ACCSCIENCE PUBLISHING

International Journal of AI for Materials and Design

ORIGINAL RESEARCH ARTICLE

Data-driven identification of functional additives and solution parameters in mixed Sn-Pb perovskite solar cells via β-VAE augmentation

Supplementary Files

S1. Data analysis section

S1.1. Construction of the dataset

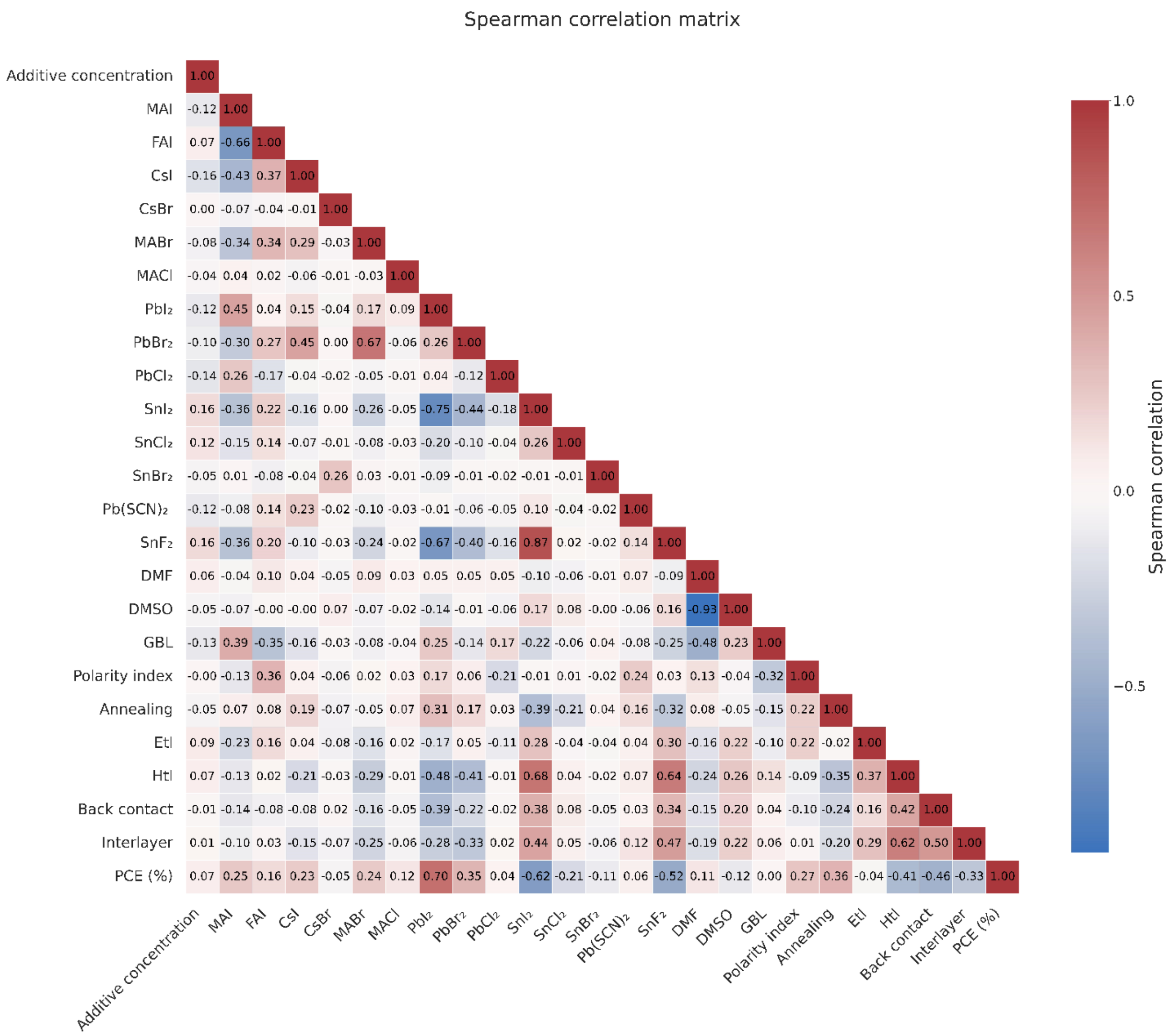


Figure S1. Spearman's rank correlation matrix between all feature pairs. The color intensity and numerical values (ranging from −1 to +1) represent the strength and direction of monotonic relationships. The sign (±) of the value means positive (+) or negative (−) correlation.

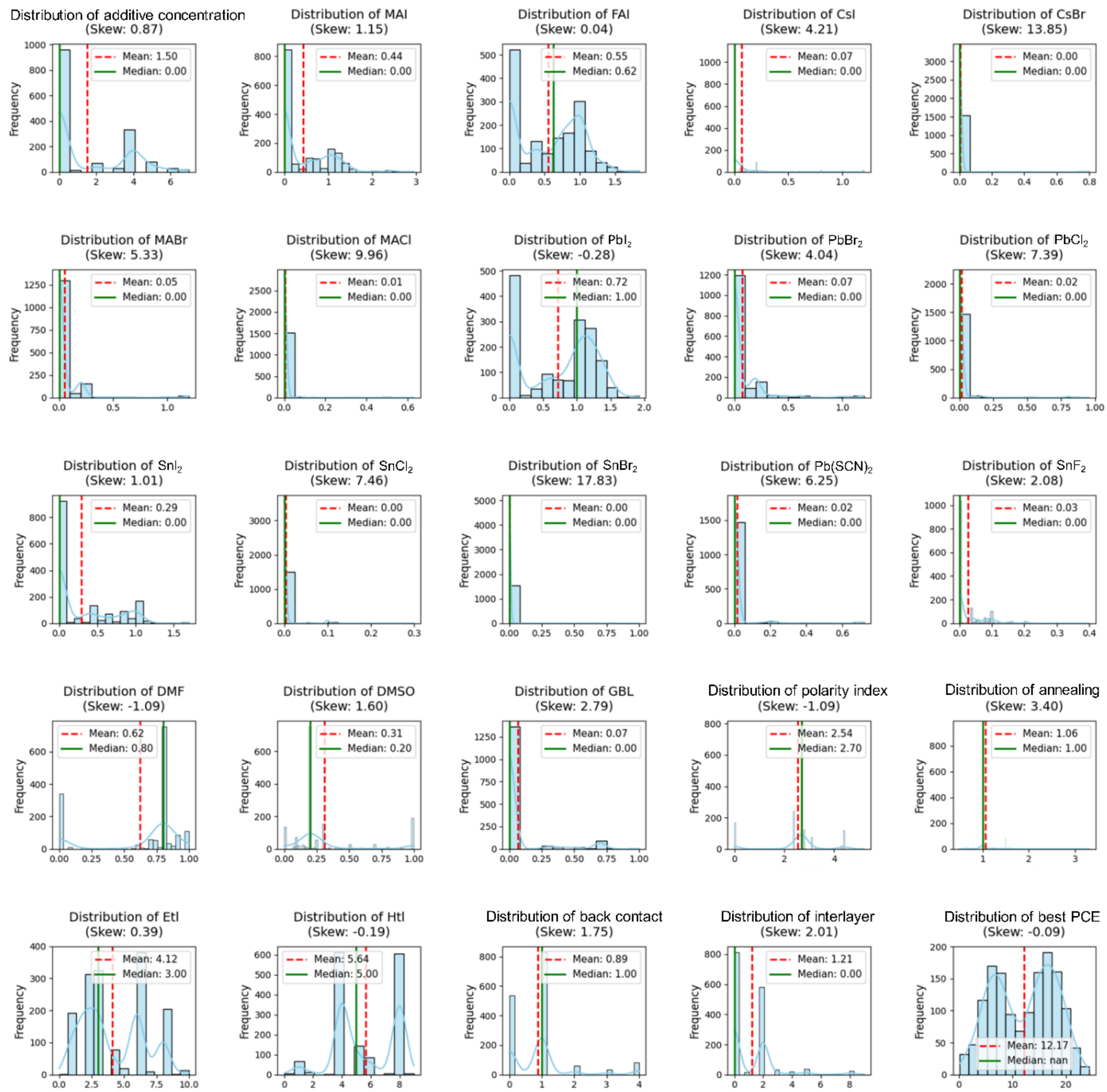


Figure S2. Distribution plots of all features with mean (red dashed line) and median (green solid line) values. Each subplot shows the frequency distribution, skewness measure, and central tendency indicators for one feature. The kernel density estimate (blue line) represents the smoothed probability density.

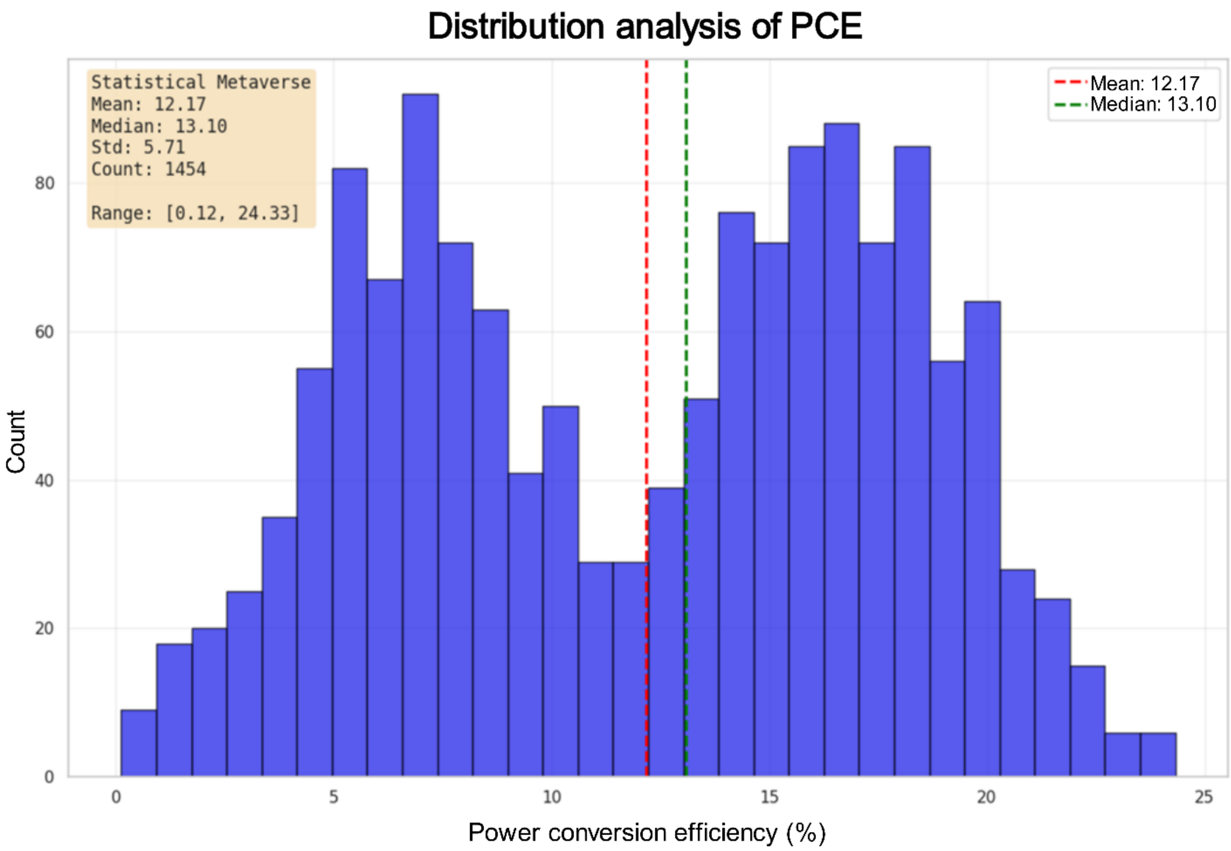


Figure S3. Quantitative distribution of PSC samples in different efficiency ranges
Abbreviations: PCE: Power conversion efficiency; PSC: Perovskite solar cell.

Table S1. Meaning of abbreviations

Abbreviation	Meaning
CGAN	Conditional Generative Adversarial Network
GAN	Generative Adversarial Network
AE	Autoencoder
PCE	Power Conversion Efficiency
ETL	Electron transfer layer
HTL	Hole transport layer
MSE	Mean Squared Error
RMSE	Root mean square error
Ai	Artificial intelligence
ML	Machine learning
DNN	Deep Neural Network
CNN	Convolutional Neural Network
ResNet	Residual Neural Network

(Cont'd...)

Table S1. *Continued*

Abbreviation	Meaning
RF	Random Forest
ETree	Extra Trees
DTree	Decision Tree
ABoost	Ada Boosting
GBoost	Gradient Boosting
MLP	Multilayer Perceptron
XGBoost	Xtreme Gradient Boosting
PCA	Principal Component Analysis
Anti	Antisolvent
EA	Ethyl acetate
Tol	Toluene
CB	Chlorobenzene
DE	Diethyl ether
ctio2	compact TiO_2
mtio2	Meso TiO_2
SHAP	Shapley Additive Explanations

Table S2. Structures utilized in this work

ID	Additive concentration	MAI	FAI	CsI	CsBr	MABr	MACl	PbI_2	$PbBr_2$	PbCl2	SnI2	SnCl2	SnBr2
1	0	1.3	0	0	0	0	0	1.3	0	0	0	0	0
2	0	1.3	0	0	0	0	0	1.4	0	0	0	0	0
3	0	1.3	0	0	0	0	0	1.3	0	0.1	0	0	0
4	0	1.3	0	0	0	0	0	1.25	0	0.15	0	0	0
5	0	2.5	0	0	0	0	0	0.7	0	0.7	0	0	0
6	0	1	0	0	0	0	0.3	1.4	0	0	0	0	0
7	0	1.3	0	0	0	0	0	1.25	0	0.15	0	0	0
8	0	2.4	0	0	0	0	0	0	0	0.8	0	0	0
9	0	2.4	0	0	0	0	0	0.08	0	0.8	0	0	0
10	0	2.4	0	0	0	0	0	0.08	0	0.85	0	0	0
11	0	2.5	0	0	0	0	0	0	0	0.83	0	0	0
12	0	2.5	0	0	0	0	0	0	0	0.875	0	0	0
13	0	2.5	0	0	0	0	0	0	0	0.916	0	0	0
14	0	2.5	0	0	0	0	0	0	0	0.958	0	0	0
15													

ID	$Pb(SCN)_2$	SnF_2	DMF	DMSO	GBL	Polarity index	Annealing temperature	ETL	HTL	Back contact	Interlayer	Best	Reference
1	0	0	0	0.3	0.7	Toluene	95	Ctio2–Mtio2	Spiro	1	0	13.3	1
2	0	0	0	0.3	0.7	Toluene	95	Ctio2–Mtio2	Spiro	1	0	14.5	1
3	0	0	0	0.3	0.7	Toluene	95	Ctio2–Mtio2	Spiro	1	0	15.4	1
4	0	0	0	0.3	0.7	Toluene	95	Ctio2–Mtio2	Spiro	1	0	16.3	1
5	0	0	0	0.3	0.7	Toluene	95	Ctio2–Mtio2	Spiro	1	0	16.4	1
6	0	0	0	0.3	0.7	Toluene	95	PCBM	Pedot	1	0	16.6	1
7	0	0	0	0.3	0.7	Toluene	95	PCBM	Pedot	1	0	18.4	1
8	0	0	1	0	0	0	140	Ctio2–Mtio2	Spiro	0	0	8.41	2
9	0	0	1	0	0	0	140	Ctio2–Mtio2	Spiro	0	0	7.76	2
10	0	0	1	0	0	0	140	Ctio2–Mtio2	Spiro	0	0	8.26	2
11	0	0	1	0	0	0	1.1	Ctio2–Mtio2	Spiro	0	0	9.9	3
12	0	0	1	0	0	0	1.1	Ctio2–Mtio2	Spiro	0	0	10	3
13	0	0	1	0	0	0	1.1	Ctio2–Mtio2	Spiro	0	0	11.6	3
14	0	0	1	0	0	0	1.1	Ctio2–Mtio2	Spiro	0	0	9.8	3

S1.2. Feature extraction and data preprocessing

In the field of machine learning, data preprocessing is a crucial step. This process includes feature encoding, handling missing values, and selecting relevant features. Generally, machine learning and deep learning algorithms cannot process string data and only accept numeric input. We employed label encoding to convert categorical features into numeric representations, enabling the models to extract valuable insights from these encoded features. Data with ambiguous molar values or incomplete feature selections, such as cases where precursor values were unspecified, and only stoichiometry was available, were removed to enhance data quality. To ensure the validity of the results, the test data were collected in a completely separate dataset. Lastly, we standardized the data using the standard scaling method to ensure that different features are equally weighted and scaled, effectively minimizing variation among them. Among the available features, certain elements, such as the electron and hole transport layers, require conversion to numeric data, with training conducted using regression models due to the critical importance of obtaining precise return values.

S2. Machine learning section

S2.1. Generative model

Table S3. Artificial data generated by the beta-variational autoencoder

ID	Additive concentration	MAI	FAI	CsI	CsBr	MABr	MACl	PbI_2	$PbBr_2$	$PbCl_2$	SnI_2	$SnCl_2$	$SnBr_2$
1	0.415153	0.975968	0.110275	0.011836	0.000898	0	0	1.104729	0.001147	0.000878	0.00914	0.004862	0
2	2.671681	0.424697	0.761896	0.006835	0.00664	0.047501	0	1.188655	0.074588	0.003838	0	0.008004	0
3	1.834574	0.059055	0.72823	0.001513	0.003898	0.015017	0	0	0.007379	0.001335	0.874447	0.042894	0.005856
4	4.365298	0.955566	0.130437	0.001492	0.001107	0	0	1.039034	0.006615	0.006691	0.017785	0.007689	0
5	0	0.48297	0.33973	0.124586	0.006998	0.243365	0.00951	0.876672	0.274141	0.002194	0.004887	0	0.005203
6	0	0.204033	1.102474	0.394121	0	0	0.06884	1.26043	0	0.004387	0	0.006261	0
7	3.852087	0.194628	0.714904	0.022343	0.022526	0.005058	0	0	0.002588	0	1.100364	0.043757	0.009699
8	4.063103	0.037132	0.488393	0	0.019922	0.022539	0	0	0.036118	0.005621	0.597227	0.005024	0
9	4.028085	0.062568	0.36366	0	0.006451	0.008079	0	0	0.008323	0.006756	0.437355	0.004478	0.002843
10	0	0.909734	0.217326	0.020649	0.004824	0	0	1.111382	0.014884	0	0.014663	0.002972	0
11	0.045575	0.79482	0.11656	0.033441	0	0.021387	0.001543	1.000966	0.003895	0.006239	0.027988	0.009329	0.002529
12	0	0.834141	0.234588	0.012253	0.000241	0.003656	0	1.079024	0.01398	0.000798	0.018045	0.000907	0
13	1.414932	0.154813	0.944103	0.043393	0	0.101757	0.003992	1.12819	0.120537	0	0	0.01681	0
14	0	0.38098	0.08406	0.687087	0	0	0	0.764029	0.367474	0.012745	0.00074	0	0.004146
15	3.976115	0.265277	0.547215	0.059741	0.010612	0.003766	0	0	0.00303	0.006264	1.100128	0.024306	0.007096
16	0	0.110601	1.252103	0.019122	0	0.030788	0.002124	1.408508	0.059731	0.007607	0	0.00365	0
17	0.068941	0.63466	0.315186	0.000608	0.007044	0.006993	0	0.740386	0.004752	0.060289	0.142123	0	0.007014
18	0	0.498903	0.543473	0.041702	0.007075	0.035802	0	1.08239	0.082692	0.000288	0	0.004452	0
19	0	0.566276	0.598029	0.023452	0.004594	0	0	0.750104	0.058942	0.000784	0.334769	0.001538	0
20	2.76945	0.075052	0.759135	0.013664	0.010703	0.002764	0	0	0	0.003369	0.861244	0.014176	0
21	0	0.806645	0.290523	0.033262	0.01103	0	0	1.091682	0.046414	0.000513	0.009552	0.005576	0
22	0	0.806586	0.272942	0.006302	0.003963	0.002765	0	1.102303	0.007185	0	0.02473	0.000279	0.00028
23	0.151171	0.348895	0.749356	0.030561	0.003616	0.098337	0	1.101848	0.123948	0.001179	0.012546	0.0041	0
24	0.030194	0.629688	0.355561	0.007169	0.008277	0.02073	0	0.507747	0.003846	0	0.490216	0.001224	0
25	2.397815	0.056003	0.648098	0.003684	0.004337	0.005723	0	0	0	0.004009	0.732456	0.008618	0
26	0	0.069968	0.052681	0.093904	0.019478	1.010534	0	0.009201	1.182513	0.006521	0	0.012591	0
27	1.737967	0.089738	1.158899	0.051557	0.004363	0.035425	0	1.29565	0.070367	0.01108	0	0.007885	0
28	3.822678	0.202469	0.682358	0.005386	0.018479	0.001792	0	0.037065	0.014248	0.004293	0.982219	0.01923	0.001768
29	2.333321	0.291541	0.249163	0.145138	0.044589	0.024295	0	0.055544	0.020542	0.000131	0.973734	0.002063	0.008118
30	0	0.401656	0.361574	0	0.002492	0.008285	0	0.15055	0	0.002163	0.508493	0.013162	0.027129
31	3.013562	0.626507	0.803303	0.00154	0.007311	0	0	1.293023	0.138346	0.00781	0	0.010816	0
32	4.200638	0.05887	0.832922	0.00647	0.008791	0.000516	0	0	0.000533	0.006692	0.853294	0.040568	0.000549
33	3.826324	0.07781	1.232324	0.047056	0.00523	0.032411	0	1.455075	0.062034	0.010251	0	0.010618	0.008221
34	2.741039	0.585203	0.632033	0.005081	0.004314	0.024098	0	1.190133	0.058946	0.007367	0.003772	0.004188	0
35	0	0.598965	0.721612	0.018971	0.007874	0	0	1.193295	0.132498	0.002724	0.008327	0.004627	0.001011
36	0.005342	0.765187	0.182293	0.025746	0.00546	0.004116	0	1.046824	0.018615	0.002207	0.015803	0.001759	0
37	0	0.723005	0.353887	0.008279	0.005411	0.010422	0	1.10857	0.024779	0	0.034527	0	0.00802
38	0	0.976784	0.134649	0.010408	0.002026	0	0	1.1162	0.005617	0	0.014088	0.002574	0
39	0	0.475747	0.803904	0.023614	0.007802	0	0.002046	1.195552	0.113967	0.006062	0	0.003337	0

(*Cont'd...*)

Table S3. *Continued*

ID	Additive concentration	MAI	FAI	CsI	CsBr	MABr	MACl	PbI_2	$PbBr_2$	$PbCl_2$	SnI_2	$SnCl_2$	$SnBr_2$
40	0	0.364981	0.490524	0.253648	0.005846	0.000979	0	0.938277	0.228637	0.002116	0.006977	0.000745	0.006452
41	0	0.795745	0.304591	0.06017	0.004562	0	0	1.08719	0.028448	0.003911	0.019561	3.2×10^{-5}	0
42	3.937571	0.048575	0.545002	0.013294	0.011833	0.001272	0	0	0.018882	0.002999	0.637792	0.006367	0.004987
43	0	0.681359	0.40116	0.023788	0.001125	0.031327	0	1.082308	0.057178	0.000336	0.015387	0.002038	0
44	3.947827	0.083926	0.755904	0.001798	0.004843	0.008263	0	0.016117	0.000673	0.006619	0.842602	0.012856	0
45	0.459402	0.066005	0.662317	0.012814	0.007297	0.018067	0	0	0	0.00375	0.831747	0.021027	0.003407
46	0.154269	0.353511	0.819123	0.020964	0.005663	0.079941	0	1.170542	0.120176	0.00365	0.006655	0.001082	0
47	2.98191	0.096335	0.052636	0.929962	0.009926	0.018821	0	0.500056	0.52484	0.007026	0.045649	0	0.013814
48	5.827664	0.567283	0.665147	0	0.00018	0.000912	0	1.188941	0.015888	0.007313	0	0.011924	0
49	0.098849	0.071179	0.616561	0.00303	0.010681	0.009564	0	0	0.000534	0.003705	0.762051	0.002327	0.002029
50	1.656439	0.093068	0.715476	0.011639	0.001326	0.005229	0	0	0	0.002662	0.806727	0.036689	0.000634
51	0	0.746266	0.347579	0.006217	0.002266	0.00757	0	1.116737	0.019932	0	0.030148	0	0.003862
52	1.658873	0.003881	0	0.551155	0.241586	0.014761	0	0.079586	0.056896	0	0.643453	0.002067	0.000984
53	0	0.720111	0.365828	0.034869	0.012574	0.014677	0	1.134003	0.060353	0.00696	0.012811	0.001235	0
54	0	0.685715	0.71422	0.001811	0.005548	0	0	1.248378	0.135873	0.003416	0.005078	0.003702	0
55	3.476727	0.172293	0.73401	0	0.005629	0.011959	0	0.119315	0	0	0.80303	0.014144	0
56	0	0.724921	0.356934	0.019103	0.001794	0.023632	0	1.087631	0.049288	0.001089	0.013854	0.000912	0
57	3.698035	0.169581	0.688913	0.006638	0.000551	0.010689	0	0.100343	0.007407	0.009258	0.71632	0.012028	0
58	3.89677	0.760134	0.76891	0	0.00439	0.003945	0	1.228552	0.140912	0.008348	0.085403	0.021958	0.002416
59	1.406781	0.406864	0.499963	0.250443	0.009654	0	0.000171	0.955012	0.225327	0.007192	0.005609	0.006003	0
60	0.81058	0.292626	0.494558	0.003878	0.000468	0.004753	0	0.104868	0	0.001629	0.659806	0.022509	0.004976
61	0.026496	0.830802	0.092853	0.045063	0.007595	0	0	0.991301	0.007216	0.00167	0	0	0
62	0	0.687236	0.385769	0.010256	0.002901	0.001376	0	1.069938	0.024649	0	0.043505	0	0.005183
63	4.021639	0	0	1.100166	0.014087	0.004233	0	0.419308	0.606865	0.013528	0	0	0
64	2.400138	0.395918	0.440428	0.021235	0.005216	0.008432	0	0	0.011219	0.000411	1.147003	0	0.005109
65	1.455832	0.474131	0.1573	0.446316	0.010682	0.010992	0	0.861775	0.283632	0.001519	0.014937	0	0
66	0	0.622096	0.489416	0.041351	0.006862	0.02654	0	1.122549	0.087746	0	0.008084	0.004659	0
67	3.723433	0.860152	0.496122	0	0.005656	0	0	1.226177	0.096031	0.008319	0.003133	0.013477	0.001083
68	4.157473	0.062742	0.692529	0.017713	0.007559	0.019055	0	0.013912	0.010378	0	0.784824	0.012205	0
69	0.175081	0.033153	0	0.66713	0	0.026298	0	0.006919	0.069862	0	0.562608	0	0
70	0.116678	0.937932	0	0.207402	0	0.002951	0	0.965466	0.097368	0	0.046376	0.000381	0
71	2.650766	0.059129	0.631989	0.002454	0.004448	0.003362	0	0	0	0.004966	0.712841	0.008402	0
72	4.299689	0.005551	1.238514	0.047612	0.004903	0.039636	0	1.380617	0.094714	0.005709	0	0.014807	0.011052
73	4.0525	0.079328	0.918929	0.006424	0.013758	0.021137	0	0	0.006962	0	1.057264	0.010371	0
74	3.995962	0.05021	0.670665	0.000537	0.005894	0.004216	0	0	0.005807	0.007184	0.719533	0.00951	0.000615
75	0.249377	0.069314	0.691883	0	0.011821	0.021616	0	0	0.003064	0.004653	0.845232	0	0.000939
76	0	0.833763	0.328538	0.002492	0.003259	0.016493	0	1.073283	0.026151	0	0.051714	0	0.008265
77	0.063967	0.085305	0.577983	0.00103	0.009753	0.008899	0	0	0	0.004362	0.715508	0.001116	0.001109
78	0	0.236945	0.571045	0.266987	0.000479	0.046661	0	0.925489	0.261281	0	0.012689	0.000119	0.011583
79	4.046864	0.082165	0.575803	0.007283	0.011836	0.023296	0	0	0.014142	0.000959	0.719075	0.002022	0

(*Cont'd...*)

Table S3. *Continued*

ID	Additive concentration	MAI	FAI	CsI	CsBr	MABr	MACl	PbI_2	$PbBr_2$	$PbCl_2$	SnI_2	$SnCl_2$	$SnBr_2$
80	0	0.999364	0.083797	0.016068	0.007809	0	0	0.881957	0	0.002857	0.159598	0.003952	4.34×10^{-6}
81	0	0.675281	0.432606	0.029381	0.002133	0.012447	0	1.068287	0.028825	0.005457	0.018151	0.001642	0
82	0.230017	0.068554	0.64208	0	0.010199	0.01329	0	0	0.002133	0.004308	0.7936	0.007307	0.001596
83	0.419379	0.911935	0.182377	0.007864	0.00084	0	0	1.081262	0.003724	0.002358	0.012436	0.004932	0
84	4.535512	0.662347	0	0.426791	0.00805	0.000849	0	0.847807	0.233449	0.001661	0	0.00481	0
85	0.132952	0.572005	0.566265	0.089468	0.010638	0	0	1.040756	0.035089	0.007573	0.051001	0.003425	0
86	4.296143	0.411156	0.808521	0.022038	0	0.007406	0	1.202132	0.124381	0.00728	0	0.011024	0
87	0	0.459492	0.769624	0.016958	0.006308	0.022854	0	1.173286	0.113649	0.006003	0.003187	0.005581	0
88	2.724993	0.070247	0.737941	0.012917	0.006006	0.002286	0	0	0.002021	0.008203	0.825512	0.009333	0
89	0	0.900796	0.135713	0.022186	0.005521	0	0	1.067342	0.007705	0.00081	0.046765	0.000245	0.002168
90	0.0729	0.456623	0.656452	0.032299	0.001295	0.074058	0	1.130622	0.09318	0.002184	0.005481	0.005154	0
91	0.27556	0.943899	0.148985	0.025427	0.001606	0	0	1.106511	0.010237	0	0.006238	0.005234	0
92	0.374962	0.508274	0.737067	0.010007	0.00577	0.00328	0.006507	1.165526	0.093223	0.007551	0.015151	0.00895	0
93	0	0.714001	0.327919	0.037446	0.009672	0.01	0	1.11749	0.041557	0.004868	0	0.002106	0
94	0.052836	0.084206	0.691466	0.00333	0.008673	0.030459	0	0	0.012027	0.006518	0.857069	0.004502	0
95	0.013127	1.255141	0	0	0.014198	0	0	1.071332	0	0	0.082268	0.000325	0
96	0	0.754279	0.275997	0.017433	0.000688	0.013817	0	1.037885	0.027102	0	0.036267	0.0058	0
97	4.415469	0.055131	0.619785	0.001689	0.009299	0.006322	0	0	0.012047	0.009292	0.713184	0.002456	0.003875
98	0	0.920491	0.135122	0.013845	0.004064	0	0	0.985281	0	0.002254	0.075227	0	0
99	4.09262	0.051991	0.672433	0	0.007966	0.003088	0	0	0.001197	0.004968	0.760615	0.013336	0

ID	$Pb(SCN)_2$	SnF_2	DMF	DMSO	GBL	Polarity index	Annealing temperature	ETL	HTL	Back contact	Interlayer	Best
1	0.004507	0.005705	0.653842	0.274364	0.004365	2.781996	0.993799	1.951343	4.113338	0.272683	0	14.58654
2	0.006514	0	0.816453	0.191269	0	2.689513	1.141532	1.467469	4.056153	0.046702	0	20.25198
3	0.001941	0.069504	0.573343	0.421187	0	2.535017	0.869422	3.03164	6.969278	1.200715	1.065184	4.764983
4	0.006094	0.007129	0.70912	0.187289	0.031602	2.671946	1.056816	1.597112	4.0613	0.151123	0	15.61034
5	0	0.008753	0.587028	0.422216	0	2.60347	1.282152	1.772297	0.842206	3.495469	0.766398	10.02643
6	0	0	0.285202	0.700093	0	2.683446	1.507121	1.966439	2.015386	0.299045	0	16.31469
7	0	0.08382	0.575558	0.411869	0	1.358734	0.724671	2.148974	5.152972	0.766942	0	5.012728
8	0.007939	0.066589	0.653038	0.357238	0	2.713806	0.971487	6.028769	7.648202	1.231145	1.945422	8.585852
9	0.006946	0.059331	0.769076	0.236272	0	2.751518	0.985815	6.200039	8.015178	0.972329	1.9571	7.253573
10	0.005283	0.00468	0.704395	0.245329	0	2.745076	0.982559	2.59217	4.47596	0.466399	0.33884	15.41595
11	0	0.009755	0.745825	0.215455	0.00762	2.665277	1.046645	2.029695	3.828478	0.558996	0.188876	13.70804
12	0.004438	0.006222	0.680557	0.264615	0	2.727282	0.981847	1.581142	4.003014	0.182513	0	14.42473
13	0.003138	0	0.81862	0.200567	0	2.718226	1.11099	1.596126	4.055007	0.064025	0	20.44584
14	0	0.030049	0.37023	0.556345	0.02814	0.821779	1.431086	4.346721	3.97955	0.375423	0.237506	9.302351
15	0.007539	0.136519	0.65666	0.340007	0	1.266383	0.78572	2.340875	6.113974	0.584544	0.115654	4.171223
16	0.006964	0	0.842115	0.145524	0	3.164644	1.037469	8.131218	3.997874	0	0	20.75144
17	0.005205	0.041696	0.305701	0.547034	0.207004	2.551908	1.225512	5.179726	8.117341	0.948284	3.898627	3.633995
18	0	0	0.691722	0.293428	0	2.688687	1.025316	1.530062	3.539953	0.344486	0	14.8958

(Cont'd...)

Table S3. *Continued*

ID	$Pb(SCN)_2$	SnF_2	DMF	DMSO	GBL	Polarity index	Annealing temperature	ETL	HTL	Back contact	Interlayer	Best
19	0.007747	0.045732	0.722169	0.259935	0	2.917987	1.0418	5.331677	7.593105	1.219052	1.842356	13.28698
20	0.00254	0.105496	0.436442	0.547133	0	2.63271	0.990295	3.449471	7.965357	1.054921	1.730472	7.158906
21	0.006563	0.001792	0.711817	0.241868	0	2.721084	0.988995	2.828916	4.358005	0.419054	0.07532	15.51248
22	0.003532	0.009398	0.677554	0.242042	0.0045	2.707007	0.989654	1.524641	3.971555	0.14332	0	12.47761
23	0.005893	0	0.774705	0.218157	0	2.739047	0.975716	1.494628	4.081084	0.0982	0	18.05435
24	0.007208	0.05545	0.588736	0.328594	0.002688	2.611898	0.974607	3.995768	8.167687	1.113322	2.182431	10.21293
25	0.005283	0.089486	0.4155	0.56399	0	2.73435	0.995179	5.932817	8.072737	0.982842	1.996629	6.11762
26	0.006679	0.013535	0.550456	0.468064	0.008966	2.432986	1.29433	0.691046	1.013779	3.930223	0	6.283865
27	0.007199	0.002161	0.802112	0.184073	0	2.773235	1.062734	7.152713	4.127144	0.18897	0	20.8913
28	0.000164	0.086947	0.76778	0.230092	0	0.934146	0.863541	1.908913	5.372181	0.759535	0.045514	6.020408
29	0	0.13708	0.864742	0.118494	0.000165	0	0.796099	1.750474	5.361304	0.836563	0.011828	4.281732
30	0	0.010522	0.909349	0.114802	0	2.498332	0.956892	2.051815	4.020161	0.443655	0.052806	3.242928
31	0.011081	0.007372	0.777044	0.216711	0	3.855973	1.181758	8.104069	3.989277	0.164804	0	19.36132
32	0.004434	0.075107	0.401846	0.607488	0	2.717194	0.926625	3.761095	7.398759	1.085982	1.370285	6.60632
33	0.010785	0.006501	0.834527	0.160657	0	3.069955	0.989686	8.112378	4.056151	0.119528	0	22.152
34	0.005877	0	0.804841	0.19266	0	2.706077	1.117581	1.829618	4.080168	0.075138	0	19.5777
35	0.003406	0.007384	0.752264	0.241501	0	3.273227	1.132348	8.013732	4.001014	0.093325	0	17.30164
36	0.00399	0.010857	0.66147	0.222871	0.025412	2.311073	1.029529	1.784899	3.831802	0.223905	0.029745	11.13605
37	0	0.008795	0.691176	0.224796	0.003681	2.724868	1.006645	1.784998	4.034126	0.136013	0.040712	11.29764
38	0.004568	0.004231	0.679175	0.257074	0.007514	2.724567	0.99118	2.239026	4.318815	0.332615	0.096332	14.57998
39	0.004773	0.004337	0.757807	0.236655	0	3.216846	1.135974	8.14013	4.004375	0.068173	0	18.04582
40	0.004191	0.015083	0.582527	0.377664	0	2.026455	1.162283	7.913386	4.008441	0.338138	0.00615	15.58517
41	0.000903	0.007504	0.770726	0.231589	0	2.89433	0.988669	2.993253	5.354783	0.935824	2.051771	15.63292
42	0.003877	0.061074	0.574742	0.426229	0	2.724416	0.979194	6.042843	7.909952	0.935351	2.016864	8.680302
43	0.006903	0.001707	0.707465	0.259911	0	2.753636	0.988163	1.64151	4.02633	0.127199	0	16.03975
44	0.01029	0.107533	0.479938	0.512182	0	2.621484	0.954453	6.225921	8.104207	0.978812	1.98604	11.03447
45	0	0.057253	0.726747	0.238452	0.002824	2.639145	0.870451	3.328886	6.916027	1.151567	1.821472	4.869885
46	0.006009	0	0.793643	0.191154	0	2.735573	1.015534	2.671584	3.969702	0	0	19.06752
47	0.002451	0.001039	0.274918	0.609436	0.006926	0.333855	1.526766	7.517873	4.379444	0.826816	0.490923	14.41609
48	0.008518	0.003067	0.873634	0.156363	0	2.708899	1.215075	1.160866	4.048873	0.054895	0	19.35637
49	0.00255	0.090412	0.461972	0.510426	0	2.706749	0.973415	5.596983	8.061931	0.966154	1.957284	4.728098
50	0	0.049364	0.810851	0.182443	0	2.685441	0.812019	3.116415	6.439332	1.11315	2.016018	6.66367
51	0.000351	0.011943	0.674179	0.219646	0.011736	2.709335	1.003446	1.396054	3.953491	0.061966	0	11.28951
52	0	0.144614	0.717024	0.282395	0.015032	0.052939	0.729357	2.801657	4.580142	3.766039	0.280731	4.013611
53	0	0.005182	0.743673	0.230926	0	2.747285	0.983162	2.441186	4.668364	0.484597	1.644247	16.93555
54	0.001258	0.009342	0.773313	0.211356	0	3.561162	1.140836	7.968908	4.040699	0.098056	0	17.31913
55	0.0154	0.099647	0.54614	0.457482	0	2.602035	0.951669	6.199013	8.145935	0.982788	2.037219	11.78433
56	0.007116	0.001793	0.693953	0.265184	0	2.739645	0.989997	1.587555	4.033373	0.118662	0	15.53584
57	0.012552	0.092289	0.506173	0.483134	0	2.716126	0.978114	5.932925	8.126912	0.985401	2.009456	10.82262
58	0.00817	0.005455	0.779295	0.21009	0	3.886987	1.150383	7.835442	4.611137	0.393191	0.526796	18.99934
59	0.010966	0.011009	0.700145	0.278978	0	2.207833	1.279939	8.023592	4.012227	0.166833	0.001596	17.25504

(*Cont'd...*)

Table S3. *Continued*

ID	$Pb(SCN)_2$	SnF_2	DMF	DMSO	GBL	Polarity index	Annealing temperature	ETL	HTL	Back contact	Interlayer	Best
60	0	0.022397	0.880104	0.043606	0.02419	2.014932	0.832416	2.903543	5.704217	0.988805	1.527616	6.290647
61	0.006924	0.001103	0.674273	0.263239	0.007739	2.604614	1.042141	2.497002	4.795558	0.589243	0.12425	12.94294
62	0.001601	0.015732	0.705005	0.215049	0	2.780181	1.00707	1.463155	3.972432	0.103031	0	11.48152
63	0.009819	0	0.410277	0.513367	0	0	1.80308	7.757024	4.019501	0.343199	0.290178	15.94089
64	0	0.146367	0.827365	0.159612	0	0.42584	0.832657	2.015358	5.732443	0.144726	0	4.467381
65	0.017808	0.023097	0.526849	0.35118	0.041909	1.461625	1.253452	6.370577	4.848197	0.537447	0.173055	16.14652
66	0.007222	0.001416	0.725724	0.245064	0	2.690129	0.991485	2.319047	4.283534	0.31759	0.224345	17.08794
67	0	0.016488	0.649016	0.232063	0	3.525677	1.086959	8.105436	4.118848	0.469648	0.166695	17.29155
68	0	0.106344	0.653902	0.369734	0	2.6792	0.940805	6.446777	8.109418	1.018125	1.876118	11.13962
69	0	0.211768	0.6381	0.387094	0.022951	0	0.805784	2.464815	0	1.334054	0.075806	1.5634
70	0	0	0.529207	0.158926	0.255973	0.368309	1.021952	2.548588	4.455439	0.34076	0.197229	8.584852
71	0.005475	0.08683	0.431756	0.548581	0	2.730258	0.994222	6.000895	8.064573	0.972023	1.989518	6.263808
72	0.012576	0.00583	0.820672	0.161118	0	2.939582	1.014945	8.092123	4.101423	0.15174	0	21.74903
73	0.002353	0.112527	0.204042	0.74693	0.012973	2.531962	0.881294	5.951871	7.930334	0.984074	1.955222	1.999704
74	0.006672	0.088651	0.515989	0.485524	0	2.738224	0.98583	6.155052	8.0447	0.981881	1.994328	8.951317
75	0	0.086947	0.392465	0.558401	0	2.693521	0.959289	6.082873	8.023767	0.973766	2.011885	3.627284
76	0	0.022469	0.693518	0.196829	0.015607	2.679809	1.021682	1.698147	4.220066	0.066949	0.064539	8.678481
77	0.003051	0.083293	0.448795	0.52334	0.002366	2.709846	0.984289	6.158834	8.05674	0.977812	2.011601	5.415419
78	0.00216	0.018094	0.563519	0.397163	0	1.934533	1.113342	7.909039	3.975561	0.686283	0	14.12412
79	0.001579	0.094327	0.601326	0.365421	0	2.751075	0.93431	6.891453	8.07478	1.058285	1.941853	10.53401
80	0.010603	0.017803	0.536682	0.25658	0.146558	2.741976	0.985058	3.054684	8.163182	0.976467	0.800261	13.48293
81	0.00288	0.002108	0.690509	0.273895	0	2.676331	1.02205	1.749531	3.731347	0.340654	0	15.34278
82	0.004392	0.091664	0.475168	0.50169	0	2.693204	0.958535	5.007619	8.031444	0.994419	1.882079	4.551345
83	0.003756	0.004855	0.669382	0.26715	0.001091	2.833208	0.985583	1.899688	3.98523	0.224003	0	15.3013
84	0.008293	0.011517	0.596801	0.214954	0.111464	0.847663	1.306359	4.199836	5.206449	0.272421	0.06333	16.0837
85	0.010565	0.009387	0.787102	0.205597	0	2.755401	0.983492	3.158863	5.858042	0.973419	2.274765	18.03472
86	0.010698	0.014957	0.790618	0.218307	0	3.093062	1.094626	6.455305	4.080322	0.338922	0	19.15299
87	0.00441	0.00068	0.771229	0.227107	0	3.051347	1.055167	6.03266	3.971569	0.046373	0	18.09297
88	0.005864	0.101478	0.346292	0.636072	0	2.712122	0.983553	5.920352	8.088679	1.001413	2.017472	8.103883
89	0.002551	0.007464	0.745008	0.214994	0	2.764964	0.990481	2.339027	4.865187	0.454256	0.786609	13.06747
90	0.006438	0	0.729598	0.248785	0	2.732813	0.997538	1.511	4.049035	0.073937	0	18.18911
91	0	0.000999	0.670698	0.285713	0	2.859819	0.998043	2.099519	4.476734	0.424558	0.548244	15.4614
92	0.009847	0.005211	0.725296	0.260756	0	3.036689	1.133963	8.047119	4.026911	0.142684	0	17.93184
93	0	0.0001	0.710324	0.272708	0	2.721387	0.976636	2.039178	4.287843	0.351812	1.522122	16.88998
94	0	0.085315	0.423801	0.499933	0.004711	2.6401	0.893735	5.31953	8.015256	0.968298	1.873674	2.30798
95	0	0.00167	0.368022	0.078701	0.517211	0.185894	0.985754	2.36178	5.153667	0.580447	0	7.633739
96	0.002983	0.008244	0.681943	0.29164	0	2.761635	0.99968	1.494435	3.55098	0.594712	0.066672	13.5318
97	0.005535	0.08553	0.488706	0.50684	0	2.754383	0.991295	6.217269	8.022775	0.976603	1.976795	6.11403
98	0.001079	0.011071	0.702345	0.238639	0.036493	2.801138	0.996788	2.885433	6.240924	0.756222	0.73084	12.70539
99	0.004619	0.086666	0.545149	0.452512	0	2.702073	0.973192	4.453375	8.020393	1.009897	1.801355	6.979408

Table S4. Optimizing the Variational Autoencoder model to enhance data generation, focusing on minimizing reconstruction loss and Kullback–Leibler (KL) divergence across diverse architectures

ID	Hidden_dims	Batch_ size	Learning Rate	Annealing_ epochs	Latent_ dim	Dropout_ rate	Validation			
							Kl loss	Beta	Reconstruction loss	Total loss
1	[512, 256, 128, 64, 32, 25]	32	0.0001	No	100	0.2	0.0873	1	0.8472	0.9345
2	[512, 256, 128, 64, 32, 25]	32	0.0001	No	50	0.2	0.1160	1	0.4409	0.5569
3	[512, 256, 128, 64, 32, 25]	32	0.0001	No	20	0.2	0.1213	1	0.7767	0.8979
4	[512,256, 128, 64, 32, 25]	32	0.0001	No	10	0.2	0.3609	1	0.420	0.7808
5	[256, 128, 64, 32, 25]	32	0.0001	No	10	0.2	0.3549	1	0.4221	0.7771
6	[1024, 512, 256,128, 64, 32]	32	0.001	No	10	0	0.1625	3	0.7853	1.2726

S2.2. Principal component analysis and *t*-distributed stochastic neighbor embedding

Dimensionality reduction techniques are essential for the analysis and visualization of high-dimensional data. Among these methods, principal component analysis (PCA) and t-distributed stochastic neighbor embedding (t-SNE) are widely used, but they differ significantly in their methodologies, objectives, and applications. PCA is a linear technique that seeks to identify axes that capture the greatest variance in the data. It projects the data onto these new axes, thereby reducing dimensionality while preserving the most informative elements. PCA is characterized by its speed, simplicity, and interpretability; however, its performance is limited when dealing with complex and nonlinear data structures. Conversely, t-SNE is a nonlinear technique that prioritizes the preservation of local relationships within the data, striving to maintain proximity among points that are close in the original space when represented in a lower-dimensional context. This approach is particularly advantageous for revealing clusters and intricate structures within the data. Nevertheless, t-SNE is generally slower than PCA, requires more parameter adjustments, and produces results that can be more challenging to interpret.

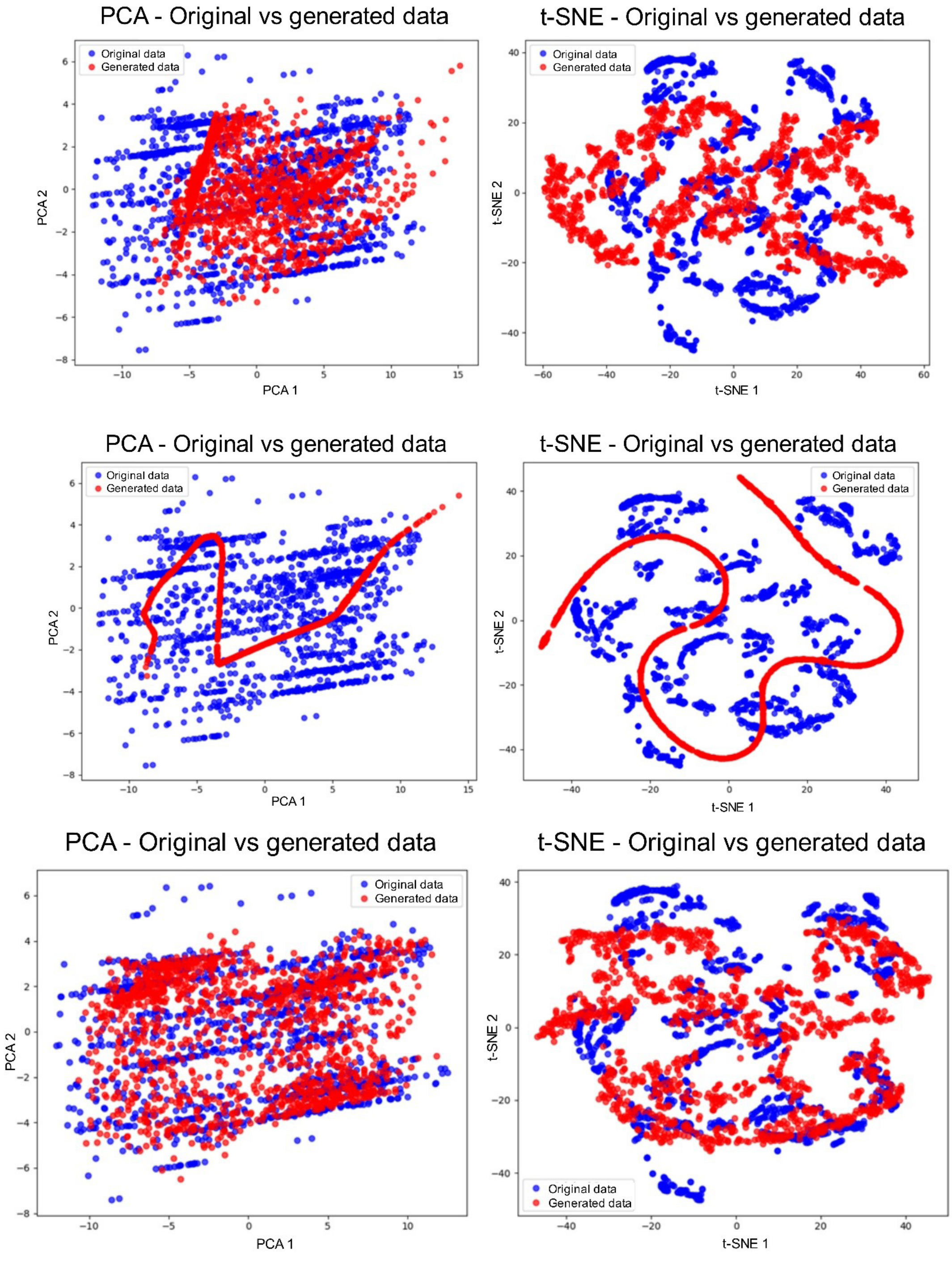


Figure S4. The process of optimizing hyperparameters to enhance the understanding of data distribution

Table S5. Results for the main dataset

Model	Train RMSE	Train MAE	Train R^2	Train Explained Variance	Test RMSE	Test MAE	Test R^2	Test Explained Variance	Model Type
GBoost	1.01891	0.765376	0.967762	0.967762	2.142974	1.593454	0.822256	0.836276	Classical
XGBoost	1.126246	0.842245	0.960613	0.960613	2.145163	1.576399	0.821892	0.839197	Classical
ETree	1.467393	1.050586	0.933137	0.933137	2.209748	1.65998	0.812345	0.823089	Classical
Wide & Deep	2.71232	1.918994	0.77156	0.772001	2.257236	1.818707	0.802796	0.803994	Neural network
Random Forest	1.316496	0.977451	0.946182	0.946195	2.282132	1.660084	0.798422	0.822253	Classical
Residual NN	2.646561	1.885426	0.782502	0.782744	2.616682	2.052683	0.734989	0.752479	Neural network
SVM	2.647754	1.817086	0.782306	0.782335	2.763939	2.059022	0.704322	0.765415	Classical
1D CNN	2.695377	1.973595	0.774405	0.774409	2.764966	2.193664	0.704103	0.727149	Neural network
ABoost	2.894045	2.310834	0.739923	0.743549	2.926813	2.302578	0.668448	0.77009	Classical
MLP	1.577639	1.115052	0.922713	0.923278	2.996738	2.077032	0.652417	0.675847	Classical
Deep NN	2.506645	1.644033	0.804891	0.805033	3.002193	2.238039	0.65115	0.679979	Neural network
TabTransformer	3.083218	2.083531	0.704811	0.714143	3.056922	2.340301	0.638315	0.660692	Neural network
Attention NN	3.82298	2.860718	0.546168	0.641394	3.181763	2.673999	0.60817	0.621095	Neural network
DTree	1.389911	0.977655	0.940012	0.940012	3.483989	2.37507	0.530198	0.553039	Classical
Simple NN	3.596028	2.565804	0.598452	0.599935	4.193914	3.279886	0.31923	0.377697	Neural network

Abbreviations: 1D CNN: One-Dimensional Convolutional Neural Network; ABoost: Adaptive Boosting; Attention NN: Attention Neural Network; Deep NN: Deep Neural Network; DTree: Decision Tree; ETree: Extra Trees; GBoost: Gradient Boosting; MAE: Mean absolute error; MLP: Multilayer Perceptron; R^2: coefficient of determination; Residual NN: Residual Neural Network; RMSE: Root mean square error; Simple NN: Simple Neural Network; SVM: Support Vector Machine; XGBoost: Extreme Gradient Boosting.

Table S6. Results for the integrated synthesis and main dataset

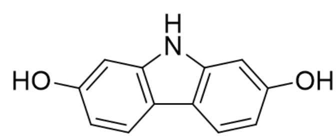
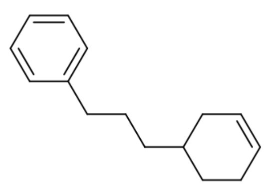
Model	Train RMSE	Train MAE	Train R^2	Train Explained Variance	Test RMSE	Test MAE	Test R^2	Test Explained Variance	Model Type
Residual NN	1.15409	0.81505	0.95675	0.957091	1.83425	1.49550	0.86977	0.870269	Neural network
XGBoost	0.78710	0.53327	0.97988	0.979887	1.85052	1.46666	0.86745	0.882549	Classical
GBoost	0.75593	0.52089	0.98144	0.981449	1.92522	1.51371	0.85654	0.870206	Classical
ETree	1.14425	0.80135	0.95749	0.957494	1.92686	1.41360	0.85629	0.871447	Classical
Wide & Deep	1.28937	0.90128	0.94602	0.946828	1.95502	1.46170	0.85206	0.85471	Neural network
Deep NN	1.24521	0.83903	0.94966	0.949689	2.14076	1.71923	0.82262	0.837063	Neural network
1D CNN	1.41560	1.01551	0.93494	0.93533	2.22560	1.76135	0.80828	0.811487	Neural network
Random Forest	0.90893	0.62888	0.97317	0.973186	2.29563	1.67722	0.79602	0.81277	Classical
Simple NN	1.08842	0.73756	0.96154	0.961547	2.41174	1.9122	0.77487	0.775628	Neural network
Attention NN	1.97613	1.40630	0.87322	0.883468	2.42617	1.84600	0.77217	0.784079	Neural network
TabTransformer	1.83549	1.20104	0.89062	0.895641	2.60854	2.03445	0.73663	0.737374	Neural network
MLP	0.87951	0.58317	0.97488	0.974951	2.82078	2.13989	0.69203	0.699237	Classical
SVM	1.68999	1.03468	0.90727	0.907816	2.97368	2.15722	0.65774	0.676371	Classical
ABoost	2.43134	1.92877	0.80808	0.810895	3.15449	2.49119	0.61485	0.760332	Classical
DTree	1.040986	0.701719	0.96482	0.96482	3.391377	2.211752	0.55484	0.590323	Classical

Abbreviations: 1D CNN: One-Dimensional Convolutional Neural Network; ABoost: Adaptive Boosting; Attention NN: Attention Neural Network; Deep NN: Deep Neural Network; DTree: Decision Tree; ETree: Extra Trees; GBoost: Gradient Boosting; MAE: Mean absolute error; MLP: Multilayer Perceptron; R^2: coefficient of determination; Residual NN: Residual Neural Network; RMSE: Root mean square error; Simple NN: Simple Neural Network; SVM: Support Vector Machine; XGBoost: Extreme Gradient Boosting.

S3. Organic additives

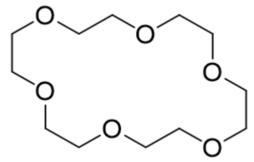
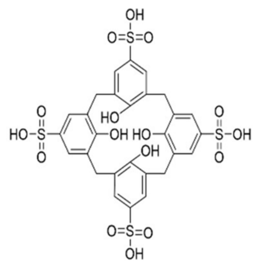
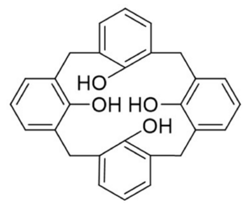
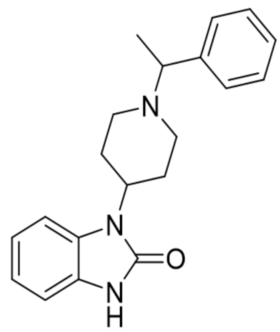
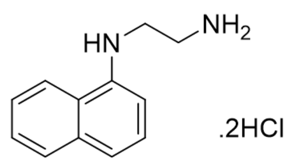
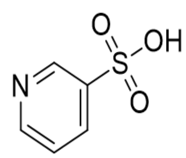
S3.1. Candidate additives identified by models

Table S7. 20 New organic additives candidates with potentially high power conversion efficiencies were identified using the XGBoost regression model, with the increase in PCE set as the target variable

Sample number	Candidate	Classification	Target	Chemical Structure
1	9H-Carbazole-2,7-diol	Aromatic alcohol	2.69	
2	[3-(Cyclohex-3-en-1-yl)propyl]benzene	Aromatic hydrocarbon	2.55	

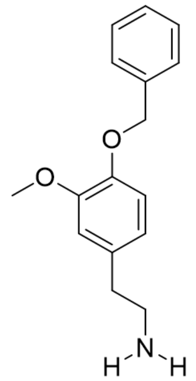
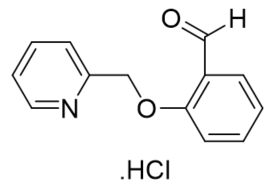
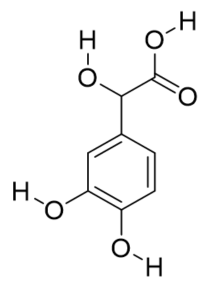
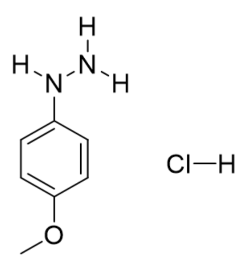
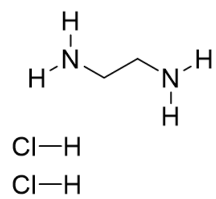
(*Cont'd...*)

Table S7. *Continued*

Sample number	Candidate	Classification	Target	Chemical Structure
3	18-Crown-6-	Polyether	2.17	
4	4-Sulfocalix[4]arene	Calixarene	1.89	
5	Calix[4]arene-25,26,27,28-tetrol	Calixarene	2.81	
6	1,3-dihydro-1-[1-(1-phenylethyl)-4-piperidinyl]-2H-benzimidazol-2-one	Heterocyclic compound	4.26	
7	N-1-Naphthylethylenediamine dihydrochloride	Aromatic amine salt	2.44	
8	Pyridine-3-sulfonic acid	Heterocyclic aromatic	1.76	

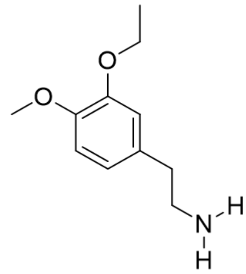
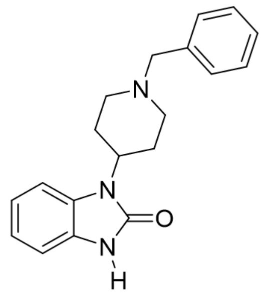
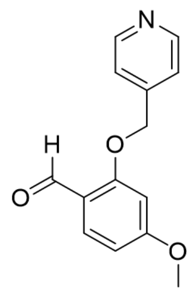
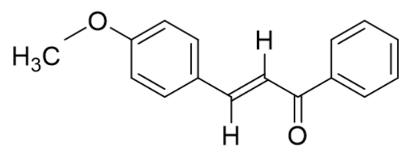
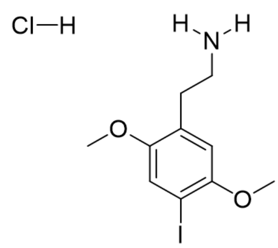
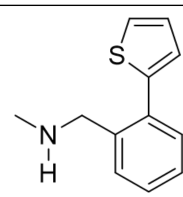
(*Cont'd...*)

Table S7. *Continued*

Sample number	Candidate	Classification	Target	Chemical Structure
9	Phenethylamine, 4-(benzyloxy)-3-methoxy-	Phenethylamine	2.16	
10	2-(4-pyridinylmethoxy)benzaldehyde hydrochloride	Aromatic aldehyde	4.15	
11	2 -hydroxy-2-(4-hydroxy-3-methoxyphenyl) acetic acid	Aromatic hydroxy acid	2.95	
12	4'-Methoxychalcone	Flavonoid / aromatic ketone	2.3	
13	Ethylenediamine dihydrochloride	Amine Salt	1.34	

(*Cont'd...*)

Table S7. *Continued*

Sample number	Candidate	Classification	Target	Chemical Structure
14	-Ethoxy-4-methoxyphenethylamine-3	Aromatic Amine	2.14	
15	1,3-Dihydro-1-[1-(phenylmethyl)-4-piperidinyl]-2H-benzimidazol-2-one	Aromatic Heterocyclic Amines	4.41	
16	4-Methoxy-2-(4-pyridinylmethoxy) benzaldehyde	Aromatic Aldehydes	3.97	
17	4-Methoxyphenylhydrazine hydrochloride	Hydrazine Salts	2.94	
18	2,5-Dimethoxy-4-iodophenethylamine Hydrochloride	Phenethylamine	2.8	
19	N-methyl-1-(2-thiophen-2-ylphenyl) methanamine	Phenylthiophene Amine	2.55	

References

1. Jiang F, Rong Y, Liu H, *et al*. Synergistic Effect of PbI2 Passivation and Chlorine Inclusion Yielding High Open-Circuit Voltage Exceeding 1.15 V in Both Mesoscopic and Inverted Planar CH3NH3PbI3 (Cl)-Based Perovskite Solar Cells. *Adv Funct Mater.* 2016; *26*(44):8119-8127.

   doi: 10.1002/adfm.201603968

2. Ueoka N, Oku T. Stability characterization of PbI2-added CH3NH3PbI3−x Cl x photovoltaic devices. *ACS Appl Mater Interfaces* 2018;*10*(51):44443-44451.

   doi: 10.1021/acsami.8b16029

3. Zhang Z.L, Men B.Q, Liu Y.F, Gao H.P, Mao Y.L. Effects of precursor solution composition on the performance and IV hysteresis of perovskite solar cells based on CH3NH3PbI3-xClx. *Nanoscale Res Lett.* 2017;12(1):84.

   doi: 10.1186/s11671-017-1872-8