Supplementary Information

Accelerated Design and Discovery of Perovskites with High Conductivity for Energy applications through Machine Learning

Pikee Priya[1,2#], N. R. Aluru[1,2]

[1]Department of Mechanical Science and Engineering, [2]Beckman Institute for Advanced Science and Technology, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, USA

[#]corresponding author: pikeepriya@gmail.com

Supplementary Notes

The following features corresponding to physical, chemical, electrical, mechanical and magnetic properties of the A, B and M (A/B site dopant) sites and ABO3 structure were used for training the model and predictions.

Supplementary Table 1: List of features used in Machine Learning Model and their detailed description

COND	Total Conductivity. The training data was extracted from Literature using WebPlotDigitizer(1).
MW	Molecular Weight of ABO3
WPA	Weight per atom
BM	Bulk Modulus: Approximated as average of the BM of the individual constituent oxides (Materials Project (2))
CV	Cell Volume: Calculated from Molar Volume =Molar Volume/Avagadro Number
RHO	Density: Approximated as average of the density of the individual constituent oxides (Materials Project(2))
EMAX	Maximum electronegativity among the constituent atoms
EMIN	Minimum electronegativity among the constituent atoms
EDIF	Difference of EMAX and EMIN
RMAX	Maximum ionic radii among the constituent atoms (Emery et al. (3))
RMIN	Minimum ionic radii among the constituent atoms (Emery et al. (3))
RDIF	Difference of RMAX and RMIN
MV	Molar Volume: Approximated as an average of the constituent oxides (Materials Project (2))
POL	Sum of the Polarizabilities of the constituent ions (4)
PERM	Permeability (Clausius Mosotti Equation.) (4)
TEMP	Temperature
PO2	Partial pressure of O2
PH2O	Partial pressure of H2O
OV	Off stoichiometry of the O atoms in ABO3
AM_A_MIN	Minimum atomic mass of the A site atoms including the M (dopants)
AM_A_MAX	Maximum atomic mass of A site atoms including the M (dopants)

AM_A_DIF	Difference of AM_A_MAX and AM_A_MIN
AM_A_AVG	Average atomic mass of A site atoms including the M (dopants)
AM_B_MIN	Minimum atomic mass of the B site atoms including the M (dopants)
AM_B_MAX	Maximum atomic mass of B site atoms including the M (dopants)
AM_B_DIF	Difference of AM_B_MAX and AM_B_MIN
AM_B_AVG	Average atomic mass of B site atoms including the M (dopants)
BG_A_MIN	Minimum band gap of the A site oxides (Materials Project (2)) including M site oxides
BG_A_MAX	Maximum band gap of A site oxides (Materials Project (2)) including M site oxides
BG_A_DIF	Difference of BG_A_MAX and BG_A_MIN
BG_A_AVG	Average atomic mass of A site oxides including M site oxides
BG_B_MIN	Minimum band gap of the B site oxides (Materials Project (2)) including M site oxides
BG_B_MAX	Maximum band gap of B site oxides (Materials Project (2)) including M site oxides
BG_B_DIF	Difference of BG_B_MAX and BG_B_MIN
BG_B_AVG	Average band gaps of B site oxides including M site oxides
BM_A_MIN	Minimum bulk modulus of the A site oxides (Materials Project (2)) including M site oxides
BM_A_MAX	Maximum bulk modulus of A site oxides (Materials Project (2)) including M site oxides
BM_A_DIF	Difference of BM_A_MAX and BM_A_MIN
BM_A_AVG	Average bulk moduli of A site oxides including M site oxides
BM_B_MIN	Minimum bulk modulus of the B site oxides (Materials Project (2)) including M site oxides
BM_B_MAX	Maximum bulk modulus of B site oxides (Materials Project (2)) including M site oxides
BM_B_DIF	Difference of BM_B_MAX and BM_B_MIN
BM_B_AVG	Average bulk moduli of B site oxides including M site oxides
CV_A_MIN	Minimum cell volume of the A site oxides (Materials Project (2)) including M site oxides
CV_A_MAX	Maximum cell volume of A site oxides (Materials Project (2)) including M site oxides
CV_A_DIF	Difference of CV_A_MAX and CV_A_MIN
CV_A_AVG	Average cell volumes of A site oxides including M site oxides
CV_B_MIN	Minimum cell volume of the B site oxides (Materials Project (2)) including M site oxides
CV_B_MAX	Maximum cell volume of B site oxides (Materials Project (2)) including M site oxides
CV_B_DIF	Difference of CV_B_MAX and CV_B_MIN
CV_B_AVG	Average cell volumes of B site oxides including M site oxides
RO_A_MIN	Minimum density of the A site oxides (Materials Project (2)) including M site oxides
RO_A_MAX	Maximum density of A site oxides (Materials Project (2)) including M site oxides
RO_A_DIF	Difference of RO_A_MAX and RO_A_MIN
RO_A_AVG	Average densities of A site oxides including M site oxides
RO_B_MIN	Minimum density of the B site oxides (Materials Project (2)) including M site oxides
RO_B_MAX	Maximum density of B site oxides (Materials Project (2)) including M site oxides
RO_B_DIF	Difference of RO_B_MAX and RO_B_MIN
RO_B_AVG	Average densities of B site oxides including M site oxides
EN_A_MIN	Minimum electronegativity of the A site atoms including M site (dopants) atoms
EN_A_MAX	Maximum electronegativity of A site atoms including M site (dopants) atoms
EN_A_DIF	Difference of EN_A_MAX and EN_A_MIN
EN_A_AVG	Average electronegativities of A site atoms including M site (dopants) atoms
EN_B_MIN	Minimum electronegativity of the B site atoms including M site (dopants) atoms
EN_B_MAX	Maximum electronegativity of B site atoms including M site (dopants) atoms

EN_B_DIF	Difference of EN_B_MAX and EN_B_MIN
EN_B_AVG	Average electronegativities of B site atoms including M site (dopants) atoms
FE_A_MIN	Minimum formation energy of the A site oxides (Materials Project (2)) including M site oxides
FE_A_MAX	Maximum formation energy of A site oxides (Materials Project (2)) including M site oxides
FE_A_DIF	Difference of FE_A_MAX and FE_A_MIN
FE_A_AVG	Average formation energies of A site oxides including M site oxides
FE_B_MIN	Minimum formation energy of the B site oxides (Materials Project (2)) including M site oxides
FE_B_MAX	Maximum formation energy of B site oxides (Materials Project (2)) including M site oxides
FE_B_DIF	Difference of FE_B_MAX and FE_B_MIN
FE_B_AVG	Average formation energies of B site oxides including M site oxides
IR_A_MIN	Minimum ionic radius of the A site atoms (Emery et al. (3)) including M site (dopants) atoms
IR_A_MAX	Maximum ionic radius of A site atoms (Emery et al. (3)) including M site (dopants) atoms
IR_A_DIF	Difference of IR_A_MAX and IR_A_MIN
IR_A_AVG	Average ionic radius of A site atoms including M site (dopants) atoms
IR_B_MIN	Minimum ionic radius of the B site atoms (Emery et al. (3)) including M site (dopants) atoms
IR_B_MAX	Maximum ionic radius of B site atoms (Emery et al. (3)) including M site (dopants) atoms
IR_B_DIF	Difference of IR_B_MAX and IR_B_MIN
IR_B_AVG	Average ionic radius of A site atoms including M site (dopants) atoms
MM_A_MIN	Minimum magnetic moment of the A site oxides (Materials Project (2)) including M site oxides
MM_A_MAX	Maximum magnetic moment of A site oxides (Materials Project (2)) including M site oxides
MM_A_DIF	Difference of MM_A_MAX and MM_A_MIN
MM_A_AVG	Average magnetic moment of A site oxides including M site oxides
MM_B_MIN	Minimum magnetic moment of the B site oxides (Materials Project (2)) including M site oxides
MM_B_MAX	Maximum magnetic moment of B site oxides (Materials Project (2)) including M site oxides
MM_B_DIF	Difference of MM_B_MAX and MM_B_MIN
MM_B_AVG	Average magnetic moment of B site oxides including M site oxides
MV_A_MIN	Minimum molar volume of the A site oxides (Materials Project (2)) including M site oxides
MV_A_MAX	Maximum molar volume of A site oxides (Materials Project (2)) including M site oxides
MV_A_DIF	Difference of MV_A_MAX and MV_A_MIN
MV_A_AVG	Average molar volume of A site oxides including M site oxides
MV_B_MIN	Minimum molar volume of the B site oxides (Materials Project (2)) including M site oxides
MV_B_MAX	Maximum molar volume of B site oxides (Materials Project (2)) including M site oxides
MV_B_DIF	Difference of MV_B_MAX and MV_B_MIN
MV_B_AVG	Average molar volume of B site oxides including M site oxides
OV_A	Off stoichiometry of A site atoms
OV_B	Off stoichiometry of B site atoms
POL_A_MIN	Minimum polarizability of the A site atoms (4) including M site (dopants) atoms
POL_A_MAX	Maximum polarizability of A site atoms (4) including M site (dopants) atoms
POL_A_DIF	Difference of POL_A_MAX and POL_A_MIN
POL_A_AVG	Average polarizability of A site atoms including M site (dopants) atoms
POL_B_MIN	Minimum polarizability of the B site atoms (4) including M site (dopants) atoms
POL_B_MAX	Maximum polarizability of B site atoms (4) including M site (dopants) atoms
POL_B_DIF	Difference of POL_B_MAX and POL_B_MIN
POL_B_AVG	Average polarizability of B site atoms including M site (dopants) atoms

TF	Tolerance Factor ($(R_A+R_O)/2^{0.5}/(R_B+R_O)$)
PERM_A	Permeability of the doped A site oxide (Clausius Mosotti Equation (4))
PERM_B	Permeability of the doped B site oxide (Clausius Mosotti Equation (4))
RRAT	Ratio of IR_A_AVG and IR_B_AVG
GS	Grain Size

The Feature Importance (%) from the XG-Boost (XGB) Regression and Classification Model are listed below. The XGB Regression Model was considered with all the training data (7230) with grain size as a feature and with partial literature data (1124) which had grain size reported. Grain size was among one of the top 35 features when it was considered highlighting the importance of microstructure in determining conductivity of solid oxide perovskites.

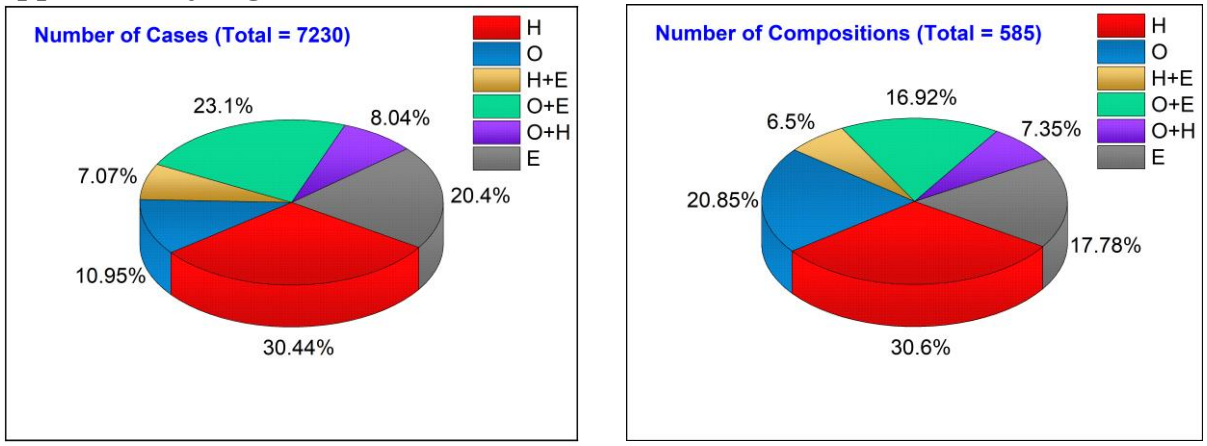
Supplementary Table 2: Features with importance % for XGB Regression and Classification Models

Feature	Feature Importance (%) XGB Classification	Feature Importance (%) XGB Regression (without grain size)	Feature Importance (%) XGB Regression (with grain size)
COND	0.2489		
MW	0.7960	0.0100	0.0433
WPA	0.2718	0.0029	0.1591
BM	0.3552	0.0516	0.1605
CV	0.4057	0.0253	0.6022
RHO	0.2381	0.0270	0.0367
EMAX	0.4212	0.0162	0.2158
EMIN	0.9652	0.3015	0.0000
EDIF	0.6147	0.0091	0.1672
RMAX	0.0000	0.0003	0.0000
RMIN	0.0475	0.0127	0.1407
RDIF	1.4638	0.1207	0.0003
MV	0.1385	0.1406	2.0243
POL	0.5045	0.0231	0.1485
PERM	0.5002	0.0157	0.0756
TEMP	0.1417	0.0149	0.3287
PO2	0.4278	0.0081	0.0728
PH2O	0.2518	0.0380	0.0649
OV	0.3552	0.0538	0.1913
AM_A_MIN	0.1829	0.3617	1.2086
AM_A_MAX	0.0116	0.0000	0.0000
AM_A_DIF	0.1048	0.0184	0.0000
AM_A_AVG	0.0048	0.0608	0.1397
AM_B_MIN	22.5070	0.0621	0.0047
AM_B_MAX	0.4331	0.0073	2.8023
AM_B_DIF	0.1921	0.0101	0.2510

AM_B_AVG	9.6536	2.6892	0.0664
BG_A_MIN	0.0229	0.0003	0.0000
BG_A_MAX	0.0000	0.0052	0.0000
BG_A_DIF	0.2105	0.7525	0.0000
BG_A_AVG	0.0108	0.0115	0.0000
BG_B_MIN	0.4024	0.0348	0.1772
BG_B_MAX	1.9063	0.5060	23.0490
BG_B_DIF	0.0685	0.1062	0.0295
BG_B_AVG	0.4154	0.0098	0.3441
BM_A_MIN	0.0241	0.0000	0.0000
BM_A_MAX	1.5234	2.1669	0.0000
BM_A_DIF	1.4716	0.0042	0.0000
BM_A_AVG	0.6355	0.1418	0.0000
BM_B_MIN	0.2893	0.1529	0.3491
BM_B_MAX	0.3606	0.0004	1.1574
BM_B_DIF	0.0862	0.0282	3.5740
BM_B_AVG	0.5958	0.0289	0.1000
CV_A_MIN	4.2911	0.0000	0.0000
CV_A_MAX	0.0047	0.0568	0.0000
CV_A_DIF	0.5007	0.0000	0.0000
CV_A_AVG	1.4906	0.1328	0.0000
CV_B_MIN	0.6971	1.9332	0.0196
CV_B_MAX	3.2778	0.0121	0.0920
CV_B_DIF	0.0583	0.9447	1.0248
CV_B_AVG	0.9981	2.9616	0.0475
RO_A_MIN	0.0000	0.0001	0.0000
RO_A_MAX	0.0000	0.0000	0.0000
RO_A_DIF	0.0022	0.0003	0.0000
RO_A_AVG	0.0272	0.1188	0.0021
RO_B_MIN	0.2534	6.6629	0.0000
RO_B_MAX	0.0275	0.0389	0.0000
RO_B_DIF	0.4288	0.0226	0.0017
RO_B_AVG	0.8594	0.3641	0.4023
EN_A_MIN	0.0000	0.0000	0.0000
EN_A_MAX	0.0000	0.0000	0.0000
EN_A_DIF	0.0493	0.0000	0.0000
EN_A_AVG	0.4088	0.1507	0.0003
EN_B_MIN	0.6363	56.0568	44.9921
EN_B_MAX	4.6990	6.3820	0.0000
EN_B_DIF	0.5244	0.0025	0.2489
EN_B_AVG	0.5372	0.0520	6.5101
FE_A_MIN	0.8008	0.0005	0.0000
FE_A_MAX	0.0365	0.0000	0.0000
FE_A_DIF	0.2146	0.0006	0.0000

FE_A_AVG	0.3901	0.0149	0.0000
FE_B_MIN	7.1677	1.2322	2.6285
FE_B_MAX	0.2432	0.0807	0.0267
FE_B_DIF	0.0206	0.0063	0.2144
FE_B_AVG	0.4633	0.0402	0.1704
IR_A_MIN	0.0456	0.0067	0.0000
IR_A_MAX	0.0000	0.0120	0.0000
IR_A_DIF	0.0112	0.0027	0.0000
IR_A_AVG	0.0134	0.0026	0.0000
IR_B_MIN	0.0029	0.0000	0.0000
IR_B_MAX	1.0453	0.0405	0.0002
IR_B_DIF	0.1625	0.0094	1.2552
IR_B_AVG	0.1966	8.8886	0.0058
MM_A_MIN	0.0000	0.0000	0.0000
MM_A_MAX	0.0000	0.0000	0.0000
MM_A_DIF	0.0000	0.0000	0.0000
MM_A_AVG	0.0000	0.0000	0.0000
MM_B_MIN	6.8975	0.0075	0.0000
MM_B_MAX	0.1125	0.0011	0.0572
MM_B_DIF	0.0183	1.8017	0.0000
MM_B_AVG	1.6422	0.0278	0.6259
MV_A_MIN	0.0000	0.0000	0.0000
MV_A_MAX	0.0000	0.0000	0.0000
MV_A_DIF	0.0118	0.0022	0.0000
MV_A_AVG	2.1563	0.0000	0.0000
MV_B_MIN	0.0158	0.0009	0.0001
MV_B_MAX	0.5795	0.0036	0.0009
MV_B_DIF	0.3503	0.0517	0.0644
MV_B_AVG	0.2909	0.0557	0.3289
OV_A	6.0344	1.0512	0.0000
OV_B	0.0764	0.0009	0.0051
POL_A_MIN	0.0000	0.0000	0.0000
POL_A_MAX	0.0612	0.0000	0.0000
POL_A_DIF	0.0000	0.0000	0.0000
POL_A_AVG	0.0163	0.0029	0.0000
POL_B_MIN	0.0040	0.0546	0.2174
POL_B_MAX	0.0037	2.0019	0.4538
POL_B_DIF	0.0457	0.0091	0.0087
POL_B_AVG	0.0447	0.0037	1.2059
TF	0.4245	0.0211	0.4551
PERM_A	0.6447	0.0745	0.0000
PERM_B	0.2695	0.5530	0.5230
RRAT	0.4528	0.0136	0.5710
GS			0.1553

Supplementary Figures

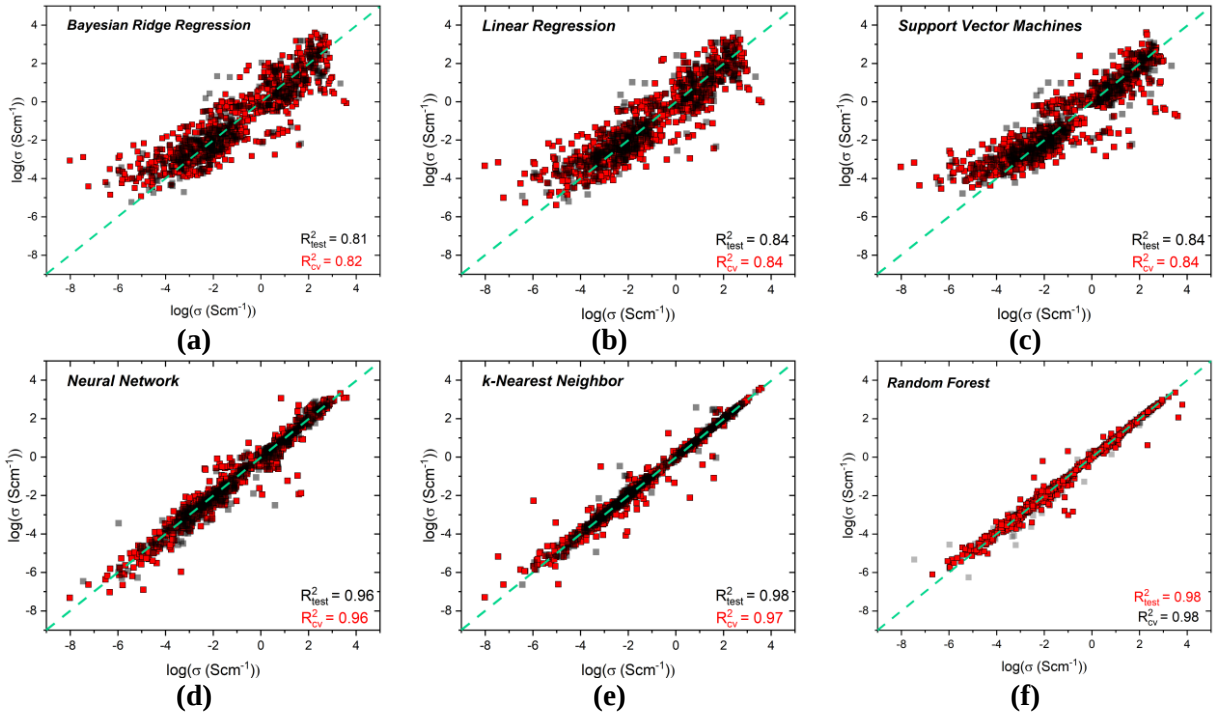


(a) **(b)**

Supplementary Figure 1. Distribution of (a) the number of total data points and (b) the number of unique compositions for different charge carrier types of the conductivity data collected from 118 journal articles.

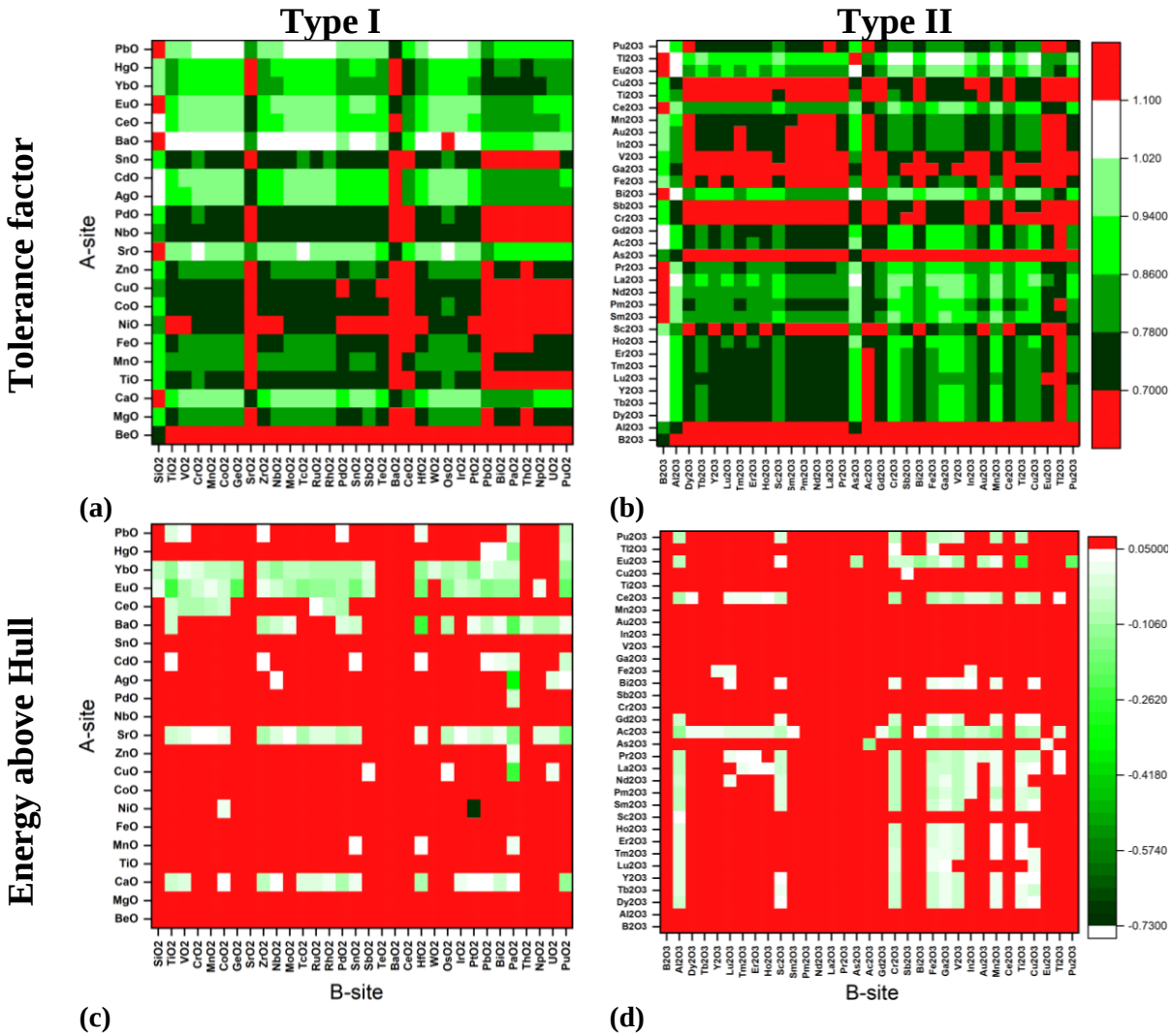
Supplementary Methods

Bayesian Ridge Regression, Linear Regression, Support Vector Machines, Neural networks, k-Nearest Neighbor and Random Forest Regression were tested for minimum errors and coefficient of determination for the testing (15%) and cross-validation (5%) sets. A comparison of the performance of different models is shown in Fig. S4. A python based Scikit-learn package (5) was used for implementation of the models. The data was normalized for all the models except Random Forest for better performance as the features varied over several orders of magnitude.



Supplementary Figure 2: Comparison of testing and cross validation predictions for different Machine Learning models (a) Bayesian Ridge (b) Linear Regression (c) Support Vector Machines (d) Neural Network (e) k-Nearest Neighbor and (f) Random Forest Regressor.

Supplementary Figures



Supplementary Figure 3.: The stable perovskites of Type I (AO+BO2) (a) and (b) and Type II (A2O3+B2O3) (c) and (d). The stabilities are shown for 0.7 < tolerance factor < 1.1 (a) and (c) and Energy above Hull < 0.05 eV (b) and (d) from Emery et al. (3).

Supplementary References

1. Rohatdi A. WebPlotDigitizer [Internet]. 2019. Available from: https://automeris.io/WebPlotDigitizer

2. Jain A, Ong SP, Hautier G, Chen W, Richards WD, Dacek S, et al. Commentary: The materials project: A materials genome approach to accelerating materials innovation. APL Mater. 2013;1(1).

3. Emery AA, Saal JE, Kirklin S, Hegde VI, Wolverton C. High-Throughput Computational Screening of Perovskites for Thermochemical Water Splitting Applications. Chem Mater. 2016;28(16):5621–34.

4. Shannon RD. Dielectric polarizabilities of ions in oxides and fluorides. J Appl Phys. 1993;73(1):348–66.

5. Varoquaux G, Buitinck L, Louppe G, Grisel O, Pedregosa F, Mueller A. Scikit-learn. GetMobile Mob Comput Commun [Internet]. 2015 Jun 1;19(1):29–33. Available from: http://dl.acm.org/citation.cfm?doid=2786984.2786995