Electronic Supplementary Material (ESI) for Journal of Materials Chemistry A.

Supplementary Information for

Machine learning discovery of medium-entropy thermoelectric materials with ultralow lattice thermal conductivity

Hangyu Zhang[a], Xianhua Nie*[a], Xinyi Zhang[a], Ruiqi Wang*[b], Shuai Deng[a], Li Zhao[a]
[a] State Key Laboratory of Engines (Tianjin University), Tianjin University, Tianjin 300350, China
[b] State key laboratory of marine thermal energy and power, Wuhan Second Ship Design and Research Institute, Wuhan 430205, China
* Corresponding author. Email: xhnie@tju.edu.cn (Xianhua Nie)
* Corresponding author. Email: qiqiD22@163.com (Ruiqi Wang)

1、Machine-Learning Methods and Model Evaluation

Table S1. The details of the database as well as the references. (See Table S1.pdf)

Table S2. Distribution of the dataset across entropy classes.

Entropy class	Threshold (ΔS_{config})	Unique materials	records
Low	< 1.0R	451	2721
Medium	1.0–1.5R	431	2549
High	≥ 1.5R	9	48

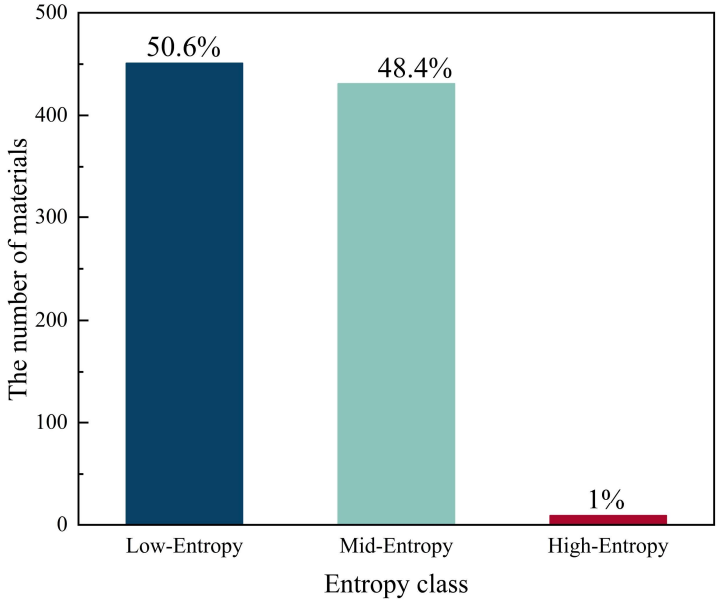

Fig. S1. Dataset counts by configurational-entropy class.

Table S3. Spearman correlation coefficients (ρ) between zT and all 135 elemental features, ranked by $|\rho|$. (See Table S3.pdf)

Table S4. Top 20 features ranked by $|\rho|$ with zT.

Rank (by |ρ|)	feature	Spearman ρ (ZT)	|ρ|
1	temperature(K)	0.4361	0.4361
2	MagpieData mode CovalentRadius	0.3047	0.3047
3	MagpieData avg_dev Electronegativity	-0.2970	0.2970
4	MagpieData mean CovalentRadius	0.2788	0.2788
5	MagpieData mean Number	0.2712	0.2712
6	MagpieData mean Row	0.2711	0.2711
7	MagpieData mean AtomicWeight	0.2595	0.2595
8	MagpieData mode Row	0.2573	0.2573
9	MagpieData maximum NdValence	0.2543	0.2543
10	MagpieData avg_dev CovalentRadius	-0.2453	0.2453
11	MagpieData mode Number	0.2420	0.2420
12	MagpieData mode AtomicWeight	0.2417	0.2417
13	MagpieData avg_dev NpValence	-0.2284	0.2284
14	MagpieData mode MeltingT	0.2282	0.2282
15	MagpieData avg_dev MeltingT	-0.2282	0.2282
16	MagpieData maximum NValence	0.2265	0.2265
17	MagpieData minimum CovalentRadius	0.2260	0.2260
18	MagpieData mode SpaceGroupNumber	0.2192	0.2192
19	MagpieData mode Electronegativity	-0.2125	0.2125
20	MagpieData avg_dev MendeleevNumber	-0.2106	0.2106

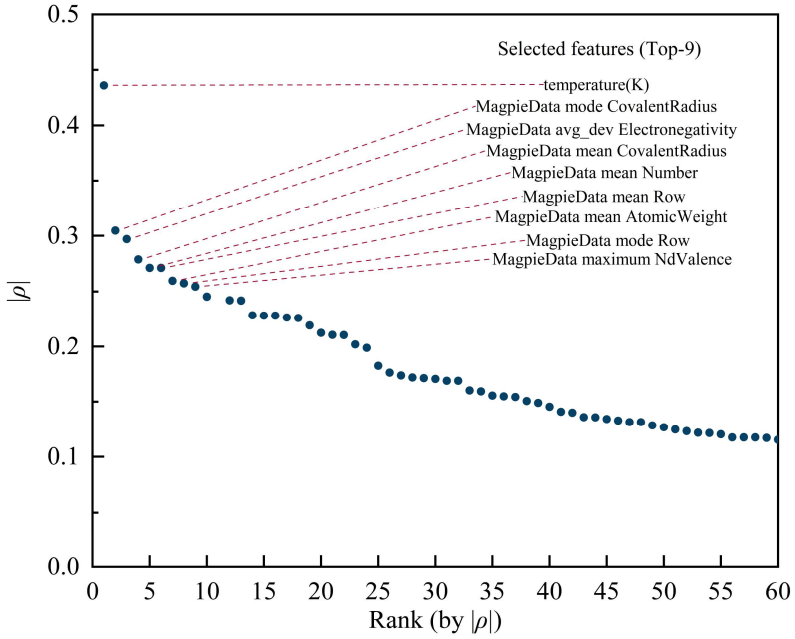

Fig. S2. Rank plot of the absolute Spearman correlation coefficient ($|\rho|$) between zT and the elemental features. Features are sorted in descending order of $|\rho|$, and the nine selected features (Top-9) are annotated.

Table S5. 10-fold cross-validated ablation and baseline comparison for the highly correlated features under the optimized XGBoost hyperparameters.

Variant	Feature setting	R^2 (mean ± sd)	MSE (mean ± sd)
Baseline_All3	Keep Row_mean + Num_mean + AW_mean	0.897 ± 0.033	0.01273 ± 0.00376
Drop_All3	Drop Row_mean + Num_mean + AW_mean	0.891 ± 0.035	0.01313 ± 0.00397
Keep_AW_mean	Keep AW_mean only	0.901 ± 0.032	0.01152 ± 0.00363
Keep_Row_mean	Keep Row_mean only	0.900 ± 0.034	0.01192 ± 0.00385
Keep_Num_mean	Keep Num_mean only	0.902 ± 0.033	0.01149 ± 0.00380

Table S6. Identification of poor thermoelectric candidates using out-of-fold predictions from 10-fold cross-validation.

Selected group (lowest predicted ZT)	Precision (fraction truly poor)	Recall (fraction of all poor captured)	Enrichment vs random
Bottom 10%	0.946	0.463	4.63×
Bottom 20%	0.837	0.819	4.10×
Bottom 30%	0.627	0.921	3.07×

**Poor candidates are defined as the lowest 20% by measured ZT. Materials are ranked by out-of-fold predicted ZT from 10-fold cross-validation, and the bottom predicted fractions are evaluated. Precision, recall, and enrichment are reported as defined in the text.*

Table S7. The optimized hyperparameters for the ML models.

Algorithm	Hyperparameters	Definition	Value
XGBoost	N_estimators	The total number of boosting trees	589
	Learning_rate	Controls the step size of each boosting iteration	0.05
	Max_depth	The maximum depth of each decision tree	5
Random Forest	N-estimators	The number of decision trees in the forest.	600
	Max-depth	The maximum depth of each decision tree	33
	Max-features	The number of features to consider when looking for the best split	Log_2
Decision Tree Regressor	Max_depth	The maximum depth of the decision tree	50
	criterion	The function to measure the quality of a split	Squared_Error
	Min_Samples_Leaf	The minimum number of samples required to be at a leaf node	1
	Min_Samples_Split	The minimum number of samples required to be at a leaf node	2
K-Nearest Neighbors	N_neighbors	The number of nearest neighbors considered when making predictions	13
	weights	The weight function used in prediction	distance
Multi-Layer Perceptron	activation	The activation function applied to each neuron	tanh
	alpha	The L2 regularization parameter	0.00001
	layer	The total number of layers in the neural network	5
	Learning_rate	The step size that controls how much the model weights are updated during each optimization iteration	0.0005
Support Vector Regression	kernel	The kernel function type	Rbf

Table S8. Extrapolation evaluation using chemical-system GroupKFold.

Split protocol	folds	R^2 (mean ± std)	MAE (mean ± std)	RMSE (mean ± std)
Chemical-system GroupKFold	5	0.44 ± 0.12	0.179 ± 0.021	0.255 ± 0.026

***** *We used a 5-fold GroupKFold split with grouping by chemical system, ensuring that each chemical system appears in only one-fold. Metrics are reported as mean ± standard deviation across the five folds.*

Table S9. The details of the predicted materials.

Formula	Temperature (K)	zT
$CuAg_3S_2$	300	0.3233152
	400	0.4329965
	500	0.5430070
	600	0.7670138
	700	0.9513484
	800	1.2365824
Ag_2SnSe_3	300	0.1487018
	400	0.4128470
	500	0.6959045
	600	0.9000177
	700	0.9065907
	800	1.2874126
$GeTe(PbSe)_2$	300	0.2393377
	400	0.4876221
	500	0.5174509
	600	0.7476122
	700	0.9307650
	800	1.3783721
$In_2Bi_4Pb_4Se_{13}$	300	0.2023237
	400	0.3813706
	500	0.4952551
	600	0.7592990
	700	0.9409911
	800	1.3703023

2、DFT Calculations and Validation Details

To justify the plane-wave cutoff energy (ENCUT) adopted in this work, we performed systematic convergence tests for both electronic properties and phonon force calculations using fixed geometries and otherwise identical computational settings. Specifically, total energies and the PBE band gap were evaluated for the primitive-cell model, and the energy difference per atom was referenced to a higher-cutoff calculation (ENCUT = 620 eV). For the phonon and force-constant workflow based on the finite-displacement method, we further benchmarked the atom-resolved forces for two representative displaced supercell configurations (disp-0019 and disp-0328). The component-wise maximum force deviations relative to ENCUT = 620 eV were quantified as :

$$\Delta F_{\alpha,max}(\mathrm{ENCUT}) = i_{max}\left| F_{i,\alpha}(\mathrm{ENCUT}) - F_{i,\alpha}(620\mathrm{eV})\right|, \alpha = x, y, z \qquad (1)$$

where i runs over all atoms in the displaced supercell and $F_{i,\alpha}$ denotes the force component. As shown in Fig. S3, $\Delta F_{\alpha,max}$ decreases with increasing ENCUT and becomes nearly constant above 520 eV for both displacement patterns.

Table S10. Convergence test of plane-wave cutoff energy (ENCUT).

ENCUT (eV)	TOTEN (eV)	ΔE vs 620 eV (meV/atom)	Eg (eV)
420	-365.70350814	0.090	0.8673
470	-365.70792949	0.042	0.8673
520	-365.71017931	0.017	0.8673
570	-365.71111084	0.007	0.8673
620	-365.71177931	0.000	0.8673

*Total energy (TOTEN), energy difference per atom relative to ENCUT = 620 eV, and the PBE band gap (Eg) were evaluated at fixed geometry (N = 92 atoms).

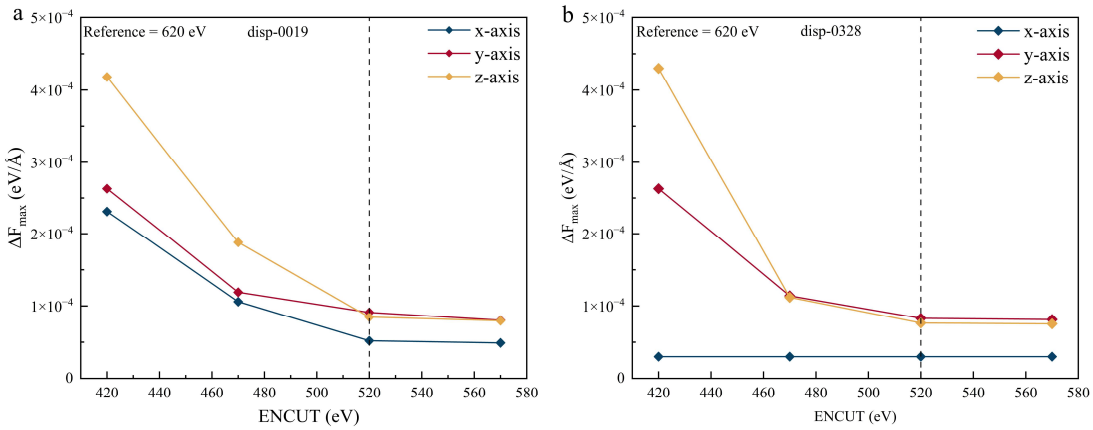

Fig. S3. ENCUT convergence of atom-resolved force deviations for two displaced supercells.

Carrier relaxation times were estimated using deformation-potential theory and used to scale the BoltzTraP transport coefficients. For each crystallographic direction i (i =a, b, c), the effective uniaxial elastic constant and the deformation potential constant were obtained from the total-energy–strain and band-edge–strain relations:

$$C_i = \frac{1}{V_0}\frac{\partial^2 E}{\partial \varepsilon_i^2} \tag{2}$$

$$E_{1,i} = \frac{\partial E_{\text{edge}}}{\partial \varepsilon_i} \tag{3}$$

where E_{edge} refers to the CBM or VBM. The relaxation time is then evaluated as:

$$\tau_i = \frac{2\sqrt{2\pi}\hbar^4 C_i}{3(k_B T)^{3/2} E_{1,i}^2 (m_i^*)^{3/2}} \quad (4)$$

The direction-resolved parameters and the derived τ_i are summarized in Table S11.

Table S11. Direction-resolved deformation-potential parameters and the derived carrier relaxation times.

x-axis	Carriers	m*/m$_0$	E$_1$ (eV)	C$_{3D}$ (GPa)	μ (cm^2·V^{-1}·s^{-1})	τ (fs)
a-axis	e	1.51	7.86	25.1	8.86	7.61
	h	1.72	7.09		7.86	7.69
b-axis	e	1.37	7.44	24.2	12.14	9.46
	h	1.87	6.43		7.47	7.94
c-axis	e	1.13	8.36	23.5	15.11	9.71
	h	1.42	7.93		9.48	7.66

To obtain a scalar relaxation time for CRTA-based transport scaling in BoltzTraP, we further defined an effective value as the arithmetic mean over the three directions:

$$\tau_{avg} = \frac{\tau_a + \tau_b + \tau_c}{3} \quad (5)$$

The resulting effective relaxation times are $\tau_{avg,e}$=8.93 fs for electrons and $\tau_{avg,h}$=7.85 fs for holes.

Table S12. Total energies, per-atom energies, and relative energies of the cation-exchange configurations.

Config	Total energy (eV)	E/atom (eV atom^{-1})	ΔE (eV/cell)	ΔE (meV atom^{-1})
a (this work)	-365.710	-3.975	0.000	0.0
b	-365.428	-3.972	0.282	3.1
c	-365.084	-3.968	0.626	6.8
d	-364.774	-3.965	0.936	10.2
e	-364.334	-3.960	1.377	15.0

***** $\Delta E(meV\cdot atom^{-1}) = 1000(E_i - E_1)/n_{atoms}$, *All configurations were structurally relaxed prior to the final single-point energy evaluation; identical computational settings were used for all cases. The simulation cell contains 92 atoms.*

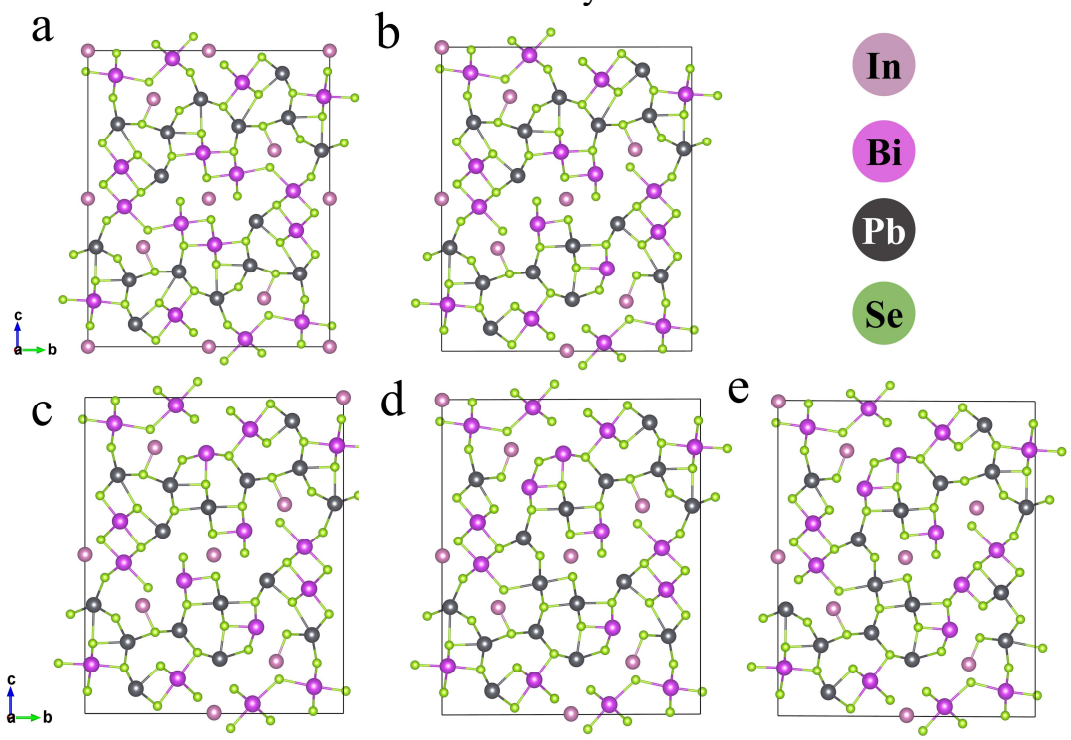

Fig. S4. The cation-exchange configurations.

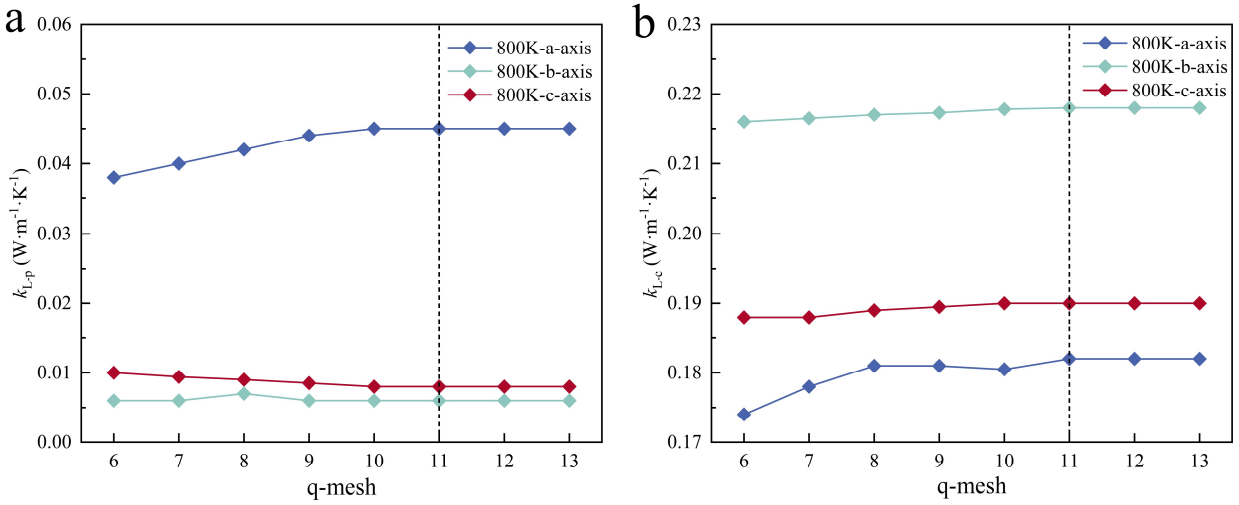

Fig. S5. Convergence test of the q mesh of the thermal conductivity calculations. (a) $k_{\text{L-p}}$, (b) $k_{\text{L-c}}$.

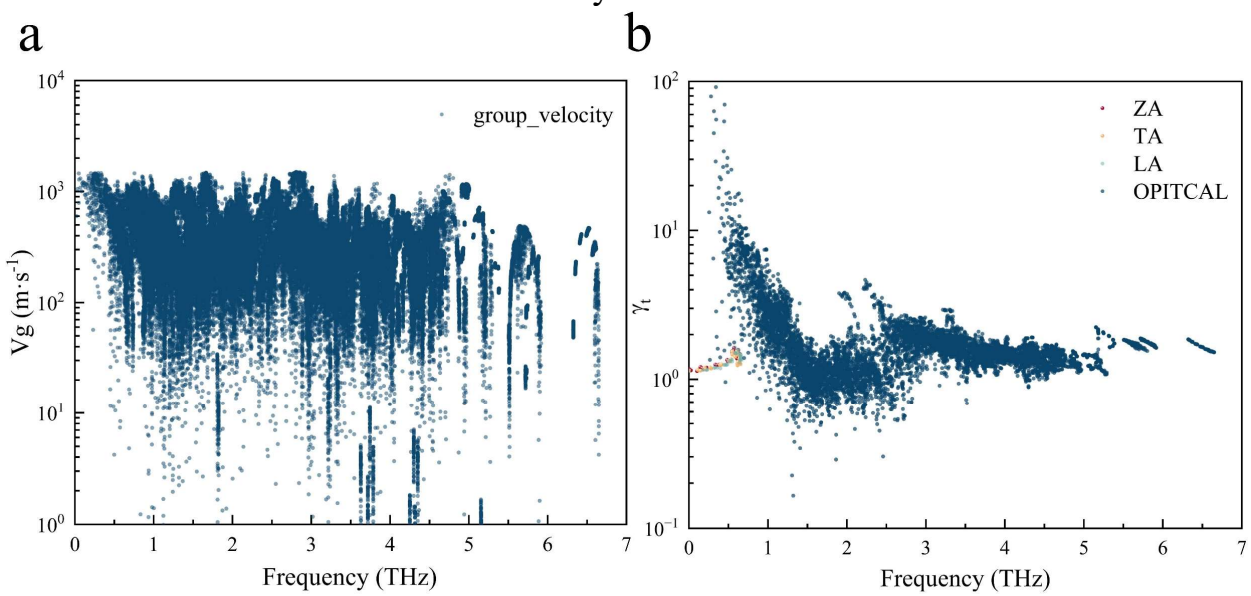

Fig. S6. (a) Phonon group velocity, (b) Grüneisen parameters of $In_2Bi_4Pb_4Se_{13}$ at 800K