Supplementary Information for

Predicting Thermodynamic Stability of Inorganic Compounds Using Ensemble Machine Learning Based on Electron Configuration

Zou *et al.*

Supplementary Notes

Supplementary Note 1

The definitions of decomposition enthalpy and energy above the hull

In this study, the decomposition energy (ΔH_d) is defined as the total energy difference between a given compound and competing compounds in a specific chemical space. It represents the magnitude of (in)stability with respect to phase separation[1].

ΔH_d differs from the commonly used “energy above the hull”. To obtain the ΔH_d of a compound, the compound needed to be excluded when constructing the energy hull of the system in which it resides. ΔH_d refers to the distance from the energy of the compound to the hull. As demonstrated in Supplementary Fig. 1, ΔH_d of A_4B is calculated as the distance from the convex hull to A_4B. If a compound is above the energy hull, the value of ΔH_d is larger than zero, corresponding to the commonly reported "energy above the hull". However, if the compound, such as AB_3, is on the energy convex hull, the ΔH_d of AB_3 represents the distance between AB_3 and a hypothetical convex hull constructed without AB_3 (indicated by the dashed line). In this case, the value of ΔH_d is negative, which quantifies the stability of the material and provides useful information for the uncertainty of stability assessment and the rationality of synthesis[1, 2].

Supplementary Note 2

Discussion about the complementarity of different base-level models.

To ensure complementarity we have selected domain knowledge from different scales: interatomic interactions, atomic properties, and EC. Models based on these kinds of domain knowledge each have their own advantages. More specifically, the input of Magpie is the statistics of various element properties, Roost considers the interaction between atoms through the attention mechanism, and ECCNN goes a step further to the level of the electron configuration of atoms. We validated this through error correlation analysis and entropy distribution of each model. The error correlation matrix is a common tool to measure the correlation between the prediction errors of multiple models on the same data set. The smaller the correlation, the stronger the complementarity. The entropy distribution is also used to analyze the difference in uncertainty of the models in prediction. If one model has high uncertainty on some samples and the other model does not, it indicates that they are complementary, as shown in Supplementary Fig. 3.

From Supplementary Fig. 3(a), the three models do not have a strong correlation in error correlation with Pearson coefficient ranging from 0.37 to 0.49. From the sample entropy distribution of the models in Supplementary Fig. 3(b), there are obvious differences in the entropy distribution of each model. Roost is more distributed in the low entropy area, Magpie is more distributed in the high entropy area, and ECCNN is more evenly distributed. In summary, it can be concluded that the three models are complementary.

Supplementary Note 3

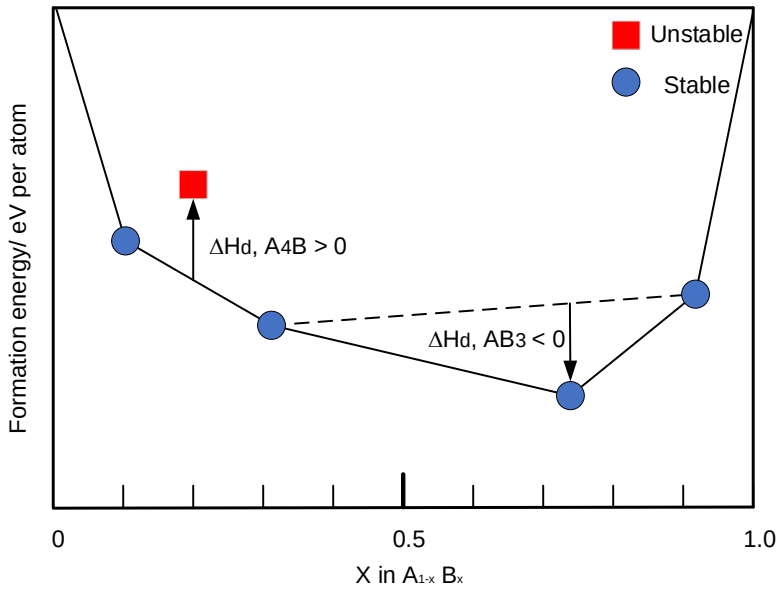
Discussion about integrating extra models and heterogeneous data.

The proposed combination method, Stacked Generalization (SG), is highly flexible because it allows the inclusion of various types of base models. This flexibility enables the seamless integration of new models designed to handle different data types, such as numerical data, spectral data, and image data.

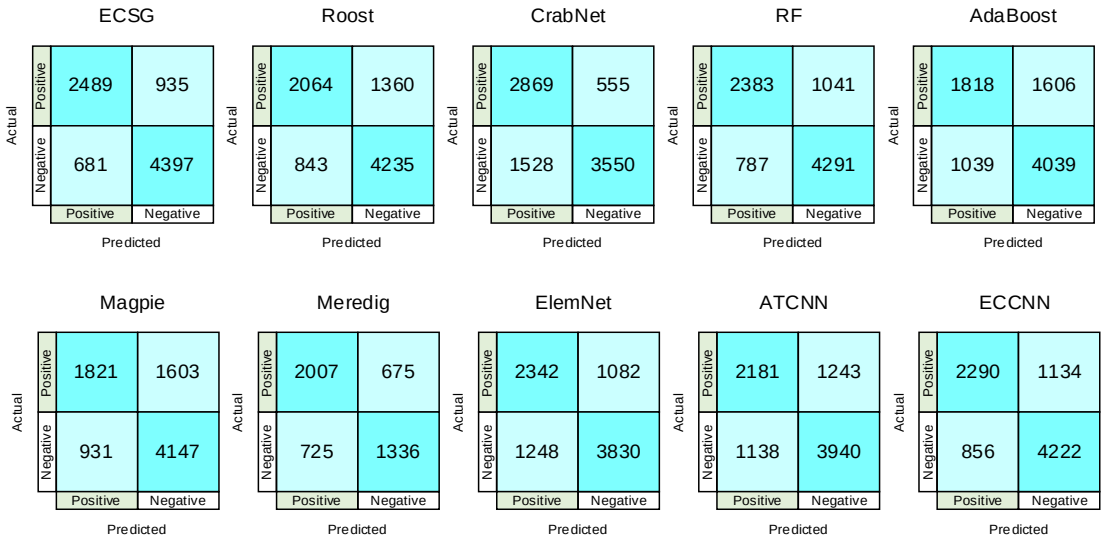
When new base models are introduced, the meta-level model dynamically adjusts the weights for all base models during training. For example, the original ECSG model combines predictions from three base models (ECCNN, Roost, and Magpie), with the meta-level model calculating the weighted prediction: $\hat{y} = \omega_1\hat{y}_1 + \omega_2\hat{y}_2 + \omega_3\hat{y}_3 + \varepsilon$, where $\hat{y}_1$, $\hat{y}_2$, $\hat{y}_3$ are predictions from the initial models, ω_1, ω_2, ω_3 are their corresponding weights, and ε is the interception. To incorporate additional data types, such as spectral data, we can introduce a new base model, K_dos_fea[3], which extracts features from spectra using one-dimensional convolution. This model generates a new prediction $\hat{y}_4$. Upon adding the spectral model, the meta-level model updates the weighted combination as follows: $\hat{y} = \omega_1^{'}\hat{y}_1 + \omega_2^{'}\hat{y}_2 + \omega_3^{'}\hat{y}_3 + \omega_4^{'}\hat{y}_4 + \varepsilon'$. Similarly, new base models using CNNs can be introduced to handle image data, with the meta-model dynamically adjusting the weights for all sub-models. This modular approach ensures that ECSG can efficiently integrate diverse data sources, assigning appropriate weights to each type for accurate predictions.

Each base-level model processes only its respective data type, and the training of these models remains independent. Since the inputs are not fused or converted, data integrity is maintained throughout the process, avoiding potential information loss.

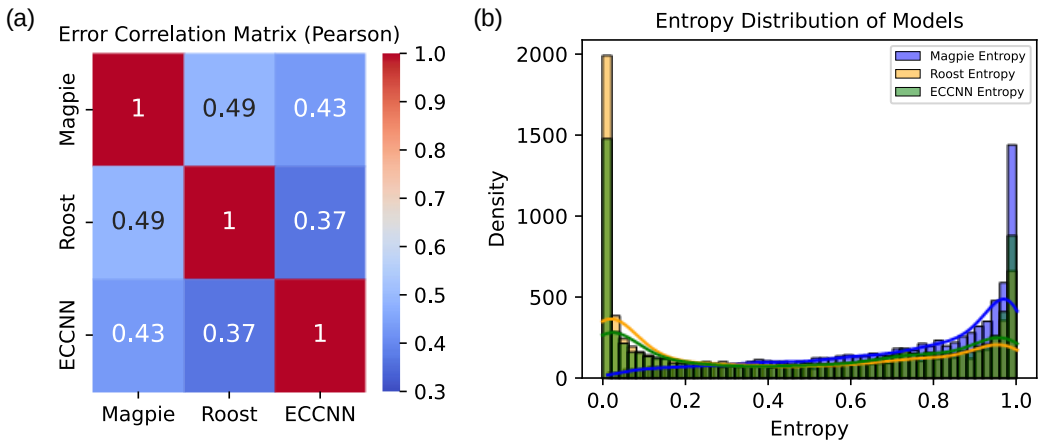
Supplementary Figures



Supplementary Fig. 1 Illustration of the definition of ΔHd by constructing a convex hull.



Supplementary Fig. 2 Confusion matrices of ECSG and other composition-based models trained on MP.



Supplementary Fig. 3 Complementarity of different base-level models. **(a)** Error correlation matrix of three base-level models. **(b)** Entropy Distribution of three base-level models.

Supplementary Tables

Supplementary Table 1 The performance comparison of our proposed ECSG and other models trained on MP. The best results are shown in bold, while the second results are underlined.

No.	Model	ACC	Precision	Recall	F1	NPV	AUC	AUPR
1	ECSG	0.807	0.778	0.728	0.752	0.824	0.886	0.834
2	Roost[4]	0.741	0.712	0.603	0.652	0.758	0.820	0.740
3	CrabNet[5]	0.755	0.646	0.838	0.729	0.869	0.852	0.781
4	RF	0.785	0.752	0.696	0.723	0.805	0.862	0.803
5	Adaboost	0.689	0.637	0.531	0.579	0.716	0.747	0.649
6	Magpie[6]	0.702	0.662	0.532	0.590	0.721	0.766	0.670
7	Meredig[7]	0.695	0.652	0.522	0.579	0.716	0.758	0.649
8	ElemNet[8]	0.726	0.653	0.684	0.667	0.780	0.719	0.732
9	ATCNN[9]	0.72	0.658	0.637	0.647	0.760	0.707	0.720
10	ECCNN	0.766	0.727	0.669	0.697	0.788	0.842	0.770

Supplementary Table 2 The performance comparison of classifiers with ΔH_d threshold = 25 meV/atom on MP.

No.	Model	ACC	Precision	Recall	F1	NPV	AUC	AUPR
1	ECSG	0.837	0.843	0.889	0.865	0.827	0.908	0.921
2	Roost	0.711	0.595	0.875	0.708	0.878	0.813	0.852
3	CrabNet	0.816	0.789	0.935	0.856	0.878	0.892	0.903
4	RF	0.819	0.825	0.877	0.850	0.808	0.891	0.912
5	Adaboost	0.714	0.727	0.821	0.771	0.687	0.772	0.814
6	Magpie	0.795	0.802	0.865	0.832	0.782	0.866	0.891
7	Meredig	0.787	0.791	0.868	0.828	0.780	0.857	0.880
8	ElemNet	0.782	0.799	0.841	0.820	0.754	0.842	0.865
9	ATCNN	0.785	0.803	0.843	0.822	0.757	0.850	0.875
10	ECCNN	0.763	0.785	0.821	0.803	0.726	0.830	0.862

Supplementary Table 3 The performance comparison of classifiers with ΔH_d threshold = 40 meV/atom on MP.

No.	Model	ACC	Precision	Recall	F1	NPV	AUC	AUPR
1	ECSG	0.848	0.857	0.919	0.887	0.827	0.911	0.953
2	Roost	0.668	0.552	0.914	0.688	0.898	0.794	0.837
3	CrabNet	0.830	0.819	0.948	0.879	0.864	0.897	0.929
4	RF	0.830	0.841	0.909	0.874	0.803	0.894	0.933
5	Adaboost	0.731	0.750	0.877	0.809	0.671	0.772	0.851
6	Magpie	0.809	0.820	0.905	0.860	0.782	0.869	0.893
7	Meredig	0.800	0.808	0.907	0.855	0.778	0.858	0.876
8	ElemNet	0.801	0.823	0.882	0.852	0.749	0.846	0.863
9	ATCNN	0.798	0.822	0.880	0.850	0.745	0.851	0.874
10	ECCNN	0.768	0.801	0.855	0.827	0.694	0.823	0.852

Supplementary Table 4 The performance comparison of regressors for ΔH_d on MP.

No.	Model	MAE/eV per atom
1	ECSG	0.064±0.001
2	Roost	0.067±0.006
3	CrabNet	0.070±0.004
4	RF	0.078±0.001
5	Adaboost	0.506±0.008
6	Magpie	0.086±0.001
7	Meredig	0.086±0.002
8	ElemNet	0.073±0.009
9	ATCNN	0.071±0.005
10	ECCNN	0.086±0.001

Note: The best results are shown in bold, while the second results are underlined.

Supplementary Table 5 The performance comparison of our proposed ECSG and other models trained on OQMD.

No.	Model	ACC	Precision	Recall	F1	NPV	AUC	AUPRC
1	ECSG	0.940	0.806	0.601	0.688	0.952	0.967	0.804
2	Roost	0.928	0.722	0.614	0.640	0.954	0.958	0.756
3	CrabNet	0.914	0.643	0.499	0.558	0.940	0.929	0.637
4	RF	0.927	0.766	0.485	0.594	0.939	0.948	0.721
5	Adaboost	0.890	0.490	0.059	0.106	0.895	0.808	0.313
6	Magpie	0.912	0.708	0.331	0.451	0.923	0.923	0.604
7	Meredig	0.905	0.721	0.223	0.341	0.912	0.904	0.552
8	ElemNet	0.907	0.618	0.424	0.499	0.932	0.923	0.584
9	ATCNN	0.904	0.616	0.349	0.441	0.924	0.908	0.543
10	ECCNN	0.902	0.604	0.328	0.423	0.922	0.895	0.520

Note: The best results are shown in bold, while the second results are underlined.

Supplementary Table 6 The performance comparison of our proposed ECSG and other models trained on JARVIS.

No.	Model	ACC	Precision	Recall	F1	NPV	AUC	AUPRC
1	ECSG	0.970	0.825	0.813	0.819	0.983	0.988	0.914
2	Roost	0.957	0.730	0.797	0.757	0.981	0.984	0.853
3	CrabNet	0.954	0.708	0.768	0.735	0.979	0.981	0.824
4	RF	0.954	0.767	0.645	0.701	0.968	0.973	0.767
5	Adaboost	0.940	0.680	0.543	0.604	0.959	0.967	0.704
6	Magpie	0.939	0.675	0.527	0.592	0.957	0.966	0.700
7	Meredig	0.950	0.746	0.606	0.669	0.964	0.979	0.798
8	ElemNet	0.944	0.636	0.806	0.705	0.982	0.972	0.629
9	ATCNN	0.945	0.673	0.681	0.675	0.971	0.973	0.748
10	ECCNN	0.944	0.719	0.553	0.624	0.960	0.969	0.721

Supplementary Table 7 The performance of integrating the three base models using different combination methods.

Methods	ACC	Precision	Recall	F1	NPV	AUC	AUPR
SG	0.807	0.778	0.728	0.752	0.824	0.886	0.834
Averaging	0.788	0.772	0.673	0.719	0.797	0.865	0.804
Voting	0.763	0.726	0.663	0.693	0.785	0.747	0.762

Supplementary Table 8 Performance of ECSG after integrating CGCNN on the MP-structure database. ECSG+C represents the model after integrating CGCNN into ECSG, and ACC_M represents the accuracy of correctly distinguishing polymorphs.

	Accuracy	Precision	Recall	F1	NPV	AUC	AUPR	ACC_M
ECSG	0.826	0.719	0.557	0.628	0.853	0.879	0.721	0
CGCNN	0.835	0.738	0.578	0.648	0.860	0.899	0.746	0.193
ECSG+C	0.844	0.753	0.607	0.672	0.869	0.905	0.769	0.121

Supplementary Table 9 The performance comparison in predicting perovskite halides in unknown space.

No.	Model	ACC	F1	AUC	AUPRC
1	ECSG	0.790	0.373	0.758	0.474
2	Roost	0.654	0.408	0.730	0.375
3	CrabNet	0.488	0.307	0.513	0.234
4	RF	0.750	0.285	0.702	0.390
5	AdaBoost	0.571	0.352	0.593	0.337
6	Magpie	0.788	0.326	0.708	0.429
7	Meredig	0.752	0.248	0.733	0.371
8	ElemNet	0.718	0.337	0.701	0.353
9	ATCNN	0.675	0.404	0.691	0.345
10	ECCNN	0.337	0.327	0.451	0.207

Supplementary Table 10 The performance comparison in predicting Li-containing oxides in unknown space.

No.	Model	ACC	F1	AUC	AUPRC
1	ECSG	0.891	0.495	0.848	0.510
2	Roost	0.813	0.408	0.802	0.417
3	CrabNet	0.778	0.403	0.779	0.407
4	RF	0.886	0.457	0.792	0.450
5	AdaBoost	0.877	0.239	0.754	0.340
6	Magpie	0.880	0.449	0.799	0.438
7	Meredig	0.880	0.302	0.789	0.385
8	ElemNet	0.802	0.219	0.736	0.385
9	ATCNN	0.863	0.315	0.741	0.346
10	ECCNN	0.880	0.063	0.749	0.314

Supplementary Table 11 The performance comparison in predicting transition metal oxides in unknown space.

No.	Model	ACC	F1	AUC	AUPRC
1	ECSG	0.843	0.202	0.819	0.507
2	Roost	0.841	0.219	0.799	0.464
3	CrabNet	0.836	0.294	0.792	0.440
4	RF	0.833	0.150	0.779	0.427
5	AdaBoost	0.828	0.058	0.736	0.358
6	Magpie	0.841	0.224	0.792	0.459
7	Meredig	0.843	0.304	0.808	0.470
8	ElemNet	0.830	0.006	0.522	0.474
9	ATCNN	0.830	0.238	0.753	0.390
10	ECCNN	0.831	0.013	0.741	0.358

Supplementary Table 12 The performance of ECSG and comparison models testing in the 2DMatpeida.

No.	Model	ACC	Precision	Recall	F1	NPV	AUC	AUPR
1	ECSG	0.737	0.763	0.775	0.769	0.701	0.790	0.786
2	Roost	0.687	0.730	0.71 4	0.718	0.644	0.752	0.769
3	CrabNet	0.691	0.671	0.892	0.766	0.755	0.752	0.768
4	RF	0.711	0.760	0.714	0.737	0.655	0.778	0.807
5	Adaboost	0.703	0.730	0.755	0.742	0.666	0.759	0.760
6	Magpie	0.706	0.750	0.721	0.735	0.654	0.777	0.790
7	Meredig	0.705	0.735	0.748	0.741	0.665	0.763	0.770
8	ElemNet	0.649	0.716	0.639	0.664	0.604	0.719	0.718
9	ATCNN	0.664	0.724	0.655	0.686	0.604	0.733	0.744
10	ECCNN	0.606	0.678	0.578	0.623	0.540	0.649	0.696

Supplementary Table 13 VASP calculation results for stable perovskite oxides recommended by ECSG. The unit of total energy is eV per unit cell, and the unit of formation energy is eV per atom.

No.	Compounds	Total energy	Formation energy	Stability on MP	Stability on OQMD	Stability on JARVIS
1	Na2WNiO6	-65.853	-2.299	TRUE	TRUE	TRUE
2	Na2MnTbO6	-62.898	-2.498	TRUE	TRUE	TRUE
3	Ba2SmWO6	-79.120	-3.611	TRUE	TRUE	TRUE
4	YbPrMnNiO6	-72.760	-3.000	TRUE	TRUE	TRUE
5	PrGdV2O6	-84.592	-2.607	FALSE	FALSE	FALSE
6	TmInMn2O6	-73.733	-2.668	TRUE	TRUE	TRUE
7	LuPmCo2O6	-73.406	-2.842	TRUE	TRUE	TRUE
8	LaGdNi2O6	-71.186	-1.912	FALSE	FALSE	TRUE
9	YPdWCrO6	-82.612	-2.686	TRUE	TRUE	TRUE
10	YBaNbCoO6	-79.564	-3.247	TRUE	TRUE	TRUE
11	CuNdVCoO6	-71.906	-2.535	TRUE	TRUE	TRUE
12	NaGeFeCoO6	-59.822	-1.682	FALSE	TRUE	TRUE
13	NaMnWAlO6	-77.405	-2.873	TRUE	TRUE	TRUE
14	ZrLiWVO6	-82.252	-2.826	TRUE	TRUE	TRUE
15	HoMgMnMoO6	-75.066	-2.738	FALSE	FALSE	TRUE
16	DyRbYMoO6	-77.132	-3.273	TRUE	TRUE	TRUE
17	KScWSrO6	-72.552	-2.896	FALSE	FALSE	FALSE
18	HoPrMnCoO6	-78.255	-3.112	TRUE	TRUE	TRUE
19	HoMnMgMoO6	-75.956	-2.827	TRUE	TRUE	TRUE
20	HoYbCrTiO6	-82.337	-3.717	TRUE	TRUE	TRUE
21	LiDyYbMoO6	-71.594	-3.119	TRUE	TRUE	TRUE
22	LiYbMoDyO6	-73.111	-3.271	TRUE	TRUE	TRUE
23	YRbMoDyO6	-77.251	-3.285	TRUE	TRUE	TRUE
24	CrGdWNbO6	-86.851	-2.770	FALSE	FALSE	TRUE
25	NaTlFeMnO6	-60.931	-1.813	TRUE	TRUE	TRUE
26	WGdCrNbO6	-82.945	-1.466	FALSE	FALSE	FALSE
27	SrKWScO6	-79.139	-3.555	TRUE	TRUE	TRUE
28	SrYbFeBiO6	-63.481	-2.640	TRUE	TRUE	TRUE
29	LuAgFeMnO6	-70.250	-2.377	TRUE	TRUE	TRUE
30	KScMnTiO6	-77.666	-2.872	FALSE	TRUE	TRUE
31	KGdCuVO6	-67.731	-1.786	FALSE	FALSE	TRUE
32	KGdWMgO6	-75.396	-2.415	FALSE	FALSE	TRUE
33	MoSrWCrO6	-81.562	-2.491	TRUE	FALSE	TRUE
34	NdCsCrZnO6	-64.256	-2.618	TRUE	TRUE	TRUE
35	ScCaMnVO6	-80.052	-3.198	TRUE	TRUE	TRUE

Supplementary Table 14 VASP calculation results for stable perovskite oxides recommended by Tala.[10]. The unit of total energy is eV per unit cell, and the unit of formation energy is eV per atom.

No.	Compounds	Total energy	Formation energy	Stability on MP	Stability on OQMD	Stability on JARVIS
1	TeRbReSrO6	-62.108	-1.959	FALSE	FALSE	TRUE
2	NdTaReSrO6	-77.074	-2.206	FALSE	FALSE	FALSE
3	FeAuReSrO6	-63.617	-1.574	FALSE	FALSE	FALSE
4	PbIrReSrO6	-65.133	-1.419	FALSE	FALSE	FALSE
5	NbLaReSrO6	-74.556	-2.113	FALSE	FALSE	FALSE
6	NbLaBaReO6	-72.698	-1.904	FALSE	FALSE	FALSE
7	NdNbReCdO6	-73.250	-2.077	FALSE	FALSE	FALSE
8	PrTaMgReO6	-78.822	-2.388	FALSE	FALSE	FALSE
9	TcAuReSrO6	-64.868	-1.284	FALSE	FALSE	FALSE
10	PrNbMgReO6	-78.459	-2.528	FALSE	FALSE	FALSE
11	MoSmReSrO6	-73.192	-2.244	FALSE	FALSE	FALSE
12	NaTeReSrO6	-61.902	-1.905	FALSE	FALSE	FALSE
13	WNdReSrO6	-76.940	-2.526	FALSE	FALSE	FALSE
14	WPdReSrO6	-67.101	-1.501	FALSE	FALSE	FALSE
15	HfNdMgReO6	-81.463	-2.844	FALSE	FALSE	FALSE
16	ZrNdReSrO6	-77.656	-2.595	FALSE	FALSE	FALSE
17	CrNdReSrO6	-77.741	-2.693	TRUE	FALSE	TRUE
18	NbLaMgReO6	-77.058	-2.372	FALSE	FALSE	FALSE
19	BeAuReSrO6	-58.267	-1.286	FALSE	FALSE	FALSE
20	NiTlReSrO6	-62.661	-1.867	FALSE	FALSE	TRUE
21	NdNbReSrO6	-76.274	-2.301	FALSE	FALSE	FALSE
22	RhPrReCdO6	-68.668	-1.894	FALSE	FALSE	TRUE
23	TeFeReSrO6	-64.136	-1.639	FALSE	FALSE	FALSE
24	RuHgReSrO6	-61.543	-1.357	FALSE	FALSE	FALSE
25	PrNbReSrO6	-75.041	-2.177	FALSE	FALSE	FALSE
26	NdTaMnTiO6	-85.703	-2.943	FALSE	FALSE	FALSE
27	TaTiReSrO6	-70.910	-1.277	FALSE	FALSE	FALSE
28	CsNiReSrO6	-62.788	-2.026	FALSE	FALSE	TRUE
29	RhVReSrO6	-67.798	-1.468	FALSE	FALSE	FALSE
30	RhLaReSrO6	-71.547	-2.088	FALSE	FALSE	FALSE
31	RhPrReSrO6	-72.595	-2.208	FALSE	FALSE	FALSE
32	NbLaReCdO6	-70.940	-1.829	FALSE	FALSE	FALSE
33	TaLaMgReO6	-77.546	-2.245	FALSE	FALSE	FALSE
34	MoPmReSrO6	-78.511	-2.772	TRUE	TRUE	TRUE
35	CrPdReSrO6	-67.525	-1.630	FALSE	FALSE	FALSE

Supplementary Table 15 The Names and the number of data-points of databases in this study.

Database	Total number	Number of positive samples	Number of negative samples
MP	85,014	33,998	51,016
OQMD	503,514	56,939	446,575
JARVIS	54,592	4,434	50,158
Li-containing oxides	6,168	750	5,418
Transition-metal oxides	7,137	1,211	5,926
Perovskite oxides	3,469	1,514	1,955
MP-structure	125,45,	33,001	92,450
C2DB	13,408	7,441	5,967
2DMatpeida	4,743	2,682	2061

References

1. Bartel CJ, Trewartha A, Wang Q, Dunn A, Jain A, Ceder G. A critical examination of compound stability predictions from machine-learned formation energies. *npj Comput Mater* **6**, 97 (2020).
2. Bartel CJ. Review of computational approaches to predict the thermodynamic stability of inorganic solids. *J Mater Sci* **57**, 10475-10498 (2022).
3. Fung V, Hu G, Ganesh P, Sumpter BG. Machine learned features from density of states for accurate adsorption energy prediction. *Nat Commun* **12**, 88 (2021).
4. Goodall REA, Lee AA. Predicting materials properties without crystal structure: deep representation learning from stoichiometry. *Nat Commun* **11**, 6280 (2020).
5. Wang AY-T, Kauwe SK, Murdock RJ, Sparks TD. Compositionally restricted attention-based network for materials property predictions. *npj Comput Mater* **7**, 77 (2021).
6. Kailkhura B, Gallagher B, Kim S, Hiszpanski A, Han TY-J. Reliable and explainable machine-learning methods for accelerated material discovery. *npj Comput Mater* **5**, 108 (2019).
7. Agrawal A, Meredig B, Wolverton C, Choudhary A. A Formation Energy Predictor for Crystalline Materials Using Ensemble Data Mining. In: *2016 IEEE 16th International Conference on Data Mining Workshops (ICDMW)*) (2016).
8. Jha D*, et al.* ElemNet: Deep Learning the Chemistry of Materials From Only Elemental Composition. *Sci Rep* **8**, 17593 (2018).
9. Zeng S, Zhao Y, Li G, Wang R, Wang X, Ni J. Atom table convolutional neural networks for an accurate prediction of compounds properties. *npj Comput Mater* **5**, 84 (2019).
10. Talapatra A, Uberuaga BP, Stanek CR, Pilania G. A Machine Learning Approach for the Prediction of Formability and Thermodynamic Stability of Single and Double Perovskite Oxides. *Chem Mater* **33**, 845-858 (2021).