Electronic Supplementary Information

Machine Learning Assisted Approximation of Descriptors Binding Energy (CO and OH) on Cu-based Bimetallic alloys

Pallavi Dandekar[a#], Aditya Singh Ambesh[a#], Tuhin Suvra Khan[b] and Shelaka Gupta[a*]

[a]MMEC Lab, Department of Chemical Engineering, Indian Institute of Technology Hyderabad, Kandi, Sangareddy-502285, Telangana (India)

[b]Climate Change and Data Science Division, CSIR – Indian Institute of Petroleum, Dehradun -248005, India

[*]Corresponding Author: shelaka@che.iith.ac.in

[#]These authors contributed equally to this work

This file includes: Figure S1 and Tables S1 to S6

Table S1: List of features included in the ML models

Sr. No.	Features	Abbreviation for main metal "A"	Abbreviation for guest metal "B"
1.	Atomic number	M-At No.	At No.
2.	Atomic weight	M-At wt.	At wt.
3.	Density	M-Density	Density
4.	Melting point	M-M. P	M. P
5.	Boiling point	M-B. P	B. P
6.	Enthalpy of fusion	M-Enth.fus	Enth.fus
7.	Enthalpy of atomization	M-Enth.atom	Enth.atom
8.	Enthalpy of vaporization	M-Enth.vap	Enth.vap
9.	Specific heat capacity	M-Sp.ht Cap	Sp.ht Cap
10.	Electronegativity	M-Elec.-ve	Elec.-ve
11.	Surface energy	M-Surface.E	Surface.E
12.	1st ionization energy	M-1st Ion E	1st Ion E
13.	Covalent radii	M-cova. radii	cova. radii
14.	Atomic radii	M-At. radii	At. radii
15.	Group	M-Group	Group

16.	Period	M-Period	Period
17.	Work function	M-Work F.	Work F.
18.	Electron affinity	M-Elec. Aff.	Elec. Aff.

Table S2: Results of Principal Component Analysis for CO Binding Energy Prediction by xGBR model

Sr. No.	**Model Configuration**	**Train RMSE Mean (min, max)**	**Test RMSE Mean (min, max)**
1.	All 36 features	0.004 (0.001, 0.014)	0.091 (0.086, 0.097)
2.	30 features (6 highly correlated removed)	0.004 (0.001, 0.015)	0.091 (0.085, 0.098)
3.	30 features (6 removed features reintroduced as 1 component using PCA)	0.013 (0.002, 0.028)	0.098 (0.087, 0.109)
4.	Top 10 features only (from first model)	0.007 (0.001, 0.017)	0.096 (0.087, 0.105)
5.	Top 20 features (from first model)	0.009 (0.001, 0.041)	0.093 (0.084, 0.097)
6.	Top 10 + PCA (remaining 26 features reintroduced as 8 components using PCA	0.002 (0.001, 0.012)	0.134 (0.119, 0.149)
7.	Top 20 + PCA (remaining 26 features reintroduced as 5 components using PCA	0.009 (0.001, 0.019)	0.107 (0.096, 0.111)

Table S3: Results of Principal Component Analysis for OH Binding Energy Prediction by xGBR model

Sr. No.	**Model Configuration**	**Train RMSE Mean (min, max)**	**Test RMSE Mean (min, max)**
1.	All 36 features	0.010 (0.001, 0.028)	0.196 (0.180, 0.216)

2.	30 features (6 highly correlated removed)	0.011 (0.001, 0.034)	0.198 (0.183, 0.213)
3.	30 features (6 removed features reintroduced as 2 components using PCA)	0.012 (0.001, 0.032)	0.213 (0.196, 0.234)
4.	Top 10 features only (from first model)	0.028 (0.002, 0.065)	0.223 (0.217, 0.233
5.	Top 20 features (from first model)	0.009 (0.002, 0.033)	0.202 (0.184, 0.218)
6.	Top 10 + PCA (remaining 26 features reintroduced as 7 components using PCA	0.002 (0.001, 0.017)	0.202 (0.173, 0.241)
7.	Top 20 + PCA (remaining 16 features reintroduced as 4 components using PCA	0.004 (0.001, 0.019)	0.218 (0.195, 0.243)

Table S4: Results of Principal Component Analysis for CO and OH binding energy prediction by ETR and RFR models.

Sr. No.	**Model Configuration**	**CO Binding Energy**		**OH Binding Energy**	
		Test RMSE Mean (min, max)	**Test RMSE Mean (min, max)**	**Test RMSE Mean (min, max)**	**Test RMSE Mean (min, max)**
		ETR Model	**RFR Model**	**ETR Model**	**RFR Model**
1.	All 36 features	0.0913	0.0896	0.2350	0.2820
2.	32 features (4 highly correlated removed)	0.0910	0.089		
3.	Top 10 features only (from first model)	0.1097	0.0909	0.2355	0.2958
4.	Top 20 features (from first model)	0.0927	0.0913	0.2481	0.2820
5.	Top 10 + PCA (remaining 26 features reintroduced as 7 components	0.1021	0.1201	0.2651	0.2735

	using PCA				
6.	Top 20 + PCA (remaining 16 features reintroduced as 4 components using PCA	0.0924	0.1003	0.2334	0.3041

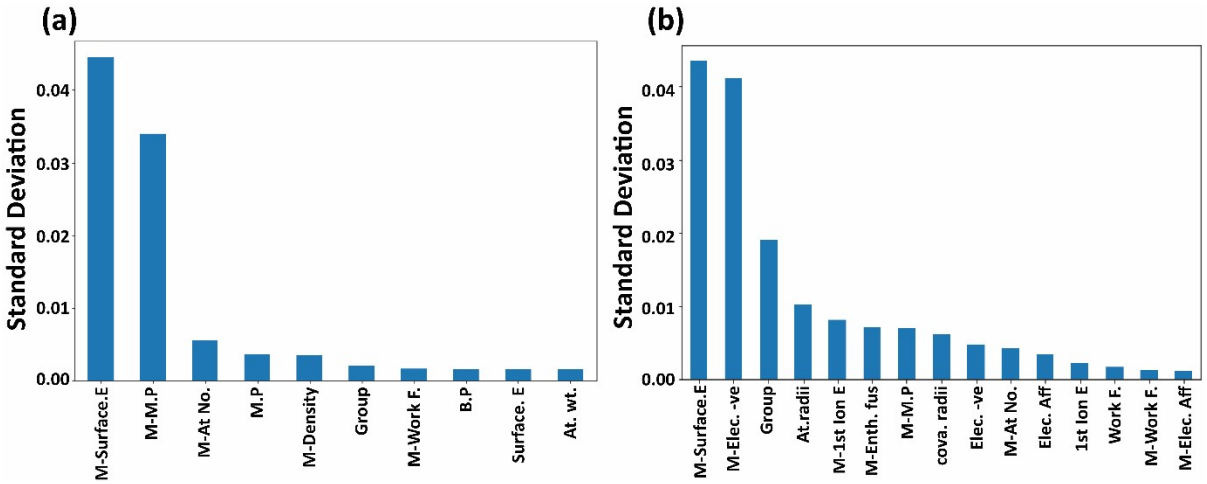

Figure S1: Standard deviation for the feature importance for predicting: (a) CO, (b) OH binding energy for the best model.

Table S5: ML Predicted CO and OH binding energies for unseen data by xGBR model for (111)-terminated Cu_3M bimetallic alloys, where Cu is the main metal and M is the guest metal.

Sr. No.	Cu-based Bimetallic	ML Predicted CO Binding Energy (eV)	ML Predicted OH Binding Energy (eV)
1.	Cu_3Sc	-0.57	-0.58
2.	Cu_3Ti	-0.35	-0.40
3.	Cu_3V	-0.27	-0.06
4.	Cu_3Cr	-0.40	0.05
5.	Cu_3Mn	-0.39	-0.01
6.	Cu_3Co	-0.34	0.38
7.	Cu_3Ni	-0.41	0.62
8.	Cu_3Zn	-0.49	0.32
9.	Cu_3Ga	-0.58	0.20
10.	Cu_3Y	-0.45	-0.78

11.	Cu_3Zr	-0.32	-0.86
12.	Cu_3Nb	-0.22	-0.22
13.	Cu_3Mo	-0.29	-0.12
14.	Cu_3Ru	-0.33	0.34
15.	Cu_3Rh	-0.36	0.57
16.	Cu_3Pd	-0.43	0.43
17.	Cu_3Ag	-0.58	0.54
18.	Cu_3Sn	-0.43	0.16
19.	Cu_3Ta	-0.15	-0.61
20.	Cu_3Re	-0.34	0.16
21.	Cu_3Os	-0.35	0.37
22.	Cu_3Ir	-0.37	0.56
23.	Cu_3Pt	-0.43	0.59
24.	Cu_3Au	-0.56	0.57

Table S6: Predictions for Carbon and Oxygen binding energies on AA-terminated A_3B type 211 surfaces.

S.No.	**Adsorbate**	**ML Model (Test/Train split)**	**Train RMSE (eV)**	**Test RMSE (eV)**	**Source**
1.	Carbon	GBR (20/80)	0.0003	0.340	*J. Mater. Chem. A*, 2020,8, 107-123
		xGBR (20/80)	0.0732	0.411	This Work
2.	Oxygen	GBR (15/85)	0.0003	0.310	*J. Mater. Chem. A*, 2020,8, 107-123
		xGBR (15/85)	0.0035	0.289	This Work