Electronic Supplementary Material (ESI) for Physical Chemistry Chemical Physics.

Supporting Information

Influence Factors of CO Adsorption on C_2N-Supported Dual-Atom Catalysts Unveiled by Machine Learning and Twofold Feature Engineering

Shikai Chang[a], Hongshuai Wang[a,b], Yujin Ji[a*], and Youyong Li[a,c*]

a. Institute of Functional Nano & Soft Materials (FUNSOM), Jiangsu Key Laboratory for Carbon-Based Functional Materials & Devices, Soochow University, Suzhou, Jiangsu 215123, China.

b. DP Technology, Beijing, 100080, China.

c. Macao Institute of Materials Science and Engineering, Macau University of Science and Technology, Taipa, Macau SAR 999078, China.

Note S1. The training process of the machine learning model

The $\Delta G_{ads}(CO)$ calculated by DFT and the characteristics of two transition metal atoms form a dataset, which is divided into training and testing sets a ratio of 9:1. Tenfold cross-validation is employed to update the best hyperparameters via a grid search (**Table S5**). To compare the performances of different ML algorithms, we repeated the training 100 times and took the average of the results. When using LightGBM to train the model in **Figure S4**, as the number of iterations increases, the errors on the training and testing sets decrease and tend to stabilize. After about 20 rounds, the error on the test set tends to stabilize, but the error on the training set is still decreasing. To prevent overfitting of the model, we stopped training where the test set error tends to stabilize.

Note S2. The datasets selected for feature engineering

When building two ML models using LightGBM, we did not separate the training set and the test set. Firstly, our dataset is relatively small, and after splitting out a portion of the test set, the model performs poorly. Secondly, this work focuses more on using machine learning methods and secondary feature engineering to screen atomic properties related to $\Delta G_{ads}(CO)$. Therefore, training with all available data can yield better results.

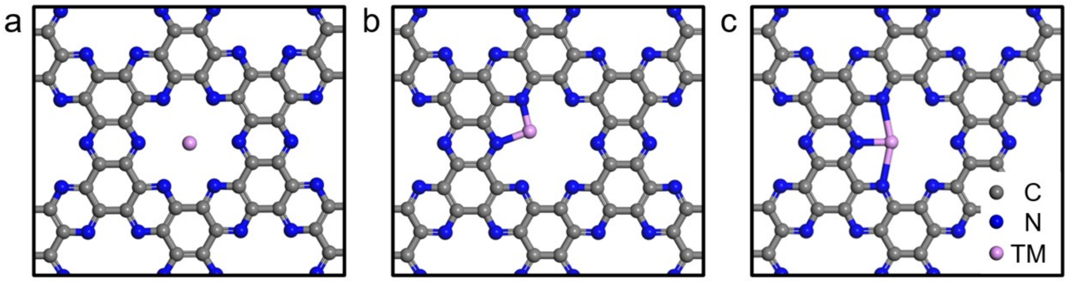

Figure S1. Three potential adsorption sites for single TM on C_2N, namely (a) site i: the center of the hole, (b) site ii: the position bonded with two N atoms; (c) site iii: the position bonded with three N atoms.

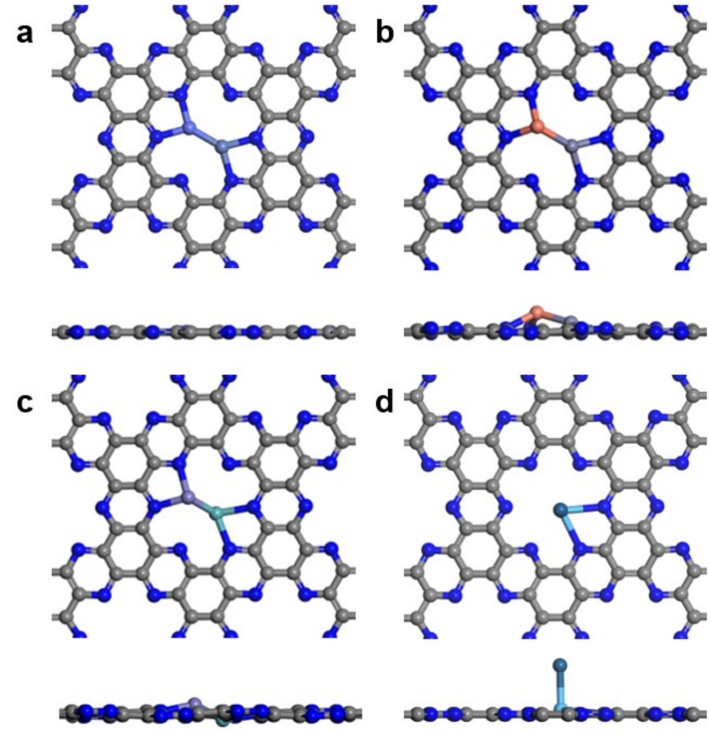

Figure S2. Optimized geometric configurations of TM1TM2@C_2N DACs: (a) CoNi@C_2N; (b) CuZn@C_2N; (c) FeHf@C_2N; (d)HfPt@C_2N.

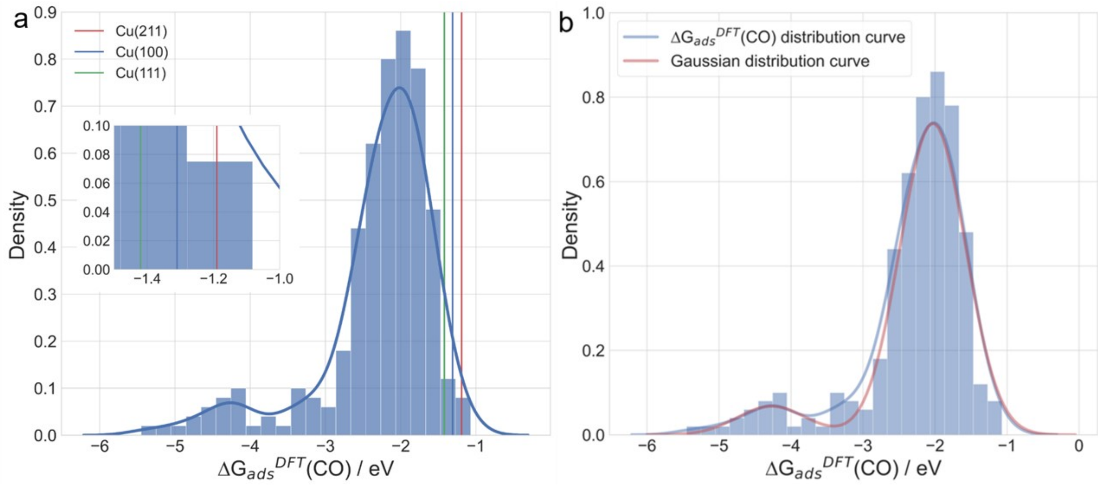

Figure S3. (a) The distribution of ΔG_{ads}(CO) on TM1TM2@C_2N; (b) A Gaussian distribution curve is used to fit the distribution curve of ΔG_{ads}(CO).

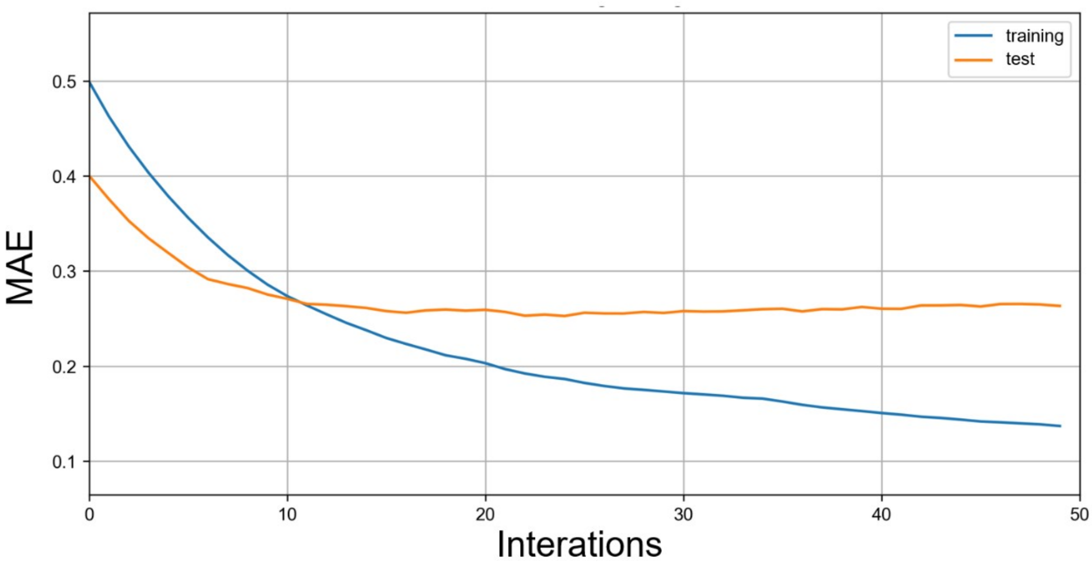

Figure S4. One of the 100 test results of the ML model built by LightGBM. The blue line represents the training set data, and the yellow line represents the testing set data.

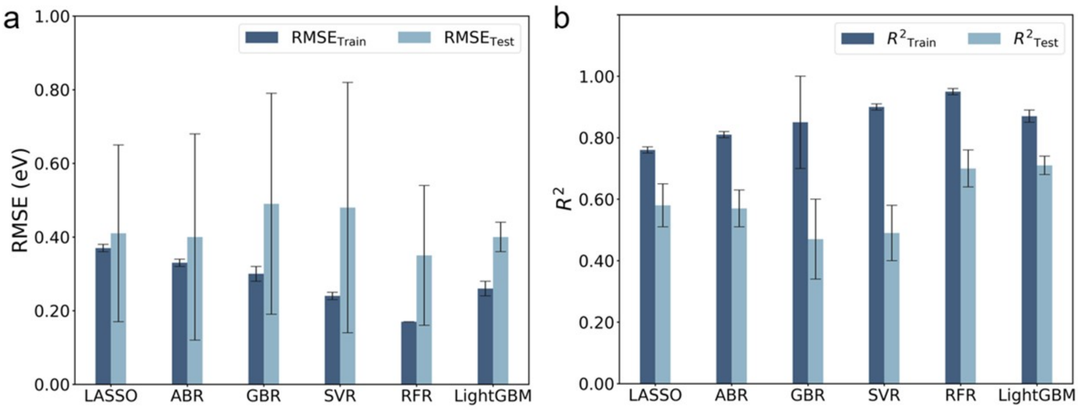

Figure S5. Performance of machine learning models using six algorithms on training and test sets.

Table S1. The energies calculated of C_2N anchoring TM atoms at three possible sites. The most stable configurations are highlighted in red typeface. i, ii, iii represents that after the structural optimization, the anchoring TM atoms move to another sites

	C_2N@TM_i	C_2N@TM_ii	C_2N@TM_iii		C_2N@TM_i	C_2N@TM_ii	C_2N@TM_iii
Sc	ii	-158.89	-158.90	**Tc**	ii	-157.05	ii
Ti	ii	-158.30	-158.27	**Ru**	ii	-156.50	ii
V	ii	-157.41	ii	**Rh**	ii	-155.40	ii
Cr	ii	-156.79	ii	**Pd**	ii	-153.46	-153.43
Mn	ii	-156.37	ii	**Ag**	-152.12	i	i
Fe	ii	-155.90	ii	**Cd**	-150.83	i	i
Co	ii	-154.90	ii	**Hf**	iii	iii	-160.55
Ni	ii	-153.77	ii	**Ta**	ii	-159.93	ii
Cu	ii	-152.63	-152.58	**W**	ii	-158.82	-158.78
Zn	ii	-150.61	ii	**Re**	ii	-157.86	ii
Y	-160.32	i	i	**Os**	ii	-157.24	ii
Zr	ii	-159.80	-159.79	**Ir**	ii	-155.87	ii
Nb	ii	-158.93	-158.92	**Pt**	ii	-153.54	-153.46
Mo	ii	-157.75	-157.71	**Au**	-151.27	i	i

Table S2. The binding energy calculated of 248 TM1TM2@C_2N DACs through the eq 1

TM1TM2	E_f **(eV)**	**TM1TM2**	E_f **(eV)**	**TM1TM2**	E_f **(eV)**
ScV	-8.35	**MnPd**	-2.01	**ZrRu**	-11.01
ScCr	-6.43	**MnHf**	-7.44	**ZrRh**	-10.93
ScMn	-6.48	**MnTa**	-7.88	**ZrHf**	-10.97
ScFe	-7.71	**MnW**	-6.52	**ZrTa**	-11.35
ScCo	-8.54	**MnRe**	-5.62	**ZrW**	-10.41
ScNi	-8.18	**MnOs**	-6.04	**ZrRe**	-9.95
ScRu	-9.90	**MnIr**	-5.55	**ZrOs**	-10.77
ScRh	-9.61	**MnPt**	-3.84	**ZrIr**	-10.57
ScHf	-10.02	**FeFe**	-4.45	**NbNb**	-8.93
ScTa	-10.13	**FeCo**	-5.04	**NbMo**	-7.76
ScRe	-8.58	**FeNi**	-5.21	**NbRu**	-10.34
ScOs	-9.58	**FeCu**	-3.85	**NbRh**	-9.50
ScIr	-9.59	**FeZn**	-2.53	**NbPd**	-7.18
TiTi	-10.04	**FeZr**	-8.89	**NbHf**	-10.45
TiV	-8.96	**FeNb**	-8.34	**NbTa**	-11.00
TiCr	-7.10	**FeMo**	-6.39	**NbW**	-10.01
TiMn	-6.55	**FeRu**	-6.74	**NbRe**	-9.53
TiFe	-8.53	**FeRh**	-5.43	**NbOs**	-10.13
TiCo	-9.17	**FePd**	-3.38	**NbIr**	-9.67
TiNi	-8.48	**FeHf**	-8.55	**NbPt**	-7.88
TiCu	-6.91	**FeTa**	-8.91	**MoMo**	-7.72
TiZr	-10.7	**FeW**	-7.48	**MoRu**	-8.36
TiNb	-10.21	**FeRe**	-6.51	**MoRh**	-7.23
TiMo	-8.80	**FeOs**	-6.84	**MoPd**	-4.76
TiRu	-10.45	**FeIr**	-6.49	**MoTa**	-9.72

TiRh	-9.86	**FePt**	-4.89	**MoW**	-8.59
TiPd	-7.42	**CoCo**	-5.89	**MoRe**	-7.76
TiHf	-10.57	**CoNi**	-6.06	**MoOs**	-8.17
TiTa	-11.00	**CoCu**	-4.54	**MoIr**	-7.63
TiW	-9.64	**CoZn**	-2.96	**MoPt**	-5.75
TiRe	-9.23	**CoZr**	-9.51	**RuRu**	-8.4
TiOs	-10.10	**CoNb**	-8.74	**RuRh**	-7.74
TiIr	-9.87	**CoMo**	-6.65	**RuPd**	-5.44
VV	-7.82	**CoRu**	-7.19	**RuHf**	-10.77
VCr	-6.15	**CoRh**	-6.64	**RuTa**	-11.01
VMn	-6.19	**CoPd**	-4.35	**RuW**	-9.46
VFe	-7.18	**CoHf**	-9.21	**RuRe**	-8.29
VCo	-7.44	**CoTa**	-9.51	**RuOs**	-8.55
VNi	-6.78	**CoW**	-7.97	**RuIr**	-8.2
VCu	-4.71	**CoRe**	-6.71	**RuPt**	-6.62
VZn	-3.74	**CoOs**	-7.32	**RhRh**	-7.16
VZr	-9.57	**CoIr**	-7.05	**RhPd**	-4.93
VNb	-9.25	**CoPt**	-5.66	**RhHf**	-10.36
VMo	-7.85	**NiNi**	-6.08	**RhTa**	-10.31
VRu	-9.15	**NiCu**	-4.51	**RhW**	-8.61
VRh	-8.25	**NiZn**	-3.00	**RhRe**	-7.25
VPd	-5.41	**NiZr**	-8.46	**RhOs**	-7.76
VHf	-9.42	**NiNb**	-8.35	**RhIr**	-7.52
VTa	-9.89	**NiMo**	-6.26	**RhPt**	-5.91
VW	-8.69	**NiRu**	-6.79	**PdPd**	-2.55
VRe	-8.18	**NiRh**	-6.56	**PdTa**	-8.16
VOs	-8.82	**NiPd**	-4.46	**PdW**	-6.27
VPt	-6.78	**NiHf**	-8.22	**PdRe**	-4.92

CrCr	-4.32	**NiTa**	-9.34	**PdOs**	-5.6
CrMn	-3.98	**NiW**	-7.69	**PdIr**	-5.39
CrFe	-4.78	**NiRe**	-6.44	**HfHf**	-10.81
CrCo	-4.67	**NiOs**	-6.85	**HfTa**	-11.22
CrNi	-4.11	**NiIr**	-6.89	**HfW**	-9.86
CrCu	-2.72	**NiPt**	-5.92	**HfRe**	-9.74
CrZr	-7.81	**CuCu**	-3.38	**HfOs**	-10.61
CrNb	-7.49	**CuZn**	-1.99	**HfIr**	-10.38
CrMo	-6.01	**CuNb**	-6.88	**TaTa**	-11.78
CrRu	-6.7	**CuMo**	-4.65	**TaW**	-10.48
CrRh	-5.66	**CuRu**	-5.97	**TaRe**	-10.23
CrPd	-2.8	**CuRh**	-5.23	**TaOs**	-10.96
CrHf	-7.55	**CuPd**	-3.12	**TaIr**	-10.6
CrTa	-7.95	**CuTa**	-7.9	**TaPt**	-8.98
CrW	-6.82	**CuW**	-5.96	**WW**	-9.54
CrRe	-6.06	**CuRe**	-4.68	**WRe**	-8.86
CrOs	-6.58	**CuOs**	-6.01	**WOs**	-9.43
CrIr	-6.08	**CuIr**	-5.96	**WIr**	-9.06
CrPt	-4.27	**CuPt**	-3.78	**WPt**	-7.66
MnMn	-3.18	**ZnRu**	-4.59	**ReRe**	-8.07
MnFe	-4.05	**ZnRh**	-3.76	**ReOs**	-8.4
MnCo	-4.05	**ZnPd**	-1.9	**ReIr**	-7.69
MnNi	-3.92	**ZnW**	-4.32	**RePt**	-6.27
MnCu	-2.81	**ZnRe**	-3.65	**OsOs**	-8.68
MnZn	-1.42	**ZnOs**	-4.52	**OsIr**	-8.24
MnZr	-7.8	**ZnIr**	-4.2	**OsPt**	-6.67
MnNb	-7.42	**ZnPt**	-2.95	**IrIr**	-7.92
MnMo	-5.65	**ZrZr**	-10.96	**IrPt**	-6.34

MnRu	-6.05	**ZrNb**	-10.61	**PtPt**	-4.42
MnRh	-5.07	**ZrMo**	-9.49		

Table S3. $\Delta G_{ads}(CO)$ of 248 TM1TM2@C_2N DACs calculated through the eq 2, eq 3, eq 4, eq 5 and eq 6 .

TM1TM2	ΔG_{ads}(CO)/eV	TM1TM2	ΔG_{ads}(CO) /eV	TM1TM2	ΔG_{ads}(CO) /eV
ScV	-2.33	**MnPd**	-2.26	**ZrRu**	-2.24
ScCr	-2.17	**MnHf**	-2.25	**ZrRh**	-3.21
ScMn	-2.40	**MnTa**	-2.02	**ZrHf**	-2.63
ScFe	-2.51	**MnW**	-1.80	**ZrTa**	-1.92
ScCo	-2.70	**MnRe**	-2.02	**ZrW**	-2.02
ScNi	-3.04	**MnOs**	-2.01	**ZrRe**	-1.96
ScRu	-2.54	**MnIr**	-1.58	**ZrOs**	-1.36
ScRh	-2.56	**MnPt**	-2.23	**ZrIr**	-3.10
ScHf	-1.84	**FeFe**	-2.24	**NbNb**	-3.63
ScTa	-2.44	**FeCo**	-2.38	**NbMo**	-2.84
ScRe	-2.83	**FeNi**	-2.31	**NbRu**	-1.76
ScOs	-3.09	**FeCu**	-2.26	**NbRh**	-2.09
ScIr	-3.34	**FeZn**	-1.96	**NbPd**	-1.82
TiTi	-2.41	**FeZr**	-2.38	**NbHf**	-1.94
TiV	-2.18	**FeNb**	-1.96	**NbTa**	-2.53
TiCr	-2.02	**FeMo**	-1.83	**NbW**	-2.24
TiMn	-2.75	**FeRu**	-1.92	**NbRe**	-2.39
TiFe	-1.94	**FeRh**	-2.18	**NbOs**	-2.45
TiCo	-1.97	**FePd**	-2.99	**NbIr**	-2.63
TiNi	-2.32	**FeHf**	-2.40	**NbPt**	-1.53
TiCu	-1.92	**FeTa**	-2.07	**MoMo**	-1.58
TiZr	-2.79	**FeW**	-1.98	**MoRu**	-1.55
TiNb	-2.38	**FeRe**	-2.07	**MoRh**	-1.85
TiMo	-1.91	**FeOs**	-2.10	**MoPd**	-1.91
TiRu	-1.99	**FeIr**	-1.78	**MoTa**	-2.14

TiRh	-1.99	**FePt**	-2.39	**MoW**	-1.57
TiPd	-1.93	**CoCo**	-2.35	**MoRe**	-1.47
TiHf	-2.80	**CoNi**	-2.20	**MoOs**	-1.53
TiTa	-2.46	**CoCu**	-2.08	**MoIr**	-1.68
TiW	-2.61	**CoZn**	-2.32	**MoPt**	-1.49
TiRe	-2.14	**CoZr**	-2.64	**RuRu**	-1.54
TiOs	-1.85	**CoNb**	-2.22	**RuRh**	-1.68
TiIr	-2.65	**CoMo**	-2.11	**RuPd**	-1.89
VV	-1.43	**CoRu**	-1.83	**RuHf**	-2.43
VCr	-1.63	**CoRh**	-1.46	**RuTa**	-2.05
VMn	-1.54	**CoPd**	-1.91	**RuW**	-1.82
VFe	-1.63	**CoHf**	-2.65	**RuRe**	-1.65
VCo	-1.91	**CoTa**	-2.26	**RuOs**	-1.69
VNi	-2.21	**CoW**	-2.10	**RuIr**	-1.71
VCu	-2.71	**CoRe**	-2.31	**RuPt**	-2.08
VZn	-1.52	**CoOs**	-1.89	**RhRh**	-1.55
VZr	-1.95	**CoIr**	-1.98	**RhPd**	-1.85
VNb	-2.36	**CoPt**	-2.26	**RhHf**	-2.47
VMo	-1.76	**NiNi**	-2.05	**RhTa**	-2.07
VRu	-1.26	**NiCu**	-2.21	**RhW**	-1.79
VRh	-1.19	**NiZn**	-2.45	**RhRe**	-1.72
VPd	-1.76	**NiZr**	-3.35	**RhOs**	-1.55
VHf	-1.91	**NiNb**	-2.32	**RhIr**	-1.61
VTa	-2.57	**NiMo**	-2.26	**RhPt**	-1.97
VW	-2.32	**NiRu**	-2.30	**PdPd**	-1.66
VRe	-1.89	**NiRh**	-1.97	**PdTa**	-1.72
VOs	-2.15	**NiPd**	-1.71	**PdW**	-1.80
VPt	-2.22	**NiHf**	-3.41	**PdRe**	-1.83

CrCr	-1.34	**NiTa**	-2.23	**PdOs**	-1.82
CrMn	-1.59	**NiW**	-2.29	**PdIr**	-1.71
CrFe	-1.74	**NiRe**	-2.31	**HfHf**	-1.39
CrCo	-2.07	**NiOs**	-2.46	**HfTa**	-1.89
CrNi	-2.63	**NiIr**	-2.18	**HfW**	-2.58
CrCu	-2.47	**NiPt**	-1.86	**HfRe**	-2.19
CrZr	-2.33	**CuCu**	-4.01	**HfOs**	-2.91
CrNb	-1.81	**CuZn**	-3.36	**HfIr**	-3.13
CrMo	-1.47	**CuNb**	-4.18	**TaTa**	-2.63
CrRu	-1.61	**CuMo**	-4.61	**TaW**	-2.51
CrRh	-1.67	**CuRu**	-4.30	**TaRe**	-2.48
CrPd	-1.78	**CuRh**	-4.07	**TaOs**	-2.48
CrHf	-2.45	**CuPd**	-4.20	**TaIr**	-2.57
CrTa	-2.42	**CuTa**	-4.06	**TaPt**	-1.34
CrW	-1.88	**CuW**	-4.56	**WW**	-1.49
CrRe	-1.82	**CuRe**	-4.97	**WRe**	-2.20
CrOs	-1.74	**CuOs**	-4.34	**WOs**	-1.76
CrIr	-1.71	**CuIr**	-3.73	**WIr**	-2.17
CrPt	-2.15	**CuPt**	-5.45	**WPt**	-2.06
MnMn	-2.03	**ZnRu**	-3.69	**ReRe**	-1.68
MnFe	-1.96	**ZnRh**	-3.34	**ReOs**	-1.85
MnCo	-2.33	**ZnPd**	-4.32	**ReIr**	-2.10
MnNi	-2.49	**ZnW**	-4.19	**RePt**	-2.00
MnCu	-2.23	**ZnRe**	-4.28	**OsOs**	-1.99
MnZn	-1.86	**ZnOs**	-4.69	**OsIr**	-1.80
MnZr	-1.97	**ZnIr**	-4.48	**OsPt**	-2.37
MnNb	-1.57	**ZnPt**	-5.19	**IrIr**	-2.02
MnMo	-1.66	**ZrZr**	-2.71	**IrPt**	-2.62

MnRu	-1.80	**ZrNb**	-1.95	**PtPt**	-2.78
MnRh	-1.22	**ZrMo**	-2.06		

The features used in this work are shown in Table S4. The last three features are combined features formed by addition, subtraction, and multiplication to describe the synergistic effect between two transition metal atoms.

Table S4. Complete feature space of individual and composite feature.

Features	Description
$\boldsymbol{Z}$	Atomic number
$\boldsymbol{G}$	Group number
$\boldsymbol{v}$	Valence
$\boldsymbol{P}$	Period number
$\boldsymbol{\phi}$	Unfilled valence orbitals
$\boldsymbol{M}$	Atomic weight
$\boldsymbol{\chi}$	Electronegativity
$\boldsymbol{I}_1$	The first ionization energy
$\boldsymbol{I}_2$	The second ionization energy
$\boldsymbol{d}_{\text{elec}}$	d electron number
$\boldsymbol{EA}$	Electronic Affinity
$\varepsilon_{\text{HOMO}}$	Highest occupied molecular orbital Kohn-Sham eigenvalue
$\varepsilon_{\text{LUMO}}$	Lowest unoccupied molecular orbital Kohn-Sham eigenvalue
$\boldsymbol{r}_{\text{M}}$	Metal atomic radius
$\boldsymbol{r}_{\text{cov}}$	Covalent radius
$\boldsymbol{r}_{\text{s}}$	s orbital radius
$\boldsymbol{r}_{\text{p}}$	p orbital radius
$\boldsymbol{r}_{\text{d}}$	d orbital radius
$\boldsymbol{r}_{\text{v}}$	The last one occupies orbital radius
$\boldsymbol{X}_{\text{TM1}}+\boldsymbol{X}_{\text{TM2}}$, $\bar{\boldsymbol{X}}$	Addition between features
$\lvert\boldsymbol{X}_{\text{TM1}}-\boldsymbol{X}_{\text{TM2}}\rvert$, $\lvert\boldsymbol{X}-\rvert$	Absolute value of difference between features
$\boldsymbol{X}_{\text{TM1}}*\boldsymbol{X}_{\text{TM2}}$, $\boldsymbol{X}*$	Product between features

Table S5 List of hyperparameters search range with five ML algorithms using GridSearchCV.

Algorithm	Hyperparameters search range
LASSO	alpha = [10^(-6 ~ 6), step = 1]
ABR	learning_rate = [01~0.9, step = 0.1] n_estimators = [50 ~ 200, step = 10]
GBR	max_depth = [5 ~ 15, step =1] learning_rate = [01~0.9, step = 0.1] n_estimators = [50 ~ 200, step = 30]
SVR	C = [10^(-6 ~ 6), step = 1] gamma = [10^(-6 ~ 6), step = 1]
RFR	n_estimators = [50~200, step = 10] max_depth = [5 ~ 15, step = 1]

Table S6. List of hyperparameters of LightGBM.

Algorithm	Hyperparameter	Value
LightGBM	learning_rate	0.1
	verbose	-1
	feature_pre_filter	FALSE
	lambda_l	0
	lambda_l2	0
	num_leaves	1
	feature_fraction	0.7
	bagging_fraction	1
	bagging_freq	1
	min_child_samples	20

Table S7. The MAE, RMSE, R^2 and their standard deviations (σ) of six machine learning models tested in this work.

Algorithm	Set	MAE (eV)	σ	RMSE (eV)	σ	R^2	σ
LASSO	Training	0.27	0.01	0.37	0.01	0.76	0.01
	Test	0.31	0.05	0.41	0.07	0.58	0.24
ABR	Training	0.27	0.01	0.01	0.01	0.81	0.01
	Test	0.32	0.05	0.40	0.06	0.57	0.28
GBR	Training	0.21	0.01	0.30	0.15	0.85	0.02
	Test	0.35	0.07	0.49	0.13	0.47	0.30
SVR	Training	0.16	0.01	0.24	0.01	0.90	0.01
	Test	0.33	0.06	0.48	0.09	0.49	0.34
RFR	Training	0.13	0.00	0.17	0.01	0.95	0.01
	Test	0.27	0.05	0.35	0.06	0.71	0.19
LightGBM	Training	0.19	0.01	0.26	0.02	0.87	0.03
	Test	0.30	0.02	0.40	0.03	0.71	0.04

Table S8. The time spent by six machine learning models tested in this work under the same computing resources.

Algorithm	Time (s)
LASSO	1.21
ABR	5.48
GBR	1.24
SVR	4.99
RFR	8.44
LightGBM	1.30

Table S9. Two machine learning models based on LightGBM are used for feature analysis, which used the same hyperparameters to compare the impact of single and combined features on machine learning models and feature analysis.

	Hyperparameters
Training params	learning_rate = 0.1
	metric = ['l1']
	verbose = -1
	feature_pre_filter = False
	lambda_l1 = 0.0
	lambda_l2 = 0.0
	num_leaves = 31
	feature_fraction = 0.86
	bagging_fraction = 1.0
	bagging_freq = 0
	min_child_samples = 5