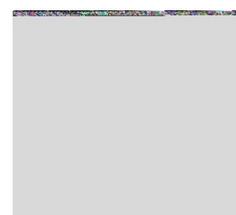
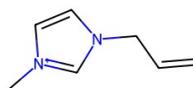


Br⁻

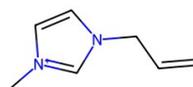
[Amim][C₆H₅COO]



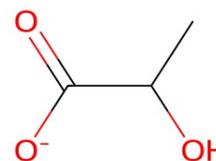
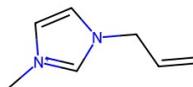
[Amim][CH₃(CH₂)₂COO]



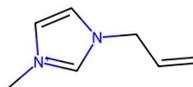
[Amim][CH₃CH₂COO]



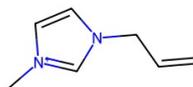
[Amim][CH₃CHOHCOO]



[Amim][CH₃COO]

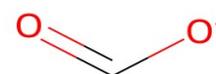
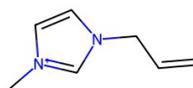


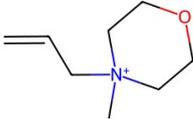
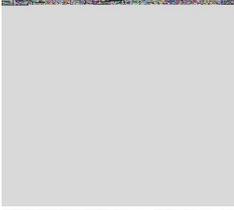
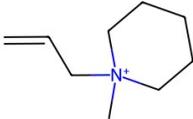
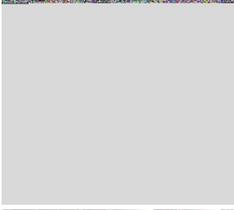
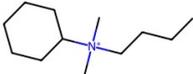
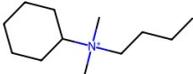
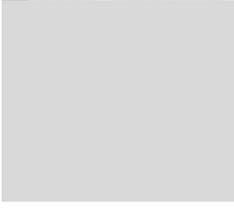
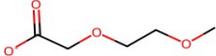
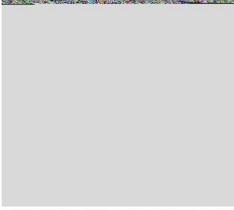
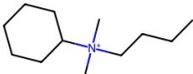
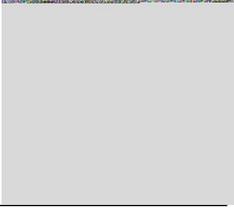
[Amim][Cl]

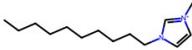
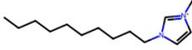
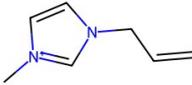
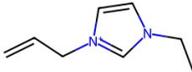
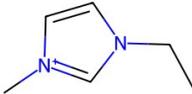
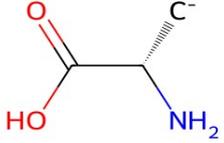
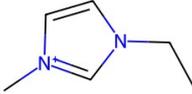
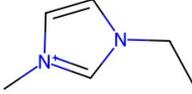
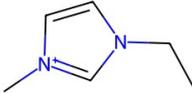


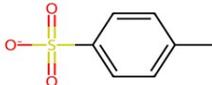
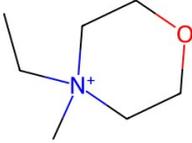
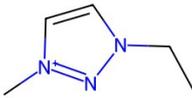
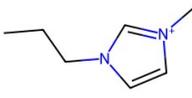
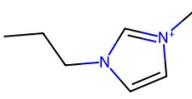
Cl⁻

[Amim][Formate]

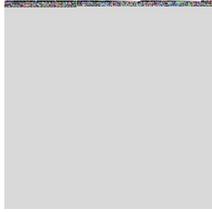
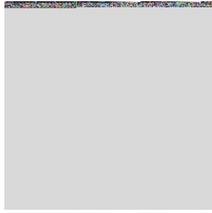
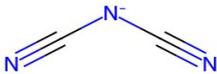
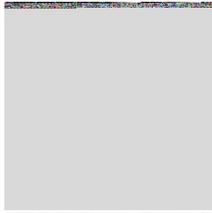
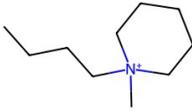
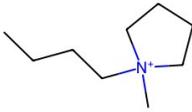


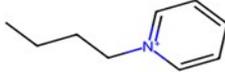
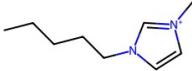
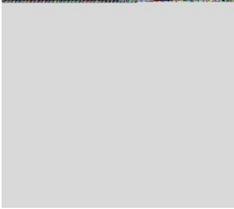
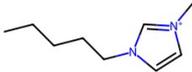
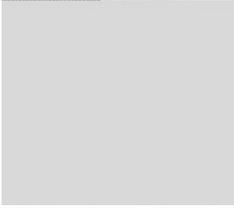
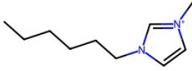
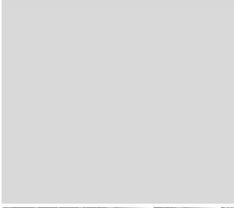
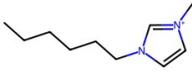
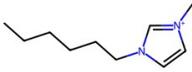
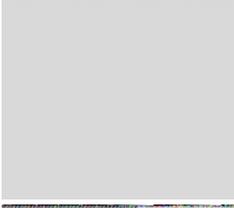
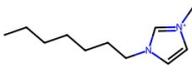
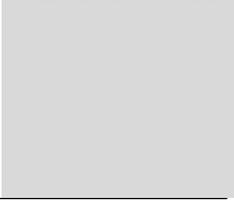
[Amim][HOCH ₂ COO]		
[Ammor][CH ₃ COO]		
[Ampip][CH ₃ COO]		
[But-CDM-N][CH ₃ CH ₂ OCH ₂ COO]		
[But-CDM-N][CH ₃ COO]		
[But-CDM-N][CH ₃ OCH ₂ CH ₂ OCH ₂ COO]		
[But-CDM-N][CH ₃ OCH ₂ COO]		
[But-CDM-N][Formate]		

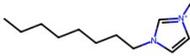
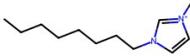
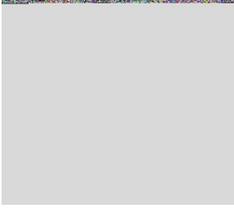
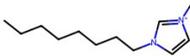
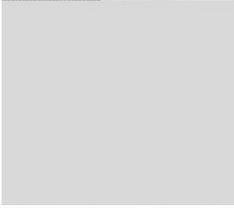
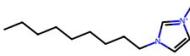
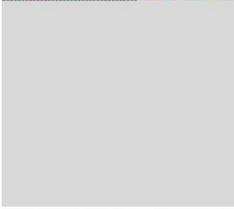
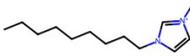
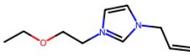
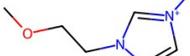
[C ₁₀ mim][CH ₃ COO]		
[C ₁₀ mim][Cl]		
[C ₁ Aim][CH ₃ OCH ₂ COO]		
[C ₂ Aim][CH ₃ OCH ₂ COO]		
[C ₂ mim][Ala]		
[C ₂ mim][Br]		
[C ₂ mim][CH ₃ COO]		
[C ₂ mim][DEP]		

$[C_2mim][F]$		F^-
$[C_2mim][MEPA]$		
$[C_2mim][CH_3CHOHCOO]$		
$[C_2mim][TsO]$		
$[C_2mmor][CH_3COO]$		
$[C_2mtr][CH_3CHOHCOO]$		
$[C_3mim][Br]$		
$[C_3mim][CH_3COO]$		

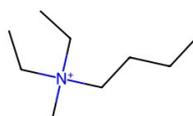
[C ₃ mim][Cl]		
[C ₄ Aim][CH ₃ OCH ₂ COO]		
[C ₄ dmim][CH ₃ COO]		
[C ₄ dmim][Cl]		
[C ₄ mim][Br]		
[C ₄ mim][CH ₃ (CH ₂) ₂ COO]		
[C ₄ mim][CH ₃ CH ₂ COO]		
[C ₄ mim][CH ₃ CHOHCOO]		

[C ₄ mim][CH ₃ COO]		
[C ₄ mim][CH ₃ OCH ₂ COO]		
[C ₄ mim][DCA]		
[C ₄ mim][Formate]		
[C ₄ mim][HOCH ₂ COO]		
[C ₄ mim][C ₆ H ₅ COO]		
[C ₄ mpip][CH ₃ COO]		
[C ₄ mPyr][CH ₃ COO]		

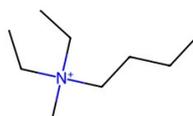
[C ₄ mPyr][CH ₃ OCH ₂ COO]		
[C ₄ Pyrd][CH ₃ OCH ₂ COO]		
[C ₅ mim][Br]		
[C ₅ mim][Cl]		
[C ₆ mim][Br]		
[C ₆ mim][CH ₃ COO]		
[C ₆ mim][Cl]		
[C ₇ mim][Br]		

[C ₇ mim][Cl]		
[C ₈ mim][Br]		
[C ₈ mim][CH ₃ COO]		
[C ₈ mim][Cl]		
[C ₉ mim][Br]		
[C ₉ mim][Cl]		
[CH ₃ CH ₂ OCH ₂ CH ₂ Aim][CH ₃ OCH ₂ COO]		
[CH ₃ OCH ₂ CH ₂ min][CH ₃ COO]		

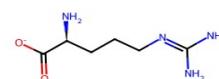
[DEMB][MEPA]



[DEMB][CH₃OCH₂COO]



[DEME][Arg]



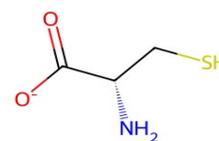
[DEME][Asp]



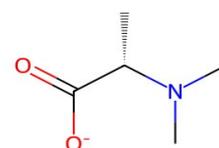
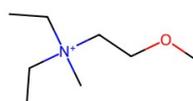
[DEME][CH₃(CH₂)₂COO]



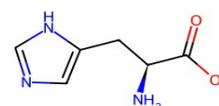
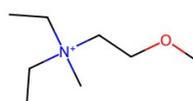
[DEME][Cys]

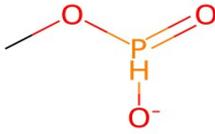
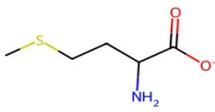
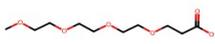
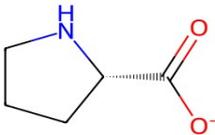


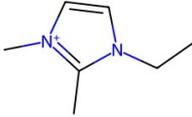
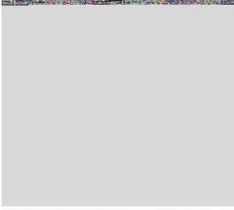
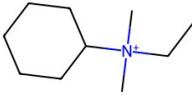
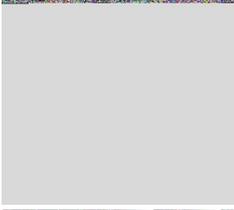
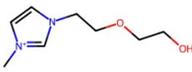
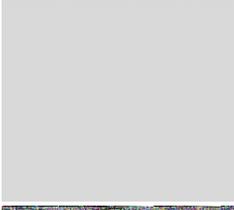
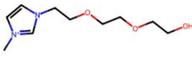
[DEME][DMA]

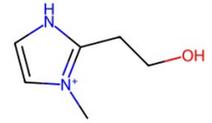
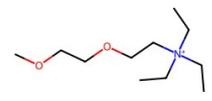
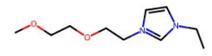
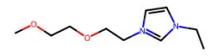


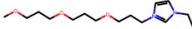
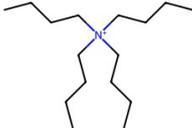
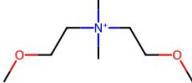
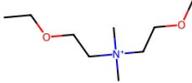
[DEME][His]

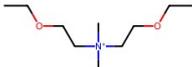
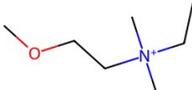
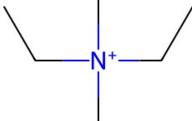
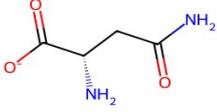
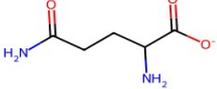
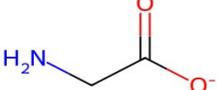


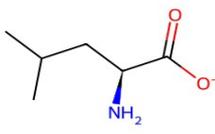
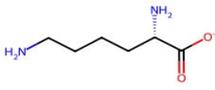
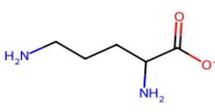
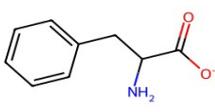
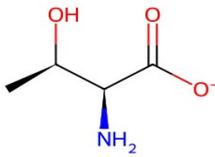
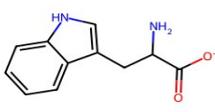
[DEME][MDEPA]	
[DEME][MeO(H)PO ₂]	
[DEME][MEPA]	
[DEME][Met]	
[DEME][CH ₃ OCH ₂ COO]	
[DEME][MTEPA]	
[DEME][OH]	HO ⁻
[DEME][Pro]	

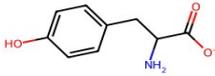
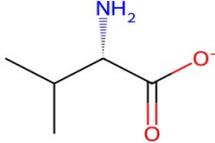
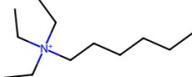
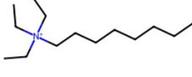
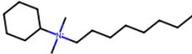
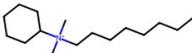
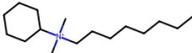
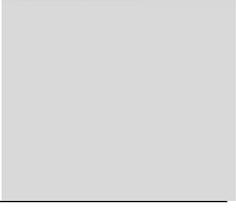
[DEME][CH ₃ CH ₂ COO]		
[Edmim][CH ₃ COO]		
[Et-CDM-N][CH ₃ COO]		
[H(OEt) ₂ -Me-Im][CH ₃ COO]		
[H(OEt) ₂ -Me-Im][Cl]		
[H(OEt) ₃ -Me-Im][CH ₃ COO]		
[Hex-CDM-N][CH ₃ CH ₂ OCH ₂ COO]		
[Hex-CDM-N][CH ₃ COO]		

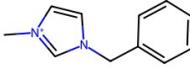
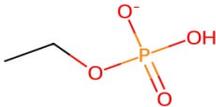
[Hex-CDM-N][CH ₃ OCH ₂ CH ₂ OCH ₂ COO]		
[Hex-CDM-N][CH ₃ OCH ₂ COO]		
[Hex-CDM-N][Formate]		
[HOCH ₂ CH ₂ mim][CH ₃ COO]		
[Me(OEt) ₂ -Et ₃ N][CH ₃ COO]		
[Me(OEt) ₂ -Et-Im][CH ₃ COO]		
[Me(OEt) ₂ -Et-Im][Cl]		
[Me(OEt) ₃ -Et ₃ N][CH ₃ COO]		

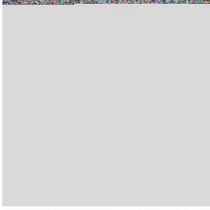
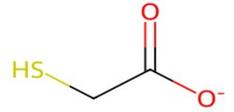
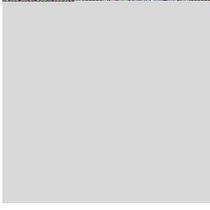
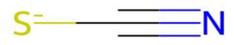
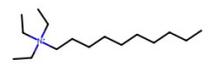
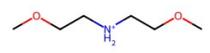
[Me(OEt) ₃ -Et-Im][CH ₃ COO]		
[Me(OEt) ₃ -MeOEtOMe-Im][CH ₃ COO]		
[Me(OEt) ₄ -Et-Im][CH ₃ COO]		
[Me(OEt) ₇ -Et-Im][CH ₃ COO]		
[Me(OPr) ₃ -Et-Im][CH ₃ COO]		
[N(CH ₂ CH ₂ CH ₂ CH ₃) ₄][Formate]		
[N _{11,102,102}][CH ₃ COO]		
[N _{11,102,202}][CH ₃ COO]		

[N _{11,202,202}][CH ₃ COO]		
[N _{112,102}][CH ₃ COO]		
[N ₂₂₁₁][Ala]		
[DEME][Ala]		
[DEME][Asn]		
[DEME][Gln]		
[DEME][Gly]		
[DEME][Ile]		

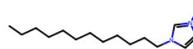
[DEME][Leu]	
[DEME][Lys]	
[DEME][CH ₃ COO]	
[DEME][Orn]	
[DEME][Phe]	
[DEME][Ser]	
[DEME][Thr]	
[DEME][Trp]	

[DEME][Tyr]		
[DEME][Val]		
[N ₂₂₂₆][CH ₃ CH ₂ COO]		
[N ₂₂₂₈][CH ₃ CH ₂ COO]		
[N ₄₂₂₁][Ala]		
[Oct-CDM-N][CH ₃ CH ₂ OCH ₂ COO]		
[Oct-CDM-N][CH ₃ COO]		
[Oct-CDM-N][CH ₃ OCH ₂ CH ₂ OCH ₂ COO]		

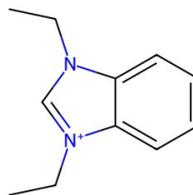
[Oct-CDM-N][CH ₃ OCH ₂ COO]		
[Oct-CDM-N][Formate]		
[P ₄₄₄₄][Formate]		
[P ₁₄₆₆₆][DCA]		
[phC ₁ mim][CH ₃ COO]		
[C ₂ mim][Cl]		
[C ₂ mim][EtPO ₄]		
[C ₄ mim][Cl]		

[C ₄ mim][Gly]		
[C ₄ mim][HSCH ₂ COO]		
[C ₄ mim][I]		
[C ₄ mim][SCN]		
[N ₂₂₂₁₀][CH ₃ COO]		
[N ₂₂₂₆][CH ₃ COO]		
[N ₂₂₂₈][CH ₃ COO]		
[(MeOEt) ₂ NH ₂][CH ₃ COO]		

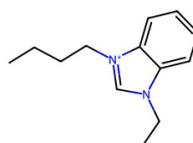
[C₁₂mim][CH₃COO]



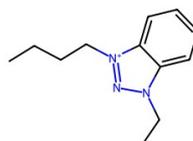
[C₂ebim][CH₃COO]



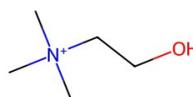
[C₄ebim][CH₃COO]



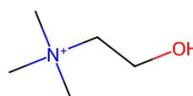
[C₄ebt][CH₃COO]



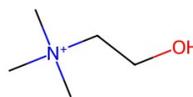
[Ch][Ala]



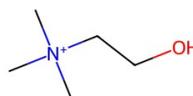
[Ch][Arg]

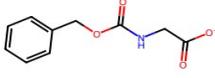
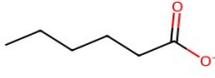
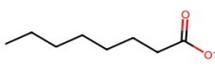


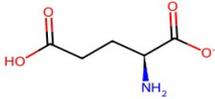
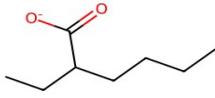
[Ch][Asn]



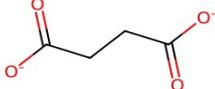
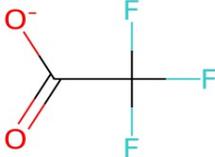
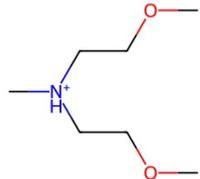
[Ch][Asp]



[Ch][Boc-Gly]	
[Ch][C ₆ H ₅ COO]	
[Ch][Cbz-Gly]	
[Ch][CH ₃ (CH ₂) ₂ COO]	
[Ch][CH ₃ (CH ₂) ₄ COO]	
[Ch][CH ₃ (CH ₂) ₆ COO]	
[Ch][CH ₃ CH ₂ OCH ₂ COO]	
[Ch][CH ₃ COO]	

[Ch][Gln]		
[Ch][Glu]		
[Ch][Gly]		
[Ch][Formate]		
[Ch][His]		
[Ch][HOCH ₂ COO]		
[Ch][Ile]		
[Ch][i-Oct]		

[Ch][Leu]		
[Ch][Lys]		
[Ch][Mal]		
[Ch][Met]		
[Ch][Nic]		
[Ch][Oxa]		
[Ch][Phe]		
[Ch][Piv]		

[Ch][Pro]		
[Ch][Ser]		
[Ch][Suc]		
[Ch][CF ₃ COO]		
[Ch][Thr]		
[Ch][Trp]		
[Ch][Val]		
[M(MeOEt) ₂ NH][CH ₃ COO]		

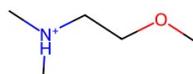
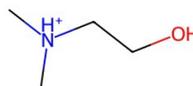


Table S2. Search range and optimized hyperparameters of each model.

Model	Searching Range
SVR	C: [1e-3, 1e3], epsilon: [1e-4, 1e-1], gamma: [scale, auto] kernel: [linear, poly, rbf, sigmoid], degree: [2, 5], coef0: [0.0, 1.0] learning_rat: [1e-6, 1e-1], max_dept: [1, 300], n_estimators: [1, 500], subsample: [0.3, 1.0], gamma: [0, 10], reg_alpha: [0, 20], colsample_bytree: [0.1, 1.0]
XGB	n_estimators: [1, 500], max_depth: [1, 500], min_samples_split: [2, 20], min_samples_leaf: [1, 20] learning_rate: [1e-5, 1e-1], hidden_dim: [16, 258], num_layers: [1, 4], dropout_rate: [.1, 0.5], weight_decay: [1e-5, 1e-1] alpha: [e-6, 1e3], fit_intercept: [True, False], solver: [auto, svd, cholesky, lsqr, saga], max_iter: [1000, 10000]
RF	
MLP	
RR	

Table S3. The optimal hyperparameter for each fold of all models.

Morgan fingerprints-based model					
	MLP	RF	RR	XGB	SVM
Fold1	learning_rate: 0.0016873, hidden_dim: 243, num_layer: 3, dropout_rate: 0.3280244, weight_decay: 2.4474057e-05	n_estimators: 144, max_depth: 296, min_samples_split: 2, min_samples_leaf: 1	alpha: 5.5221181, fit_intercept: True, solver: svd, max_iter: 8768	learning_rate: 0.0379317, max_depth: 23, n_estimators: 168, subsample: 0.7335811, gamma: 0.0055280, reg_alpha: 2.7623027, colsample_bytree: 0.3312467	C: 595.2285716, epsilon: 0.0005693, gamma: scale, kernel: rbf, degree: 2, coef0: 0.2786464
Fold2	learning_rate: 0.0219386, hidden_dim: 93, num_layers: 1, dropout_rate: 0.1163100, weight_decay: 0.0023097	n_estimators: 98, max_depth: 35, min_samples_split: 3, min_samples_leaf: 1	alpha: 16.5878299, fit_intercept: True, solver: saga, max_iter: 2237	learning_rate: 0.0584128, max_depth: 57, n_estimators: 255, subsample: 0.8220806, gamma: 0.0379699, reg_alpha: 3.7153948, colsample_bytree: 0.9291234	C: 1.5686880, epsilon: 0.0001430, gamma: scale, kernel: poly, degree: 3, coef0: 0.6815027
Fold3	learning_rate: 0.0219386, hidden_dim: 93, num_layers: 1, dropout_rate: 0.1163100, weight_decay: 0.0023097	n_estimators: 8, max_depth: 59, min_samples_split: 2, min_samples_leaf: 1	alpha: 5.5221181, fit_intercept: True, solver: svd, max_iter: 8768	learning_rate: 0.0998334, max_depth: 91, n_estimators: 497, subsample: 0.6970545, gamma: 0.0111517, reg_alpha: 1.7465061, colsample_bytree: 0.3397500	C: 1.0442814, epsilon: 0.0371081, gamma: scale, kernel: poly, degree: 5, coef0: 0.5754741
Fold4	learning_rate: 0.0042702 hidden_dim: 154, num_layers: 1, dropout_rate: 0.2470863, weight_decay: 0.0001150	n_estimators: 8, max_depth: 59, min_samples_split: 2, min_samples_leaf: 1	alpha: 5.5221181, fit_intercept: True, solver: svd, max_iter: 8768	learning_rate: 0.0709182, max_depth: 222, n_estimators: 78, subsample: 0.7848065, gamma: 0.0013194, reg_alpha: 0.5131404, colsample_bytree: 0.8773729	C: 16.7678974, epsilon: 0.0004354, gamma: scale, kernel: poly, degree: 5, coef0: 0.3482554
Fold5	learning_rate: 0.0044677, hidden_dim: 91, num_layers: 3, dropout_rate: 0.3186841, weight_decay: 5.4880470e-05	n_estimators: 98, max_depth: 35, min_samples_split: 3, min_samples_leaf: 1	alpha: 5.5221181, fit_intercept: True, solver: svd, max_iter: 8768	learning_rate: 0.0362867, max_depth: 285, n_estimators: 474, subsample: 0.6347476, gamma: 0.0022492, reg_alpha: 1.2489579, colsample_bytree: 0.3036714	C: 9.5946026, epsilon: 0.0646381, gamma: scale, kernel: rbf, degree: 3, coef0: 0.5968428
RDKit descriptors-based model					
	MLP	RF	RR	XGB	SVM

Fold1	learning_rate: 0.0005110, hidden_dim: 214, num_layers: 4, dropout_rate: 0.1027808, weight_decay: 0.0011040	n_estimators: 144, max_depth: 296, min_samples_split: 2, min_samples_leaf: 1	alpha: 215.8210282, fit_intercept: True, solver: saga, max_iter: 9783	learning_rate: 0.0234841, max_depth: 64, n_estimators: 342, subsample: 0.7715868, gamma: 0.0013762, reg_alpha: 0.0514792, colsample_bytree: 0.2573586	C: 4.0096010, epsilon: 0.0098880, gamma: auto, kernel: rbf, degree: 2, coef0: 0.7380336
Fold2	learning_rate: 0.0015665, hidden_dim: 184, num_layers: 3, dropout_rate: 0.1897077, weight_decay: 0.0070585	n_estimators: 144, max_depth: 296, min_samples_split: 2, min_samples_leaf: 1	alpha: 400.7630614, fit_intercept: True, solver: cholesky, max_iter: 6796	learning_rate: 0.0441347, max_depth: 106, n_estimators: 377, subsample: 0.6909028, gamma: 0.0450555, reg_alpha: 1.5644745, colsample_bytree: 0.4439324	C: 9.5946026, epsilon: 0.0646381, gamma: scale, kernel: rbf, degree: 3, coef0: 0.5968428
Fold3	learning_rate: 0.0015736, hidden_dim: 189, num_layers: 3, dropout_rate: 0.2119735, weight_decay: 0.0659874	n_estimators: 446, max_depth: 169, min_samples_split: 9, min_samples_leaf: 2	alpha: 85.3496689, fit_intercept: True, solver: saga, max_iter: 2846	learning_rate: 0.0596474, max_depth: 99, n_estimators: 370, subsample: 0.8668984, gamma: 0.3871157, reg_alpha: 0.1856635, colsample_bytree: 0.7065030	C: 0.5971435, epsilon: 0.0008019, gamma: scale, kernel: poly, degree: 5, coef0: 0.6242379
Fold4	learning_rate: 0.0029067, hidden_dim: 170, num_layers: 1, dropout_rate: 0.2498450 weight_decay: 0.0031874	n_estimators: 8, max_depth: 59, min_samples_split: 2, min_samples_leaf: 1	alpha: 69.3613018, fit_intercept: True, solver: saga, max_iter: 1209	learning_rate: 0.0281348, max_depth: 30, n_estimators: 154, subsample: 0.7725355, gamma: 0.0093372, reg_alpha: 2.3919887, colsample_bytree: 0.8000371	C: 43.0000158, epsilon: 0.0003945, gamma: auto, kernel: rbf, degree: 3, coef0: 0.1158690
Fold5	learning_rate: 0.0032930, hidden_dim: 185, num_layers: 2, dropout_rate: 0.3510232, weight_decay: 0.0021739	n_estimators: 144, max_depth: 296, min_samples_split: 2, min_samples_leaf: 1	alpha: 69.3613018, fit_intercept: True, solver: saga, max_iter: 1209	learning_rate: 0.0261145, max_depth: 257, n_estimators: 237, subsample: 0.8181197, gamma: 0.0012596, reg_alpha: 0.0330590, colsample_bytree: 0.8509245	C: 9.5946026, epsilon: 0.0646381, gamma: scale, kernel: rbf, degree: 3, coef0: 0.5968428

Table S4. Supplementary Metrics for All Models.

Model	Morgan fingerprints			RDKit descriptors		
	% LogS \pm	% LogS \pm	% LogS \pm	% LogS \pm	% LogS \pm	% LogS \pm
	0.5	0.7	1.0	0.5	0.7	1.0
RF	80.85	90.27	95.74	80.85	88.15	95.14
SVR	82.37	90.88	95.14	80.85	89.36	95.14
XGB	79.33	90.88	95.74	79.94	88.45	95.14
MLP	83.59	92.71	97.57	83.28	91.49	95.74
RR	77.51	89.36	93.62	75.99	86.32	92.41

Table S5. Experimental validation results.

ILs	T (°C)	Pred_S (g/mol)	Exp_S (g/mol)
[Amim][DMP]	60	41.57	30.94
	70	47.67	33.72
	80	55.81	40.86
	90	66.70	50.31
	100	80.57	63.37
[C ₂ mim][DMP]	60	16.56	24.21
	70	17.50	30.87
	80	18.83	35.08
	90	20.85	41.36
	100	23.92	49.99
[Dmim][OAc]	60	10.94	8.42
	70	13.13	11.54
	80	15.95	14.98
	90	19.59	16.94
	100	24.35	17.92
[C ₂ dmim][OAc]	60	3.33	2.16
	70	3.44	3.48
	80	3.64	6.06
	90	3.92	8.86
	100	4.25	11.93
[C ₆ dmim][Cl]	60	1.39	2.61
	70	1.36	3.83
	80	1.32	5.07
	90	1.30	6.81
	100	1.34	8.01