Supplementary Information

Machine-learning-assisted high-throughput screening for accelerating the discovery of diatomic catalysts in N_2 electroreduction

Keyuan Chen[1#], Yongzhi Wu[1#], Hanqing Li[1], Li Ma[1], Jueyi Ye[1],

Xingkao Zhang[1], Ju Rong[1*], Xiaohua Yu[1,2*], Yongqiang Yang[3*]

[1]Faculty of Materials Science and Engineering, Kunming University of Science and Technology, Kunming, 650093, China

[2]Yunnan Key Laboratory of Integrated Computational Materials Engineering for Advanced Light Metals, Kunming, 650093, China

[3]Shenyang National Laboratory for Materials Science, Institute of Metal Research, Chinese Academy of Sciences, Shenyang, 110016, China

[#]These authors contributed equally to this work.
Corresponding authors. E-mail: jrong_kmust@163.com (Ju Rong), Xiaohua_y@163.com(Xiaohua Yu)

1. SHAP interpretation of the distal pathway (Path 1) ML model.

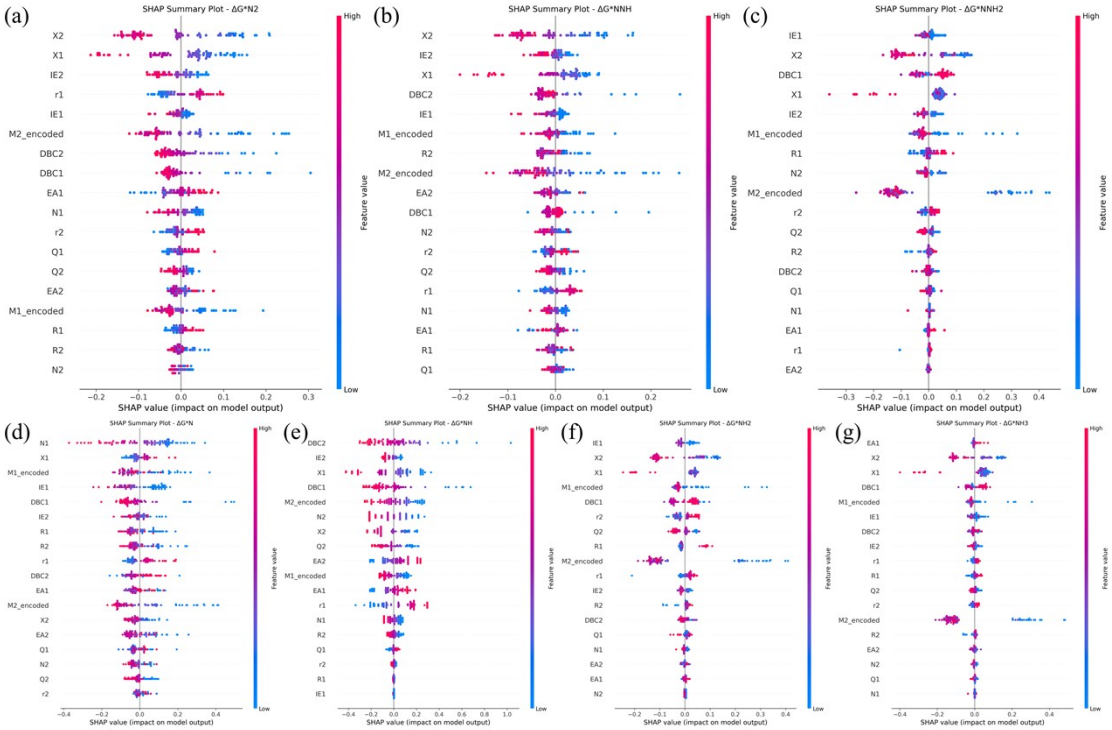


Fig. S1. SHAP summary plot of the MLP model for the distal pathway (Path 1): (a) ΔG^*N_2, (b) ΔG^*NNH, (c) ΔG^*NNH_2, (d) ΔG^*N, (e) ΔG^*NH, (f) ΔG^*NH_2 and (g) ΔG^*NH_3.

2. The Prediction of six machine learning models for the distal pathway (Path 1)

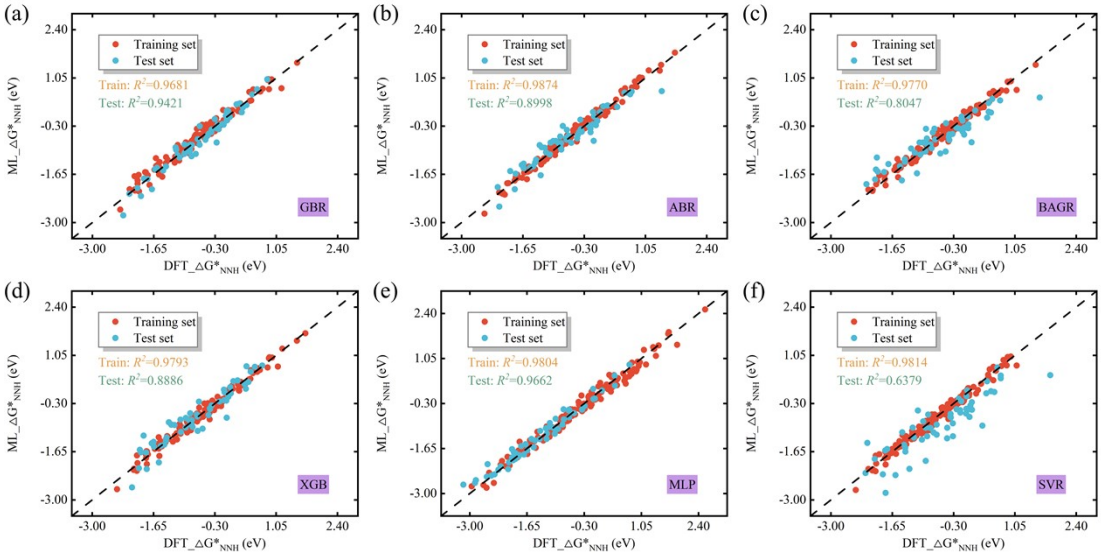


Fig. S2. The machine learning models were used to predict the adsorption free energy in the first hydrogenation step of the distal pathway (Path 1): (a) GBR, (b) ABR, (c) BAGR, (d) XGB, (e) MLP, and (f) SVR.

3. AIMD simulation results for the distal pathway (Path 1)

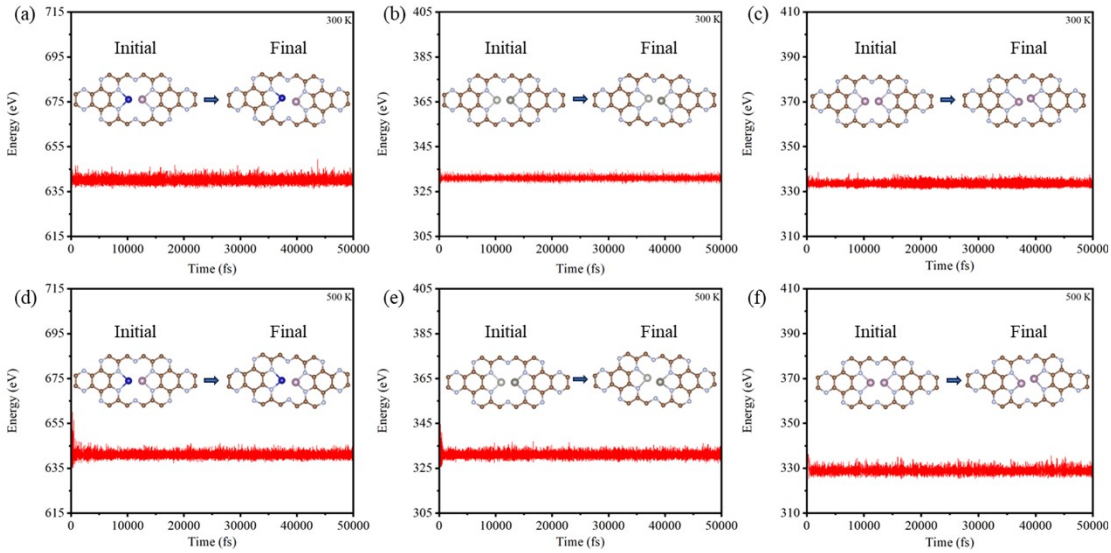


Fig. S3. AIMD simulations of PdW–C_2N (a, d), WW–C_2N (b, e), and CoMo–C_2N (c, f) at 300 K and 500 K (50 ps, 1 fs step). All systems exhibit stable energy fluctuations and intact structures, confirming their thermal robustness as catalytic sites.

4. The Prediction of six machine learning models for the alternating pathway (Path 2)

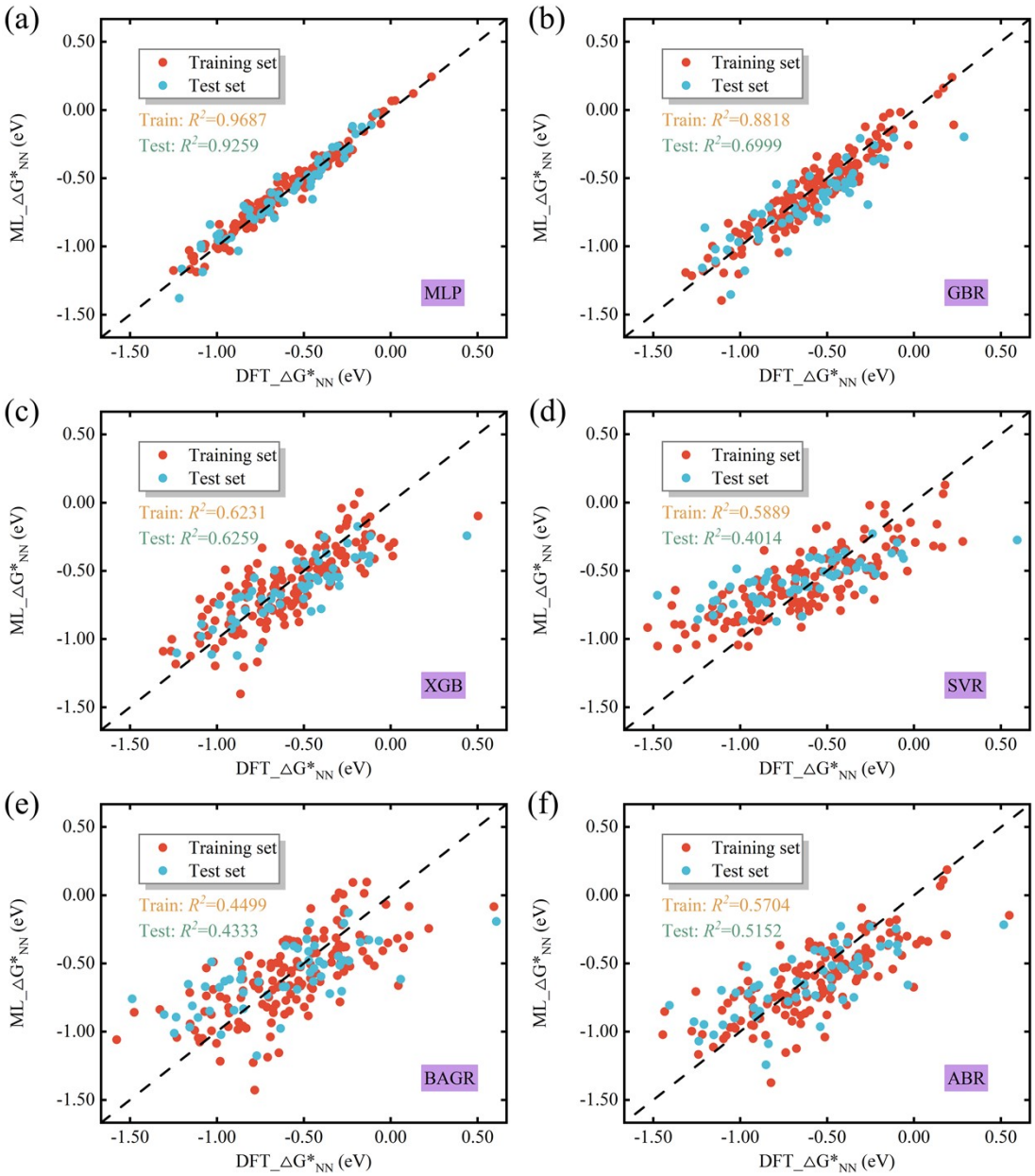


Fig. S4. Fitting and prediction results of all six machine learning models—(a) MLP, (b) GBR, (c) XGR, (d) SVR, (e) BAGR, (f) ABR on the alternating pathway dataset (**Path 2**), with performance comparison.

5. The Prediction of six machine learning models for the enzymatic pathway (Path 3)

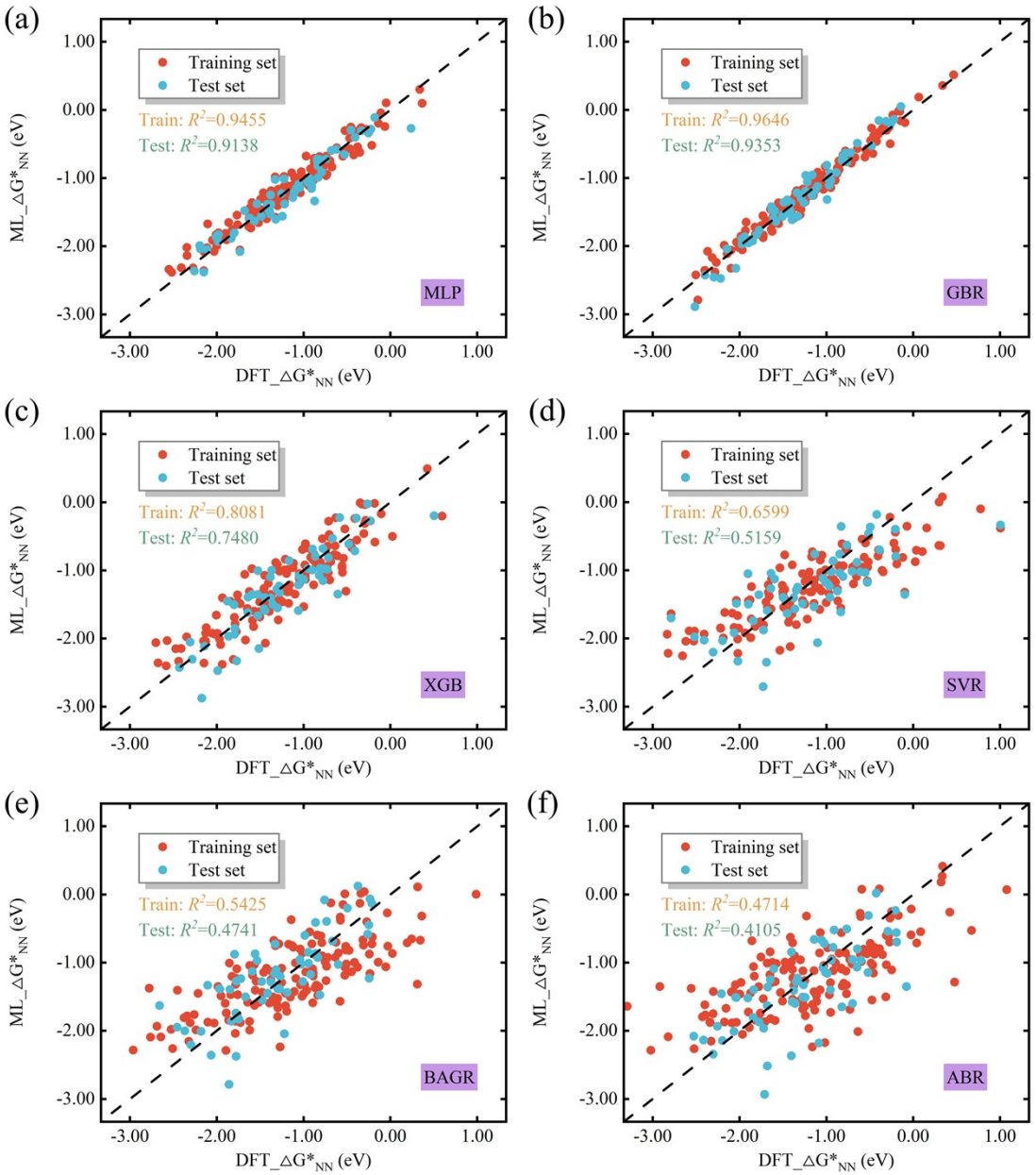


Fig. S5. Fitting and prediction results of all six machine learning models—(a) MLP, (b) GBR, (c) XGR, (d) SVR, (e) BAGR, (f) ABR on the enzymatic pathway dataset (**Path 3**), with performance comparison.

6. Comparison between DFT-calculated and MLP-predicted for the alternating pathway (Path 2)

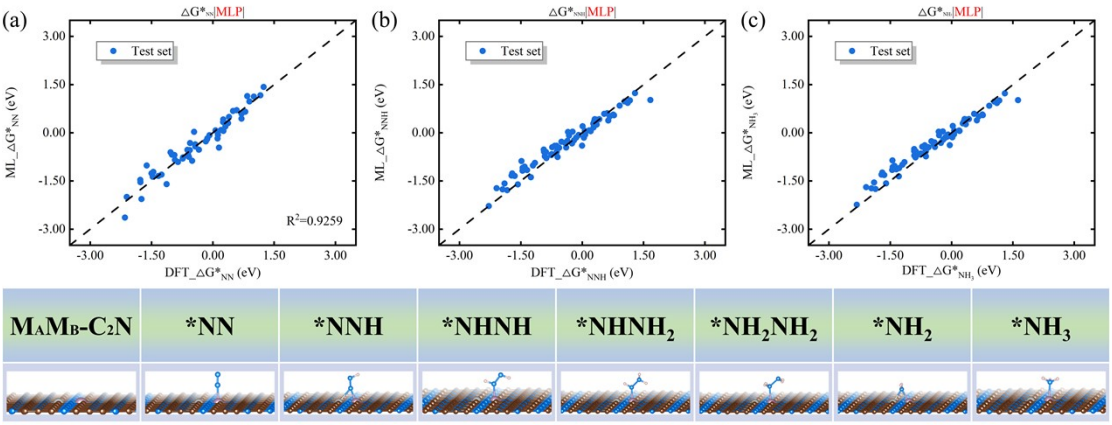


Fig.S6. Comparison between DFT-calculated and MLP-predicted ΔG values for the alternating pathway (Path 2): (a) ΔG*NN, (b) ΔG*NNH, and (c) ΔG**NH_3. The sub-figures display the optimized geometries of the corresponding reaction intermediates involved in the alternating pathway.

7. Comparison between DFT-calculated and MLP-predicted for the enzymatic pathway (Path 3)

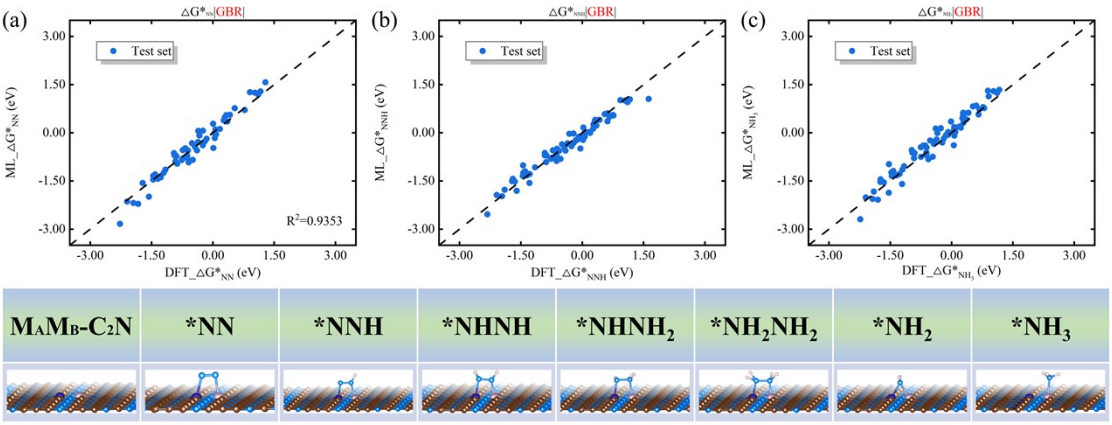


Fig.S7. Comparison between DFT-calculated and MLP-predicted ΔG values for the distal pathway (Path 3): (a) ΔG*NN, (b) ΔG*NNH, and (c) ΔG**NH$_3$. The sub-figures display the optimized geometries of the corresponding reaction intermediates involved in the enzymatic pathway.

8. **Error analysis of MLP model**

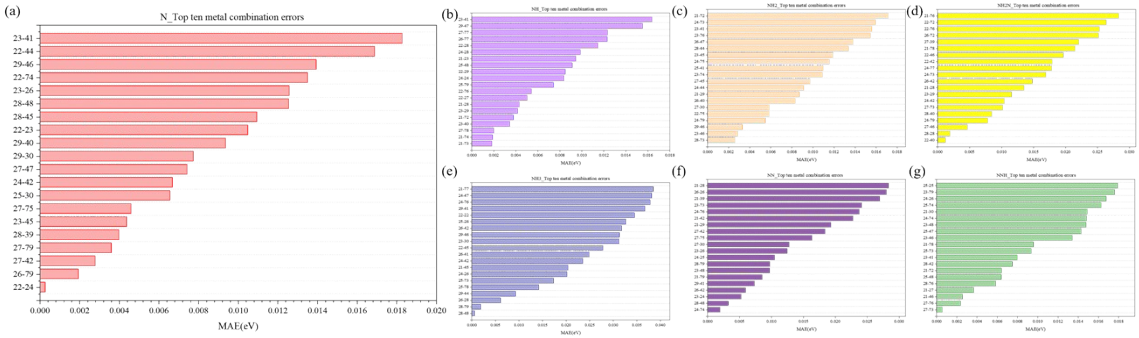


Fig. S8. Error distribution of the top 20 metal combinations with the largest MLP model error.

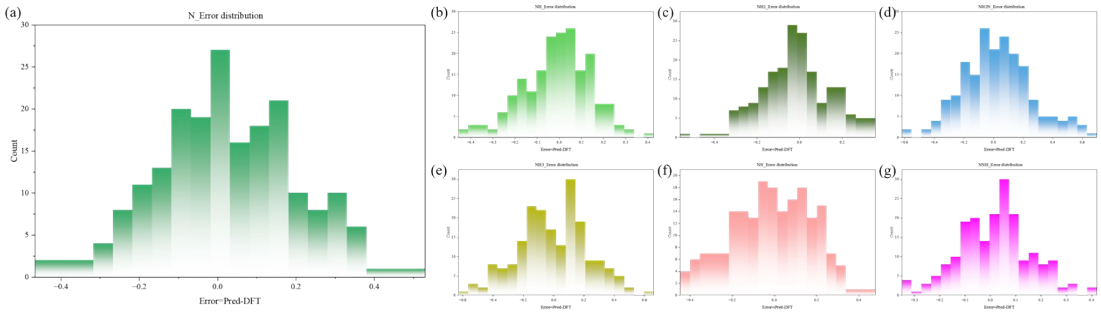


Fig. S9. MLP model error histogram

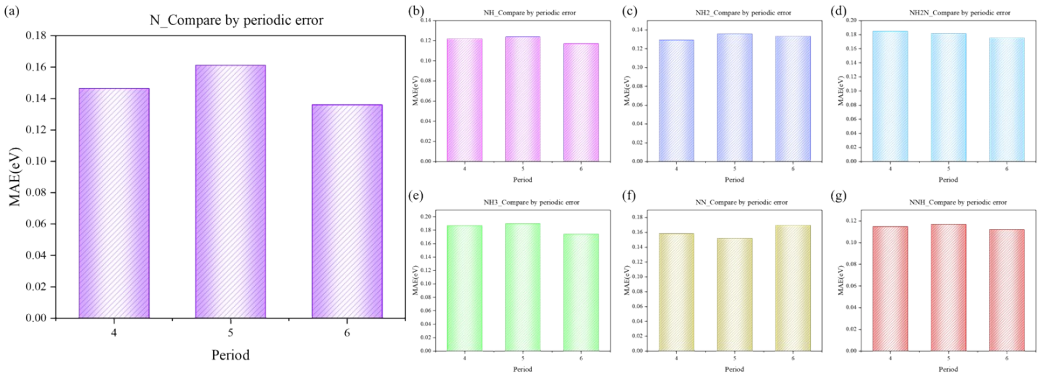


Fig.S10. Error distribution of MLP model by periodic.

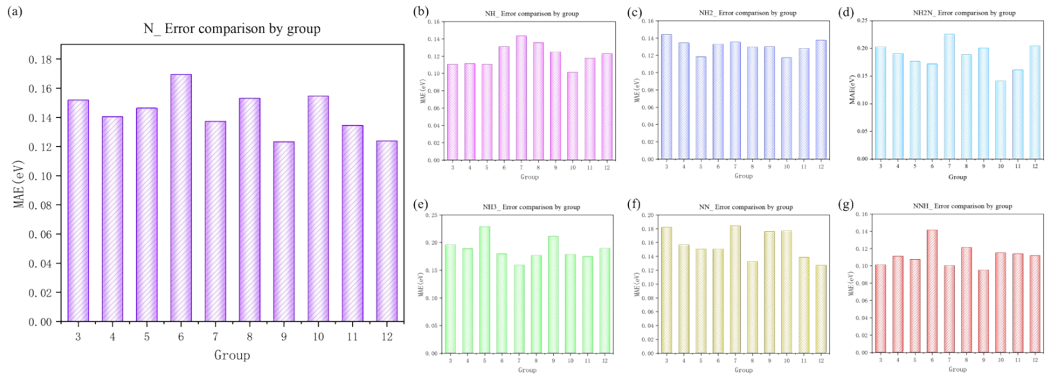


Fig.S11. Error distribution of MLP model by group.

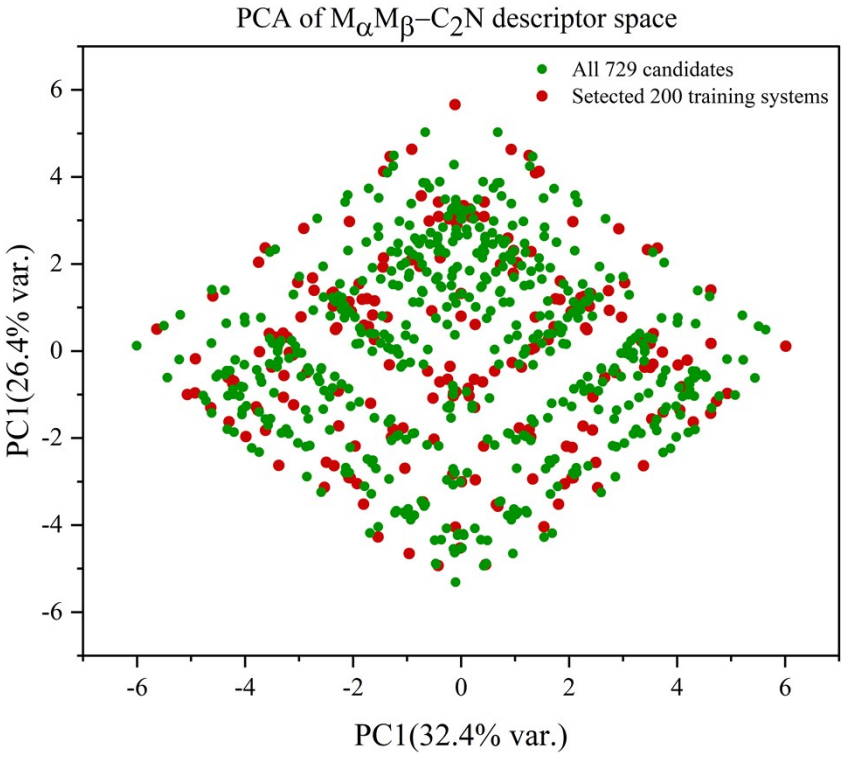
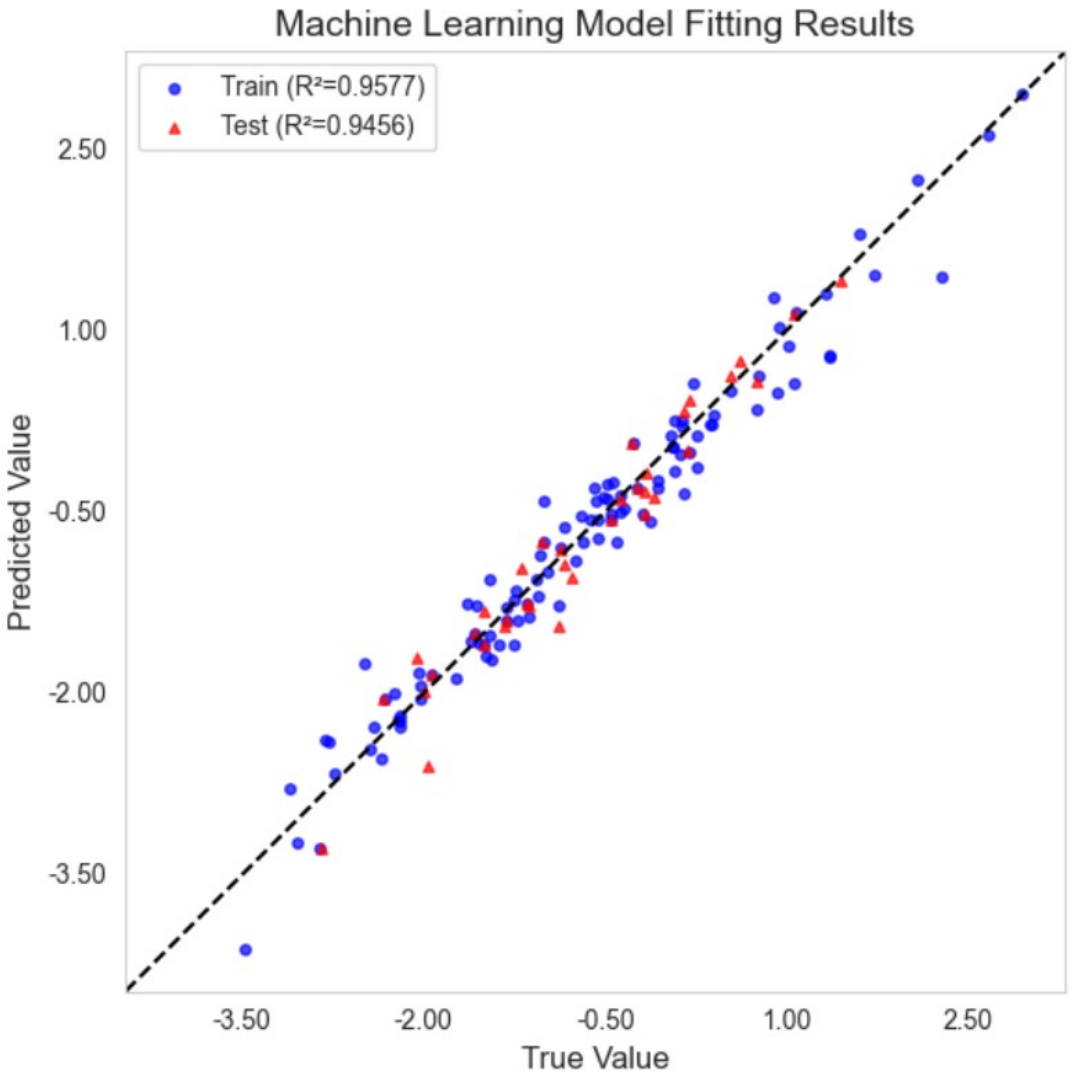


Fig.S12. Principal component analysis of initial data set




```
train R² = 0.9577
test R² = 0.9456
```

Fig.S13. The fitting effect of the structure after MAMB replacement in the initial data set.

9. The detailed thermodynamic corrections and Gibbs free energies under both PBE and PBE+U.

Table S1. Zero-point energies and entropy correction values of species in the gas phase and adsorbed on the Pd-W @C_2N catalyst (PBE), used to describe the distal pathway of electrochemical nitrogen reduction to ammonia on this catalyst.

Name	E /eV	E_{ZPE}/eV	TS/eV	G/eV
Pd–W@C_2N	-325.95249	0	0	-325.95249
**NN*	-342.60573	0.50724	0.63399	-342.73248
**NNH*	-346.43342	0.97631	1.01877	-346.47588
**NNH_2*	-350.68853	1.25861	1.03270	-350.46262
**N*	-334.63767	0.34346	0.06301	-334.35722
**NH*	-338.06459	1.03354	0.62614	-337.65719
**NH_2*	-342.00134	1.68443	0.41045	-340.72736
**NH_3*	-345.16821	1.47817	0.97739	-344.66743

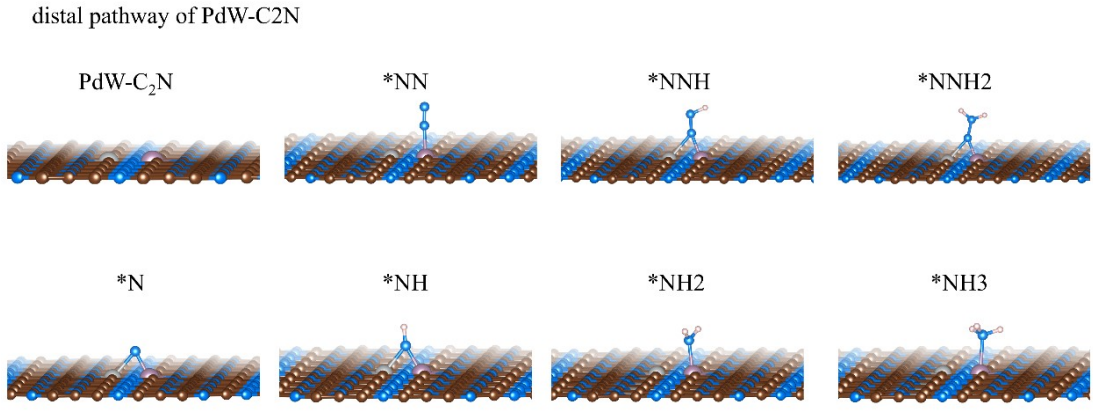


Table S2 Zero-point energies and entropy correction values of species in the gas phase and adsorbed on the W-W @C_2N catalyst (PBE), used to describe the distal pathway of electrochemical nitrogen reduction to ammonia on this catalyst.

Name	E /eV	E_{ZPE}/eV	TS/eV	G/eV
W–W@C_2N	-329.73065	0	0	-346.45065
**NN*	-346.96716	0.80412	0.28754	-346.45058
**NNH*	-349.59703	0.73043	1.11369	-349.98029
**NNH_2*	-354.12377	1.34717	1.31401	-354.09061
**N*	-339.03926	0.32891	0.19533	-338.90568
**NH*	-342.81529	0.76427	0.44462	-342.49564
**NH_2*	-346.8983	1.69425	0.71166	-345.91571
**NH_3*	-349.01547	1.29234	1.07263	-348.79576

distal pathway of WW-C2N

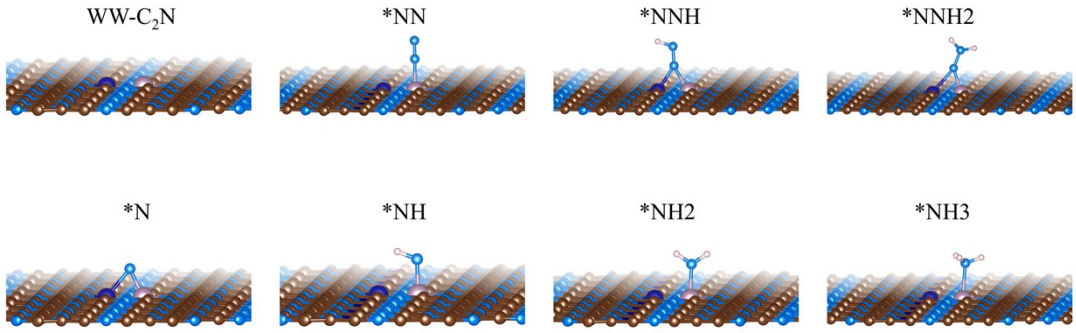


Table S3 Zero-point energies and entropy correction values of species in the gas phase and adsorbed on the Co-Mo @C_2N catalyst (PBE), used to describe the distal pathway of electrochemical nitrogen reduction to ammonia on this catalyst.

Name	E /eV	E_{ZPE}/eV	TS/eV	G/eV
Co–Mo@C_2N	-327.70296	0	0	-327.70296
**NN*	-345.44309	0.42992	0.55978	-345.57295
**NNH*	-348.76964	1.16321	0.79651	-348.40294
**NNH_2*	-352.17425	1.15427	0.86291	-351.88289
**N*	-336.62397	0.18953	0.18355	-336.61799
**NH*	-340.74512	0.95221	0.50505	-340.29796
**NH_2*	-345.19159	1.52588	0.5623	-344.22801
**NH_3*	-347.56132	1.36741	0.80353	-346.99744

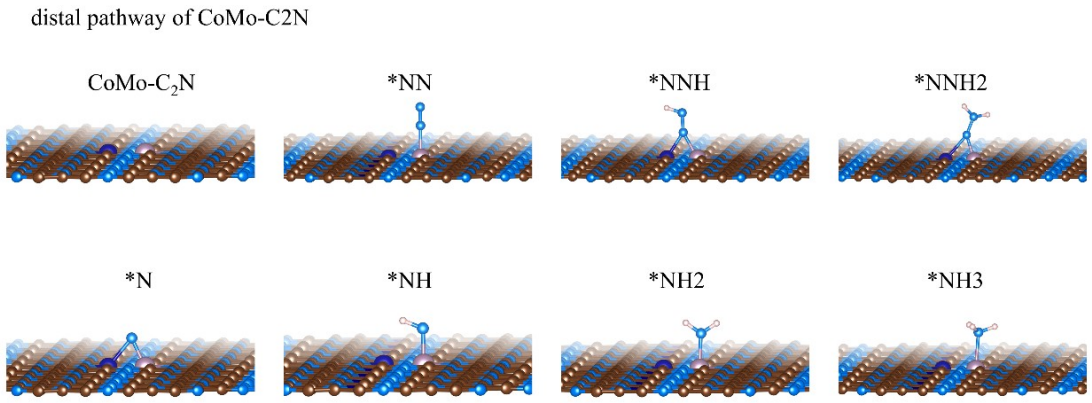


Table S4 Zero-point energies and entropy correction values of species in the gas phase and adsorbed on the Pd-W @C_2N catalyst (PBE+U), used to describe the distal pathway of electrochemical nitrogen reduction to ammonia on this catalyst.

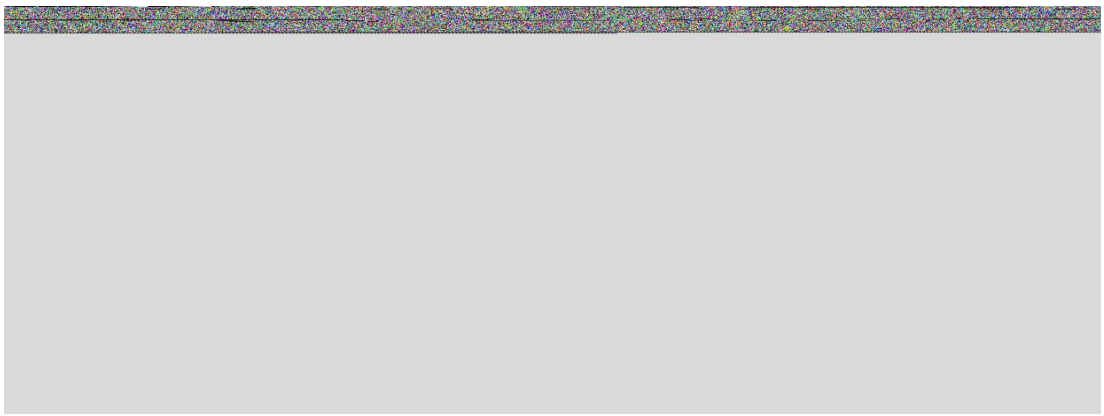
Name	E /eV	E_{ZPE}/eV	TS/eV	G/eV
Pd–W@C₂N	-322.85594	0	0	-322.85594
**NN*	-339.24819	0.36521	0.46281	-339.34579
**NNH*	-343.24723	0.66630	0.70295	-343.28388
**NNH₂*	-347.28685	0.93148	0.69191	-347.04728
**N*	-331.52551	0.24024	0.04726	-331.33253
**NH*	-334.81916	0.75427	0.43830	-334.50319
**NH₂*	-338.78652	1.11164	0.29552	-337.97041
**NH₃*	-341.75104	1.04947	0.64508	-341.34665

Table S5 Zero-point energies and entropy correction values of species in the gas phase and adsorbed on the W-W @C_2N catalyst (PBE+U), used to describe the distal pathway of electrochemical nitrogen reduction to ammonia on this catalyst.

Name	E /eV	E_{ZPE}/eV	TS/eV	G/eV
W–W@C_2N	-326.59821	0	0	-326.59821
**NN*	-343.56688	0.56283	0.20990	-343.21395
**NNH*	-346.38045	0.52565	0.74617	-346.60097
**NNH_2*	-350.68878	0.91604	0.98551	-350.75825
**N*	-335.88629	0.24342	0.13673	-335.7796
**NH*	-339.52436	0.50441	0.31568	-339.33563
**NH_2*	-343.63746	1.16907	0.46970	-342.93809
**NH_3*	-345.52532	0.91764	0.72939	-345.33707

Table S6 Zero-point energies and entropy correction values of species in the gas phase and adsorbed on the Co-Mo @C_2N catalyst (PBE+U), used to describe the distal pathway of electrochemical nitrogen reduction to ammonia on this catalyst.

Name	E /eV	E_{ZPE}/eV	TS/eV	G/eV
Co–Mo@C_2N	-324.58978	0	0	-324.58978
**NN*	-342.05775	0.30094	0.41424	-342.17105
**NNH*	-345.52377	0.83750	0.53367	-345.21994
**NNH_2*	-348.75816	0.78486	0.64718	-348.62048
**N*	-333.45937	0.13838	0.12849	-333.44948
**NH*	-337.47395	0.62845	0.36869	-337.21419
**NH_2*	-342.01483	1.08342	0.38236	-341.31377
**NH_3*	-344.12046	0.94316	0.57051	-343.74781



10. The parameters of six machine learning models

According to the given scope of the parameters search, the best parameter combination can be obtained through GridSearchCV.

Table S7

The optimal hyperparameters of the MLP model (*ΔG*NN*).

models	The scope of theparameters search	Best parameters
XGB	n_estimators = [50, 100, 150, 200] learning_rate = [0.01, 0.05, 0.1, 0.2] max_depth = [3, 5, 7, 9] subsample = [0.5, 0.7, 1.0] colsample_bytree = [0.5, 0.7, 1.0] random_state = [0, 2, 5, 10]	{'n_estimators': 100, 'learning_rate': 0.1, 'max_depth': 5, 'subsample': 1.0, 'colsample_bytree': 1.0, 'random_state': 0}
SVR	gamma = [0.001, 0.01, 0.02, 0.05, 0.08, 0.1, 0.2, 0.3, 0.5, 0.8, 1, 5, 8] epsilon = [0.01, 0.05, 0.1, 0.5, 1, 5, 8]	{'epsilon': 0.1, 'gamma': 0.001}
MLP	hidden_layer_sizes = [(50,), (100,), (50, 50), (100, 50), (100, 100)] activation = ['relu', 'tanh'] solver = ['adam', 'sgd'] alpha = [0.0001, 0.001, 0.01] learning_rate_init = [0.001, 0.01, 0.1] random_state = [0, 1, 2, 5]	{'hidden_layer_sizes': (100,), 'activation': 'relu', 'solver': 'adam', 'alpha': 0.0001, 'learning_rate_init': 0.001, 'random_state': 1}
GBR	n_estimators = [10, 20, 40, 60, 80, 100, 150, 200] learning_rate = [0.001, 0.003, 0.005, 0.01, 0.03, 0.05, 0.1] max_depth = [1, 2, 3, 4, 5] loss = ['squared_error', 'absolute_error', 'huber'] random_state = [0, 2, 5, 10, 20, 30]	{'n_estimators': 50, 'learning_rate': 0.1, 'max_depth': 3, 'loss': 'huber', 'random_state': 0}
ABR	n_estimators = [10, 30, 50, 60, 70, 80, 90, 100, 125, 150], learning_rate = [0.001, 0.008, 0.01, 0.02, 0.03, 0.04, 0.05, 0.08,	{'learning_rate':0.1, 'n_estimators': 125, 'random_state': 0}

	0.09, 0.1], random_state = [0, 5, 10, 15, 20, 30]	
BAGR	n_estimators = [2, 5, 8, 10, 20, 30, 40, 50, 60, 80, 100, 150, 200] random_state = [0, 2, 5, 10, 15, 20, 30]	{‘n_estimators’:150, ‘random_state’:2}

Table S8

The optimal hyperparameters of the MLP model (*ΔG*NNH*).

models	The scope of theparameters search	Best parameters
XGB	n_estimators = [50, 100, 150, 200] learning_rate = [0.01, 0.05, 0.1, 0.2] max_depth = [3, 5, 7, 9] subsample = [0.5, 0.7, 1.0] colsample_bytree = [0.5, 0.7, 1.0] random_state = [0, 2, 5, 10]	{'n_estimators': 121, 'learning_rate': 0.3, 'max_depth': 3, 'subsample': 1.0, 'colsample_bytree': 1.0, 'random_state': 40}
SVR	gamma = [0.001, 0.01, 0.02, 0.05, 0.08, 0.1, 0.2, 0.3, 0.5, 0.8, 1, 5, 8] epsilon = [0.01, 0.05, 0.1, 0.5, 1, 5, 8]	{'epsilon': 0.001, 'gamma': 0.3}
MLP	hidden_layer_sizes = [(50,), (100,), (50, 50), (100, 50), (100, 100)] activation = ['relu', 'tanh'] solver = ['adam', 'sgd'] alpha = [0.0001, 0.001, 0.01] learning_rate_init = [0.001, 0.01, 0.1] random_state = [0, 1, 2, 5]	{'hidden_layer_sizes': (130,), 'activation': 'relu', 'solver': 'adam', 'alpha': 0.0003, 'learning_rate_init': 0.002, 'random_state': 33}
GBR	n_estimators = [10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 125, 150, 175, 200], random_state = [0, 5, 10, 20, 30], learning_rate = [0.001, 0.003, 0.005, 0.008, 0.01, 0.03, 0.05, 0.08, 0.1], max_depth = [1, 2, 3, 4, 5], loss = ['huber']	{'learning_rate': 0.02, 'loss': 'huber', 'max_depth': 4, 'n_estimators': 125, 'random_state': 0}
ABR	n_estimators = [10, 30, 50, 60, 70,	{'learning_rate': 0.05, 'n_estimators': 10,

	80, 90, 100, 125, 150], learning_rate = [0.001, 0.008, 0.01, 0.02, 0.03, 0.04, 0.05, 0.08, 0.09, 0.1], random_state = [0, 5, 10, 15, 20, 30]	'random_state': 15}
BAGR	n_estimators = [2, 5, 8, 10, 20, 30, 40, 50, 60, 80, 100, 150, 200] random_state = [0, 2, 5, 10, 15, 20, 30]	{'n_estimators':8, 'random_state':5}

Table S9

The optimal hyperparameters of the MLP model (ΔG^*NH_3).

models	The scope of theparameters search	Best parameters
XGB	n_estimators = [50, 100, 150, 200] learning_rate = [0.01, 0.05, 0.1, 0.2] max_depth = [3, 5, 7, 9] subsample = [0.5, 0.7, 1.0] colsample_bytree = [0.5, 0.7, 1.0] random_state = [0, 2, 5, 10]	{'n_estimators': 100, 'learning_rate': 0.5, 'max_depth': 7, 'subsample': 1.0, 'colsample_bytree': 1.0, 'random_state': 32}
SVR	gamma = [0.001, 0.01, 0.02, 0.05, 0.08, 0.1, 0.2, 0.3, 0.5, 0.8, 1, 5, 8] epsilon = [0.01, 0.05, 0.1, 0.5, 1, 5, 8]	{'epsilon': 0.1, 'gamma': 0.001}
MLP	hidden_layer_sizes = [(50,), (100,), (50, 50), (100, 50), (100, 100)] activation = ['relu', 'tanh'] solver = ['adam', 'sgd'] alpha = [0.0001, 0.001, 0.01] learning_rate_init = [0.001, 0.01, 0.1] random_state = [0, 1, 2, 5]	{'hidden_layer_sizes': (110,), 'activation': 'relu', 'solver': 'adam', 'alpha': 0.0006, 'learning_rate_init': 0.006, 'random_state': 66}
GBR	n_estimators = [10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 125, 150, 175, 200], random_state = [0, 5, 10, 20, 30], learning_rate = [0.001, 0.003, 0.005, 0.008, 0.01, 0.03, 0.05, 0.08, 0.1],	{'learning_rate': 0.09, 'loss': 'huber', 'max_depth': 4, 'n_estimators': 20, 'random_state': 10}

	max_depth = [1, 2, 3, 4, 5], loss = ['huber']	
ABR	n_estimators = [10, 30, 50, 60, 70, 80, 90, 100, 125, 150], learning_rate = [0.001, 0.008, 0.01, 0.02, 0.03, 0.04, 0.05, 0.08, 0.09, 0.1], random_state = [0, 5, 10, 15, 20, 30]	{'learning_rate': 0.04, 'n_estimators': 10, 'random_state': 20}
BAGR	n_estimators = [2, 5, 8, 10, 20, 30, 40, 50, 60, 80, 100, 150, 200] random_state = [0, 2, 5, 10, 15, 20, 30]	{'n_estimators':5, 'random_state':15}

Table S10

The optimal hyperparameters of the MLP model (ΔG^*NH_2N).

models	The scope of theparameters search	Best parameters
XGB	n_estimators = [50, 100, 150, 200] learning_rate = [0.01, 0.05, 0.1, 0.2] max_depth = [3, 5, 7, 9] subsample = [0.5, 0.7, 1.0] colsample_bytree = [0.5, 0.7, 1.0] random_state = [0, 2, 5, 10]	{'n_estimators': 100, 'learning_rate': 0.1, 'max_depth': 5, 'subsample': 1.0, 'colsample_bytree': 1.0, 'random_state': 1}
SVR	gamma = [0.001, 0.01, 0.02, 0.05, 0.08, 0.1, 0.2, 0.3, 0.5, 0.8, 1, 5, 8] epsilon = [0.01, 0.05, 0.1, 0.5, 1, 5, 8]	{'epsilon': 0.1, 'gamma': 0.001}
MLP	hidden_layer_sizes = [(50,), (100,), (50, 50), (100, 50), (100, 100)] activation = ['relu', 'tanh'] solver = ['adam', 'sgd'] alpha = [0.0001, 0.001, 0.01] learning_rate_init = [0.001, 0.01, 0.1] random_state = [0, 1, 2, 5]	{'hidden_layer_sizes': (100,), 'activation': 'relu', 'solver': 'adam', 'alpha': 0.0001, 'learning_rate_init': 0.001, 'random_state': 1}
GBR	n_estimators = [10, 20, 40, 60, 80, 100, 150, 200] learning_rate = [0.001, 0.003, 0.005, 0.01, 0.03, 0.05, 0.1] max_depth =	{'n_estimators': 50, 'learning_rate': 0.1, 'max_depth': 3, 'loss': 'huber', 'random_state': 0}

	[1, 2, 3, 4, 5] loss = ['squared_error', 'absolute_error', 'huber'] random_state = [0, 2, 5, 10, 20, 30]	
ABR	n_estimators = [10, 30, 50, 60, 70, 80, 90, 100, 125, 150], learning_rate = [0.001, 0.008, 0.01, 0.02, 0.03, 0.04, 0.05, 0.08, 0.09, 0.1], random_state = [0, 5, 10, 15, 20, 30]	{'learning_rate':0.1, 'n_estimators': 125, 'random_state': 0}
BAGR	n_estimators = [2, 5, 8, 10, 20, 30, 40, 50, 60, 80, 100, 150, 200] random_state = [0, 2, 5, 10, 15, 20, 30]	{‘n_estimators’:150, ‘random_state’:2}

Table S11

The optimal hyperparameters of the MLP model (*ΔG*N*).

models	**The scope of theparameters search**	**Best parameters**
XGB	n_estimators = [50, 100, 150, 200] learning_rate = [0.01, 0.05, 0.1, 0.2] max_depth = [3, 5, 7, 9] subsample = [0.5, 0.7, 1.0] colsample_bytree = [0.5, 0.7, 1.0] random_state = [0, 2, 5, 10]	{'n_estimators': 100, 'learning_rate': 0.1, 'max_depth': 5, 'subsample': 1.0, 'colsample_bytree': 2.0, 'random_state': 2}
SVR	gamma = [0.001, 0.01, 0.02, 0.05, 0.08, 0.1, 0.2, 0.3, 0.5, 0.8, 1, 5, 8] epsilon = [0.01, 0.05, 0.1, 0.5, 1, 5, 8]	{'epsilon': 0.1, 'gamma': 0.001}
MLP	hidden_layer_sizes = [(50,), (100,), (50, 50), (100, 50), (100, 100)] activation = ['relu', 'tanh'] solver = ['adam', 'sgd'] alpha = [0.0001, 0.001, 0.01] learning_rate_init = [0.001, 0.01, 0.1] random_state = [0, 1, 2, 5]	{'hidden_layer_sizes': (150,), 'activation': 'relu', 'solver': 'adam', 'alpha': 0.0001, 'learning_rate_init': 0.001, 'random_state': 21}
GBR	n_estimators = [10, 20, 40, 60, 80, 100, 150, 200]	{'n_estimators': 50, 'learning_rate': 0.2,

	learning_rate = [0.001, 0.003, 0.005, 0.01, 0.03, 0.05, 0.1] max_depth = [1, 2, 3, 4, 5] loss = ['squared_error', 'absolute_error', 'huber'] random_state = [0, 2, 5, 10, 20, 30]	'max_depth': 3, 'loss': 'huber', 'random_state': 0}
ABR	n_estimators = [10, 30, 50, 60, 70, 80, 90, 100, 125, 150], learning_rate = [0.001, 0.008, 0.01, 0.02, 0.03, 0.04, 0.05, 0.08, 0.09, 0.1], random_state = [0, 5, 10, 15, 20, 30]	{'learning_rate':0.1, 'n_estimators': 125, 'random_state': 0}
BAGR	n_estimators = [2, 5, 8, 10, 20, 30, 40, 50, 60, 80, 100, 150, 200] random_state = [0, 2, 5, 10, 15, 20, 30]	{‘n_estimators’:150, ‘random_state’:2}

Table S12

The optimal hyperparameters of the MLP model (*ΔG*NH*).

models	**The scope of theparameters search**	**Best parameters**
XGB	n_estimators = [50, 100, 150, 200] learning_rate = [0.01, 0.05, 0.1, 0.2] max_depth = [3, 5, 7, 9] subsample = [0.5, 0.7, 1.0] colsample_bytree = [0.5, 0.7, 1.0] random_state = [0, 2, 5, 10]	{'n_estimators': 100, 'learning_rate': 0.1, 'max_depth': 5, 'subsample': 1.0, 'colsample_bytree': 1.0, 'random_state': 0}
SVR	gamma = [0.001, 0.01, 0.02, 0.05, 0.08, 0.1, 0.2, 0.3, 0.5, 0.8, 1, 5, 8] epsilon = [0.01, 0.05, 0.1, 0.5, 1, 5, 8]	{'epsilon': 0.1, 'gamma': 0.001}
MLP	hidden_layer_sizes = [(50,), (100,), (50, 50), (100, 50), (100, 100)] activation = ['relu', 'tanh'] solver = ['adam', 'sgd'] alpha = [0.0001, 0.001, 0.01] learning_rate_init = [0.001, 0.01, 0.1] random_state = [0, 1, 2, 5]	{'hidden_layer_sizes': (100,), 'activation': 'relu', 'solver': 'adam', 'alpha': 0.0001, 'learning_rate_init': 0.001, 'random_state': 1}

GBR	n_estimators = [10, 20, 40, 60, 80, 100, 150, 200] learning_rate = [0.001, 0.003, 0.005, 0.01, 0.03, 0.05, 0.1] max_depth = [1, 2, 3, 4, 5] loss = ['squared_error', 'absolute_error', 'huber'] random_state = [0, 2, 5, 10, 20, 30]	{'n_estimators': 50, 'learning_rate': 0.1, 'max_depth': 3, 'loss': 'huber', 'random_state': 0}
ABR	n_estimators = [10, 30, 50, 60, 70, 80, 90, 100, 125, 150], learning_rate = [0.001, 0.008, 0.01, 0.02, 0.03, 0.04, 0.05, 0.08, 0.09, 0.1], random_state = [0, 5, 10, 15, 20, 30]	{'learning_rate':0.1, 'n_estimators': 125, 'random_state': 0}
BAGR	n_estimators = [2, 5, 8, 10, 20, 30, 40, 50, 60, 80, 100, 150, 200] random_state = [0, 2, 5, 10, 15, 20, 30]	{‘n_estimators’:150, ‘random_state’:2}

Table S13

The optimal hyperparameters of the MLP model (ΔG^*NH_2).

models	The scope of theparameters search	Best parameters
XGB	n_estimators = [50, 100, 150, 200] learning_rate = [0.01, 0.05, 0.1, 0.2] max_depth = [3, 5, 7, 9] subsample = [0.5, 0.7, 1.0] colsample_bytree = [0.5, 0.7, 1.0] random_state = [0, 2, 5, 10]	{'n_estimators': 100, 'learning_rate': 0.1, 'max_depth': 5, 'subsample': 1.0, 'colsample_bytree': 1.0, 'random_state': 0}
SVR	gamma = [0.001, 0.01, 0.02, 0.05, 0.08, 0.1, 0.2, 0.3, 0.5, 0.8, 1, 5, 8] epsilon = [0.01, 0.05, 0.1, 0.5, 1, 5, 8]	{'epsilon': 0.1, 'gamma': 0.001}
MLP	hidden_layer_sizes = [(50,), (100,), (50, 50), (100, 50), (100, 100)] activation = ['relu', 'tanh'] solver = ['adam', 'sgd'] alpha = [0.0001, 0.001, 0.01] learning_rate_init = [0.001, 0.01, 0.1] random_state =	{'hidden_layer_sizes': (150,), 'activation': 'relu', 'solver': 'adam', 'alpha': 0.0001, 'learning_rate_init': 0.001, 'random_state': 29}

	[0, 1, 2, 5]	
GBR	n_estimators = [10, 20, 40, 60, 80, 100, 150, 200] learning_rate = [0.001, 0.003, 0.005, 0.01, 0.03, 0.05, 0.1] max_depth = [1, 2, 3, 4, 5] loss = ['squared_error', 'absolute_error', 'huber'] random_state = [0, 2, 5, 10, 20, 30]	{'n_estimators': 50, 'learning_rate': 0.1, 'max_depth': 3, 'loss': 'huber', 'random_state': 0}
ABR	n_estimators = [10, 30, 50, 60, 70, 80, 90, 100, 125, 150], learning_rate = [0.001, 0.008, 0.01, 0.02, 0.03, 0.04, 0.05, 0.08, 0.09, 0.1], random_state = [0, 5, 10, 15, 20, 30]	{'learning_rate':0.1, 'n_estimators': 125, 'random_state': 0}
BAGR	n_estimators = [2, 5, 8, 10, 20, 30, 40, 50, 60, 80, 100, 150, 200] random_state = [0, 2, 5, 10, 15, 20, 30]	{‘n_estimators’:150, ‘random_state’:2}

Table S14

106 catalysts exhibiting high NRR activity

1	Ag+Fe	21	Pt+Fe
2	Ag+Ru	22	Ru+Fe
3	Ag+W	23	Ru+Zr
4	Au+W	24	Rh+Mo
5	Au+Ru	25	Rh+Fe
6	Zn+W	26	W+Mo
7	Ta+Ta	27	W+W
8	Fe+Mo	28	Mn+Fe
9	Fe+W	29	Co+Cu
10	Cu+Ru	30	Co+Mo
11	Ir+Mo	31	Co+Y
12	Ir+Ru	32	Zr+W
13	Mo+Fe	33	Zr+Mo
14	Mo+Mo		
15	Mo+Co		
16	Mo+B		
17	Nb+W		
18	Nb+Zr		
19	Pd+Fe		
20	Pd+W		

Table S15 Zero-point and entropic corrections to the free energy of the gas phase and the adsorbed species on the water solvation catalyst PdW-C2N via the distal pathway.

Name	E /eV	E_{ZPE}/eV	TS/eV	G/eV
Pd–W@C_2N	-603.06655	4.56408	0.87242	-619.84655
**NN*	-623.53821	4.50568	1.03621	-623.77655
**NNH*	-627.24602	4.99415	0.95301	-627.42655
**NNH_2*	-631.46769	4.71161	0.78634	-611.32155
**N*	-615.24682	4.71120	0.88476	-614.38155
**NH*	-618.20799	4.77137	0.82771	-617.37155
**NH_2*	-621.28521	4.67049	0.95644	-617.34155
**NH_3*	-625.47560	4.56408	0.87242	-619.84655

Table S16 The optimized geometry of screened catalysts PdW-C_2N, WW-C_2N and CoMo-C_2N.

CONTCAR for PdW-C2N

1.00000000000000

14.4149999618999995 0.0000000000000000 0.0000000000000000

0.0000000000000000 8.3225002288999992 0.0000000000000000

0.0000000000000000 0.0000000000000000 15.0001001358000003

N C Pd W

12 24 1 1

Direct

-0.0005466553803268 0.6702156169338599 0.5000000000000000

0.5014361879606904 0.1710104907138636 0.5000000000000000

0.6693105466322429 0.6680838568968854 0.5000000000000000

0.1624000155235557 0.1645643465661867 0.5000000000000000

0.3271913135962669 0.6718361763435627 0.5000000000000000

0.8373827622754477 0.1625786497812114 0.5000000000000000

0.5014361879606904 0.8289895242861343 0.5000000000000000

-0.0005466553803268 0.3297843830661404 0.5000000000000000

0.8373827622754477 0.8374213502187887 0.5000000000000000

0.3271913135962669 0.3281638236564376 0.5000000000000000

0.1624000155235557 0.8354356534338133 0.5000000000000000

0.6693105466322429 0.3319161431031145 0.5000000000000000

0.5814222600206186 0.9135255347650708 0.5000000000000000

0.0774836558116920 0.4140524599551643 0.5000000000000000

0.7555500543845531 0.9129832037970511 0.5000000000000000

0.2469584871724834 0.4124406334048626 0.5000000000000000

0.1634769383689767 0.6752133316058188 0.5000000000000000

0.6686357729644313	0.1665266010431958	0.5000000000000000
0.9211608326427001	0.5860911766766217	0.5000000000000000
0.4202678730189635	0.0858677784081145	0.5000000000000000
0.7544378046151033	0.5853121238336561	0.5000000000000000
0.2436584633033227	0.0862166975668948	0.5000000000000000
0.3307164022242993	0.8318694559559854	0.5000000000000000
0.8357430363367041	0.3239942711609395	0.5000000000000000
0.4202678730189635	0.9141322365918833	0.5000000000000000
0.9211608326427001	0.4139087943233796	0.5000000000000000
0.2436584633033227	0.9137833254331036	0.5000000000000000
0.7544378046151033	0.4146878761663439	0.5000000000000000
0.8357430363367041	0.6760056988390578	0.5000000000000000
0.3307164022242993	0.1681305440440149	0.5000000000000000
0.0774836558116920	0.5859475110448366	0.5000000000000000
0.5814222600206186	0.0864744802349270	0.5000000000000000
0.2469584871724834	0.5875593665951373	0.5000000000000000
0.7555500543845531	0.0870168192029473	0.5000000000000000
0.6686357729644313	0.8334733989568041	0.5000000000000000
0.1634769383689767	0.3247866383941788	0.5000000000000000
0.4288861373673394	0.5000000000000000	0.5000000000000000
0.5777424346892353	0.5000000000000000	0.5000000000000000

CONTCAR for WW-C2N

1.00000000000000

14.4149999618999995 0.0000000000000000 0.0000000000000000

0.0000000000000000 8.3225002288999992 0.0000000000000000

0.0000000000000000 0.0000000000000000 15.0001001358000003

N C W

12 24 2

Direct

0.0000000000000000 0.6701329956525859 0.5000000000000000

0.5000000000000000 0.1779757659140085 0.5000000000000000

0.6689393042724069 0.6702780143236650 0.5000000000000000

0.1634153145793425 0.1637149765898654 0.5000000000000000

0.3310607247275921 0.6702780143236650 0.5000000000000000

0.8365847154206599 0.1637149765898654 0.5000000000000000

0.5000000000000000 0.8220242490859891 0.5000000000000000

0.0000000000000000 0.3298670043474140 0.5000000000000000

0.8365847154206599 0.8362850234101348 0.5000000000000000

0.3310607247275921 0.3297219856763349 0.5000000000000000

0.1634153145793425 0.8362850234101348 0.5000000000000000

0.6689393042724069 0.3297219856763349 0.5000000000000000

0.5802377485991147 0.9123841453535121 0.5000000000000000

0.0784047560077350 0.4139598208315983 0.5000000000000000

0.7553773610074695 0.9137204999305291 0.5000000000000000

0.2463630564368253 0.4146900455495640 0.5000000000000000

0.1644080790284265 0.6753734553263030 0.5000000000000000

0.6673999058942028 0.1652914648870965 0.5000000000000000

0.9215952369922680 0.5860401501684027 0.5000000000000000

0.4197622214008899	0.0876158696464857	0.5000000000000000
0.7536369145631759	0.5853099544504363	0.5000000000000000
0.2446226389925305	0.0862795230694691	0.5000000000000000
0.3326001241057997	0.8347085351129035	0.5000000000000000
0.8355919499715725	0.3246265146736947	0.5000000000000000
0.4197622214008899	0.9123841453535121	0.5000000000000000
0.9215952369922680	0.4139598208315983	0.5000000000000000
0.2446226389925305	0.9137204999305291	0.5000000000000000
0.7536369145631759	0.4146900455495640	0.5000000000000000
0.8355919499715725	0.6753734553263030	0.5000000000000000
0.3326001241057997	0.1652914648870965	0.5000000000000000
0.0784047560077350	0.5860401501684027	0.5000000000000000
0.5802377485991147	0.0876158696464857	0.5000000000000000
0.2463630564368253	0.5853099544504363	0.5000000000000000
0.7553773610074695	0.0862795230694691	0.5000000000000000
0.6673999058942028	0.8347085351129035	0.5000000000000000
0.1644080790284265	0.3246265146736947	0.5000000000000000
0.4212505466979786	0.5000000000000000	0.5000000000000000
0.5787494243020226	0.5000000000000000	0.5000000000000000

CONTCAR for CoMo-C2N

1.00000000000000

14.4149999618999995 0.0000000000000000 0.0000000000000000

0.0000000000000000 8.3225002288999992 0.0000000000000000

0.0000000000000000 0.0000000000000000 15.0001001358000003

N C Co Mo

12 24 1 1

Direct

0.0003937666274036 0.6694126007731116 0.5000000000000000

0.5026683864650783 0.1728239410687804 0.5000000000000000

0.6656236540363307 0.6624219606758446 0.5000000000000000

0.1655912769368192 0.1646021028961620 0.5000000000000000

0.3336390105261560 0.6692197736119384 0.5000000000000000

0.8359789113241045 0.1634253650136134 0.5000000000000000

0.5026683864650783 0.8271760889312220 0.5000000000000000

0.0003937666274036 0.3305873692268859 0.5000000000000000

0.8359789113241045 0.8365746499863844 0.5000000000000000

0.3336390105261560 0.3307801973880630 0.5000000000000000

0.1655912769368192 0.8353979121038356 0.5000000000000000

0.6656236540363307 0.3375780103241566 0.5000000000000000

0.5819996481199858 0.9132394419773345 0.5000000000000000

0.0792973993729224 0.4140800237277608 0.5000000000000000

0.7540196232650671 0.9118771864822711 0.5000000000000000

0.2504910369970253 0.4133315271564214 0.5000000000000000

0.1667236623944320 0.6751559572145662 0.5000000000000000

0.6678362010226189 0.1718085937824137 0.5000000000000000

0.9212057024020807 0.5861004848470330 0.5000000000000000

0.4221473168449349	0.0860897612498281	0.5000000000000000
0.7511974159566693	0.5855159622708153	0.5000000000000000
0.2473435665643866	0.0869010545742274	0.5000000000000000
0.3342180508105457	0.8309774814108917	0.5000000000000000
0.8340853254811904	0.3247419375904045	0.5000000000000000
0.4221473168449349	0.9139102387501717	0.5000000000000000
0.9212057024020807	0.4138995151529667	0.5000000000000000
0.2473435665643866	0.9130989454257726	0.5000000000000000
0.7511974159566693	0.4144840377291846	0.5000000000000000
0.8340853254811904	0.6752580624095957	0.5000000000000000
0.3342180508105457	0.1690225185891083	0.5000000000000000
0.0792973993729224	0.5859199762722392	0.5000000000000000
0.5819996481199858	0.0867605580226655	0.5000000000000000
0.2504910369970253	0.5866684728435785	0.5000000000000000
0.7540196232650671	0.0881228135177288	0.5000000000000000
0.6678362010226189	0.8281914062175864	0.5000000000000000
0.1667236623944320	0.3248440427854338	0.5000000000000000
0.4167035987726272	0.5000000000000000	0.5000000000000000
0.5543765209318791	0.5000000000000000	0.5000000000000000

S1. SISSO Model Computational Details

Using universal descriptors such as atomic number (M), atomic radius (R), d-electron count (N), d-band center (DBC), Bader charge (Q), Pauling electronegativity (X), electron affinity (EA), first ionization energy (IE), and van der Waals radius (r), an empirical formula was derived through the SISSO method.

SISSO (Sure Independence Screening and Sparsifying Operator) is an efficient high-dimensional data modeling method. Its computational process consists of two steps: first, it rapidly selects important features through correlation-based screening (SIS) to reduce data dimensionality; then, it constructs a concise linear or nonlinear model using the sparsifying operator (SO). This method ensures computational efficiency while achieving a model with strong interpretability and good predictive performance.

Feature Selection Using Linear Correlation Analysis

Consider a dataset containing n observations and p explanatory variables $x_1, x_2, \ldots, x_p$, paired with a target variable y. For continuous predictors x_j, their association strength with y is measured through Pearson's correlation coefficient w_j:

$$w_j = \frac{\sum_{i=1}^{n} (x_{ij} - \bar{x}_j)(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n} (x_{ij} - \bar{x}_j)^2 \sum_{i=1}^{n} (y_i - \bar{y})^2}}$$

Where $\bar{x}_j$ and $\bar{y}$ represent the mean values of x_j and y, respectively. Variables are then sorted by the magnitude of $|w_j|$, and the d most strongly correlated features ($d < p$) are retained for subsequent model development.

Model Formulation and Optimization Process

Following the SIS screening step, we establish the design matrix $X \in R^{n*d}$, the response vector $y \in R^{n*1}$, and the sparse coefficient vector $\beta \in R^{d*1}$. The optimization problem is defined by the following objective function:

$$L(\beta) = \frac{1}{2n} \|y - X\beta\|_2^2 + \lambda \|\beta\|_1$$

The first component quantifies the prediction error using the squared Euclidean norm:

$$\|y - X\beta\|_2^2 = \sum_{i=1}^{n} \left(y_i - \sum_{j=1}^{d} x_{ij} \beta_j \right)^2$$

The second component introduces sparsity through L1 regularization:

$$\|\beta\|_1 = \sum_{j=1}^{d} |\beta_j|$$

Here, λ serves as a tuning parameter that balances model accuracy and sparsity. The optimization procedure iteratively refines the coefficients until convergence, yielding a model that maintains predictive performance while promoting interpretability through feature sparsity.

Represented by the following regression equation:

$$\Delta G = -0.45(N2 \cdot \cos(Q1)) - \left(\frac{0.21N1 \cdot R2}{(IE1)^3}\right) + \frac{26R1^6 - 8 \times 10^{-17} M2 \cdot X2}{r1 \cdot DBC2} - 0.05$$

S2. Data set sampling principle

To obtain a structurally diverse and representative subset of dual-atom catalyst (DAC) configurations, we adopted a structure-based farthest-point sampling (FPS) strategy, following the principles introduced by Bartók et al. [1] and adapted here for DAC configuration selection. The full set of 729 DAC structures exhibits uneven coverage of the structural configuration space, with certain bonding motifs and coordination geometries more heavily represented than others. FPS aims for uniform coverage by iteratively selecting structures that are maximally dissimilar from those already chosen.

Given a selected set $S_m = \{A_i\}_{i=1}^{m}$ of DAC configurations, the next configuration A_{m+1} is determined by:

$$A_{m+1} = arg\max_{A \in D}\left[\min_{A' \in S_m} D(A,A')\right]$$

where Dd enotes the full dataset and $D(A,A')$ is a structural dissimilarity metric computed from low-cost structural fingerprints (e.g., metal pair identity, relative positions, local coordination environment). This procedure ensures that each newly selected configuration lies as far as possible from all previously chosen ones, thus maximizing geometric diversity in the absence of full descriptor information. Since this FPS variant relies solely on structural information, it provides a practical and efficient way to determine where to allocate high-cost DFT calculations while ensuring comprehensive coverage of the relevant structural space. For more details, please refer to [1].

S3. Details of Dataset Splitting, Model Training

1.Train/Test Split and Random Seed Settings

All machine-learning models in this study were constructed using the same dataset. For each reaction step, the available samples were randomly divided into training and test sets with an 8:2 ratio after shuffling the data. A fixed random seed (random_state = 42) was used during dataset splitting to ensure reproducibility.

2.Modeling Strategy for Reaction Pathways and Individual Steps

The three reaction pathways considered in this work were modeled separately. For each pathway, the adsorption free energies of intermediates or transition states were treated as independent learning targets. Accordingly, a standalone MLP regression model was trained for each reaction step (NN, NNH, NH_2N, N, NH, NH_2, NH_3).

Hyperparameter settings, including random seeds used during model training, are listed in the Supplementary Information. We emphasize that the random seed used for dataset splitting and the random seed used for MLP training are independent: the former controls data partitioning, whereas the latter governs the stochastic processes within the neural network during optimization. We believe this clarification, along with the detailed hyperparameter reporting, will substantially improve the reproducibility of our work.

3.Hyperparameter Search and Selection of Optimal Settings

A grid-search strategy was employed for hyperparameter optimization. The search space included variations in hidden-layer architecture, activation functions, regularization coefficients, and initial learning rates (e.g., hidden_layer_sizes = [(50,), (100,), (50, 50), (100, 50), (100, 100)], activation $\in$ {relu, tanh}, etc.). Cross-validation was used to select the hyperparameter set that yielded the best performance on the validation dataset.

References

[1] A. P. Bartok, S. De, C. Poelking, N. Bernstein, J. R. Kermode, G. Csanyi and M. Ceriotti, Sci Adv, 2017, 3, e1701816.