Supplementary materials

Performance of Nitrogen Reduction Reaction on Metal Bound g-C_6N_6: Combined Approach of Machine Learning and DFT

Moumita Mukherjee[a], Sayan Dutta[a], Madhusudan Ghosh[b], Partha Basuchowdhuri[b], Ayan Datta[a*]*

[a]*School of Chemical Sciences, Indian Association for the Cultivation of Science, Kolkata.*

[b]*School of Mathematical and Computational Science, Indian Association for the Cultivation of Science, Kolkata*

Table S1: List of Hyperparameters for the Regression Model and Classification model.

Models		Hyperparameters
Regression model	**GBR**	n_estimators = 100 learning_rate = 0.1 max_depth = 3 min_sample_split = 2 min_sample_leaf = 1
	RFR	n_estimators = 100 min_sample_split = 2 min_sample_leaf = 1
Classification Model	**Logistic Regression**	regularization parameter = 0.01, 0.1, 1.0, 10.0, 100.0

	Support Vector Machine	regularization parameter = 2, 3, 4, 5 Kernal = rbf, linear
	Random Forest Classifier	Number of trees = 10, 50, 100, 200 Max_depth = 2, 3, 4, 5

Here, n_estimators determine the maximum number of trees; max_depth estimates the depth of each individual trees in the forest; min_sample_splits is the representative of minimum number of samples required to split an internal node; min_samples_leaf is the minimum number of samples required to be at a leaf node.

In this study, we have used two different strategies, namely, Bootstrap aggregating or bagging and boosting for machine learning.

Bagging: Given a training set, bagging generates a number of new training subsets, by sampling from the original training set uniformly and with replacement. Suppose m training subsets are generated. Then m models are fitted using these m samples, and the final prediction is done by averaging the results. Consider the case of Random Forest (RF). Decision trees that are very deep often overfit the training set, *i.e*. they have low bias, but high variance (i.e., they are flexible). An RF averages multiple (decorrelated) deep decision trees, trained on different parts of the same training set. This should lead to decreased variance because some of the variance should cancel out (at least in theory). This can come at the cost of a slight increase in bias.

Boosting: In case of boosting, we build a strong classifier by combining multiple "weak learners" that have high bias but low variance (i.e., they are inflexible). The weak learners are added sequentially to the strong classifier. After a weak learner is added, the misclassified data in the training set are given higher weights and the

correctly classified data are given smaller weights so that future weak learners can concentrate on the misclassified data in the training set. Boosting, thus, reduces bias without increasing the variance. In fact, it has been observed that boosting can also reduce the variance.

Table S2: Adsorption energy of Metals in C_6N_6 substrate at different position.

Catalyst	Metal adsorption energy (eV)	
	Position_1	**Position_2**
Sc-C_6N_6	-416.426	-417.928
Ti-C_6N_6	-414.603	-417.779
Cr-C_6N_6	-414.273	-418.342
Mn-C_6N_6	-414.435	-417.860
Co-C_6N_6	-414.667	-414.678
Ni-C_6N_6	-413.577	-413.581
Cu-C_6N_6	-411.760	-411.761
Nb-C_6N_6	-415.012	-419.126
Mo-C_6N_6	-414.452	-418.375
Tc-C_6N_6	-416.824	-417.319
Ru-C_6N_6	-412.561	-416.149
Hf-C_6N_6	-415.954	-420.643
Ta-C_6N_6	-416.143	-420.345
W-C_6N_6	-415.973	-419.483
Re-C_6N_6	-415.524	-418.317
Ir-C_6N_6	-415.745	-415.933
Pt-C_6N_6	-413.454	-413.497
Au-C_6N_6	-410.916	-410.920

There are four possible binding sites available on M@C6N6. After optimization, the metal atoms on position_ 3 and 4 coincide with position_2. Therefore, comparing the energy values obtained for position_1 and position_2, we find that the adsorption energy is higher in position_2 than position_1.

Table S3: Adsorption energy of Nitrogen on metal catalyst in two modes.

Catalyst	N_2 adsorption energy (eV) (E_{ads})	
	End-On	Side-On
Sc-C_6N_6	-0.80	-0.48
Ti-C_6N_6	-1.05	-0.61
Cr-C_6N_6	-0.47	Not stable
Mn-C_6N_6	-0.43	-0.09
Co-C_6N_6	-0.98	-0.42
Ni-C_6N_6	-0.67	-0.64
Cu-C_6N_6	-0.74	-0.73
Nb-C_6N_6	-1.31	-0.82
Mo-C_6N_6	-0.92	-0.40
Tc-C_6N_6	-0.90	-0.40
Ru-C_6N_6	-1.05	-0.33
Hf-C_6N_6	-1.33	-0.88
Ta-C_6N_6	-1.36	-0.84
W-C_6N_6	-1.33	-0.71
Re-C_6N_6	-0.96	-0.21
Ir-C_6N_6	-1.07	-0.90
Pt-C_6N_6	-0.70	-0.61
Au-C_6N_6	-0.21	-0.21

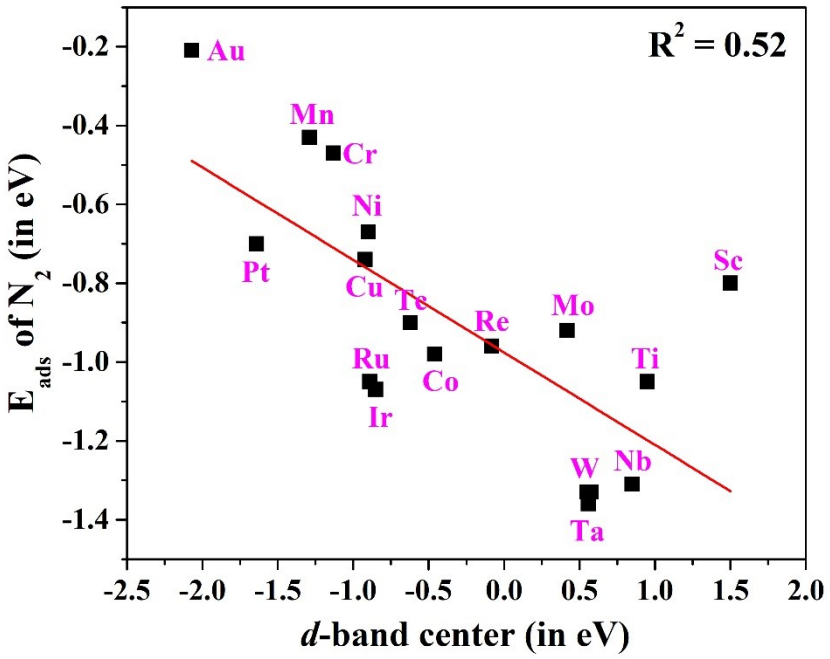

Figure S1: The relation between *d*-band center and adsorption energy of N_2.

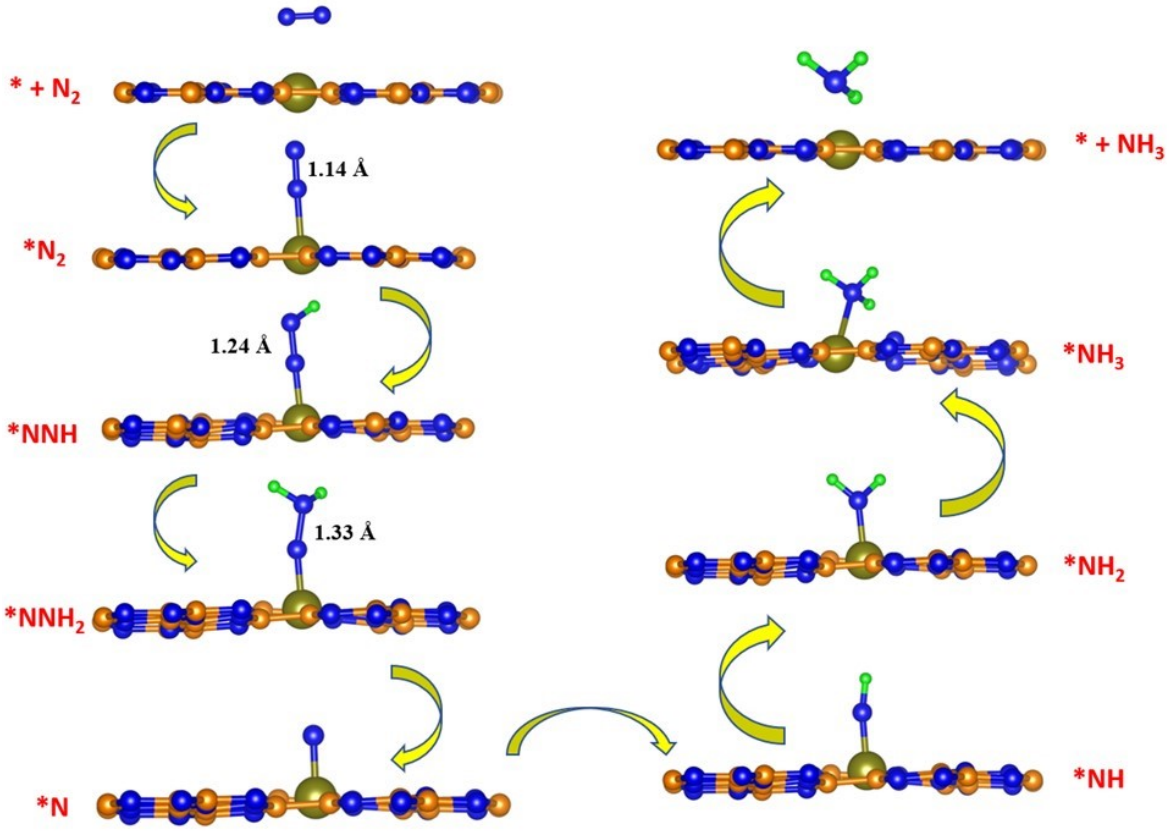

Figure S2: Optimized structures of different intermediates on $Ta@C_6N_6$.

Table S4: Feature statistics of the dataset used for machine learning method.

Metals doped in Substrate	r_{cov}	d_{M-N}	ε_d	q	EA	IE	EN	d_z	θ_d
Sc	1.7	2.13	1.50	1.33	0.188	6.54	1.36	0.31	1
Ti	1.6	2.06	0.95	2.6	0.075	6.82	1.54	0.28	2
Cr	1.39	1.93	-1.13	4.75	0.675	6.76	1.66	0.25	5
Mn	1.39	2.08	-1.29	5.72	0.500	7.435	1.55	0.23	5
Co	1.26	1.94	-0.46	8.13	0.662	7.86	1.88	0.21	7
Ni	1.24	1.93	-0.90	9.21	1.157	7.635	1.91	0.195	8
Cu	1.32	2.04	-0.92	10.25	1.24	7.73	1.9	0.19	10
Nb	1.64	2.12	0.85	11.33	0.917	6.31	1.6	0.51	4
Mo	1.54	2.01	0.42	4.77	0.747	7.099	2.16	0.49	5
Tc	1.47	0.83	-0.62	5.9	0.55	7.28	1.99	0.455	5
Ru	1.42	1.88	-0.89	7.24	1.046	7.37	2.22	0.45	7
Hf	1.75	2.16	0.55	2.35	0.178	6.824	1.3	0.63	2
Ta	1.7	2.077	0.56	3.42	0.323	7.887	1.5	0.605	3
W	1.62	1.99	0.58	4.63	0.816	7.98	2.36	0.59	5
Re	1.51	1.95	-0.08	5.79	0.06	7.877	1.9	0.565	5
Ir	1.41	1.87	-0.85	8.32	1.564	9.121	2.22	0.526	7
Pt	1.36	1.96	-1.64	9.39	2.125	9.0	2.28	0.51	8
Au	1.36	2.53	-2.07	10.49	2.308	9.225	2.54	0.488	10

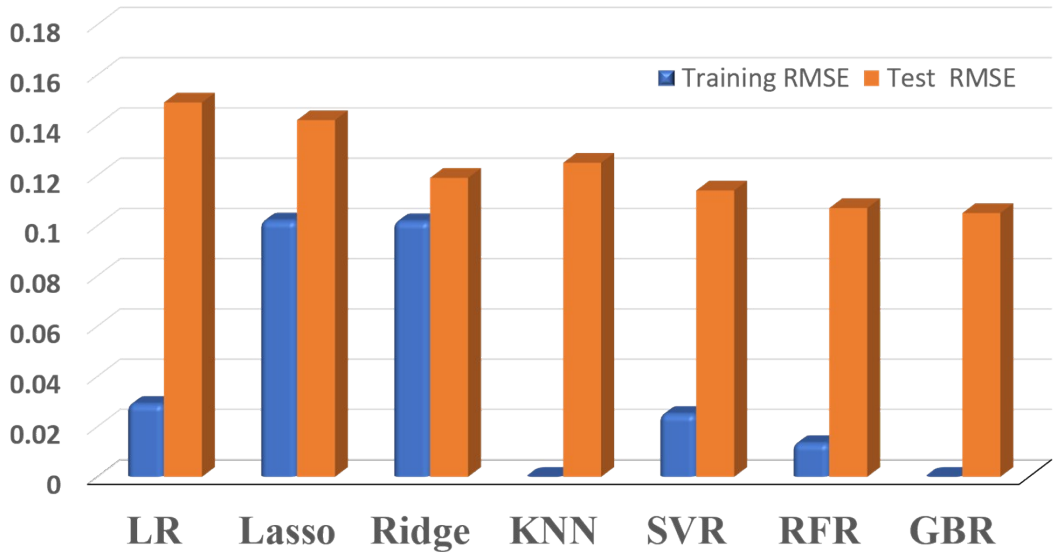

Figure S3. Comparison of Training and Test RMSE over average of 50 cross validations

LR: Linear regression

Lasso: Lasso regression

Ridge: Ridge regression

KNN: K-nearest neighbor regression

SVR: Support vector regression

RFR: Random Forest regression

GBR: Gradient boost regression

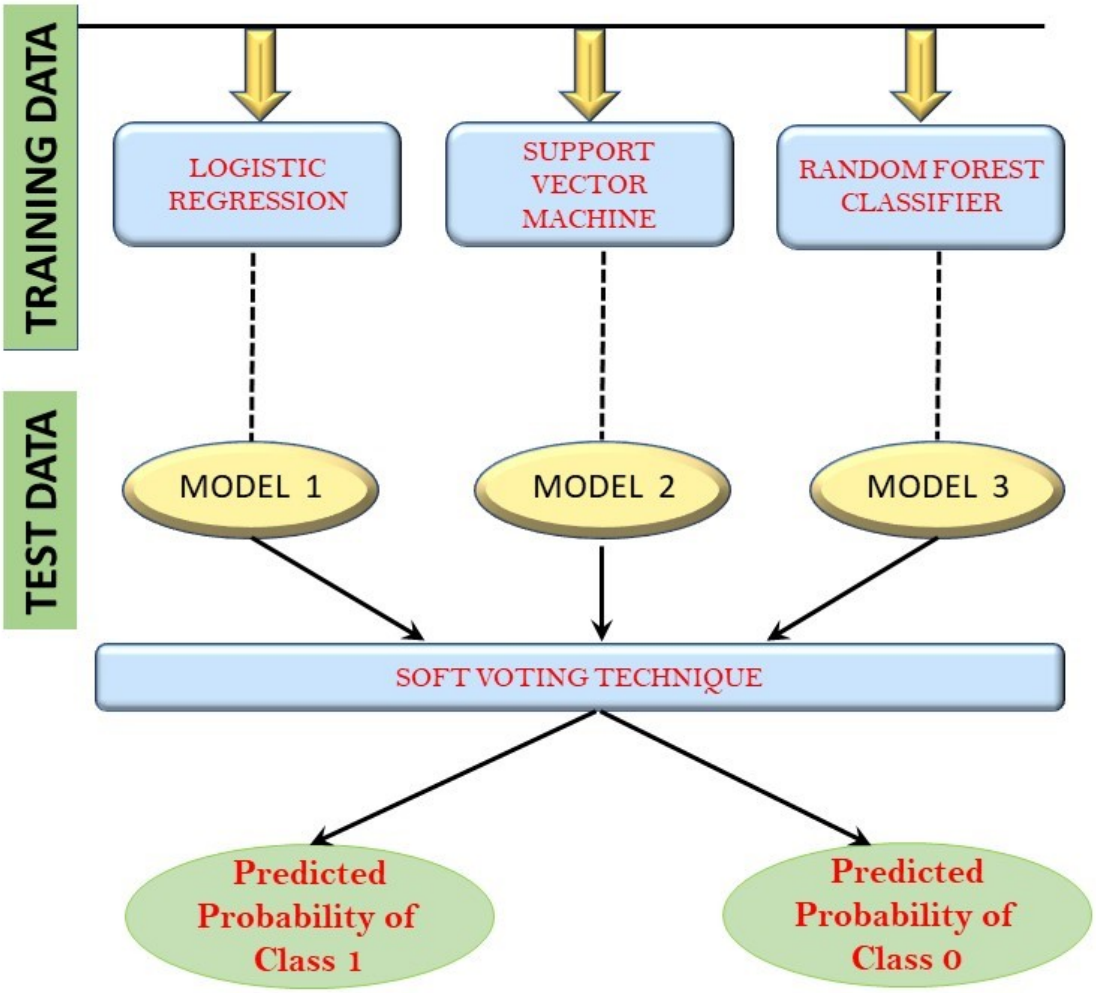

Figure S4. Schematic diagram of work-flow for the classification model.

Table S5: Cross validation result for different classification models.

No. of Distribution	Metals	Predicted probability by			
		Voting Classifier	Logistic Regression	Support Vector Machine	Random Forest Classifier
1	**Ta**	0.710	0.419	0.810	0.900
	Ti	0.357	0.416	0.254	0.400
	Ni	0.243	0.431	0.198	0.100
	Pt	0.361	0.415	0.270	0.400
2	**Ta**	0.713	0.335	0.904	0.900
	Nb	0.643	0.334	0.695	0.900
	Pt	0.217	0.242	0.209	0.200
	Ni	0.146	0.430	0.105	0.100
3	**Ta**	0.520	0.251	0.609	0.700
	Co	0.127	0.248	0.134	0.100
	W	0.506	0.251	0.367	0.900
	Hf	0.426	0.252	0.526	0.500
4	**Cr**	0.243	0.332	0.099	0.300
	Pt	0.227	0.332	0.149	0.200
	Ta	0.587	0.335	0.826	0.600
	Re	0.417	0.334	0.519	0.400
5	**Ti**	0.327	0.417	0.064	0.500
	Nb	0.691	0.419	0.754	0.900
	Tc	0.554	0.398	0.747	0.500
	Ru	0.404	0.416	0.598	0.200
6	**Re**	0.505	0.417	0.597	0.500
	Co	0.197	0.411	0.082	0.100
	Ni	0.206	0.410	0.107	0.100
	Sc	0.442	0.419	0.309	0.600

7	**Re**	0.526	0.334	0.545	0.700
	Sc	0.336	0.336	0.174	0.500
	Nb	0.557	0.336	0.735	0.600
	Mn	0.106	0.172	0.046	0.100
8	**Mo**	0.323	0.334	0.437	0.200
	W	0.461	0.335	0.549	0.500
	Pt	0.205	0.332	0.185	0.100
	Ru	0.183	0.317	0.217	0.100
9	**Tc**	0.331	0.333	0.361	0.300
	W	0.597	0.334	0.555	0.900
	Re	0.520	0.344	0.516	0.700
	Ti	0.328	0.364	0.120	0.500
10	**Nb**	0.587	0.335	0.626	0.800
	Re	0.477	0.334	0.597	0.500
	Cr	0.156	0.325	0.043	0.100
	Pt	0.274	0.332	0.091	0.400
11	**Co**	0.277	0.468	0.037	0.300
	Ru	0.453	0.480	0.410	0.400
	Mn	0.281	0.498	0.047	0.300
	Cr	0.315	0.480	0.047	0.400
12	**Hf**	0.726	0.420	0.657	0.900
	Mn	0.157	0.414	0.030	0.100
	Ni	0.570	0.412	0.029	0.000
	Cr	0.258	0.414	0.022	0.100
13	**Ni**	0.134	0.365	0.056	0.100
	Ti	0.625	0.462	0.837	0.700
	Mo	0.586	0.423	0.732	0.600
	Cr	0.413	0.434	0.726	0.100
14	**Tc**	0.486	0.415	0.445	0.600
	Ni	0.300	0.383	0.516	0.100
	Hf	0.458	0.446	0.429	0.500
	Cr	0.386	0.424	0.436	0.300
15	**Tc**	0.394	0.334	0.556	0.500
	Mn	0.262	0.331	0.076	0.100
	Nb	0.466	0.335	0.651	0.700
	Hf	0.437	0.336	0.621	0.600
16	**Mn**	0.237	0.414	0.096	0.200
	Pt	0.294	0.413	0.168	0.300

	Mo	0.411	0.417	0.417	0.400
	Cr	0.193	0.414	0.414	0.100
17	**Mo**	0.463	0.417	0.472	0.500
	Pt	0.266	0.415	0.183	0.200
	Co	0.170	0.412	0.097	0.100
	Ti	0.391	0.416	0.157	0.600
18	**Ni**	0.163	0.411	0.077	0.100
	Ti	0.408	0.417	0.108	0.700
	Mo	0.491	0.417	0.457	0.600
	Cr	0.191	0.414	0.060	0.100
19	**Re**	0.331	0.251	0.441	0.300
	Au	0.182	0.248	0.298	0.100
	Nb	0.343	0.251	0.484	0.300
	Mo	0.181	0.250	0.293	0.100
20	**Tc**	0.586	0.416	0.744	0.600
	Ni	0.157	0.412	0.059	0.100
	Nb	0.570	0.420	0.789	0.500
	Cr	0.258	0.414	0.059	0.300