

Supporting Information: Many-body physics and machine learning enabled discovery of promising solar materials

Tathagata Biswas, Adway Gupta, Arunima K. Singh

I. PERFORMANCE OF VARIOUS ML REGRESSION AND CLASSIFICATION ALGORITHMS

Algorithm	QP gap		EBE	
	R ²	RMSE (eV)	R ²	RMSE (eV)
RF	0.98	0.36	0.86	0.29
MLP	0.94	0.64	0.71	0.39
SVM	0.94	0.73	0.45	0.59
KRR	0.92	0.76	0.34	0.86

TABLE S1 Performance of various ML regression algorithms in predicting QP gap and EBE. The accuracy is determined in terms of R² score and RMSE in eV. Abbreviations-RF=Random Forest, MLP=Multilayer Perceptron, SVM=Support Vector Machine, KRR=Kernel Ridge Regression

One of the important features used by the ML model for predicting several properties is the electronegativity difference (*END*) between cations and anions in the material. This feature is undefined for single-element compounds, which keeps them excluded from the list of materials included in the QP gap prediction. However, *END* is not in the list of 5 features that are found to be necessary for our QPG ML model. We tested our ML model applicability to predict the QP gap of both single-element as well as multi-element compounds. We found the accuracy of our ML model decreased slightly for single elements and we found an RMSE value of 0.61 eV and an R² score of 0.95. In contrast to QP gap prediction, EBE prediction requires *END* therefore we can not test our ML model's accuracy in predicting EBE of single-element compounds. However, one can set the *END* to zero for all the single-element compounds arbitrarily and include them in the training set. We find that this approach produces the same accuracy in the cross-validation RMSE score (0.43 eV) and R² score (0.75) as in the case of not including the *END* as a feature in the ML model.

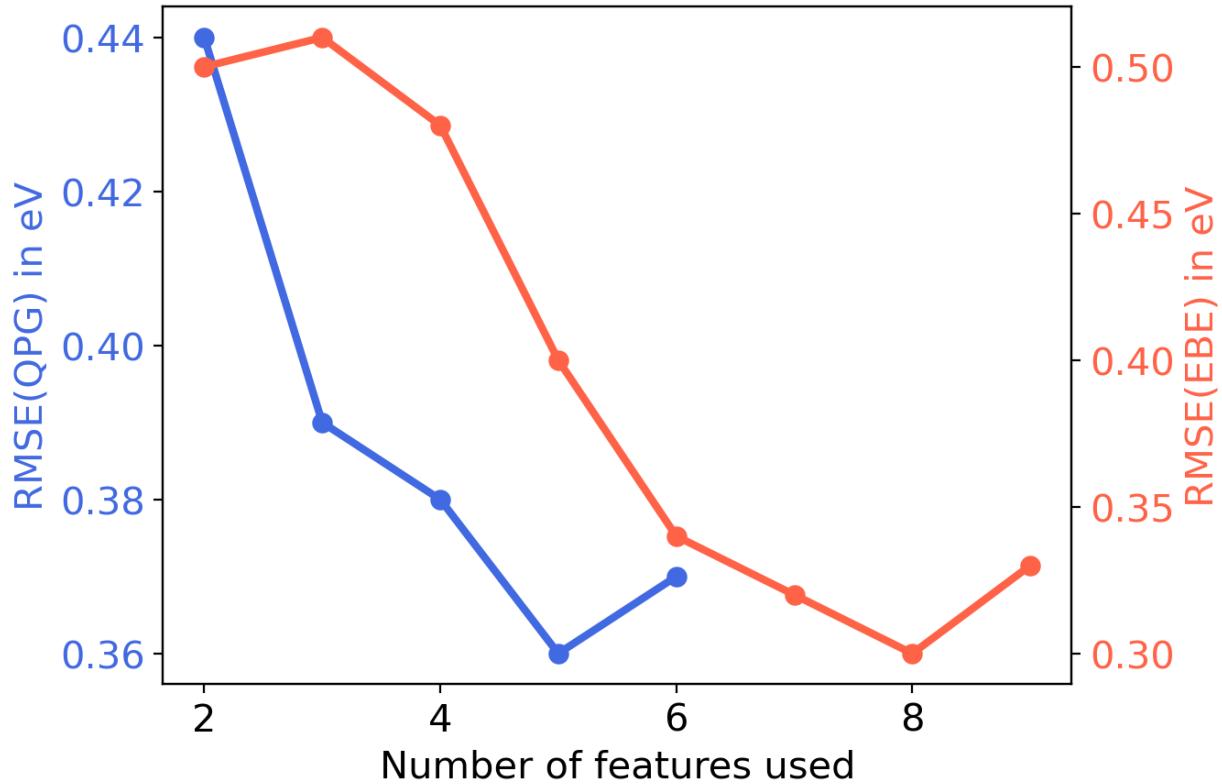


FIG. S1 Performance of Random Forest algorithm with different numbers of (most important) features used in predicting quasiparticle gap, QPG, and exciton binding energy, EBE. These results show that using 6 and 8 features for QPG and EBE prediction respectively provides optimal accuracy by avoiding the risk of overfitting.

Algorithm	EBE	IAC	AAC
RFC	93	94	83
RFC(single)	90	94	85
MLP	85	91	84
SGD	78	74	68
ADA	92	92	80

TABLE S2 Performance of various ML classification algorithms in classifying materials based on their EBE, IAC, and AAC. The accuracy is determined in terms of % of truly classified materials. Abbreviations-RFC=Random Forest Classifier, RFC(single)=Random Forest Classifier which includes single element materials by setting electronegativity difference as zero, MLP=Multilayer Perceptron, SGD=Stochastic Gradient Descent, ADA=AdaBoost

II. DESCRIPTION OF THE FEATURES CONSIDERED FOR ML ALGORITHMS

Features derived from the crystal structure and elemental composition of the materials were considered. They were obtained using *featurizers* as implemented in the *matminer* (1) package. The *matminer* package retrieves large datasets from different databases and then extracts features to transform the raw data into representations suitable for machine learning without the need to familiarize oneself with the API for each data source and preprocessing retrieved data. Some of the crystal structure-dependent features generated using *matminer* were volume per atom (VPA), and packing fraction. Several composition-dependent features were also considered, including the mean, maximum, and minimum of various properties such as melting point, electronegativity, atomic weight, and Mendeleev number of the constituent elements of the materials. These features were generated using *matminer* which retrieves data from The Materials Agnostic Platform for Informatics and Exploration (Magpie) database (2). These elemental features are sufficient for ML prediction of diverse properties of crystalline and amorphous materials, such as band gap energy and glass-forming ability (2).

A. Magpie elemental features

Feature	Description
Number (N_{atomic})	Atomic number
MendeleevNumber	Mendeleev Number (position on the periodic table, counting columnwise from H)
AtomicWeight	Atomic weight
MeltingT (T_{melt})	Melting temperature of element in K
Column	Column on periodic table
Row	Row on periodic table
CovalentRadius	Covalent radius of each element in pm
Electronegativity (χ)	Pauling electronegativity
NValence (N_{val})	Number of filled orbitals
NsValence (Ns_{val})	Number of filled s orbitals

NpValence (Np _{val})	Number of filled p orbitals
NdValence (Nd _{val})	Number of filled d orbitals
NfValence (Nf _{val})	Number of filled f orbitals
NUnfilled (N _{emp})	Number of unfilled orbitals
NsUnfilled (Ns _{emp})	Number of unfilled s orbitals
NpUnfilled (Np _{emp})	Number of unfilled p orbitals
NdUnfilled (Nd _{emp})	Number of unfilled d orbitals
NfUnfilled (Nf _{emp})	Number of unfilled f orbitals
GSvolume_pa (GSVPA)	DFT volume per atom of T=0K ground state in $\text{\AA}^3 \text{atom}^{-1}$
GSbandgap	DFT bandgap energy of T=0K ground state in eV
GSmagmom	DFT magnetic momenet of T=0K ground state
SpaceGroupNumber	Space group of T=0K ground state structure

TABLE S3: Table shows the elemental features derived using

Matminer from the Magpie database

B. Other composition and structure-dependent features

Feature	Description
oxidation state	Concentration-weighted statistics (minimum, maximum, range, and standard deviation) of the oxidation states of constituent atoms.
HOMO_energy, LUMO_energy	The highest occupied molecular orbital (HOMO) and lowest unoccupied molecular orbital (LUMO) estimated from the atomic orbital energies of the composition. The atomic orbital energies are from NIST: https://www.nist.gov/pml/data/atomic-reference-data-electronic-structure-calculations
gap_AO	Estimated bandgap from HOMO and LUMO energies
EN difference	Concentration-weighted statistics (minimum, maximum, range, and standard deviation) of electronegativity difference between cations and anions.
density	The density of the material in units of g cm ⁻³
vpa	Volume per atom of the material crystal structure in Å ³ atom ⁻¹
packing fraction	Ratio of the sum of all atomic volume ($4\pi r_{atom}^3$, r_{atom} is the atomic radius) and unit cell of volume
HOMO_character_encoded, LUMO_character_encoded	Estimated orbital character of HOMO and LUMO ('s', 'p', 'd', or 'f') encoded using 'LabelEncoder' of <i>scikit – learn</i>
ϵ	DFT computed the electronic part of the static dielectric constants. It is a list which contains ϵ_{xx} , ϵ_{yy} and ϵ_{zz} .
μ_e/μ_h	List of all electron (e) and hole (h) effective masses. The list contains all the effective masses computed for different degenerate bands at CBM/VBM and along all the high-symmetry directions available at those k-points.
E_g^{DFT}	DFT computed bandgaps of the materials.

TABLE S4: Table shows the composition and structure dependent features derived using Matminer

C. Statistical operations and the abbreviations used for them

Abbreviation	Operation
min	Minimum value in a list
max	Maximum value in a list
range	Range of a list
mean (avg)	Arithmetic mean of list
avg-dev (σ)	Mean absolute deviation of a list of element data. This is computed by first calculating the mean of the list and then computing the average absolute difference between each value and the mean.
std_dev ($s\sigma$)	Standard deviation of a list of element data
mode	Mode of a list of data. If multiple elements occur equally frequently (or the same weight, if weights are provided), this function will return the minimum of those values.

TABLE S5 Different statistical operations performed by matminer to combine multiple properties to compose features used in the ML models. We have listed the corresponding abbreviations used in figures and tables in this article.

III. MOST IMPORTANT FEATURES IN PREDICTING IAC AND AAC

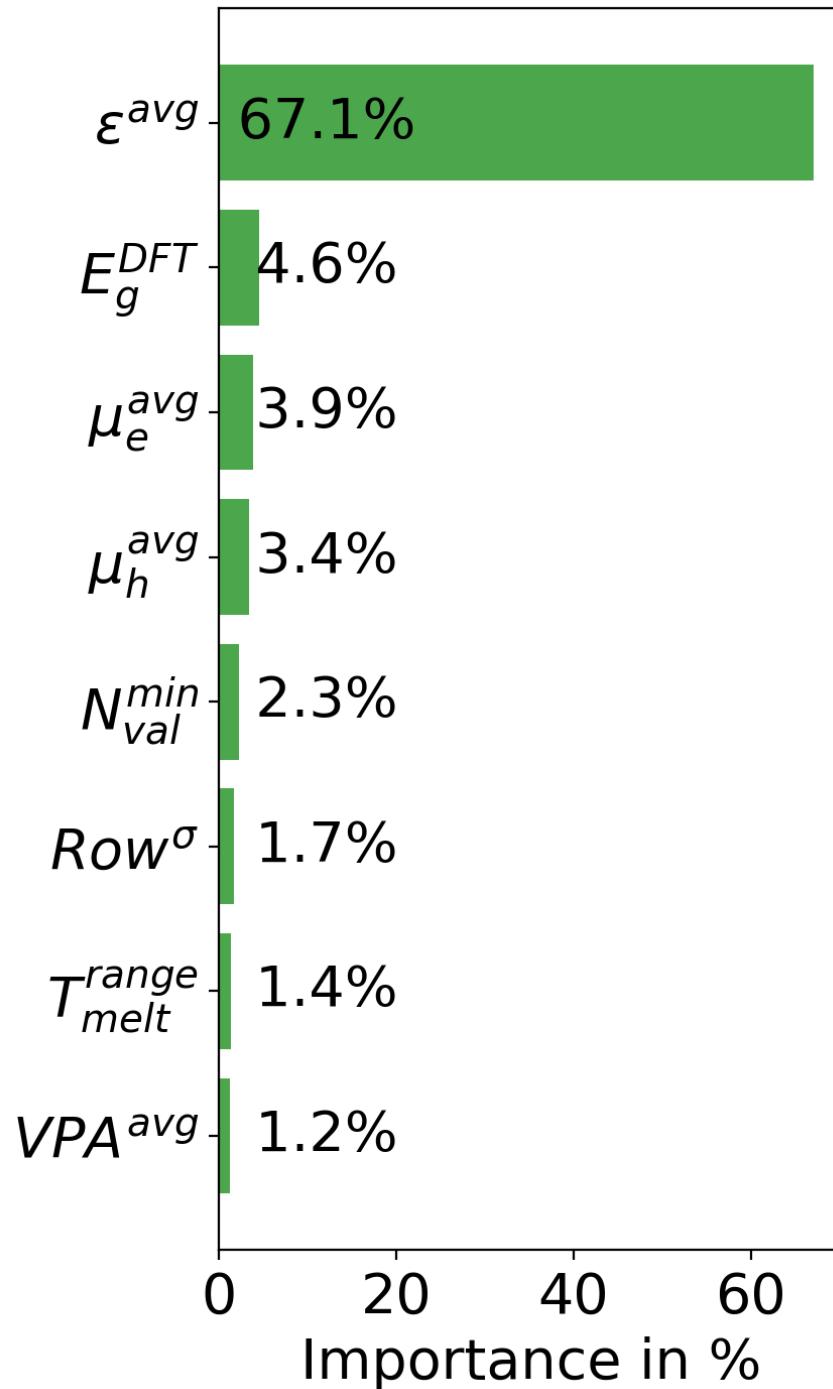


FIG. S2 The most important features in predicting the IAC.

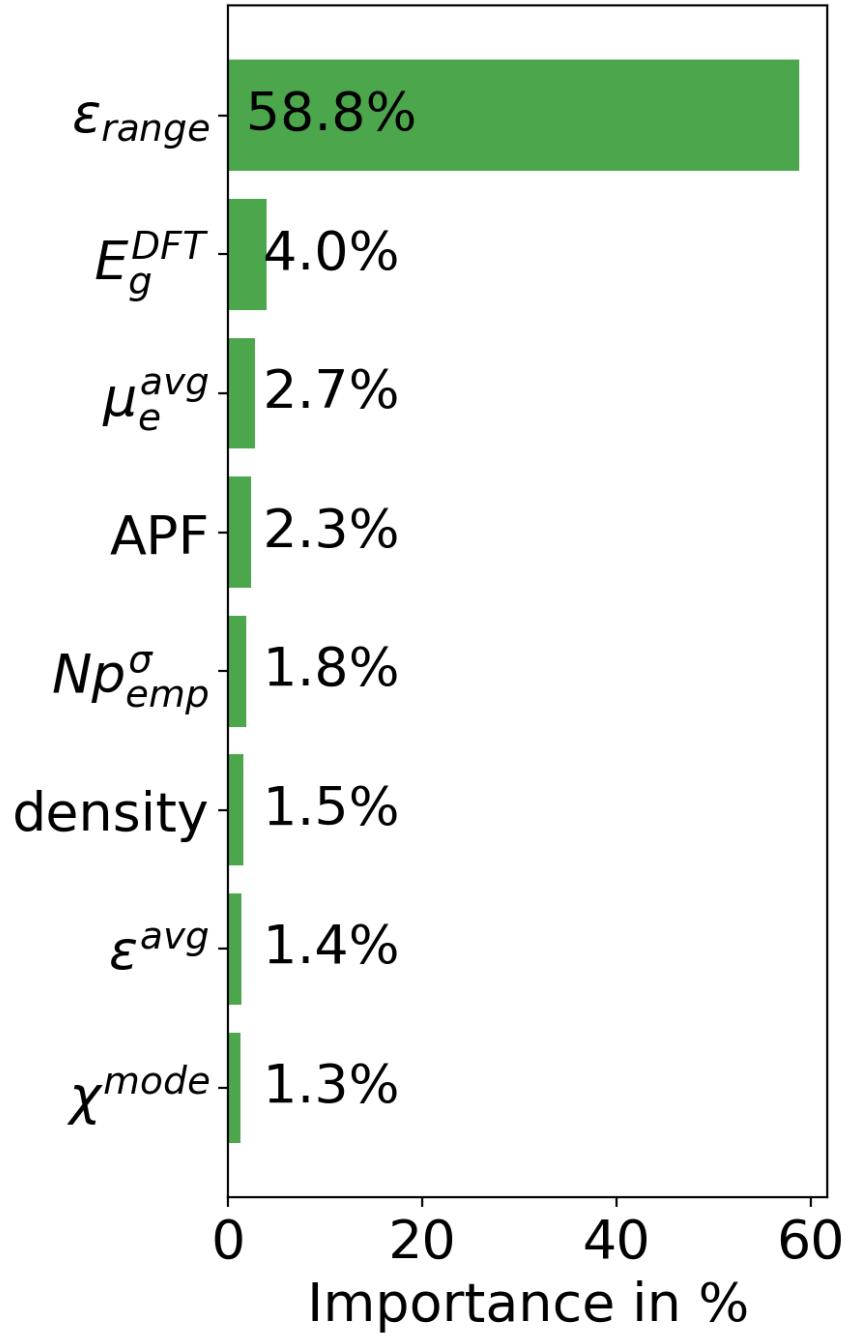


FIG. S3 The most important features in predicting the AAC.

IV. LIST SCREENED MATERIALS FOR UV/VISIBLE (VIS) LIGHT APPLICATION OBTAINED FROM MP DATABASE

TABLE S6: List of materials screened for UV/Visible (VIS) light application. The table lists their Materials Project identifier (mp-id), chemical formula, whether they have an ICSD id, i.e. if they have been synthesized before, Spacegroup number, Crystal system they belong to, Energy above hull (in eV per atom), DFT computed bandgap, ML predicted QPG, and ML predicted EBE (in eV).

No.	mp-id	Formula	Synth.	Spg No.	Crystal system	N_{atom}	E_{hull}	E_g^{DFT}	QPG	EBE	UV/VIS
1	mp-1002124	HfC	True	216	cubic	2	0.64	0.58	0.93	0.13	VIS
2	mp-1009894	ZrC	True	216	cubic	2	0.61	0.55	0.74	0.12	VIS
3	mp-1013548	Ca ₃ SbN	False	221	cubic	5	0.00	0.78	1.23	0.15	UV+VIS
4	mp-10155	P ₂ Ir	True	14	monoclinic	12	0.00	0.78	1.07	0.11	VIS
5	mp-10161	KZnSb	True	187	hexagonal	3	0.01	0.35	1.12	0.15	UV+VIS
6	mp-1018093	CaGaGeH	True	156	trigonal	4	0.00	0.50	1.10	0.27	UV
7	mp-1018100	AlSb	True	186	hexagonal	4	0.01	0.91	1.57	0.08	UV+VIS
8	mp-1018118	TmSbPd	True	216	cubic	3	0.23	0.32	0.97	0.11	UV+VIS
9	mp-1018766	Li ₃ LaAs ₂	True	164	trigonal	6	0.00	0.61	1.29	0.17	VIS
10	mp-10288	LaCuTeS	True	14	monoclinic	16	0.00	0.58	1.34	0.17	VIS
11	mp-1029267	CaZrN ₂	False	166	trigonal	4	0.00	0.42	1.14	0.16	VIS
12	mp-1029312	RbVN ₂	False	122	tetragonal	8	0.00	0.71	1.21	0.13	UV
13	mp-1029313	NaRuN	False	186	hexagonal	6	0.08	0.46	1.01	0.29	VIS
14	mp-1029479	Ca(InN) ₂	False	166	trigonal	5	0.01	0.55	0.82	0.19	UV
15	mp-1029619	Ba(ReN ₂) ₂	False	205	cubic	84	0.09	0.47	0.71	0.09	UV+VIS
16	mp-10623	ThSbRh	True	216	cubic	3	0.00	0.99	1.34	0.17	UV+VIS
17	mp-1078366	Bi ₂ S ₂ O	True	129	tetragonal	10	0.00	1.17	1.97	0.17	UV
18	mp-1078709	Na ₂ BiAu	True	63	orthorhombic	8	0.00	0.57	1.22	0.27	UV+VIS

19	mp-1079772	Ca(InP) ₂	True	194	hexagonal	10	0.00	0.41	1.14	0.18	UV
20	mp-1094088	NbCoSn	False	216	cubic	3	0.00	1.08	1.52	0.16	UV+VIS
21	mp-1095507	SbIrSe	True	198	cubic	12	0.00	1.27	1.79	0.12	UV+VIS
22	mp-1095603	YAsSe	True	62	orthorhombic	12	0.00	0.51	0.67	0.31	UV+VIS
23	mp-1100405	ScTeRh	False	216	cubic	3	0.00	0.49	0.96	0.13	UV+VIS
24	mp-1100408	TaSnRh	False	216	cubic	3	0.00	1.14	1.84	0.08	UV+VIS
25	mp-1100415	ZrSbRh	False	216	cubic	3	0.00	1.18	1.82	0.18	UV+VIS
26	mp-1101755	GePtS	True	29	orthorhombic	12	0.00	0.99	1.34	0.14	UV+VIS
27	mp-1101765	BiRhSe	True	198	cubic	12	0.00	0.32	0.74	0.11	UV+VIS
28	mp-1101917	SnBi ₄ Te ₇	True	164	trigonal	12	0.00	0.64	0.89	0.08	UV
29	mp-1102430	SbTeIr	True	198	cubic	12	0.00	0.97	1.24	0.11	UV+VIS
30	mp-1102531	PRhSe	True	198	cubic	12	0.00	1.19	1.85	0.10	UV+VIS
31	mp-1103228	BiIrSe	True	198	cubic	12	0.00	0.69	0.89	0.17	UV+VIS
32	mp-1103261	SiPtSe	True	29	orthorhombic	12	0.00	0.97	1.31	0.13	UV+VIS
33	mp-1104130	Ba(P ₂ Au) ₂	True	70	orthorhombic	14	0.00	0.37	1.08	0.28	UV
34	mp-1105594	La ₃ Cu ₃ Sb ₄	True	220	cubic	20	0.00	0.43	0.74	0.12	UV+VIS
35	mp-1106	Sr ₂ Si	True	62	orthorhombic	12	0.00	0.34	0.93	0.14	UV
36	mp-1106051	Y ₃ Cu ₃ Sb ₄	True	220	cubic	20	0.00	0.46	0.74	0.12	UV+VIS
37	mp-11520	YNiSb	True	216	cubic	3	0.00	0.35	0.99	0.20	UV+VIS
38	mp-11661	Zr ₃ N ₄	True	220	cubic	14	0.05	0.69	1.21	0.09	VIS
39	mp-1189341	HfTl ₂ PbSe ₄	True	15	monoclinic	16	0.00	0.96	1.54	0.24	UV+VIS
40	mp-1189464	ZrTl ₂ PbSe ₄	True	15	monoclinic	16	0.00	1.02	1.62	0.14	UV+VIS
41	mp-1190030	La ₃ Sb ₄ Au ₃	True	220	cubic	20	0.00	0.60	0.84	0.13	UV+VIS
42	mp-1191203	CuNiSbS ₃	True	19	orthorhombic	24	0.00	0.42	0.84	0.09	UV
43	mp-1217120	Ti ₂ FeNiSb ₂	False	160	trigonal	6	0.00	0.96	1.11	0.16	VIS
44	mp-1226003	CoPS	False	198	cubic	12	0.00	1.13	1.95	0.12	UV+VIS
45	mp-12418	SrCaGe	True	62	orthorhombic	12	0.00	0.46	0.95	0.27	UV+VIS
46	mp-1244	SrGe ₂	True	62	orthorhombic	24	0.00	0.40	0.94	0.27	UV+VIS
47	mp-12908	ScAgSe ₂	True	164	trigonal	4	0.00	0.73	1.50	0.19	UV

48	mp-13097	NaZnAs	True	129	tetragonal	6	0.00	0.36	1.08	0.15	UV+VIS
49	mp-1317	CoSb ₃	True	204	cubic	16	0.00	0.58	0.72	0.12	UV+VIS
50	mp-13654	Y ₃ Sb ₄ Au ₃	True	220	cubic	20	0.00	0.58	0.77	0.13	UV+VIS
51	mp-137	Ge	True	96	tetragonal	12	0.24	0.87	0.93	0.08	UV+VIS
52	mp-14501	Ba(As ₃ Pt ₂) ₂	True	15	monoclinic	22	0.00	0.33	0.72	0.26	UV
53	mp-1477	BaSi ₂	True	62	orthorhombic	24	0.00	0.79	1.34	0.12	UV
54	mp-14790	Ge(Te ₂ As) ₂	True	166	trigonal	7	0.00	0.73	0.88	0.08	UV
55	mp-14791	Ge ₂ Te ₅ As ₂	True	164	trigonal	9	0.01	0.59	0.81	0.03	UV
56	mp-149	Si	True	227	cubic	2	0.00	0.85	1.57	0.15	UV
57	mp-15074	Sr(PIr) ₂	True	154	trigonal	15	0.00	0.38	0.73	0.27	VIS
58	mp-15661	Sc ₄ C ₃	True	220	cubic	14	0.00	0.47	0.88	0.13	UV+VIS
59	mp-15953	P ₂ Rh	True	14	monoclinic	12	0.00	0.54	0.95	0.15	VIS
60	mp-15988	Li ₂ CuSb	True	216	cubic	4	0.00	0.53	1.08	0.11	UV+VIS
61	mp-16363	CoAsS	True	198	cubic	12	0.00	0.88	0.98	0.11	UV+VIS
62	mp-165	Si	True	194	hexagonal	4	0.01	0.51	1.38	0.14	UV
63	mp-16608	Si ₃ Os ₂	True	60	orthorhombic	40	0.00	0.86	1.15	0.12	UV+VIS
64	mp-16609	Si ₃ Os ₂	True	116	tetragonal	20	0.03	0.72	1.11	0.12	UV+VIS
65	mp-16610	Ge ₃ Os ₂	True	60	orthorhombic	40	0.00	0.74	0.91	0.09	UV+VIS
66	mp-17123	Si ₂ Os	True	64	orthorhombic	24	0.00	0.65	1.05	0.08	UV
67	mp-17568	Ag ₃ Sn ₂ (GeP ₂) ₃	True	217	cubic	28	0.00	0.34	0.78	0.10	UV+VIS
68	mp-17833	Ba ₃ GeO	True	62	orthorhombic	20	0.00	0.45	1.02	1.04	UV
69	mp-17862	Ag ₃ Ge ₅ P ₆	True	217	cubic	28	0.00	0.58	1.07	0.10	UV+VIS
70	mp-17926	Zr ₃ Ni ₃ Sb ₄	True	220	cubic	20	0.00	0.54	0.73	0.10	UV+VIS
71	mp-18500	K ₃ Sn ₄ Au	True	59	orthorhombic	16	0.00	0.61	0.93	0.46	UV
72	mp-19	Te	True	152	trigonal	3	0.00	0.58	0.80	0.08	UV
73	mp-1922	RuSe ₂	True	205	cubic	12	0.00	0.48	0.99	0.14	UV+VIS
74	mp-19717	TePb	True	225	cubic	2	0.00	1.06	1.47	0.10	UV+VIS
75	mp-19886	ThSnPt	True	216	cubic	3	0.00	0.88	0.98	0.12	UV+VIS
76	mp-20191	Si ₃ Ru ₂	True	116	tetragonal	20	0.02	0.38	0.89	0.10	UV+VIS

77	mp-2030	RuS ₂	True	205	cubic	12	0.00	0.68	1.10	0.10	UV+VIS	
78	mp-2071	Zn ₃ P ₂	True	137	tetragonal	40	0.00	0.61	1.24	0.15	VIS	
79	mp-20817	GePtSe	True	29	orthorhombic	12	0.00	0.63	0.85	0.13	UV+VIS	
80	mp-21272	ErNiSb	True	216	cubic	3	0.00	0.33	0.99	0.20	UV+VIS	
81	mp-21276	PbS	True	225	cubic	2	0.00	0.99	1.49	0.13	UV+VIS	
82	mp-2139	BaGe ₂	True	62	orthorhombic	24	0.00	0.57	1.04	0.27	UV	
83	mp-2201	PbSe	True	225	cubic	2	0.00	0.91	1.39	0.11	UV+VIS	
84	mp-22192	Si ₃ Ru ₂	True	60	orthorhombic	40	0.00	0.55	0.91	0.10	UV+VIS	
85	mp-2231	SnS	True	62	orthorhombic	8	0.00	1.11	1.85	0.18	UV+VIS	
86	mp-226	FeS ₂	True	205	cubic	12	0.01	0.46	1.02	0.13	UV+VIS	
87	mp-22786	ThNiSn	True	216	cubic	3	0.00	0.35	0.78	0.17	UV+VIS	
88	mp-22982	CuBiS ₂	True	62	orthorhombic	16	0.00	0.50	1.14	0.15	VIS	
89	mp-23116	CuBiSeO	True	129	tetragonal	8	0.00	0.34	1.09	0.18	UV	
90	mp-2455	As ₂ Os	True	58	orthorhombic	6	0.00	0.76	1.14	0.11	UV	
91	mp-2488	SiOs	True	198	cubic	8	0.01	0.62	0.81	0.11	UV+VIS	
92	mp-2576	Sr ₂ Ge	True	62	orthorhombic	12	0.00	0.37	1.00	0.20	UV	
93	mp-2624	AlSb	True	216	cubic	2	0.00	1.26	1.92	0.09	UV	
94	mp-267	Te ₂ Ru	True	58	orthorhombic	6	0.00	0.32	0.93	0.10	UV	
95	mp-2693	SnSe	True	225	cubic	2	0.00	0.82	1.00	0.09	UV+VIS	
96	mp-2695	Sb ₂ Os	True	58	orthorhombic	6	0.00	0.40	0.90	0.11	UV	
97	mp-27438	TlBiTe ₂	True	166	trigonal	4	0.00	0.41	0.79	0.12	UV+VIS	
98	mp-27736	Re ₂ P ₅	True	2	triclinic	28	0.00	0.49	0.99	0.12	UV	
99	mp-27809	Ba(GeP) ₂	True	105	tetragonal	20	0.00	0.52	1.37	0.26	UV	
100	mp-27910	Bi ₂ Te ₂ S	True	166	trigonal	5	0.00	0.63	1.22	0.07	UV	
101	mp-27948	Ge(BiTe ₂) ₂	True	166	trigonal	7	0.01	0.51	0.80	0.10	UV+VIS	
102	mp-28029	TcP ₃	True	62	orthorhombic	16	0.00	0.39	1.02	0.09	UV	
103	mp-28266	P ₂ Pd	True	15	monoclinic	6	0.00	0.41	0.81	0.09	UV+VIS	
104	mp-28413	Na ₂ (GaSb) ₃	True	62	orthorhombic	32	0.00	0.32	1.17	0.13	UV+VIS	
105	mp-28921	Tl ₅ Se ₂ Br	True	140	tetragonal	16	0.00	0.62	1.36	0.17	VIS	

106	mp-29157	Si ₃ P ₂ Pt	True	1	triclinic	12	0.00	0.58	1.02	0.06	UV		
107	mp-29208	Ca(MgBi) ₂	True	164	trigonal	5	0.00	0.39	1.03	0.11	UV+VIS		
108	mp-29209	Ba(MgBi) ₂	True	164	trigonal	5	0.00	0.40	1.00	0.27	UV+VIS		
109	mp-29238	TlAgSe	True	62	orthorhombic	12	0.00	0.67	1.11	0.13	UV		
110	mp-29397	SiTePt	True	61	orthorhombic	24	0.00	0.34	0.74	0.11	UV		
111	mp-29643	CuAsSe ₂	True	160	trigonal	4	0.11	0.52	1.05	0.10	UV		
112	mp-29644	GeBi ₄ Te ₇	True	164	trigonal	12	0.00	0.72	0.92	0.07	UV		
113	mp-29662	TlBiSe ₂	True	166	trigonal	4	0.00	0.35	1.12	0.14	UV+VIS		
114	mp-30252	Na ₁₀ CaSn ₁₂	True	217	cubic	23	0.00	0.76	1.25	0.20	UV+VIS		
115	mp-30253	Na ₁₀ SrSn ₁₂	True	217	cubic	23	0.00	0.77	1.24	0.20	UV+VIS		
116	mp-30275	BaP ₈	True	2	triclinic	18	0.01	0.92	1.53	0.39	VIS		
117	mp-304	Ca ₂ Ge	True	62	orthorhombic	12	0.00	0.31	0.94	0.12	UV		
118	mp-30456	SrLiBi	True	62	orthorhombic	12	0.00	0.55	1.19	0.30	UV		
119	mp-3078	CdSiAs ₂	True	122	tetragonal	8	0.00	0.89	1.27	0.11	UV+VIS		
120	mp-30847	TiSnPt	True	216	cubic	3	0.00	0.83	1.08	0.17	UV+VIS		
121	mp-30949	Sr ₃ SiO	True	62	orthorhombic	20	0.00	0.66	1.39	0.16	UV		
122	mp-30950	Sr ₃ GeO	True	62	orthorhombic	20	0.00	0.58	1.13	0.97	UV+VIS		
123	mp-31149	Ca ₃ BiN	True	221	cubic	5	0.00	0.78	1.30	0.96	UV+VIS		
124	mp-31243	BaAs ₂	True	7	monoclinic	18	0.00	0.43	1.00	0.13	VIS		
125	mp-31451	ZrCoBi	True	216	cubic	3	0.00	1.01	1.35	0.16	UV+VIS		
126	mp-31454	TaSbRu	True	216	cubic	3	0.00	0.90	1.18	0.11	UV+VIS		
127	mp-31493	RbGa ₃	True	119	tetragonal	12	0.00	0.62	0.83	0.12	UV+VIS		
128	mp-31507	Sb ₂ Te ₄ Pb	True	166	trigonal	7	0.01	0.30	0.80	0.09	UV		
129	mp-33526	LiBiS ₂	False	141	tetragonal	8	0.01	0.81	1.27	0.23	UV		
130	mp-34126	Cu ₄ SnP ₁₀	True	160	trigonal	15	0.00	0.37	1.08	0.07	UV+VIS		
131	mp-34202	Bi ₂ Te ₃	True	166	trigonal	5	0.00	0.39	0.79	0.10	UV+VIS		
132	mp-3432	ScNiSb	True	216	cubic	3	0.00	0.32	0.94	0.17	UV+VIS		
133	mp-358	HgTe	True	152	trigonal	6	0.06	0.49	1.11	0.11	UV		
134	mp-35907	La ₅ TlTe ₈	False	82	tetragonal	14	0.00	1.08	1.82	0.14	UV		

135	mp-3595	ZnSiAs ₂	True	122	tetragonal	8	0.00	0.89	1.52	0.13	UV+VIS
136	mp-3668	CdGeP ₂	True	122	tetragonal	8	0.00	0.65	1.37	0.10	UV+VIS
137	mp-3716	TbNiSb	True	216	cubic	3	0.00	0.38	0.99	0.30	UV+VIS
138	mp-3785	TlGaTe ₂	True	140	tetragonal	8	0.00	0.58	1.21	0.16	UV
139	mp-3839	GaCuTe ₂	True	122	tetragonal	8	0.00	0.55	1.08	0.13	UV+VIS
140	mp-38605	Sn(BiTe ₂) ₂	False	166	trigonal	7	0.00	0.46	0.80	0.10	UV+VIS
141	mp-4008	ZnGeAs ₂	True	122	tetragonal	8	0.00	0.57	0.99	0.09	UV+VIS
142	mp-4025	TmNiSb	True	216	cubic	3	0.00	0.31	0.99	0.30	UV+VIS
143	mp-4174	HoNiSb	True	216	cubic	3	0.00	0.35	0.99	0.20	UV+VIS
144	mp-4510	DyNiSb	True	216	cubic	3	0.00	0.37	0.99	0.20	UV+VIS
145	mp-4524	ZnGeP ₂	True	122	tetragonal	8	0.00	1.20	2.02	0.11	UV
146	mp-4627	CoAsS	True	29	orthorhombic	12	0.00	0.86	0.96	0.11	UV+VIS
147	mp-484	Te ₃ As ₂	True	12	monoclinic	10	0.00	0.58	0.73	0.08	UV+VIS
148	mp-4840	GaCuSe ₂	True	122	tetragonal	8	0.00	0.46	0.79	0.07	UV+VIS
149	mp-487	MnP ₄	True	2	triclinic	10	0.00	0.45	1.16	0.08	VIS
150	mp-505297	NbSbRu	True	216	cubic	3	0.00	0.62	0.82	0.11	UV+VIS
151	mp-5213	CdSnP ₂	True	122	tetragonal	8	0.00	0.67	1.32	0.10	UV+VIS
152	mp-540696	Ca ₃ (SiAs ₂) ₂	True	14	monoclinic	36	0.00	0.82	1.55	0.33	UV
153	mp-541032	Te ₇ As ₅ I	True	8	monoclinic	13	0.03	0.56	0.93	0.09	UV
154	mp-541837	Bi ₂ Se ₃	True	166	trigonal	5	0.00	0.54	1.12	0.10	UV
155	mp-541844	Ge ₅ Ir ₄	True	116	tetragonal	36	0.00	0.41	0.72	0.11	UV+VIS
156	mp-557619	Bi ₁₄ Te ₁₃ S ₈	False	148	trigonal	35	0.02	0.55	1.10	0.09	UV+VIS
157	mp-567313	Te	True	154	trigonal	3	0.00	0.58	0.80	0.08	UV
158	mp-567636	VFeSb	True	216	cubic	3	0.00	0.53	0.73	0.11	UV+VIS
159	mp-567817	HfGeTe ₄	True	36	orthorhombic	12	0.00	0.39	1.09	0.11	UV
160	mp-568177	CaAlSiH	True	156	trigonal	4	0.01	0.41	1.03	0.16	UV
161	mp-568890	Tl ₂ CdGeTe ₄	True	121	tetragonal	8	0.00	0.43	1.13	0.13	UV
162	mp-569267	Ca ₂ CdSb ₂	True	62	orthorhombic	20	0.00	0.65	1.18	0.14	UV+VIS
163	mp-569501	LiCaBi	True	62	orthorhombic	12	0.00	0.58	1.25	0.22	UV

164	mp-569522	MnP ₄	True	2	triclinic	30	0.00	0.52	1.18	0.14	VIS
165	mp-569571	Nb ₃ Sb ₂ Te ₅	True	217	cubic	20	0.00	1.09	1.76	0.09	UV+VIS
166	mp-569815	Si ₂ Ru	True	64	orthorhombic	24	0.00	0.48	0.92	0.09	UV
167	mp-570213	LiMgBi	True	216	cubic	3	0.00	0.36	1.02	0.24	UV
168	mp-570485	SrAlSiH	True	156	trigonal	4	0.00	0.62	1.39	0.28	UV
169	mp-570744	Si ₃ As ₄	True	215	cubic	7	0.06	0.41	0.92	0.11	UV+VIS
170	mp-570844	Ga ₃ Os	True	136	tetragonal	16	0.00	0.97	1.24	0.11	VIS
171	mp-574169	TiGeTe ₆	True	12	monoclinic	16	0.00	0.43	0.73	0.07	UV
172	mp-5874	TlAgTe	True	62	orthorhombic	12	0.00	0.72	1.18	0.12	UV
173	mp-5881	CoSbS	True	61	orthorhombic	24	0.00	0.41	0.80	0.13	UV
174	mp-5967	TiCoSb	True	216	cubic	3	0.00	1.07	1.51	0.17	UV+VIS
175	mp-640885	Er(CuTe) ₃	True	31	orthorhombic	28	0.00	0.38	1.25	0.11	VIS
176	mp-640889	Tm(CuTe) ₃	True	31	orthorhombic	28	0.00	0.40	1.20	0.13	UV
177	mp-643683	Ba ₅ (Ga ₃ H) ₂	True	158	trigonal	26	0.00	0.39	1.11	0.25	UV+VIS
178	mp-650442	Tl ₃ AgTe ₂	True	14	monoclinic	24	0.00	0.54	1.12	0.13	UV+VIS
179	mp-672204	Ga ₃ Ru	True	136	tetragonal	16	0.00	0.72	1.19	0.17	UV+VIS
180	mp-675531	NaBiS ₂	False	43	orthorhombic	8	0.02	1.43	1.94	0.27	UV
181	mp-675989	Tl ₉ BiSe ₆	False	87	tetragonal	16	0.00	0.83	1.21	0.19	VIS
182	mp-676274	Tl ₉ SbSe ₆	False	87	tetragonal	16	0.00	0.76	1.16	0.19	VIS
183	mp-691	SnSe	True	62	orthorhombic	8	0.00	1.33	1.96	0.09	UV
184	mp-7173	ScSbPt	True	216	cubic	3	0.00	0.66	1.08	0.12	UV+VIS
185	mp-730	P ₂ Pt	True	205	cubic	12	0.00	1.13	1.76	0.10	UV+VIS
186	mp-7372	ZnAs	True	61	orthorhombic	16	0.02	0.45	1.06	0.11	UV+VIS
187	mp-7438	KZnSb	True	194	hexagonal	6	0.00	0.44	0.80	0.14	UV+VIS
188	mp-755363	Ba ₃ SiO	False	62	orthorhombic	20	0.00	0.51	0.98	0.14	UV+VIS
189	mp-756006	LiSbS	False	1	triclinic	6	0.02	0.37	0.81	0.24	UV
190	mp-7623	MgAs ₄	True	92	tetragonal	20	0.00	0.84	1.23	0.14	VIS
191	mp-766	As ₂ Ru	True	58	orthorhombic	6	0.00	0.56	0.75	0.11	UV
192	mp-7936	LiNbS ₂	True	194	hexagonal	8	0.00	0.81	1.50	0.18	UV

193	mp-7944	NaSb	True	14	monoclinic	16	0.00	0.65	1.22	0.12	VIS
194	mp-7956	Na ₃ Sb	True	194	hexagonal	8	0.00	0.39	1.13	0.20	VIS
195	mp-8195	Ba ₂ GeAs ₂	True	14	monoclinic	20	0.00	0.56	1.37	0.35	VIS
196	mp-8200	AgP ₂	True	14	monoclinic	12	0.02	0.66	1.24	0.09	UV
197	mp-8311	Si ₃ NiP ₄	True	121	tetragonal	8	0.00	0.34	1.03	0.13	UV
198	mp-8507	Sr(P ₃ Pt ₂) ₂	True	15	monoclinic	22	0.00	0.45	1.00	0.26	UV
199	mp-8612	Sb ₂ TeSe ₂	True	160	trigonal	5	0.00	0.48	0.82	0.10	UV
200	mp-861937	Ba ₂ AsAu	False	225	cubic	4	0.00	0.61	1.14	0.15	UV+VIS
201	mp-862631	Ba ₂ SbAu	False	225	cubic	4	0.00	0.62	1.17	0.15	UV+VIS
202	mp-862947	Ba ₂ BiAu	False	225	cubic	4	0.00	0.43	0.76	0.28	UV+VIS
203	mp-8630	SbIrS	True	198	cubic	12	0.00	1.51	1.92	0.12	VIS
204	mp-865280	NbAlFe ₂	False	225	cubic	4	0.00	0.33	0.76	0.24	UV+VIS
205	mp-866062	HfSiRu ₂	False	225	cubic	4	0.00	0.33	0.76	0.20	UV+VIS
206	mp-866132	Na ₂ TlSb	False	225	cubic	4	0.00	0.48	0.79	0.15	UV+VIS
207	mp-866141	TiFe ₂ Si	False	225	cubic	4	0.00	0.40	0.77	0.29	UV+VIS
208	mp-867113	Ca ₂ AsAu	False	225	cubic	4	0.02	0.32	0.70	0.16	UV+VIS
209	mp-867168	Sr ₂ SbAu	False	225	cubic	4	0.00	0.60	0.81	0.15	UV+VIS
210	mp-867169	Sr ₂ SnHg	False	225	cubic	4	0.00	0.36	0.76	0.18	UV+VIS
211	mp-867192	Sr ₂ AsAu	False	225	cubic	4	0.01	0.56	0.80	0.17	UV+VIS
212	mp-867193	Sr ₂ BiAu	False	225	cubic	4	0.00	0.45	0.71	0.28	UV+VIS
213	mp-867249	TaAlFe ₂	False	225	cubic	4	0.00	0.32	0.77	0.30	UV+VIS
214	mp-867339	CsK ₂ Bi	False	225	cubic	4	0.00	0.57	1.21	0.31	VIS
215	mp-9198	Fe(SiP) ₄	True	1	triclinic	9	0.00	1.10	1.91	0.15	VIS
216	mp-924129	ZrNiSn	False	216	cubic	3	0.00	0.52	0.82	0.17	UV+VIS
217	mp-9270	SbIrS	True	29	orthorhombic	12	0.00	1.32	1.86	0.11	UV+VIS
218	mp-938	GeTe	True	160	trigonal	2	0.00	0.82	0.94	0.05	UV
219	mp-9383	SrHfN ₂	True	166	trigonal	4	0.03	0.35	1.04	0.19	VIS
220	mp-9437	NbFeSb	True	216	cubic	3	0.00	0.74	0.87	0.09	UV+VIS
221	mp-9548	GeAs	True	12	monoclinic	12	0.00	0.64	1.16	0.11	UV

222	mp-9570	Ca(CdP) ₂	True	164	trigonal	5	0.00	0.82	1.47	0.16	UV
223	mp-9571	Ca(ZnAs) ₂	True	164	trigonal	5	0.00	0.39	1.07	0.23	UV+VIS
224	mp-9575	LiBeSb	True	186	hexagonal	6	0.04	0.91	1.40	0.19	UV
225	mp-961673	TiFeTe	False	216	cubic	3	0.00	1.12	1.78	0.11	UV+VIS
226	mp-961675	ScNiP	False	216	cubic	3	0.17	0.71	1.06	0.33	VIS
227	mp-961693	ZrInAu	False	216	cubic	3	0.00	0.50	0.76	0.34	UV+VIS
228	mp-961706	TiSiPt	False	216	cubic	3	0.06	0.97	1.21	0.11	UV+VIS
229	mp-961713	ZrSnPt	False	216	cubic	3	0.00	1.01	1.31	0.17	UV+VIS
230	mp-978847	SrGaGeH	True	156	trigonal	4	0.00	0.43	0.76	0.27	UV+VIS
231	mp-978852	SrGaSnH	True	156	trigonal	4	0.00	0.59	0.84	0.27	UV+VIS
232	mp-980057	SrAlGeH	True	156	trigonal	4	0.00	0.70	1.25	0.34	UV
233	mp-984350	Dy ₃ Sb ₄ Au ₃	True	220	cubic	20	0.00	0.57	0.77	0.13	UV+VIS
234	mp-9919	LiZnSb	True	186	hexagonal	6	0.00	0.42	1.14	0.11	UV

REFERENCES

- [1] Ward L, Dunn A, Faghaninia A, et al. Matminer: An open source toolkit for materials data mining. *Computational Materials Science* 2018; 152: 60–69.
- [2] Ward L, Agrawal A, Choudhary A, Wolverton C. A general-purpose machine learning framework for predicting properties of inorganic materials. *npj Computational Materials* 2016; 2(1): 1–7.