Electronic Supplementary Information

Machine learning of atomic force microscopy images of organic solar cells

Yasuhito Kobayashi,[a,b] Yuta Miyake,[c] Fumitaka Ishiwari,[c,d,e] Shintaro Ishiwata[a] and Akinori Saeki[,c,d]*

[a] Division of Materials Physics, Graduate School of Engineering Science, Osaka University, 1-3 Machikaneyama, Toyonaka, Osaka 560-8531, Japan.

[b] Interactive Materials Science CADET, Osaka University, 1-3 Machikaneyama, Toyonaka, Osaka 560-8531, Japan.

[c] Department of Applied Chemistry, Graduate School of Engineering, Osaka University, 2-1 Yamadaoka, Suita, Osaka 565-0871, Japan.

[d] Innovative Catalysis Science Division, Institute for Open and Transdisciplinary Research Initiatives (ICS-OTRI), Osaka University, 1-1 Yamadaoka, Suita, Osaka 565-0871, Japan.

[e] PRESTO, Japan Science and Technology Agency (JST), Kawaguchi, Saitama 332-0012, Japan.

AUTHOR INFORMATION

Corresponding Author

* saeki@chem.eng.osaka-u.ac.jp (A.S.)

This supporting information presents the following contents.

Supplementary Code

The following is a python code for LASSO analysis. Ridge and RF analyses were also performed in the same manner. "Mord+Prop+Proc_890.csv" is a data file.

```
================================================================
from sklearn.linear_model import Lasso, Ridge
from sklearn.model_selection import train_test_split, cross_val_score
from sklearn.ensemble import RandomForestRegressor
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import glob, tqdm, math, optuna, json, pickle
from machine_learning import calc_metrics, data_split, reconstruct_linearmodel, plot_yy, set_plot_params,
plot_coef

df_original = pd.read_csv('./dataset/Mord+Prop+Proc_890.csv').set_index('Unnamed: 0')

name = 'LASSO_Original'
dataframe = df_original

dataset = data_split(dataframe,'PCE_max(%)',0.3,random_state=0)

model = Lasso()
def objective(trial):
    params = {'alpha': trial.suggest_loguniform('alpha',0.00001,100)}
    model.set_params(**params)
    scores = cross_val_score(
        model,
        dataset['x_train'],
        dataset['y_train'],
        cv=5,
        scoring='r2',
        n_jobs=-1)
    val = scores.mean()
    return val

study = optuna.create_study(
    study_name = name,
    direction = 'maximize',
    sampler = optuna.samplers.TPESampler(seed=0),
    storage = f'sqlite:///./optuna_{name}.db',
    load_if_exists = True
    )

study.optimize(objective,n_trials=100,n_jobs=-1)
with open(f'./Bayese_{name}.pickle','wb') as f:
    pickle.dump(study,f,protocol=2)

### re-learning with tuned params
re_model = Lasso(**study.best_params)
metrics, df_coef = reconstruct_linearmodel(Lasso(**study.best_params),dataset,name)
================================================================
```

Supplementary Tables

Table S1. List of material properties and process parameters of OPVs.

Material properties			Process parameters	
Material	Parameter	Description	Parameter	Description
NFA	–HOMO (eV)	Absolute HOMO	n/(p+n)	Blend ratio of polymer (p) and NFA (n)
	–LUMO (eV)	Absolute LUMO	Solvent_CF	chloroform
	E_g (eV)	Bandgap	Solvent_OX	*o*-xylene
	M (g/mol)	Molecular weight	Solvent_TMB	1,2,4-trimethylbenzene
Polymer	–HOMO (eV)	Absolute HOMO	Solvent_CB	chlorobenzene
	–LUMO (eV)	Absolute LUMO	Solvent_THF	tetrahydrofuran
	E_g (eV)	Bandgap	Solvent_AS	anisole
	M_w (kg/mol)	Weight-averaged molecular weight	Solvent_TL	toluene
	M_n (kg/mol)	Number-averaged molecular weight	Solvent_DCB	*o*-dichlorobenzene
	PDI	Polydispersity index	Additive_NA	No additive
Polymer: NFA	ΔHOMO (eV)	Difference of HOMOs of NFA and polymer	Additive_CN	1-chloronaphthalene
	ΔLUMO (eV)	Difference of LUMOs of NFA and polymer	Additive_DIO	1,8-diiodoocatane
	Effective E_g (eV)	Difference of HOMO of polymer and LUMO of NFA	Additive_Bipy	Bipyridine
	$E_{g,min}$ (eV)	Minimum E_g value of polymer and NFA	Additive_PN	1-phenylnapthalene
			Additive_CBA	*o*-chlorobenzaldehyde
			Additive_pyr	pyridine
			Additive_NMP	*N*-methylpyrrolidone
			Additive_BPO	4,4-biphenol
			Additive_DPE	diphenylether
			Additive_ODT	1,8-octanedithiol
			Additive_DBE	dibenzyl ether

Table S2. List of 2D Mordred descriptors.

No.		#	Module	Description
From	to			
1	2	2	ABCIndex	Connectivity index
3	14	12	AcidBase	Acidic/basic group count
15	16	2	AdjacencyMatrix	SpAbs of adjacency matrix
17	18	2	Aromatic	Aromatic atoms, bond count
19	35	17	AtomCount	Atom count (C, H, ,,,)
36	134	99	Autocorrelation	Autocorrelation of Topological Structure descriptor (ATS)
135	234	100	Autocorrelation	Averaged ATS descriptor (AATS)
235	341	107	Autocorrelation	Centered ATS descriptor (ATSC)
342	449	108	Autocorrelation	Averaged ATSC descriptor (AATSC)
450	545	96	Autocorrelation	Moran coefficient descriptor (MATS)
546	641	96	Autocorrelation	Geary coefficient descriptor (GATS)
642	665	24	BCUT	Burden matrix (BCUT)
666	666	1	BalabanJ	Balaban's J index descriptor
667	770	104	BaryszMatrix	Barysz matrix descriptor (Sp)
771	771	1	BertzCT	Bertz CT descriptor
772	780	9	BondCount	Number of double bonds, etc
781	782	2	CPSA	relative negative/positive charge
783	793	11	CarbonTypes	SP carbon bound to 1 other carbon, etc
794	849	56	Chi	3-ordered Chi chain weighted by sigma electrons, etc
850	865	16	Constitutional	Sum of constitutional descriptor
866	879	14	DetourMatrix	detour matrix descriptor.
880	891	12	DistanceMatrix	distance matrix descriptor
892	1207	316	EState	atom type e-state descriptor.
1208	1208	1	EccentricConnectivityIndex	eccentric connectivity index
1209	1253	45	ExtendedTopochemicalAtom	Extended Topochemical Atom(ETA) descriptor
1254	1254	1	FragmentComplexity	fragment complexity descriptor.
1255	1255	1	Framework	molecular framework ratio
1256	1257	2	HydrogenBond	number of hydrogen bond acceptor/donor
1258	1299	42	InformationContent	neighborhood information content descriptor.
1300	1302	3	KappaShapeIndex	Kappa shape index 1 descriptor.
1303	1304	2	Lipinski	Lipinski rule of 5 descriptor
1305	1305	1	LogS	Filter-it™ LogS descriptor
1306	1306	1	McGowanVolume	McGowan volume
1307	1359	53	MoeType	Labute's Approximate Surface Area descriptor
1360	1378	19	MolecularDistanceEdge	molecular distance edge descriptor
1379	1390	12	MolecularId	molecular id descriptor
1391	1411	21	PathCount	path count descriptor
1412	1413	2	Polarizability	atomic polarizability descriptor.
1414	1551	138	RingCount	ring count descriptor
1552	1553	2	RotatableBond	rotatable bonds count descriptor
1554	1555	2	SLogP	Wildman-Crippen LogP/MR descriptor
1556	1557	2	TopoPSA	topological polar surface area descriptor
1558	1578	21	TopologicalCharge	topological charge descriptor
1579	1582	4	TopologicalIndex	topological diameter, radius, etc
1583	1583	1	VdwVolumeABC	ABC van der waals volume
1584	1584	1	VertexAdjacencyInformation	vertex adjacency information descriptor
1585	1605	21	WalkCount	walk count descriptor
1606	1607	2	Weight	exact molecular weight
1608	1609	2	WienerIndex	Wiener index
1610	1613	4	ZagrebIndex	Zagreb index descriptor

https://mordred-descriptor.github.io/documentation/master/descriptors.html

Table S3. List of LASSO coefficients (Mordred descriptors and material and process parameters as the inputs) sorted in descend order.

	Negative coefficient			Positive coefficient		
	Descriptor	Coefficient	Category[a]	Descriptor	Coefficient	Category[a]
1	n_GATS6s	–4.01	N	p_AATSC0i	3.11	P
2	n_MATS4s	–2.92	N	n_AATSC2d	2.32	N
3	n_GATS3se	–2.09	N	n_SaaaC	2.02	N
4	p_AATSC8dv	–1.98	N	p_MATS2se	1.87	P
5	PDI	–1.78	P	–LUMO_n(eV)	1.74	N
6	n_SdssC	–1.72	N	p_AATS1v	1.63	P
7	n_GATS5v	–1.68	N	Solvent_TMB	1.47	Process
8	p_MATS8s	–1.62	P	p_n9FRing	1.30	P
9	n_MATS8pe	–1.47	N	n_nCl	1.27	N
10	p_GATS2dv	–1.44	P	Eg_p(eV)	1.17	P
11	ΔHOMO(eV)	–1.33	P/N	p_GATS6p	1.13	P
12	n_VSA_EState6	–1.20	N	n_AATSC4Z	0.99	N
13	p_AATSC8s	–1.08	P	n_TopoPSA	0.91	N
14	n_AATSC8m	–0.87	N	n_GATS7Z	0.72	N
15	n_MATS8s	–0.80	N	p_SssS	0.60	P
16	n_AATS8s	–0.76	N	n_nFARing	0.47	N
17	n_BCUTd–1l	–0.75	N	p_EState_VSA8	0.38	P
18	Additive_Nan	–0.69	Process	n_BCUTpe–1h	0.36	N
19	p_MDEC–33	–0.67	Process	Additive_DIO	0.26	Process
20	Solvent_CB	–0.61	Process	p_C1SP2	0.19	P
21	n_GATS3i	–0.60	N	n_NsCl	0.17	N
22	n_GATS5i	–0.53	N	p_MATS5se	0.11	P
23	Solvent_TL	–0.50	Process	n_MDEC–23	0.10	N
24	Eg_n(eV)	–0.50	N	n_MATS3p	0.06	N
25	p_n8FARing	–0.47	P	Solvent_CF	0.02	Process
26	p_AATSC8Z	–0.43	P	p_n9FHRing	0.00	P
27	n_GATS3Z	–0.39	N			
28	n_BCUTs–1l	–0.35	N			
29	p_nAHRing	–0.32	P			
30	n_n9FHRing	–0.28	N			

[a] Parameter category: P = polymer, N = NFA, P/N= difference of P and N.

Table S4. Multi–colinearity of AATSC0i (polymer, p). The descriptors and their correlation coefficients (*r*) in top 30th are sorted in decent order for the positive and negative category.

Rank	Mordred descriptor	Positive *r*	Mordred descriptor	Negative *r*
1	**p_AMID_X**	**0.881**	p_MINaasC	–0.849
2	p_AATSC0are	0.871	p_GATS1dv	–0.730
3	p_AATSC0pe	0.868	p_ETA_dEpsilon_C	–0.723
4	p_AATSC0s	0.867	p_GATS1pe	–0.705
5	**p_nF**	**0.863**	p_SM1_Dzp	–0.699
6	p_NsF	0.863	p_BCUTp–1l	–0.692
7	p_AATS0s	0.856	p_ETA_psi_1	–0.679
8	p_SlogP_VSA10	0.853	p_GATS1are	–0.676
9	p_SsF	0.843	p_MINaaCH	–0.673
10	p_BCUTi–1h	0.837	p_GATS2dv	–0.663
11	p_AATSC0se	0.835	p_BCUTv–1l	–0.652
12	p_nX	0.812	p_MINsCH3	–0.646
13	p_MID_X	0.812	p_GATS4s	–0.621
14	p_BCUTpe–1h	0.806	p_GATS2pe	–0.616
15	p_AATS0are	0.803	p_SM1_Dzv	–0.613
16	p_BCUTare–1h	0.801	p_MINaaS	–0.612
17	p_PEOE_VSA3	0.796	p_AMID_C	–0.602
18	p_AATSC3se	0.780	p_AXp–0dv	–0.587
19	p_VSA_EState1	0.775	p_GATS2are	–0.587
20	p_AATS0pe	0.775	p_GATS3pe	–0.584
21	p_AATSC3pe	0.773	p_GATS1s	–0.572
22	p_Mare	0.770	p_MAXaasC	–0.552
23	p_AATS3s	0.761	p_GATS4dv	–0.550
24	p_BCUTse–1h	0.759	p_GATS2s	–0.544
25	p_AETA_beta_ns_d	0.759	p_MAXaaS	–0.543
26	p_BCUTdv–1h	0.758	p_GATS3se	–0.542
27	p_AATSC2s	0.751	p_MINssCH2	–0.534
28	p_AATS5s	0.749	p_GATS1se	–0.518
29	p_AATSC0dv	0.747	p_MATS6i	–0.514
30	p_AATS0se	0.743	p_VSA_EState5	–0.513

Table S5. Multi-colinearity of GATS6s (NFA, n). The descriptors and their correlation coefficients (*r*) in top 30th are sorted in decent order for the positive and negative category.

Rank	Mordred descriptor	Positive *r*	Mordred descriptor	Negative *r*
1	n_GATS6are	0.960	n_MATS6s	–0.872
2	n_GATS6se	0.953	n_ATSC6s	–0.840
3	n_GATS6pe	0.953	n_AATSC6s	–0.836
4	n_GATS5s	0.895	n_MATS6se	–0.822
5	n_GATS7s	0.877	n_MATS6are	–0.819
6	n_GATS8s	0.870	n_AATSC6se	–0.818
7	n_GATS5are	0.854	n_ATSC6se	–0.813
8	n_GATS6dv	0.844	n_MATS6pe	–0.803
9	n_GATS5pe	0.838	n_AATSC6are	–0.802
10	n_GATS5se	0.831	n_ATSC6are	–0.785
11	n_GATS7pe	0.828	n_AATSC6pe	–0.783
12	n_GATS4s	0.826	**n_NtN**	**–0.753**
13	n_GATS7are	0.823	**n_C1SP1**	**–0.753**
14	n_GATS8are	0.818	n_SMR_VSA2	–0.753
15	n_GATS4are	0.817	n_StN	–0.751
16	n_GATS4se	0.815	n_EState_VSA10	–0.744
17	n_GATS8pe	0.814	n_ATSC6pe	–0.733
18	n_GATS5dv	0.812	n_nBondsT	–0.728
19	n_GATS8se	0.803	n_PEOE_VSA10	–0.722
20	n_GATS4pe	0.803	n_ATSC7se	–0.720
21	n_GATS7se	0.798	n_PEOE_VSA4	–0.712
22	n_BCUTd–1l	0.752	n_AATSC5s	–0.698
23	n_C1SP2	0.716	n_MATS5s	–0.692
24	n_AMID_O	0.714	n_n9FRing	–0.691
25	n_MATS2are	0.708	n_ATSC7s	–0.681
26	n_PEOE_VSA13	0.700	n_ATSC5s	–0.679
27	n_GATS7dv	0.697	n_AATSC7se	–0.669
28	n_MATS2pe	0.669	n_NtsC	–0.668
29	n_GATS8dv	0.665	n_ATSC8se	–0.665
30	n_GATS3i	0.659	n_MATS3s	–0.665

Table S6. List of LASSO coefficients (Mordred descriptors, material and process parameters, GLCM, and HA as the inputs) sorted in descend order. GLSM and HA parameters are highlighted by bold.

	Negative coefficient			Positive coefficient		
	Descriptor	Coefficient	Category[a]	Descriptor	Coefficient	Category[a]
1	n_GATS6s	–4.54	N	p_AATSC0i	2.83	P
2	**Homogeneity_2_0**	**–3.15**	**GLCM**	n_AATSC2d	2.60	N
3	n_MATS4s	–3.06	N	n_SaaaC	1.97	N
4	n_GATS3se	–2.53	N	p_MATS2se	1.45	P
5	p_AATSC8dv	–2.37	P	Solvent_TMB	1.40	Process
6	**Maximum Height**	**–2.13**	**HA**	p_n9FRing	1.26	P
7	n_GATS5v	–1.75	N	p_AATS2v	1.15	P
8	PDI	–1.75	P	Eg_p(eV)	1.12	P
9	p_MATS8s	–1.69	P	n_nCl	1.04	N
10	n_SdssC	–1.65	N	-LUMO_n(eV)	0.89	N
11	n_VSA_EState6	–1.36	N	n_MDEC–23	0.77	N
12	Eg_n(eV)	–1.29	N	n_GATS7Z	0.72	N
13	**Correlation_4_0**	**–1.20**	**GLCM**	p_C1SP2	0.60	P
14	n_MATS8s	–0.97	N	p_GATS6p	0.57	P
15	n_MATS8pe	–0.95	N	p_SssS	0.51	P
16	p_GATS2dv	–0.94	P	p_EState_VSA8	0.44	P
17	ΔHOMO(eV)	–0.92	P/N	n_AATSC4Z	0.37	N
18	n_BCUTd–1l	–0.86	N	p_AATS1v	0.37	P
19	n_AATSC8m	–0.81	N	p_SlogP_VSA10	0.33	P
20	p_GATS3dv	–0.78	P	n_nFARing	0.33	N
21	Additive_Nan	–0.66	Process	n_AATSC4m	0.31	N
22	p_AATSC8Z	–0.56	P	Additive_DIO	0.23	Process
23	p_MDEC–33	–0.54	P	n_TopoPSA	0.13	N
24	Solvent_CB	–0.47	Process	n_BCUTse–1h	0.11	N
25	n_AATS8s	–0.46	N	n_NsCl	0.07	N
26	p_GATS1i	–0.44	P	Solvent_DCB	0.05	Process
27	p_n8FARing	–0.37	P	p_NaaN	0.04	P
28	n_nBridgehead	–0.35	N	Additive_CN	0.01	Process
29	p_n10FHRing	–0.33	P	p_n9FHRing	0.00	P
30	n_BCUTs–1l	–0.31	N			

[a] Parameter category: P = polymer, N = NFA, P/N= difference of P and N.

Table S7. List of LASSO coefficients (Mordred descriptors, material and process parameters, FFT, and HA as the inputs) sorted in descend order. FFT and HA parameters are highlighted by bold.

	Negative coefficient			Positive coefficient		
	Descriptor	Coefficient	Category[a]	Descriptor	Coefficient	Category[a]
1	n_GATS6s	–4.259	N	p_AATSC0i	2.780	P
2	**Maximum Height**	**–3.218**	**HA**	**65.22**	**2.727**	**FFT**
3	n_MATS4s	–3.089	N	n_AATSC2d	2.511	N
4	n_GATS3se	–2.567	N	n_SaaaC	1.916	N
5	PDI	–2.295	P	**40.54**	**1.880**	**FFT**
6	p_AATSC8dv	–2.191	**P**	p_MATS2se	1.717	P
7	p_MATS8s	–1.990	P	Solvent_TMB	1.413	Process
8	n_SdssC	–1.984	N	p_n9FRing	1.254	P
9	n_GATS5v	–1.879	N	p_AATS1v	1.233	P
10	n_VSA_EState6	–1.566	N	-LUMO_n(eV)	1.208	N
11	n_MATS8pe	–1.553	N	Eg_p(eV)	1.196	P
12	ΔHOMO(eV)	–1.179	P/N	n_nCl	1.163	N
13	p_GATS2dv	–1.058	P	n_MDEC–23	0.943	N
14	Eg_n(eV)	–1.031	N	n_GATS7Z	0.832	N
15	n_MATS8s	–0.935	N	p_GATS6p	0.796	P
16	n_BCUTd–1l	–0.916	N	n_TopoPSA	0.661	N
17	p_MDEC–33	–0.798	P	p_EState_VSA8	0.626	P
18	p_GATS3dv	–0.770	P	n_AATSC4m	0.595	N
19	n_AATSC8m	–0.701	N	p_C1SP2	0.559	P
20	Additive_Nan	–0.619	Process	p_SssS	0.517	P
21	Solvent_CB	–0.490	Process	p_SlogP_VSA10	0.477	P
22	p_AATSC8Z	–0.487	P	n_nFARing	0.455	N
23	n_AATS8s	–0.446	N	**34.88**	**0.451**	**FFT**
24	n_BCUTs–1l	–0.446	N	**51.72**	**0.319**	**FFT**
25	p_NdssC	–0.433	P	p_AATS2v	0.299	P
26	p_MATS6p	–0.417	P	Additive_DIO	0.277	Process
27	n_GATS3Z	–0.407	N	Solvent_DCB	0.114	Process
28	p_AATSC8s	–0.380	P	n_BCUTse–1h	0.088	N
29	n/(p+n)	–0.374	Process	n_NsCl	0.075	N
30	p_n10FHRing	–0.361	P	p_n9FHRing	0.000	P

[a] Parameter category: P = polymer, N = NFA, P/N= difference of P and N. Descriptors with FFT category represents the spatial wavelength.

Table S8. List of Ridge regression coefficients (Mordred descriptors, material and process parameters, FFT, and HA as the inputs) sorted in descend order. FFT and HA parameters are highlighted by bold.

	Negative coefficient			Positive coefficient		
	Descriptor	Coefficient	Category[a]	Descriptor	Coefficient	Category[a]
1	**Maximum Height**	**–0.746**	**HA**	Solvent_TMB	0.840	Process
2	n_MATS4s	–0.724	N	n_MDEC–23	0.658	N
3	p_MDEC–33	–0.646	P	–LUMO_n(eV)	0.593	N
4	n_AATSC4s	–0.606	N	Eg_p(eV)	0.531	P
5	Eg_n(eV)	–0.583	N	ΔLUMO(eV)	0.529	P/N
6	ΔHOMO(eV)	–0.566	P/N	p_C1SP2	0.479	P
7	**Mean**	**–0.565**	**HA**	n_NsCl	0.431	N
8	**Skewness**	**–0.551**	**HA**	n_nCl	0.431	N
9	Additive_Nan	–0.548	Process	n_EState_VSA9	0.418	N
10	n/(p+n)	–0.534	P/N	p_MATS2se	0.411	P
11	n_GATS6s	–0.523	N	Additive_DIO	0.406	Process
12	PDI	–0.523	P	n_SsCl	0.404	N
13	p_AATSC8dv	–0.520	P	p_SssS	0.399	P
14	n_MATS8s	–0.517	N	–HOMO_p(eV)	0.393	P
15	Solvent_CB	–0.504	Process	n_PEOE_VSA4	0.372	N
16	n_SdssC	–0.477	N	n_MATS3p	0.370	**N**
17	p_MATS8s	–0.461	P	n_GATS7Z	0.367	N
18	n_ETA_dPsi_B	–0.443	N	p_GATS6p	0.355	P
19	n_BCUTd–1l	–0.441	N	Additive_CBA	0.354	Process
20	**Variance**	**–0.435**	**HA**	p_AATSC0i	0.352	P
21	p_EState_VSA7	–0.433	P	n_GATS7m	0.352	N
22	p_AATSC8s	–0.413	P	p_n9FAHRing	0.351	P
23	n_SsBr	–0.400	N	p_n9FARing	0.351	P
24	n_NsBr	–0.399	N	n_MATS6s	0.349	N
25	n_nBr	–0.399	N	n_BCUTdv–1h	0.343	**N**
26	n_JGI4	–0.393	N	n_NdsssP	0.334	N
27	p_n10FHRing	–0.386	P	n_nP	0.334	N
28	p_n10FaHRing	–0.386	P	p_SlogP_VSA10	0.328	P
29	n_EState_VSA6	–0.380	N	Solvent_THF	0.325	Process
30	p_MATS8dv	–0.376	P	**65.22**	**0.319**	**FFT**

[a] Parameter category: P = polymer, N = NFA, P/N= difference of P and N. Descriptors with FFT category represents the spatial wavelength.

Table S9. List of random forest feature importance (Mordred descriptors, material and process parameters as the inputs) sorted in descend order.

	Descriptor	Coefficient	Category[a]
1	n_StN	0.11425	N
2	n_VSA_EState3	0.05838	N
3	n_GATS5pe	0.02768	N
4	n_AATSC3m	0.01495	N
5	p_AATSC8se	0.01317	P
6	p_ETA_shape_y	0.01315	P
7	ΔHOMO(eV)	0.01288	P/N
8	n_GATS4se	0.01161	N
9	Eg_n(eV)	0.01126	N
10	–HOMO_p(eV)	0.01070	P
11	p_ATSC8s	0.01053	P
12	n_GATS3i	0.01015	N
13	n_AATSC3p	0.00932	N
14	n_BCUTs-1l	0.00896	N
15	p_MATS5dv	0.00850	P
16	n_AATSC3v	0.00761	N
17	ΔLUMO(eV)	0.00728	P/N
18	–HOMO_n(eV)	0.00719	N
19	p_AATSC5dv	0.00708	P
20	–LUMO_n(eV)	0.00603	N
21	p_GATS1i	0.00583	P
22	n_GATS6pe	0.00571	N
23	p_ETA_dEpsilon_B	0.00551	P
24	n_GATS6i	0.00515	N
25	n_GATS5are	0.00500	N
26	n_AATSC3Z	0.00482	N
27	Mn (kg/mol)	0.00461	P
28	p_AATSC8dv	0.00460	P
29	n_MATS5are	0.00453	N
30	p_AATSC4dv	0.00452	P

[a] Parameter category: P = polymer, N = NFA, P/N= difference of P and N.

Table S10. List of random forest feature importance (Mordred descriptors, material and process parameters, GLCM, and HA as the inputs) sorted in descend order. GLSM and HA parameters are highlighted by bold.

	Descriptor	Coefficient	Category[a]
1	n_StN	0.14443	N
2	n_VSA_EState3	0.05358	N
3	n_GATS5pe	0.02300	N
4	n_AATSC3m	0.01880	N
5	p_AATSC8se	0.01538	P
6	p_ETA_shape_y	0.01372	P
7	n_GATS4se	0.01284	N
8	p_ATSC8s	0.00959	P
9	n_BCUTs-1l	0.00946	N
10	Eg_n(eV)	0.00934	N
11	ΔHOMO(eV)	0.00928	P/N
12	–HOMO_p(eV)	0.00909	P
13	n_GATS3i	0.00879	N
14	n_AATSC3v	0.00828	N
15	n_AATSC3p	0.00819	N
16	p_MATS5dv	0.00695	P
17	–HOMO_n(eV)	0.00670	N
18	p_AATSC5dv	0.00629	P
19	ΔLUMO(eV)	0.00614	P/N
20	n_GATS6pe	0.00597	N
21	n_GATS6i	0.00570	N
22	**Correlation_2_0**	**0.00564**	**GLCM**
23	p_ETA_dEpsilon_B	0.00556	P
24	p_GATS1i	0.00506	P
25	p_AATSC8dv	0.00487	P
26	n_GATS5are	0.00481	N
27	–LUMO_n(eV)	0.00479	N
28	p_AETA_beta	0.00472	P
29	n_AATSC3Z	0.00470	N
30	n_GATS4s	0.00452	N

[a] Parameter category: P = polymer, N = NFA, P/N= difference of P and N.

Table S11. List of random forest feature importance (Mordred descriptors, material and process parameters, FFT, and HA as the inputs) sorted in descend order. FFT and HA parameters are highlighted by bold.

	Descriptor	Coefficient	Category[a]
1	n_StN	0.06983	N
2	n_VSA_EState3	0.04794	N
3	n_GATS5pe	0.02383	N
4	n_ATSC6s	0.01130	N
5	p_ETA_shape_y	0.01107	P
6	n_GATS4s	0.01049	N
7	n_GATS5are	0.01011	N
8	p_AATSC8se	0.00960	P
9	n_AATSC3m	0.00857	N
10	n_ATSC1s	0.00854	N
11	p_AATSC5dv	0.00834	P
12	n_BCUTs-1l	0.00820	N
13	p_ATSC8s	0.00813	P
14	p_MATS5dv	0.00773	P
15	n_GATS4se	0.00735	N
16	n_GATS3i	0.00730	N
17	ΔHOMO(eV)	0.00711	P/N
18	Eg_n(eV)	0.00707	N
19	n_GATS3s	0.00694	N
20	n_AATSC3p	0.00653	N
21	p_AATSC8dv	0.00636	P
22	n_GATS5s	0.00610	N
23	n_AATSC3Z	0.00601	N
24	p_ETA_dEpsilon_B	0.00598	P
25	n_AATSC3v	0.00528	N
26	–HOMO_p(eV)	0.00520	P
27	n_AATSC6s	0.00512	N
28	n_MATS5are	0.00502	N
29	n_AATSC1s	0.00478	N
30	n_BCUTd-1l	0.00475	N

[a] Parameter category: P = polymer, N = NFA, P/N= difference of P and N. Descriptors with FFT category represents the spatial wavelength.

Supplementary Figures

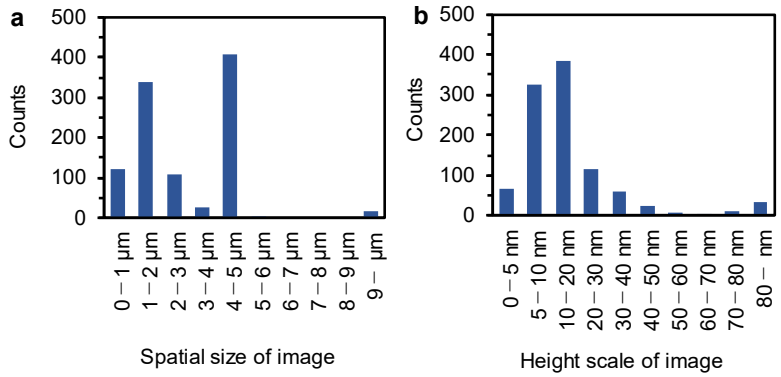

Figure S1. Histogram of AFM topological images collected from literatures (the total number is 1062). (a) The spatial size of image and (b) height scale of image.

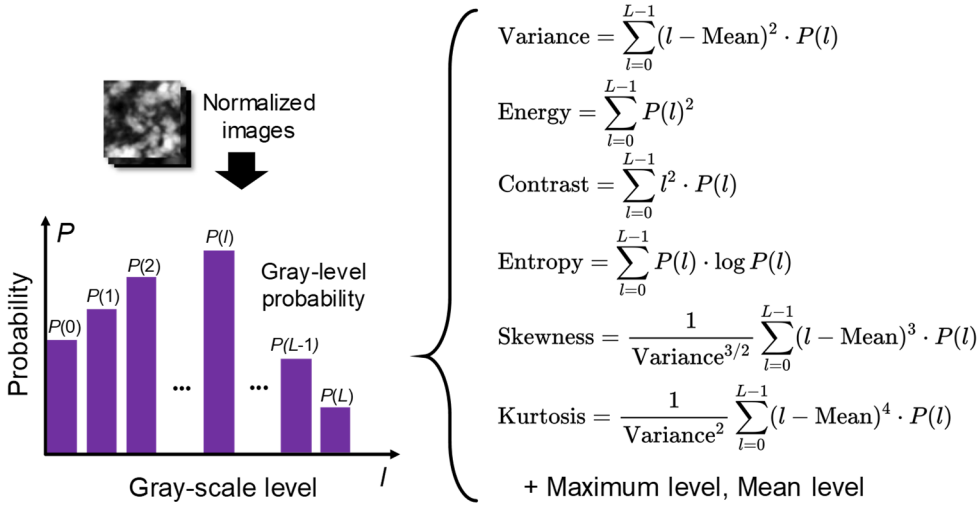

Figure S2. Schematic of histogram analysis (HA). Equations for HA (variance, energy, contrast, entropy, skewness, and kurtosis) are shown. The maximum and mean levels are also considered in ML modeling.

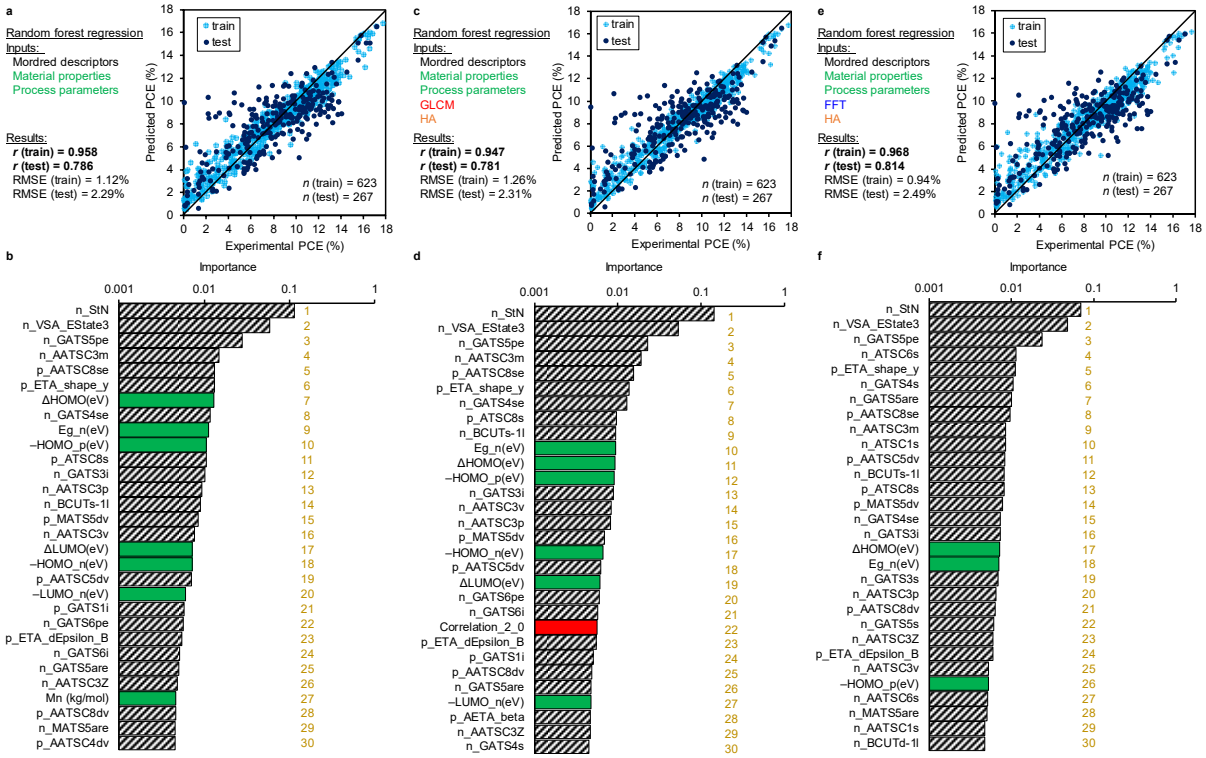

Figure S3. Results of RF regression. (a)(b) The explanatory variables are chemical structures (Mordred descriptors), material properties (bandgap, etc), and process parameters (solvent, etc). (c)(d) The aforementioned parameters plus GLCM and HA data. (e)(f) The aforementioned parameters plus FFT and HA data. The upper panels (a, c, e) are the regression plots of experimental (horizontal) and predicted (vertical) PCE. The white blue (n = 623) and dark (n = 267) circles are train and test data, respectively. The correlation coefficient (r) values of the train and test data are appended. The lower panels (b, d, f) are the feature importance in decent order. The green, red, orange, and blue bars correspond to the material properties and process parameters, GLCM parameters, HA parameters, and FFT parameters, respectively. A complete list of ranking is provided in **Tables S9–S11**.

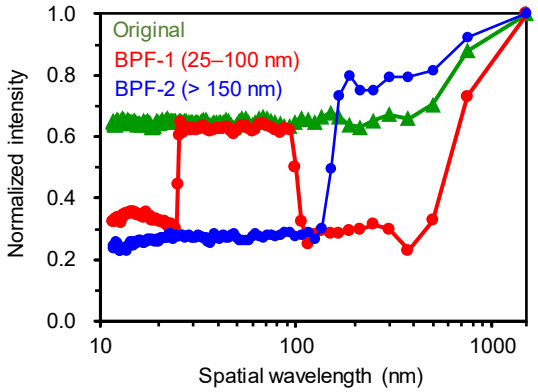

Figure S4. IFPS of gray-scale images generated by applying noise (green), that processed with BPF-1 (25–100 nm, red), and that processed with BPF-2 (> 150 nm, blue).

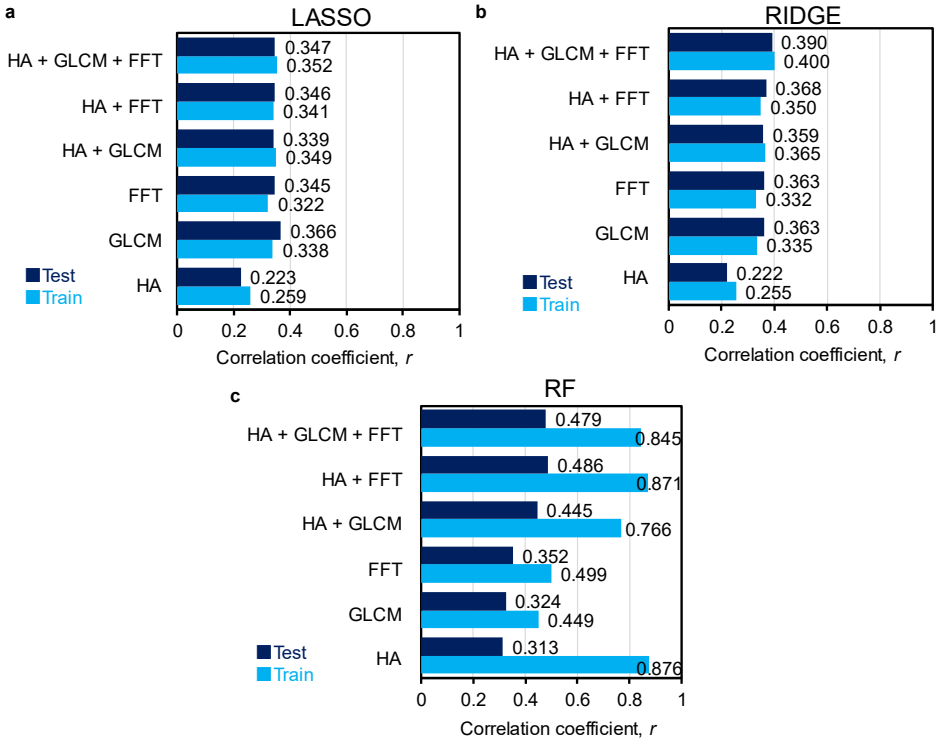

Figure S5. Correlation coefficients of (a) LASSO, (b) Ridge, and (c) RF model constructed solely using AFM data (HA, GLCM, FFT, and their combination) as the explanatory variables. Chemical structures (Mordred descriptors), material properties, and process parameters were not used. The objective variable is PCE. The dark and white blue bars are test and train data, respectively.