Supplementary Materials

Discovering polyimides and their composites with targeted mechanical properties through explainable machine learning

Weilong Hu[1,2], Enzhe Jing[1,2], Haoke Qiu[1,2,*], Zhao-Yan Sun[1,2,*]

[1]State Key Laboratory of Polymer Physics and Chemistry & Key Laboratory of Polymer Science and Technology, Changchun Institute of Applied Chemistry, Chinese Academy of Sciences, Changchun 130022, Jilin, China.

[2]School of Applied Chemistry and Engineering, University of Science and Technology of China, Hefei 230026, Anhui, China.

[*]**Correspondence to:** Haoke Qiu, Prof. Zhao-Yan Sun, State Key Laboratory of Polymer Physics and Chemistry & Key Laboratory of Polymer Science and Technology, Changchun Institute of Applied Chemistry, Chinese Academy of Sciences, Renmin Street 5625, Changchun 130022, Jilin, China. E-mail: hkqiu@ciac.ac.cn; zysun@ciac.ac.cn

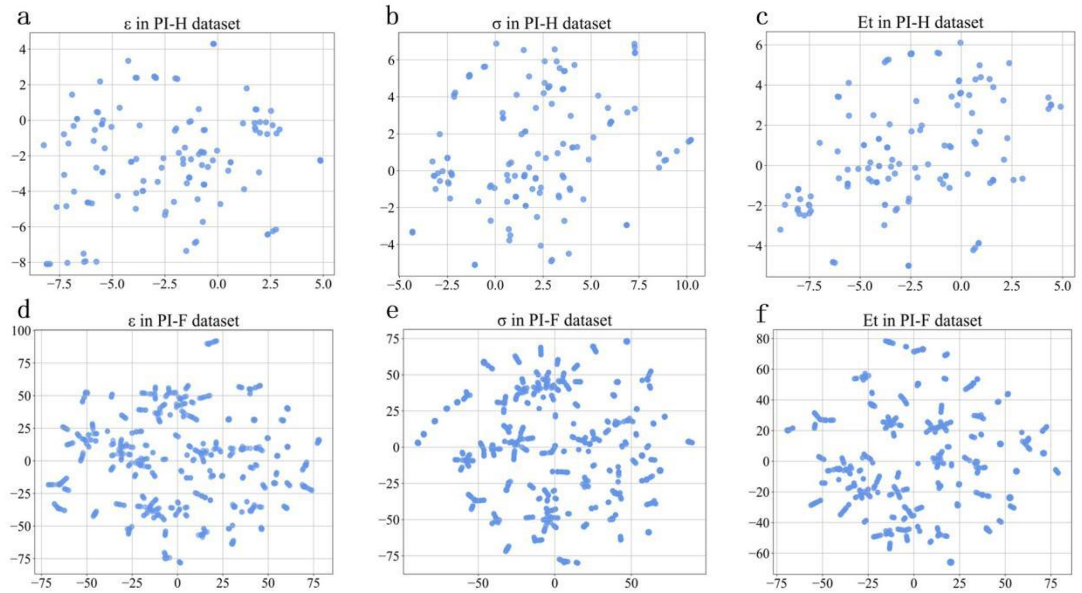
Supplementary Table 1. The mechanical properties collected for various PI structures, including elastic modulus (E), elongation at break (ε), tensile strength (σ) and tensile modulus (E_t)

	PI-H				PI-F			
	E	**ε**	**σ**	**E_t**	**E**	**ε**	**σ**	**E_t**
Number of available data points	15	106	128	125	66	671	731	540
Property range	2.08-2500	0.11-156	4.5-367.6	0.0106-8.7	0.1-6500	0.58-80.8	11-1197	0.27-28.13
unit	MPa	%	MPa	GPa	MPa	%	MPa	GPa

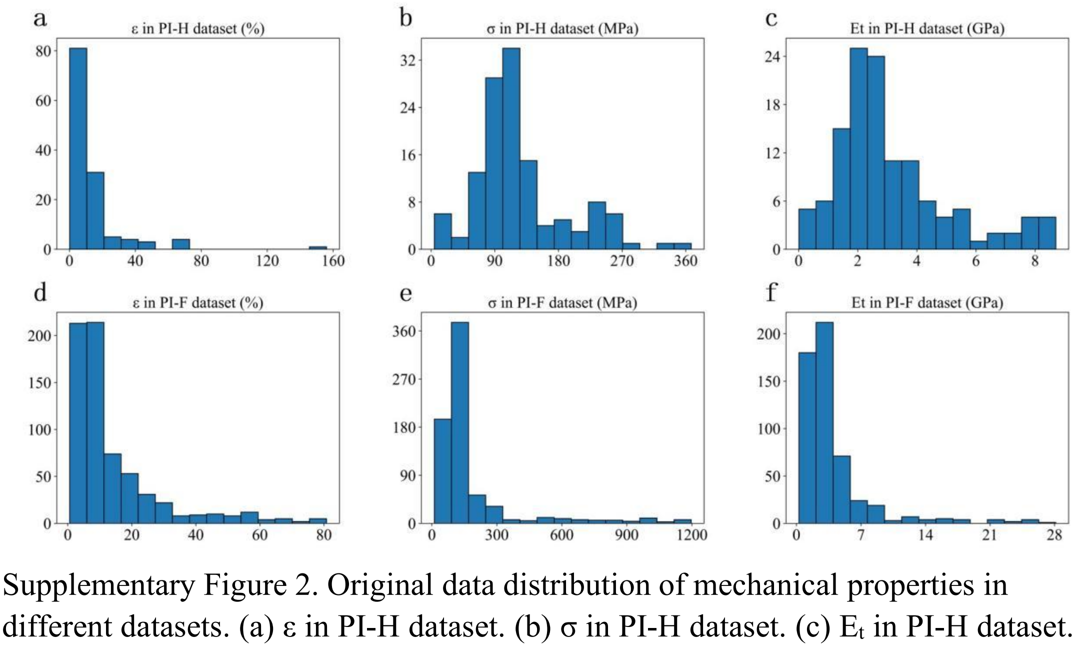
Supplementary Table 2. The DOI of our data source literature

DOI	
10.1016/j.compscitech.2013.07.005	10.1002/app.35292
10.1039/c4ra07716d	10.1155/2010/354364
10.1021/cm200909x	10.1177/096739111402200205
10.1002/adv.21886	10.1002/polb.23515
10.1002/pat.1565	10.1002/pola.29172
10.1021/am4005094	10.1016/j.polymdegradstab.2018.01.006
10.1002/pc.22637	10.1016/j.apsusc.2020.147654
10.1021/am300999g	10.1002/app.50869
10.1007/s10965-011-9564-z	10.1016/j.compstruct.2020.113305
10.1007/s10854-018-0045-6	10.1007/s12221-021-9142-x
10.1021/am302494c	10.1021/acsami.1c01992
10.1002/app.43677	10.1016/j.matchemphys.2020.123972
10.1021/am504342j	10.1016/j.cjche.2020.09.066
10.1002/pc.22895	10.1016/j.compstruct.2021.114340
10.1021/am2002229	10.1016/j.polymer.2021.124113
10.1039/c8ra01965g	10.1007/s10443-021-09878-y
10.1002/pi.5555	10.3390/polym13234057
10.1016/j.polymer.2018.02.017	10.1007/s10934-020-01028-2
10.1007/s10904-017-0641-6	10.1021/acsami.1c10055
10.1016/j.polymer.2011.09.033	10.3390/nano11030562
10.1002/app.41544	10.2478/msp-2021-0043
10.1016/j.compositesb.2011.11.071	10.1016/j.apsusc.2020.148833
10.1016/j.apsusc.2014.06.130	10.1002/pc.26165
10.1002/app.39881	10.1016/j.polymer.2021.123957
10.1007/s10965-014-0463-y	10.1039/d1na00415h
10.1016/j.compscitech.2011.06.013	10.7503/cjcu20200689
10.1007/s10118-016-1813-5	10.3390/polym13183175
10.1002/app.31523	10.1002/pc.26099
10.1016/j.memsci.2015.12.057	10.1002/macp.202000376
10.1016/j.compositesa.2015.09.009	10.1016/j.cej.2021.129341
10.1080/14658011.2018.1493274	10.1039/d1ra07500d
10.1002/pc.22854	10.1016/j.polymer.2021.123963
10.1002/pi.2889	10.3390/polym13234174
10.1002/pc.23249	10.3390/coatings11010025
10.1007/s11595-016-1502-9	10.1108/ILT-08-2021-0309
10.1002/pc.23815	10.1021/acsami.0c21842
10.1002/app.46826	10.3390/cryst11111383
10.1016/j.matlet.2014.11.003	10.1007/s12221-021-0148-1
10.1007/s13233-009-0048-5	10.7503/cjcu20200482
10.1021/am400635x	10.1002/pc.26174
10.1002/app.45168	10.1016/j.polymer.2021.124098

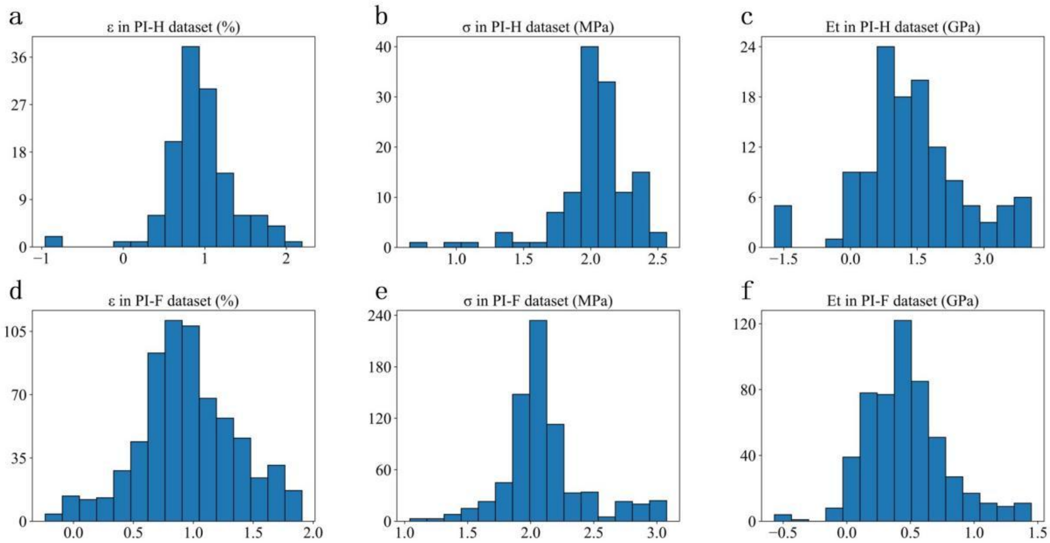
10.1016/j.coco.2016.11.002	10.3390/molecules26082143
10.1016/j.clay.2016.12.012	10.1016/j.micromeso.2021.111074
10.1021/acs.iecr.7b02183	10.1016/j.apsusc.2021.150051
10.1016/j.matdes.2016.12.082	10.1021/acsnano.0c09391
10.1002/app.31255	10.1016/j.jpcs.2021.110180
10.1016/j.compscitech.2017.10.014	10.1080/10601325.2021.1934011
10.1002/pat.3444	10.3390/nano11030812
10.1039/c8ra02479k	10.1016/j.coco.2021.100773
10.1016/j.mtcomm.2017.10.010	10.1002/mame.202000705
10.1016/j.compositesb.2013.08.042	10.1088/2053-1591/ac161f
10.3390/ijms14058698	10.3390/ma14061402
10.1080/00914037.2013.812089	10.1021/acsami.1c15018
10.1016/j.polymdegradstab.2013.03.003	10.1007/s10853-021-06201-9
10.1002/app.35412	10.3390/ma15134656
10.1007/s10965-015-0806-3	10.1007/s11771-022-5047-0
10.1007/s13233-009-0120-1	10.1016/j.compositesb.2021.109405
10.1002/pi.2939	10.1016/j.carbon.2022.07.020
10.1002/app.42358	10.1039/d2na00078d
10.1016/j.polymer.2015.04.015	10.3390/polym14214565
10.7503/cjcu20170011	10.3390/polym14153154
10.1007/s40242-014-4242-4	10.1515/rams-2022-0044
10.1016/j.apsusc.2017.06.087	10.1007/s10965-022-03071-w
10.2494/photopolymer.27.131	10.3390/polym14061269
10.1016/j.eurpolymj.2012.05.015	10.3390/polym14183803
10.1016/j.apsusc.2015.11.201	10.1016/j.apsusc.2022.152558
10.1007/s11814-011-0053-1	10.3390/membranes12100961
10.1142/S1793292018500984	10.1016/j.polymer.2022.124792
10.1002/pc.23402	10.1177/09673911221087806
10.1177/0967391111019002-321	10.1002/pat.5564
10.1179/1743289811Y.0000000053	10.3390/polym14050965
10.1016/j.compositesb.2016.12.057	10.1016/j.seppur.2021.119993
10.1016/j.polymer.2016.02.056	10.3390/polym14030389
10.1002/pola.27242	10.1016/j.triboint.2021.107427
10.1002/app.35054	10.3390/polym14122453
10.1016/j.compositesb.2013.08.065	10.1007/s42765-021-00093-9
10.1080/10601325.2018.1476823	10.1016/j.eurpolymj.2022.111260
10.1007/s11595-015-1098-5	10.1016/j.petsci.2021.10.027
10.1002/app.44000	10.1016/j.polymer.2022.125358
10.1002/mame.201600113	10.1016/j.matchemphys.2022.125716
10.1016/j.matdes.2017.11.041	10.1021/acssuschemeng.1c07621
10.1177/09540083211044054	



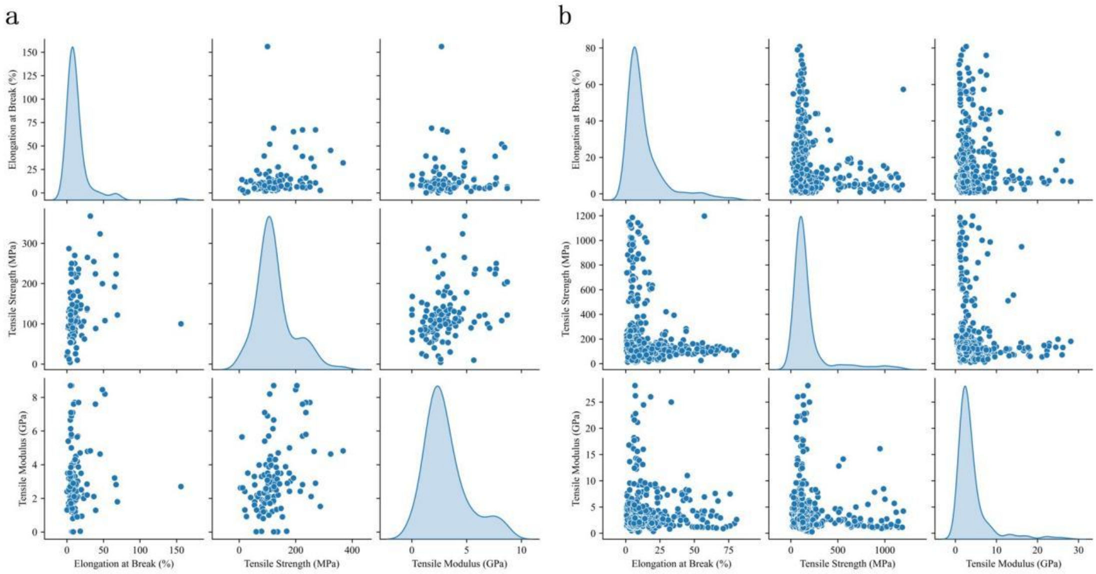
Supplementary Figure 1. Structure distribution of mechanical properties in different datasets. (a) ε in PI-H dataset. (b) σ in PI-H dataset. (c) E_t in PI-H dataset. (d) ε in PI-F dataset. (e) σ in PI-F dataset. (f) E_t in PI-F dataset.



Supplementary Figure 2. Original data distribution of mechanical properties in different datasets. (a) ε in PI-H dataset. (b) σ in PI-H dataset. (c) E_t in PI-H dataset. (d) ε in PI-F dataset. (e) σ in PI-F dataset. (f) E_t in PI-F dataset.



Supplementary Figure 3. Normalized data distribution of mechanical properties in different datasets. (a) ε in PI-H dataset. (b) σ in PI-H dataset. (c) E_t in PI-H dataset. (d) ε in PI-F dataset. (e) σ in PI-F dataset. (f) E_t in PI-F dataset.



Supplementary Figure 4. Pairwise relationships (a) between ε, σ and E_t in the PI-H dataset, and (b) between ε, σ and E_t in the PI-F dataset.

Supplementary Table 3. Values of R^2 in the test set for 7 mechanical properties by using 11 common models without tuning parameters of each model.

	PI-H dataset			PI-F dataset		
	ε	**σ**	**E_t**	**ε**	**σ**	**E_t**
RFR	-0.2	0.51	0.28	0.36	0.56	0.33
LR	-10248.66	-5983.75	-4039.05	0.33	0.36	-0.52
SVR	-0.32	-0.03	-0.18	-0.11	-0.1	-0.12
RR	-15.27	-350.23	-101.64	0.28	0.43	-0.36
GPR	-6.52	-3.06	-2.57	0.49	0.65	0.43
AR	-0.26	0.37	0.13	0.15	0.37	-0.65
GBR	-0.22	0.49	0.2	0.53	0.71	0.53
BR	0.43	0.35	0.18	0.5	0.74	0.45
ETR	-0.39	0.45	0.34	0.36	0.32	0.26
DTR	-0.96	-0.14	0.13	0.29	0.36	0.04
MLP	-320.28	-32.4	-60920.5	0.29	0.27	0.27

Supplementary Table 4. Values of R^2 in the test set after Grid Search for different models that we selected

	PI-H dataset			PI-F dataset		
	ε	**σ**	**E_t**	**ε**	**σ**	**E_t**
RFR	0.22	0.53	0.4	0.64	0.75	0.64
ETR	0.47	0.53	0.48	0.56	0.8	0.56
GBR	0.35	0.57	0.39	0.59	0.74	0.59
BR	0.14	0.22	0.25	0.59	0.78	0.59

Supplementary Table 5. Values of R^2 in the test set after hyperparameter optimization by using Optuna

	PI-H dataset			PI-F dataset		
	ε	σ	E_t	ε	σ	E_t
RFR	0.68	0.69	0.63	0.74	0.79	0.75
ETR	0.53	0.52	0.6	0.74	0.82	0.64
GBR	0.39	0.83	0.78	0.79	0.85	0.83
BR	/	/	/	0.73	0.8	0.78

Supplementary Table 6. Values of R^2 in the test set after repeated training

Rando	PI-H dataset			PI-F dataset		
m seed	ε	σ	E_t	ε	σ	E_t
0	0.68	0.83	0.78	0.79	0.85	0.83
1	0.66	0.78	0.72	0.81	0.87	0.79
10	0.69	0.81	0.75	0.83	0.83	0.78
25	0.63	0.86	0.76	0.77	0.81	0.82
50	0.72	0.81	0.74	0.75	0.85	0.83
Average	**0.68±0.05**	**0.82±0.04**	**0.75±0.03**	**0.79±0.05**	**0.84±0.03**	**0.81±0.03**

Supplementary Table 7. Values of R^2, MAE, MSE, RMSE and MAPE in the test set

	PI-H dataset			PI-F dataset		
	ε	σ	E_t	ε	σ	E_t
R^2	0.68	0.83	0.78	0.79	0.85	0.83
MAE	1.5	18.7	1.2	5.3	41.1	0.9
MSE	7.3	650.8	2.1	60.2	5878.7	1.6
RMSE	1.9	25.5	1.6	7.8	76.7	1.3
MAPE	19.3	11.9	14.7	17.1	12.1	14.8

Supplementary Table 8. Some related information of our 7 models

		Training Time (ms)	Prediction Speed (ms)	Hyperparameters
PI-H dataset	**ε**	600	0.18	n_estimators=231, criterion='friedman_mse', max_depth=98, min_samples_split=3, max_features=39, max_leaf_nodes=132
	σ	300	0.17	Loss='squared_error', learning_rate=0.6648117460909637, n_estimators=242, criterion='friedman_mse', min_samples_split=77, max_depth=83, max_features=11, max_leaf_nodes=231
	E_t	100	0.17	Loss='squared_error', learning_rate=0.8683573694740211, n_estimators=257, criterion='friedman_mse', min_samples_split=4, max_depth=48, max_features=30, max_leaf_nodes=31
PI-F dataset	**ε**	5700	0.24	Loss='huber', learning_rate=0.20863039129499694, n_estimators=115, criterion='friedman_mse', min_samples_split=18, max_depth=64, max_features=84, max_leaf_nodes=54
	σ	2200	0.15	Loss='huber', learning_rate=0.16382086787423966, n_estimators=269, criterion='friedman_mse', min_samples_split=9, max_depth=36, max_features=80, max_leaf_nodes=25
	E_t	1600	0.21	Loss='squared_error', learning_rate=0.2999766914200086, n_estimators=280, criterion='friedman_mse', min_samples_split=16, max_depth=85, max_features=94, max_leaf_nodes=43