Experimentally Validated Machine Learning Predictions of Ultralow Thermal Conductivity for SnSe Materials

N.K. Barua,[a§] A. Golabek,[a§] A.O. Oliynyk,[b] and H. Kleinke[a]*

*Corresponding author

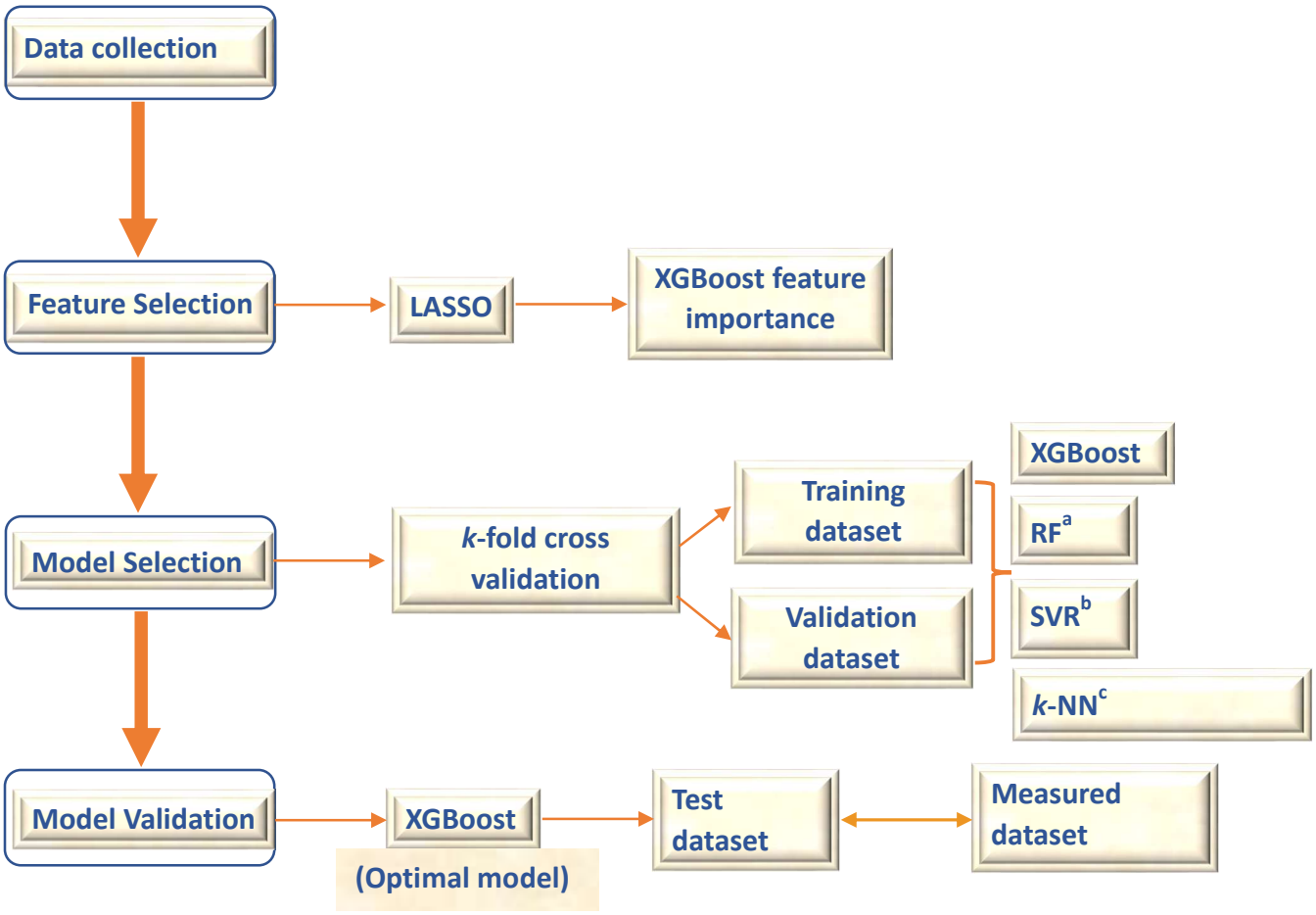
§Authors made equal contributions

[a] Department of Chemistry, University of Waterloo, 200 University Ave W, Waterloo, ON, Canada

E-mail: kleinke@uwaterloo.ca

[b] Department of Chemistry, Hunter College, City University of New York, New York 10065, NY, United States

Table S0. Table of contents in the supplementary information.



[a] *Random Forest*
[b] *Support Vector Regression*
[c] *k-Nearest Neighbor*

Figure S1. Schematic diagram of the machine learning workflow for this study.[1]

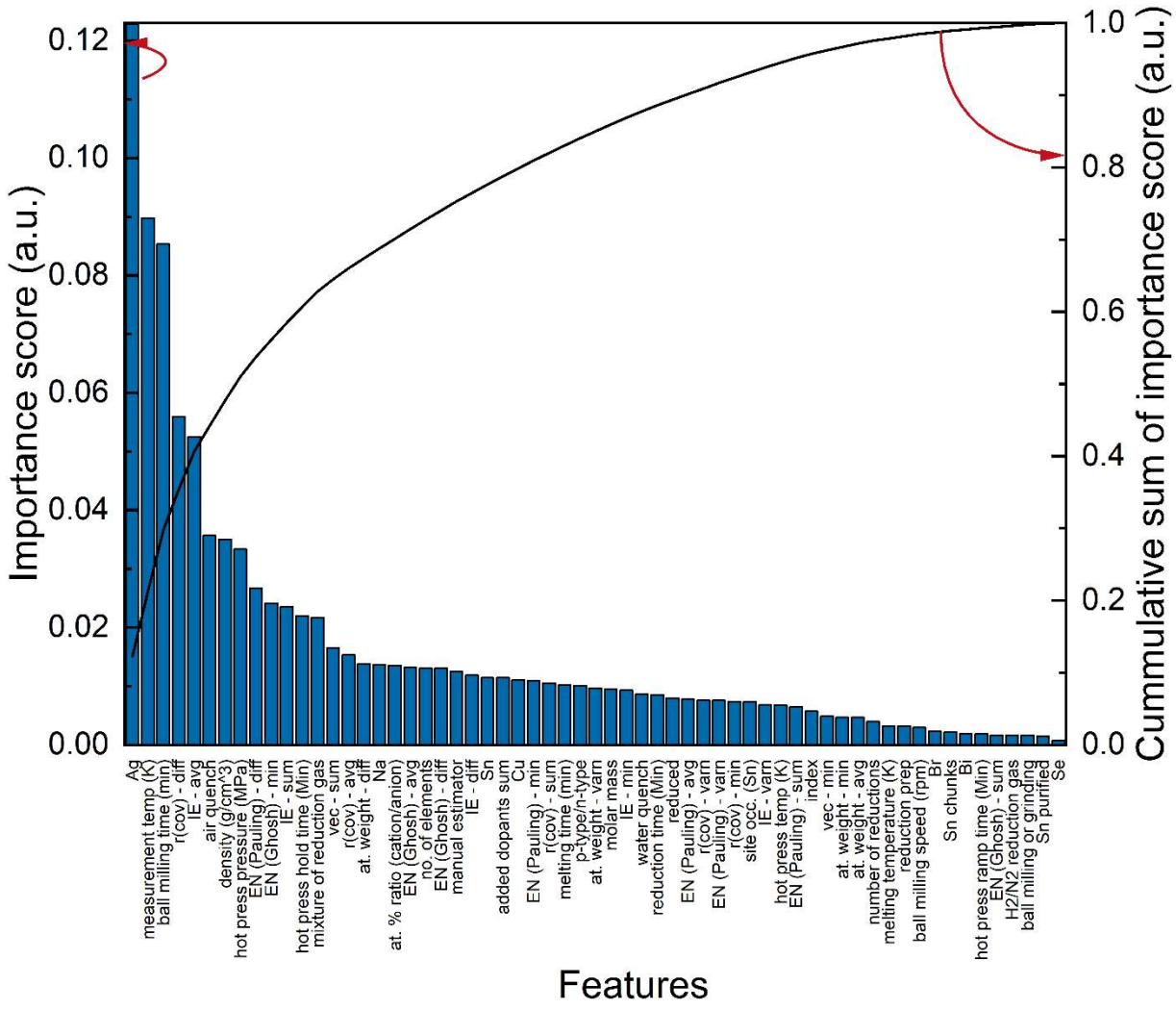


Figure S2. Extended list of features ranking based on importance score as suggested by XGBoost.

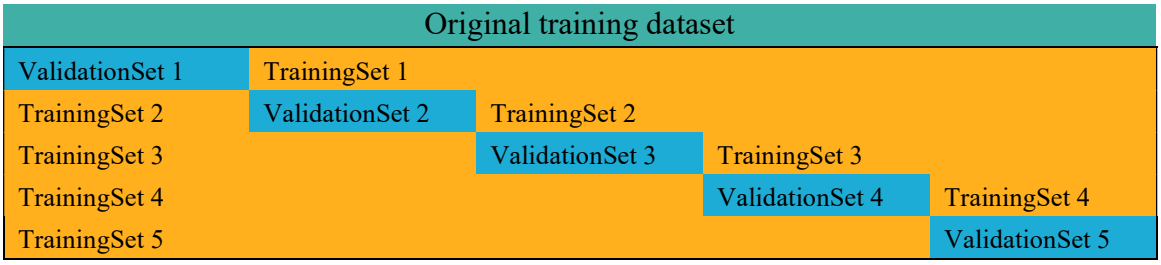


Figure S3. Schematic diagram of five k-fold cross validation.[2]

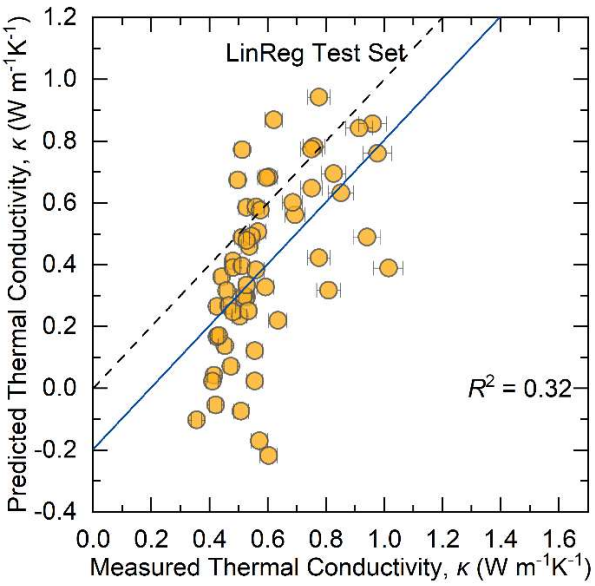


Figure S4. Scatter plot of the linear regression model prediction of the test set.

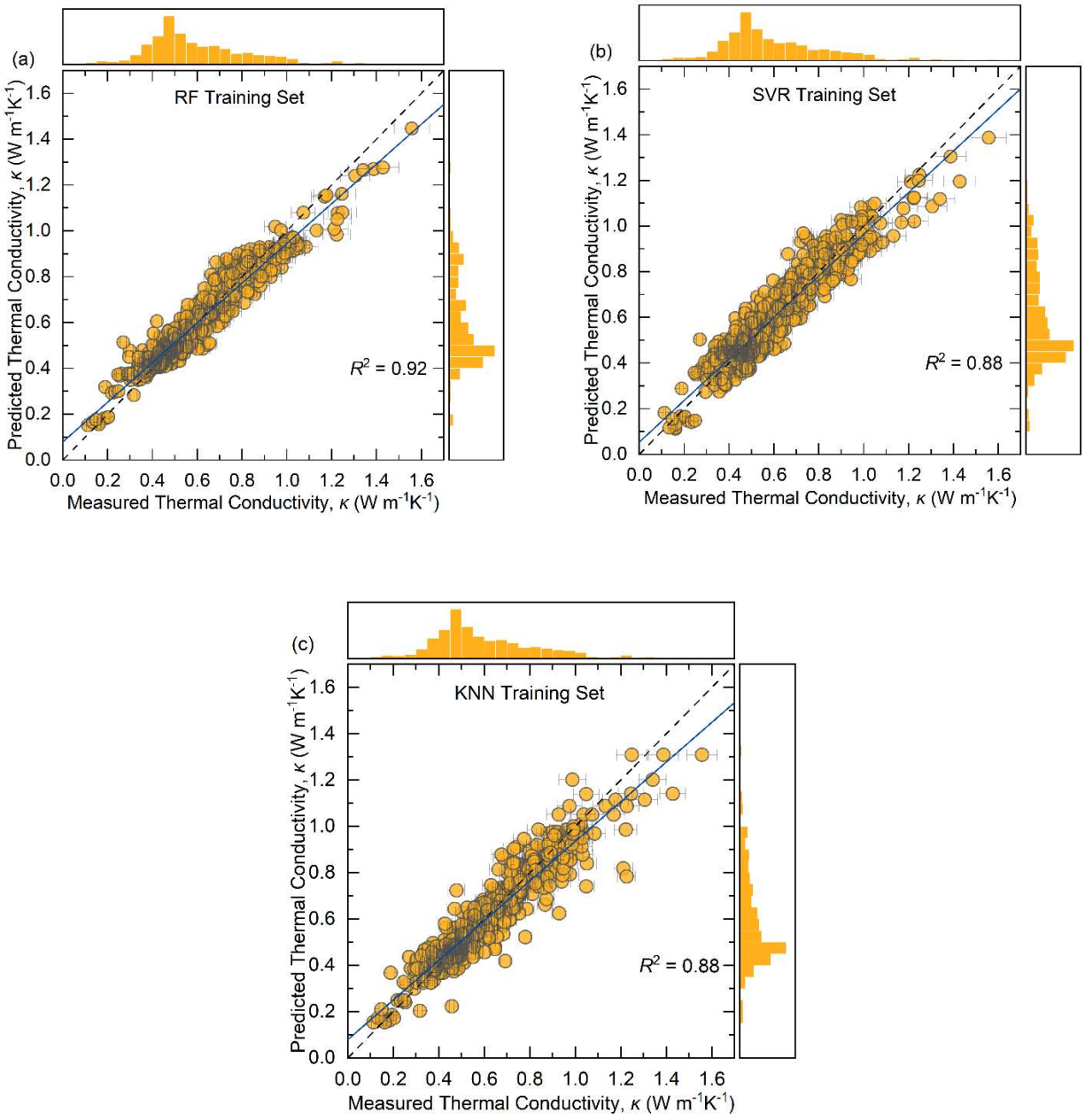


Figure S5. Scatter plots of ML models prediction of the training set.

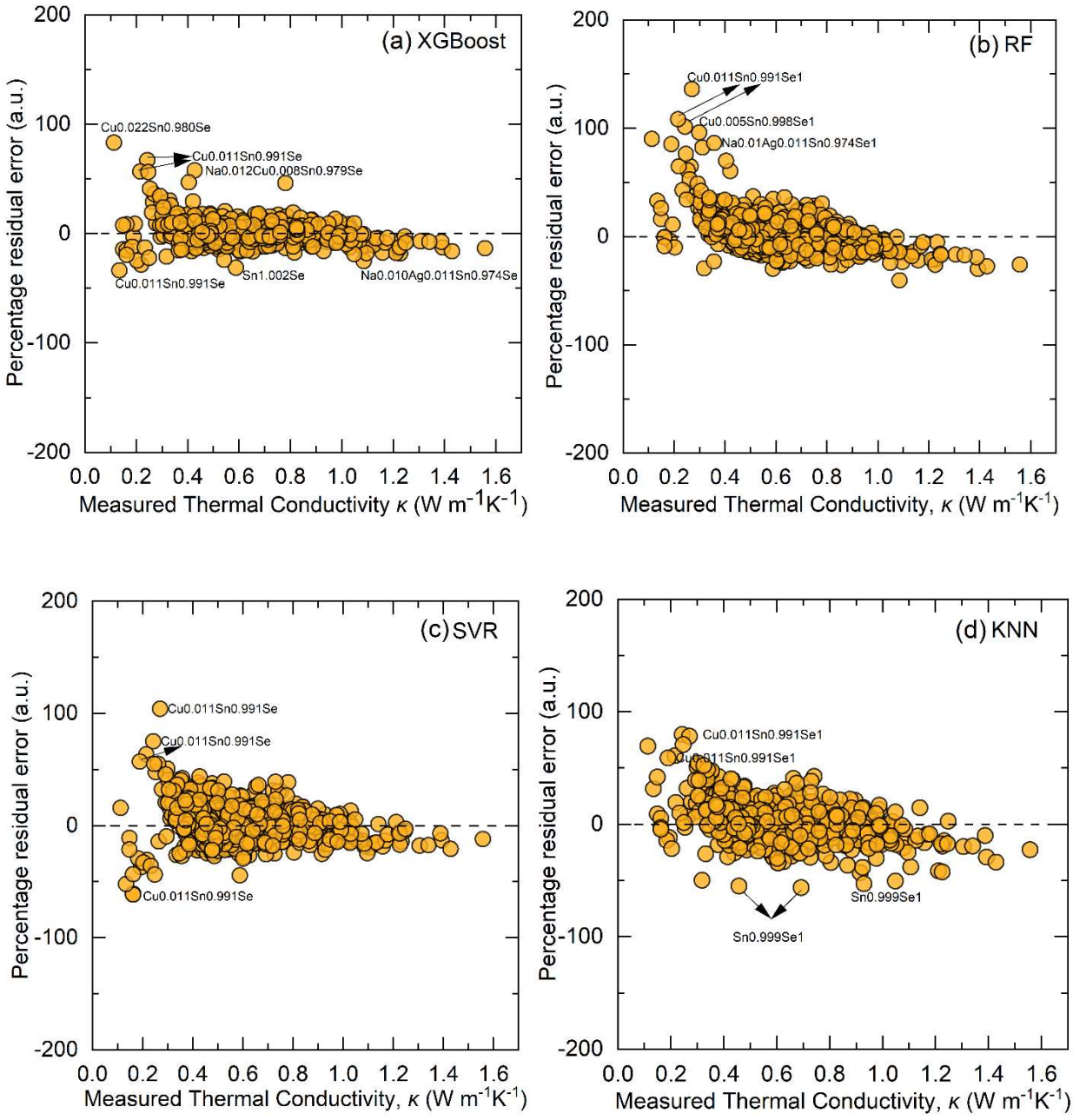


Figure S6. Percent residual error plots of ML models in combined five validation sets.[3]

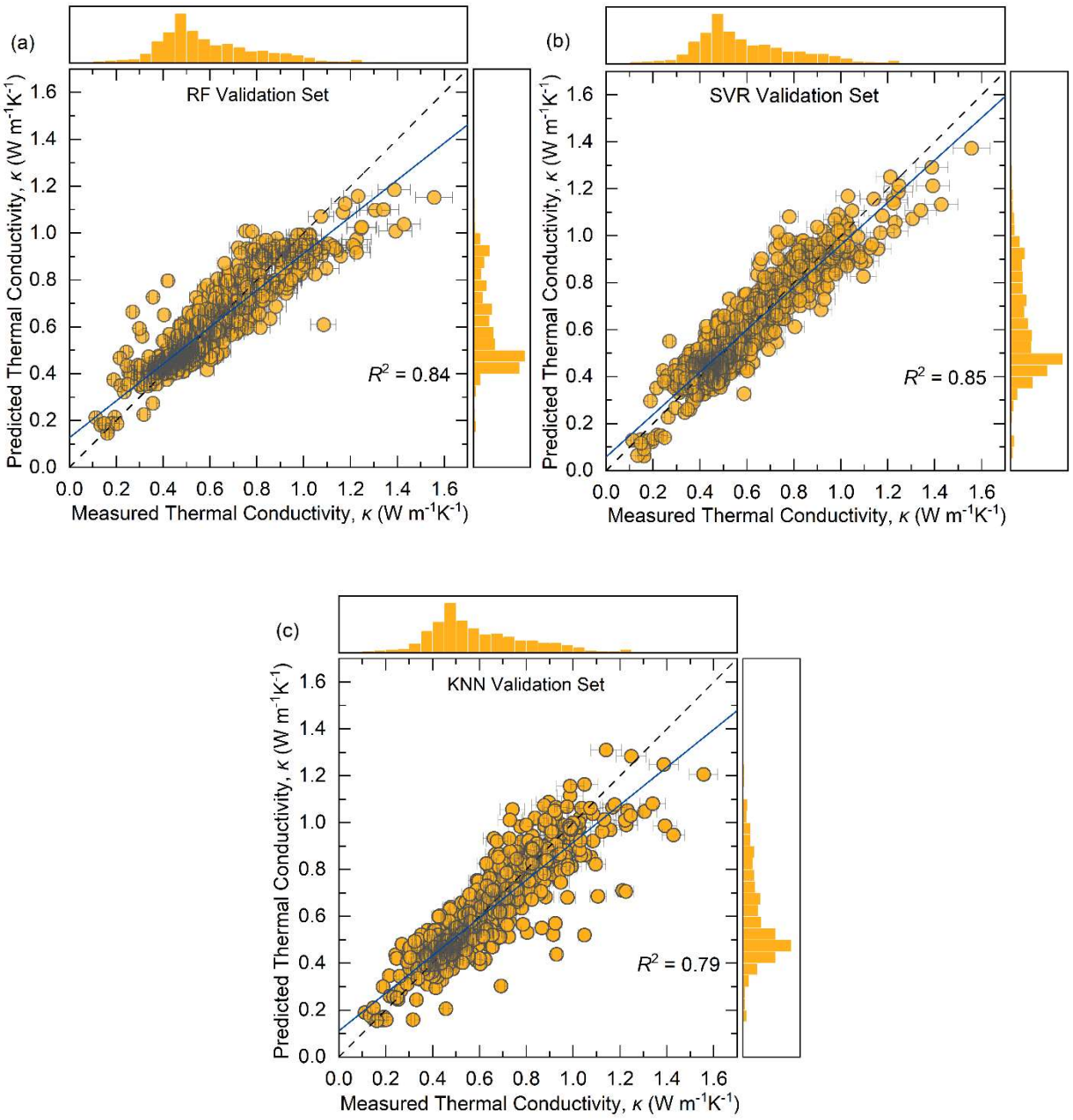


Figure S7. Scatter plots of ML models prediction of the overall validation set.

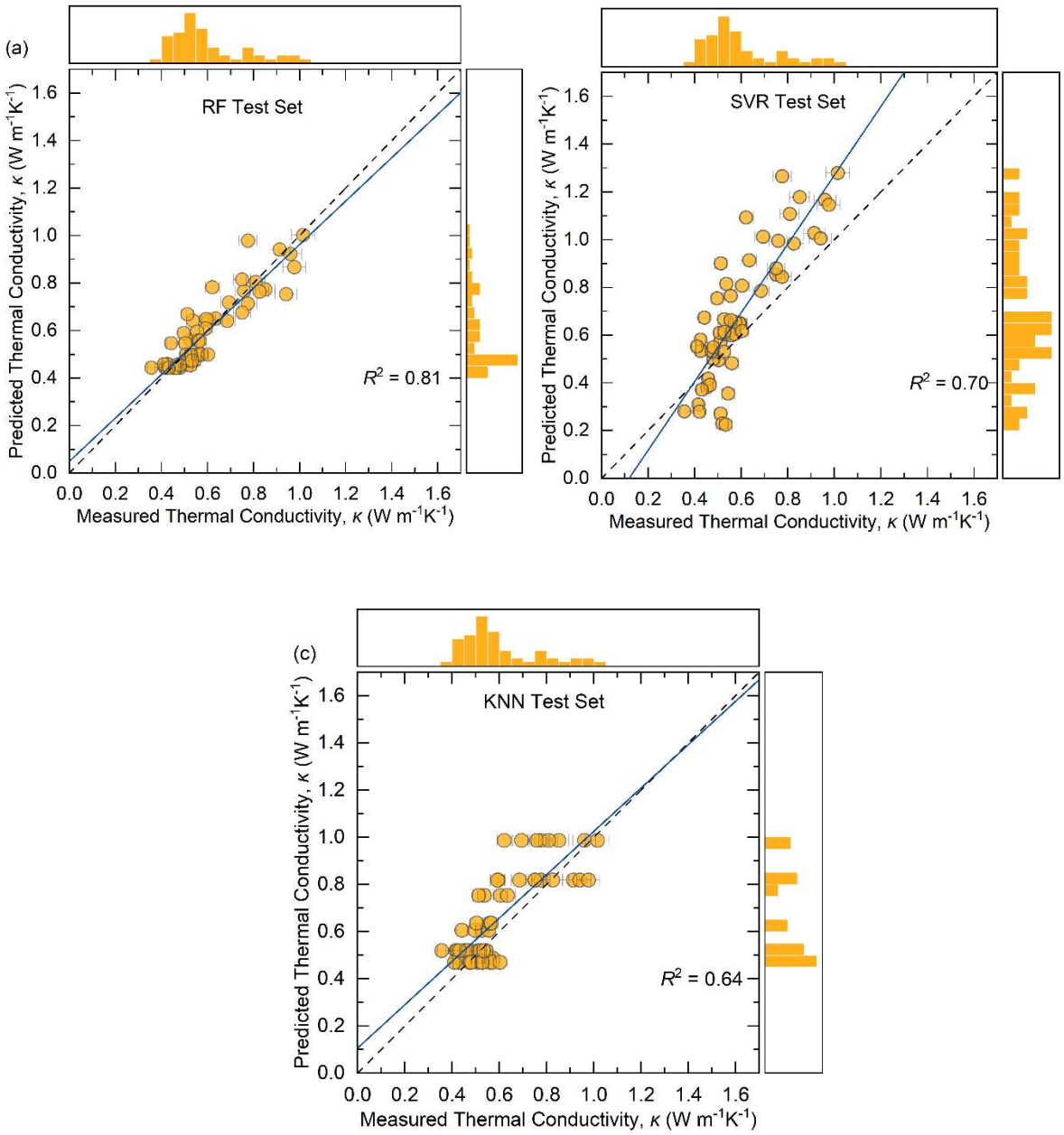


Figure S8. Scatter plots of ML models prediction of the experimental test set.

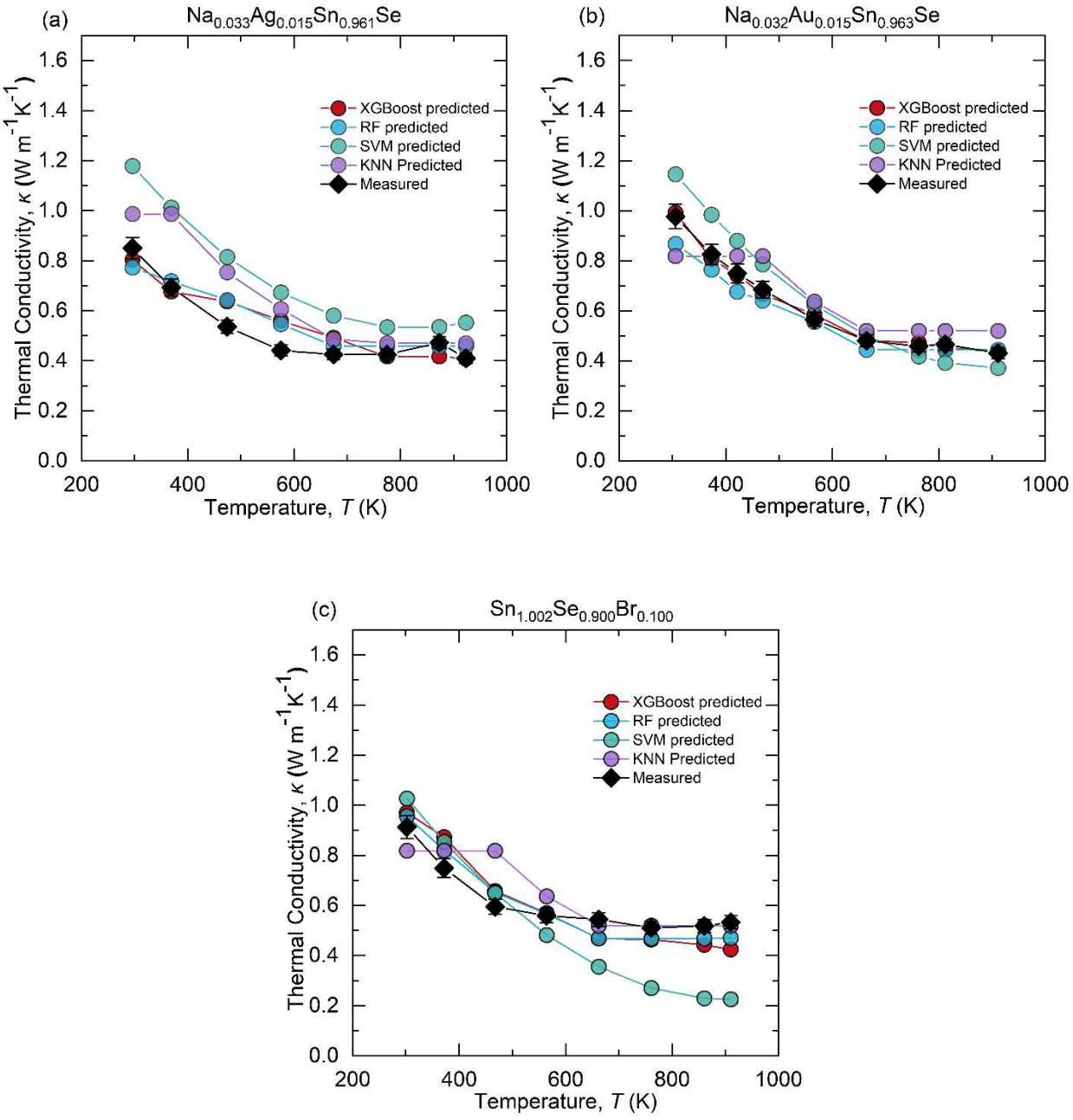


Figure S9. Scatter plots of temperature vs. measured and predicted κ values of (a) $Na_{0.033}Ag_{0.015}Sn_{0.961}Se$, (b) $Na_{0.032}Au_{0.015}Sn_{0.963}Se$, and (c) $Sn_{1.002}Se_{0.900}Br_{0.100}$ in the experimental test set.

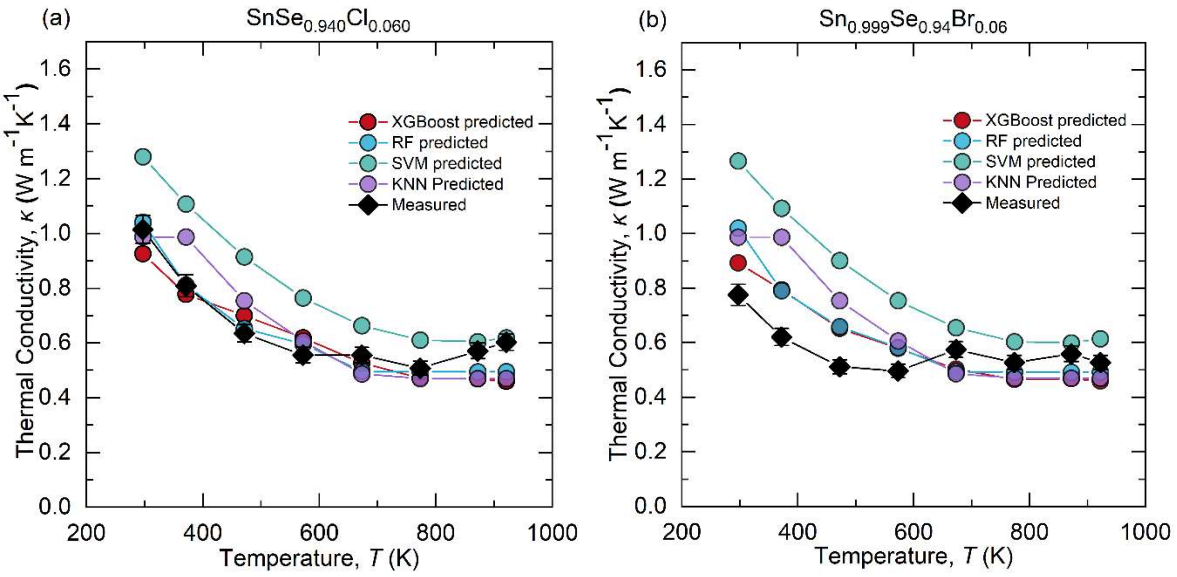


Figure S10. Scatter plots of temperature vs. measured and predicted κ values of (a) Cl- and (b) Br-doped SnSe in the experimental test set.

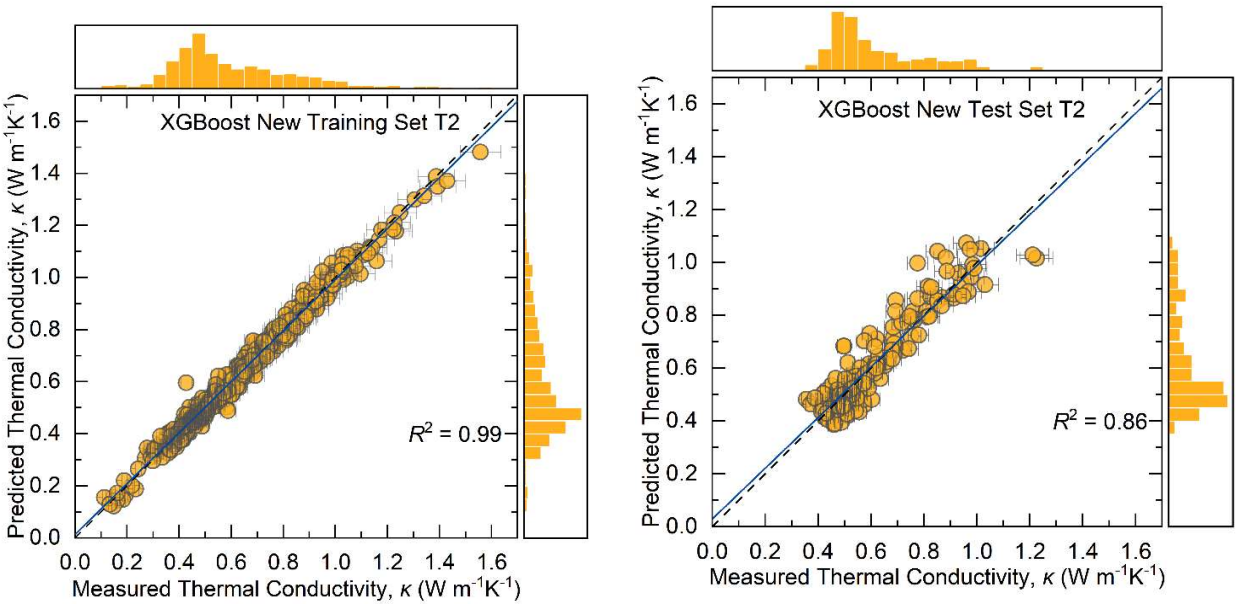


Figure S11. Scatter plots of XGBoost for prediction of the revised training set and the revised test set (T2).

Table S1. List of compositions used in the training set.

Composition	Type of Dopant
Sn0.997Se1	p-type
Cu0.021Sn0.983Se1	p-type
Sn1.006Se1	p-type
Sn1.002Se1	p-type
Cu0.005Sn0.998Se1	p-type
Cu0.01Sn0.989Se1	p-type
Cu0.019Sn0.979Se1	p-type
Cu0.041Sn0.958Se1	p-type
Cu0.011Sn0.991Se1	p-type
Cu0.022Sn0.98Se1	p-type
Cu0.062Sn0.941Se1	p-type
Cu0.078Sn0.918Se1	p-type
Cu0.08Sn0.919Se1	p-type
Sn0.999Se1	p-type
Cu0.009Sn0.999Se1	p-type
Na0.012Cu0.008Sn0.994Ge0.025Se0.95S0.05	p-type
Na0.01Cu0.011Sn0.978Se1	p-type
Li0.01Cu0.009Sn0.974Se1	p-type
Na0.01Ag0.01Sn0.998Se1	p-type
Na0.01Ag0.011Sn0.974Se1	p-type
Na0.01Sn0.976Au0.01Se1	p-type
Na0.025Cu0.005Sn0.969Se1	p-type
Na0.03Sn0.969Se1	p-type
Na0.09Cu0.012Ag0.01Sn0.971Se1	p-type
Na0.021Cu0.005Sn0.974Se1	p-type
Na0.03Sn0.97Se1	p-type
Na0.01Ag0.01Sn0.981Se1	p-type
Na0.01Cu0.01Sn0.979Se1	p-type
Na0.01Cu0.01Ag0.01Sn0.98Se1	p-type
Na0.01Cu0.01Sn0.98Se1	p-type
Na0.029Cu0.002Sn0.969Se1	p-type
Na0.028Cu0.01Sn0.963Se1	p-type
Na0.033Cu0.016Sn0.961Se1	p-type
Na0.025Cu0.006Sn0.968Se1	p-type
Na0.026Cu0.01Sn0.961Se1	p-type
Na0.021Cu0.003Sn0.978Se1	p-type
Na0.019Cu0.009Sn0.972Se1	p-type
Na0.02Cu0.015Sn0.965Se1	p-type

Na0.014Cu0.006Sn0.98Se1	p-type
Na0.014Cu0.012Sn0.973Se1	p-type
Na0.011Cu0.003Sn0.99Se1	p-type
Na0.011Cu0.008Sn0.978Se1	p-type
Na0.007Cu0.015Sn0.977Se1	p-type
Sn1.001Se1	p-type
Na0.01Ag0.01Sn0.98Se1	p-type
Na0.01Cu0.002Ag0.006Sn0.97Ge0.001Se1	p-type
Cu0.015Sn0.985Se1	p-type
Cu0.007Sn0.992Se1	p-type
Na0.015Sn0.985Se1	p-type
Na0.013Cu0.002Sn0.986Se1	p-type
Na0.03Cu0.008Sn0.962Se1	p-type
Sn0.996Bi0.006Se0.974Br0.026	n-type
Sn0.997Bi0.005Se0.97Br0.03	n-type
Sn0.988Bi0.01Se0.98Br0.02	n-type
Sn0.984Bi0.021Se0.99Br0.01	n-type
Sn0.982Bi0.026Se0.994Br0.006	n-type
Sn0.965Bi0.03Se0.994Br0.006	n-type
Sn0.987Bi0.007Se0.984Br0.016	n-type
Sn0.991Bi0.004Se0.992Br0.008	n-type
Sn0.983Bi0.006Se0.977Br0.023	n-type
Sn0.985Bi0.005Se0.98Br0.02	n-type
Sn0.99Bi0.005Se0.989Br0.011	n-type
Sn0.995Bi0.004Se0.992Br0.008	n-type
Sn0.976Bi0.011Se0.977Br0.023	n-type
Sn0.995Se0.959Br0.041	n-type
Sn0.964Bi0.041Se1	n-type
Sn0.999Se0.94Br0.06	n-type
Sn0.94Bi0.06Se1	n-type
Sn0.972Bi0.029Se0.97Br0.03	n-type
Sn0.982Bi0.02Se0.955Br0.045	n-type
Sn0.952Bi0.045Se0.98Br0.02	n-type
Sn0.953Bi0.046Se0.955Br0.045	n-type
Sn0.94Bi0.061Se0.939Br0.061	n-type
Na0.012Cu0.008Sn0.979Se1	p-type
Sn0.997Se1	p-type

75 samples, 10 temperatures -> 750 data

Table S2. Occurrences of elements in the heterogeneous compositions in training dataset.

Composition / Elements	Binary	Ternary	Quaternary	Quinary	Senary	Total
Sn	55	205	471	18	27	776
Se	55	205	471	18	27	776
Li	—	—	9	—	—	9
Na	—	48	288	18	27	381
Cu	—	118	238	18	27	401
Ag	—	—	49	18	8	75
Au	—	—	10	—	—	10
Ge	—	—	—	—	27	27
Bi	—	19	174	—	—	193
S	—	—	—	—	19	19
Br	—	20	174	—	—	194
Cl	—	—	—	—	—	—

Table S3. Features used in the training dataset prior to the implementation of feature selection methods.

Sl. No	Feature	Feature explanation
1	composition	chemical formula
2	*p*-type/*n*-type	type of semiconductor
3	at. % ratio (cation/anion)	ratio of all the cations with the anions in a composition
4	added dopants sum	sum of the dopant elements % atomic. For example, for $Na_{0.01}Cu_{0.02}Sn_{0.97}Se$: 0.01 + 0.02 = 0.03; $Sn_{0.9}Bi_{0.1}Se_{0.95}Br_{0.05}$: 0.1 + 0.05 = 0.15. Alloying elements like Ge on the Sn site are not included.
5	site occ. (Sn)	occupancy of the cations in the Sn site subtracting the selenium in the composition
6	molar mass	molar mass of the composition
7	sum alloying	summation of the alloying elements. For example: $Na_{0.012}Cu_{0.008}Sn_{0.994}Ge_{0.025}Se_{0.950}S_{0.050}$: 0.025 + 0.05 = 0.075. Doping elements are not included.
8	Cp (3R)	calculated heat capacity based on Dulong-Petit
9	relative mass (theor.)	molar mass of the composition divided by the molar mass of pure SnSe
10	manual estimator of performance	formula (Sn type × # of reductions + cooling type)/(sum of added dopants + % theoretical density). Sn type values are 1 for powder, 2 for chunks, 3 for purified; cooling type is 3 for furnace, 2 for air, 1 for water, and 0 for NA (not applicable).
11	no. of elements	number of the elements in the material
12	Na	atomic amount of a given element in the formula. For example: $Na_{0.030}Sn_{0.969}Se_1$ implies 0.03 for the Na column 0.03, 0.969 for Sn, and 1 for Se. The rest of the elements are set to 0.
13	Li	
14	Cu	
15	Ag	
16	Sn	
17	Ge	
18	Au	
19	Bi	
20	Se	
21	S	
22	Br	
23	Cl	
24	Sn purified	physical form of Sn. 1 means Sn was purified.

		0 means Sn was not purified. Purified means chunks of Sn have been subjected to the purification process[4] (repeated melting and cleaning).
25	Sn chunks	chunks mean Sn chunks that have not been purified.
26	Sn powder	powder is powder form of Sn.
27	melting temperature (K)	temperature of the furnace for melting Sn
28	melting time (min)	time taken for melting Sn
29	no quench	quenching of the reaction tube not performed
30	furnace quench	quenched reaction tube in the furnace
31	air quench	quenched reaction tube by air cooling on a firebrick
32	water quench	quenched by placing the reaction tube in water bath
33	melt- and ball milling-synthesis (1/0)	composition synthesized using melting or by ball milling
34	ball milling speed (rpm)	speed of the ball milling process
35	ball milling time (min)	time taken for the ball milling process
36	reduced	whether SnSe was reduced by reduction process
37	number of reductions	number of reduction cycles undertaken for SnSe to remove oxides
38	reduction prep (1=mortar/pestle/0=ball milled)	whether SnSe was crushed using mortar/pestle or ball milled before the reduction process
39	mixture of reduction gas	mixture of gas (2=5%H_2/Ar, 1=7%H_2/N_2, 0=NA) used for the reduction process
40	H_2/Ar reduction gas	mixture of gas used during the reduction process
41	H_2/N_2 reduction gas	
42	no reduction gas	
43	reduction temp (*K*)	maximum temperature of the furnace used for the reduction process
44	reduction time (min)	the time used for the reduction process
45	annealing	If the sample was annealed
46	annealing (1=before hot press/0=after hot press)	whether the annealing process was completed before hot pressing in the powdered form—1 or after hot pressing in the pellet form—0
47	annealing time (min)	the time used for annealing at the maximum temperature
48	annealing temp (K)	temperature at which the sample was annealed
49	ball milling or grinding before hot press	whether the sample was ball milled or ground by hand using the mortar and pestle

50	sieved before hot press	whether the sample was sieved to < 63 μm before hot pressing
51	hot press temp (K)	maximum temperature of the hot-pressing process.
52	hot press ramp time (min)	time for the sample to reach the maximum temperature in the hot press
53	hot press hold time (min)	time for the sample at the maximum temperature in the hot press
54	hot press pressure (MPa)	pressure used for sintering the sample in the hot press
55	density (g/cm^3)	density of the sample measured in the pellet form
56	lowest total thermal cond (W/m/K)	minimal thermal conductivity of the sample
57	measurement temp (K)	temperature points when measuring thermal diffusivity
58	EN (Pauling) - Na	the value is calculated by multiplying the atomic amount of an element in the formula with the Pauling electronegativity of the element.
59	EN (Pauling) - Li	
60	EN (Pauling) - Cu	
61	EN (Pauling) - Ag	
62	EN (Pauling) - Sn	
63	EN (Pauling) - Ge	
64	EN (Pauling) - Au	
65	EN (Pauling) - Bi	
66	EN (Pauling) - Se	
67	EN (Pauling) - S	
68	EN (Pauling) - Br	
69	EN (Pauling) - Cl	
70	EN (Pauling) - varn	the variation is calculated by applying variance to all the elements with weighted Pauling electronegativity calculation in the composition (Table S3 (b))
71	EN (Pauling) - sum	The sum is calculated by adding to all the elements with weighted Pauling electronegativity calculation in the composition (Table S3 (b)).
72	EN (Pauling) - max	The maximum is calculated by taking the maximum weighted value from all the elements with weighted Pauling electronegativity in the composition (Table S3 (b)).
73	EN (Pauling) - avg	The average is calculated by taking the average weighted value from all the elements with weighted Pauling electronegativity present in the composition (Table S3 (b)).
74	EN (Pauling) - min	The minimum is calculated by taking the minimum weighted value from all the elements with weighted Pauling electronegativity in the composition (Table S3 (b)).

75	EN (Pauling) - diff	The difference is calculated by subtracting EN (Pauling) – max with EN (Pauling) - min (Table S3 (b)).
76	EN (Ghosh) - Na	The value is calculated by multiplying the atomic amount of an element in composition with the Ghosh electronegativity of the element.
77	EN (Ghosh) - Li	
78	EN (Ghosh) - Cu	
79	EN (Ghosh) - Ag	
80	EN (Ghosh) - Sn	
81	EN (Ghosh) - Ge	
82	EN (Ghosh) - Au	
83	EN (Ghosh) - Bi	
84	EN (Ghosh) - Se	
85	EN (Ghosh) - S	
86	EN (Ghosh) - Br	
87	EN (Ghosh) - Cl	
88	EN (Ghosh) - varn	
89	EN (Ghosh) - sum	
90	EN (Ghosh) - max	
91	EN (Ghosh) - avg	
92	EN (Ghosh) - min	
93	EN (Ghosh) - diff	
94	IE - Na	The value is calculated by multiplying the atomic amount of an element in composition with the Ionization energy of the element.
95	IE - Li	
96	IE - Cu	
97	IE - Ag	
98	IE - Sn	
99	IE - Ge	
100	IE - Au	
101	IE - Bi	
102	IE - Se	
103	IE - S	
104	IE - Br	
105	IE - Cl	
106	IE - varn	
107	IE - sum	
108	IE - max	
109	IE - avg	
110	IE - min	
111	IE - diff	
112	at. weight - Na	

113	at. weight - Li	The value is calculated by multiplying the atomic amount of an element in composition with the atomic weight of the element.
114	at. weight - Cu	
115	at. weight - Ag	
116	at. weight - Sn	
117	at. weight - Ge	
118	at. weight - Au	
119	at. weight - Bi	
120	at. weight - Se	
121	at. weight - S	
122	at. weight - Br	
123	at. weight - Cl	
124	at. weight - varn	
125	at. weight - sum	
126	at. weight - max	
127	at. weight - avg	
128	at. weight - min	
129	at. weight - diff	
130	vec - Na	The value is calculated by multiplying the atomic amount of an element in composition with the number of valence electron of the element.
131	vec - Li	
132	vec - Cu	
133	vec - Ag	
134	vec - Sn	
135	vec - Ge	
136	vec - Au	
137	vec - Bi	
138	vec - Se	
139	vec - S	
140	vec - Br	
141	vec - Cl	
142	vec - varn	
143	vec - sum	
144	vec - max	
145	vec - avg	
146	vec - min	
147	vec - diff	
148	r(cov) - Na	The value is calculated by multiplying the atomic amount of an element in composition with the covalent radius of the element.
149	r(cov) - Li	
150	r(cov) - Cu	
151	r(cov) - Ag	

152	r(cov) - Sn	
153	r(cov) - Ge	
154	r(cov) - Au	
155	r(cov) - Bi	
156	r(cov) - Se	
157	r(cov) - S	
158	r(cov) - Br	
159	r(cov) - Cl	
160	r(cov) - varn	
161	r(cov) - sum	
162	r(cov) - max	
163	r(cov) - avg	
164	r(cov) - min	
165	r(cov) - diff	

Table S4. Weighted features mathematical calculations.[5]

Example	Weighted elemental stoichiometry features (WESF)	Formula applied	Weighted value
$Sn_{0.997}Se$	EN (Pauling) – Sn	Electronegativity of Sn × 0.9973[a]	1.954719
$Na_{0.021}Cu_{0.005}Sn_{0.974}Se$	r(cov) – Sn	Covalent radius of Sn × 0.9741[a]	1.373453
$Na_{0.010}Ag_{0.010}Sn_{0.980}Se$	vec - Ag	Valence electrons of Ag × 0.0100[a]	0.109357
$Sn_{0.996}Bi_{0.006}Se_{0.974}Br_{0.026}$	IE – Br	Ionization energy of Br × 0.026	0.306568

[a]Calculated stoichiometry after weighing.

Example	NumPy Operation	Formula applied*	Weighted value
$Sn_{0.972}Bi_{0.029}Se_{0.97}Br_{0.03}$	Variance	Var([WESF(Sn),…..,WESF(Br)])	0.670371
$Na_{0.013}Cu_{0.002}Sn_{0.986}Se$	Sum	Sum([WESF (Na),.., WESF(Se)])	4.496973
$Cu_{0.015}Sn_{0.985}Se$	Maximum	Max([WESF(Cu),..,WESF(Se)])	2.55
$Na_{0.026}Cu_{0.01}Sn_{0.961}Se$	Minimum	Min([WESF(Na),…,WESF(Se)])	0.019547
$Na_{0.012}Cu_{0.008}Sn_{0.994}Ge_{0.025}Se_{0.95}S_{0.05}$	Average	Ave([WESF(Na),….,WESF(S)])	0.762899
$Na_{0.03}Sn_{0.969}Se$	Difference	Diff(Max – Min)	2.52202

*With respect to Pauling electronegativity scale.

[] = array

Table S5. Hyperparameters used in feature selection methods.

ML Algorithm	Hyperparameter	Value used
Lasso	alpha	0.001
	max_iter	1000
	random_state	none
	selection	cyclic
XGBoost (XGBRegressor)	base_score	0.5
	booster	gbtree
	callbacks	none
	colsample_bylevel	1
	colsample_bynode	1
	colsample_bytree	1
	early_stopping_rounds	none
	enable_categorical	false
	eval_metric	none
	gamma	0
	gpu_id	1
	grow_policy	depthwise
	importance_type	none
	interaction_constraints	none
	learning_rate	0.3
	max_bin	256
	max_cat_to_onehot	4
	max_delta_step	0
	max_depth	3
	max_leaves	0
	min_child_weight	1
	missing	nan (not a number)
	monotone_constraints	none
	n_estimators	100
	n_jobs	0
	num_parallel_tree	1
	predictor	auto
	random_state	0
	reg_alpha	0
	reg_lambda	3

Table S6. Hyperparameters used in different ML models during model selection.

ML Algorithm	Hyperparameter	Value used
XGBoost (XGBRegressor)	base_score	0.5
	booster	gbtree
	callbacks	none
	colsample_bylevel	1
	colsample_bynode	1
	colsample_bytree	1
	early_stopping_rounds	none
	enable_categorical	false
	eval_metric	none
	gamma	0
	gpu_id	1
	grow_policy	depthwise
	importance_type	none
	interaction_constraints	none
	learning_rate	0.3
	max_bin	256
	max_cat_to_onehot	4
	max_delta_step	0
	max_depth	3
	max_leaves	0
	min_child_weight	1
	missing	nan (not a number)
	monotone_constraints	none
	n_estimators	100
	n_jobs	0
	num_parallel_tree	1
	predictor	auto
	random_state	0
	reg_alpha	0
	reg_lambda	3
Random Forest (RandomForestRegressor)	bootstrap	true
	n_estimators	100
	max_depth	6
	random_state	0
	verbose	0
	n_jobs	none
	max_features	1.0

	min_weight_fraction_leaf	0.0
	min_samples_split	2
	min_samples_leaf	1
k-NN (neighbors)	n_neighbors	4
	algorithm	auto
	weights	uniform
	leaf_size	30
	p	2
	metric	minkowski
	n_jobs	none
SVM (SVR)	kernel	rbf
	C	80
	epsilon	0.1
	gamma	scale
	degree	3

Table S7. Metrics of ML models of validation sets from k-fold.[6]

Metrics / **Model**	**Validation Set**					
	***R*²**					
	k-fold set 1	*k*-fold set 2	*k*-fold set 3	*k*-fold set 4	*k*-fold set 5	*k*-fold average
XGboost	0.93	0.95	0.95	0.92	0.95	0.94
RF	0.81	0.83	0.87	0.82	0.83	0.83
SVR	0.85	0.85	0.87	0.84	0.84	0.85
k-NN	0.79	0.81	0.81	0.8	0.72	0.79
	RMSE (W $m^{-1}K^{-1}$)					
XGboost	0.06	0.05	0.05	0.06	0.04	0.05
RF	0.10	0.10	0.08	0.09	0.08	0.09
SVR	0.09	0.09	0.08	0.08	0.08	0.07
k-NN	0.10	0.10	0.10	0.10	0.10	0.10
	MAE (W $m^{-1}K^{-1}$)					
XGboost	0.04	0.04	0.04	0.04	0.03	0.04
RF	0.07	0.06	0.06	0.06	0.06	0.06
SVR	0.07	0.07	0.06	0.06	0.06	0.06
k-NN	0.07	0.07	0.07	0.07	0.07	0.07

Table S8. Metrics of ML models in training, validation, and test set.[6]

Metrics / **Model**	**Training**			**Validation**			**Test of new data**		
	R^2	*RMSE* (W $m^{-1}K^{-1}$)	*MAE* (W $m^{-1}K^{-1}$)	R^2	*RMSE* (W $m^{-1}K^{-1}$)	*MAE* (W $m^{-1}K^{-1}$)	R^2	*RMSE* (W $m^{-1}K^{-1}$)	*MAE* (W $m^{-1}K^{-1}$)
XGboost	0.99	0.03	0.02	0.94	0.05	0.04	0.84	0.07	0.05
RF	0.92	0.06	0.05	0.84	0.09	0.06	0.81	0.07	0.05
SVR	0.88	0.07	0.06	0.85	0.09	0.07	0.70	0.19	0.15
k-NN	0.88	0.07	0.05	0.79	0.10	0.07	0.64	0.13	0.10

References:

1 J. Wei, X. Chu, X. Y. Sun, K. Xu, H. X. Deng, J. Chen, Z. Wei and M. Lei, *InfoMat*, 2019, **1**, 338–358.

2 T. T. Wong, *Pattern Recognit.*, 2015, **48**, 2839–2846.

3 A. Y. T. Wang, R. J. Murdock, S. K. Kauwe, A. O. Oliynyk, A. Gurlo, J. Brgoch, K. A. Persson and T. D. Sparks, *Chem. Mater.*, 2020, **32**, 4954–4965.

4 C. Zhou, Y. K. Lee, Y. Yu, S. Byun, Z. Z. Luo, H. Lee, B. Ge, Y. L. Lee, X. Chen, J. Y. Lee, O. Cojocaru-Mirédin, H. Chang, J. Im, S. P. Cho, M. Wuttig, V. P. Dravid, M. G. Kanatzidis and I. Chung, *Nat. Mater.*, 2021, **20**, 1378–1384.

5 C. R. Harris, K. J. Millman, S. J. van der Walt, R. Gommers, P. Virtanen, D. Cournapeau, E. Wieser, J. Taylor, S. Berg, N. J. Smith, R. Kern, M. Picus, S. Hoyer, M. H. van Kerkwijk, M. Brett, A. Haldane, J. F. del Río, M. Wiebe, P. Peterson, P. Gérard-Marchant, K. Sheppard, T. Reddy, W. Weckesser, H. Abbasi, C. Gohlke and T. E. Oliphant, *Nature*, 2020, **585**, 357–362.

6 D. Chicco, M. J. Warrens and G. Jurman, *PeerJ Comput. Sci.*, 2021, **7**, e623/1-24.