Electronic Supplementary Information

Synergistic Enhancement of Hydrogen Evolution on MoS_2 by Sulfur Vacancy and Transition Metal Doping: A DFT and Machine Learning Study

Xin Fan, Xiaocheng Zhou, and Yafei Li**

Jiangsu Key Laboratory of New Power Batteries, Jiangsu Collaborative Innovation Centre of Biomedical Functional Materials, School of Chemistry and Materials Science, Nanjing Normal University, Nanjing 210023, China

*E-mail: xczhou@njnu.edu.cn (XZ) and liyafei@njnu.edu.cn (YL)

Contents:

1. Computational Details
2. Supporting Fig. S1−S13
3. Supporting Table S1–S8

1. Computational Details

1.1 The charge-extrapolation method

The charge-extrapolation method for determining constant potential barriers from a constant charge calculation is based upon two assumptions: (1) The chemical and electrostatic contributions to reaction energetics are separable. (2) The electrostatic contributions to proton−electron transfer are described by a basic capacitor model.[1]

The total energy change between two states, e.g., 1 for the initial state (IS) and 2 for the transition state (TS), can be defined as:

$$E_2(\Phi_1) - E_1(\Phi_1) = E_2(\Phi_2) - E_1(\Phi_1) + \frac{(q_2 - q_1)(\Phi_2 - \Phi_1)}{2} \quad \text{(S1)}$$

where q represents the interface charge and $E(\Phi)$ is the DFT energy of the state at the electrode potential. The Φ term is the work function, which is calculated from the local potential and is given by the following equation:

$$\Phi = E_{vac} - E_F \quad \text{(S2)}$$

where E_{vac} and E_F represent the vacuum level and the Fermi level, respectively. The value of q_2-q_1 is derived from the Bader analysis. The term $E_2(\Phi_2) - E_1(\Phi_1)$, taking 1 for IS and 2 for TS as an example, represents the DFT energy difference between IS and TS.

1.2 SISSO algorithm

SISSO (Sure Independence Screening and Sparsifying Operator) is a data-driven approach that combines symbolic regression and compressed sensing techniques.[2] This methodology aims to extract key information from large datasets that effectively characterises target attributes, thereby establishing an accurate and interpretable explicit mathematical model.

The key to this method lies in constructing a vast initial feature space $\{\phi_i\}$. This space is generated by recursively applying a series of mathematical operators to the initial physicochemical features of the input. The introduction of these operators—including addition, subtraction, multiplication, division, exponentiation, and root extraction—aims to systematically explore nonlinear combinations among the original features, thereby providing a rich pool of candidate descriptors for the subsequent screening stages.

To identify the most effective descriptors from a vast feature space, the SISSO algorithm primarily executes through two core steps. First, the Sure Independence Screening (SIS) establishes the correlation between each candidate feature and the target property. Through rapid sorting and preliminary screening, the feature space dimension is compressed to a manageable scale commensurate with the dataset size. Subsequently, the Sparsifying Operator

(SO) refines the selected feature subset by solving an L0-regularised sparse regression problem. This process identifies the minimal number of key descriptors required, ensuring model sparsity and high predictive accuracy.

In this work, we employ key features obtained through feature engineering as inputs and the hydrogen adsorption free energy (ΔG_{H*}) as the target variable to train the SISSO framework. Ultimately, the algorithm outputs a linear model composed of these selected descriptors. This model not only demonstrates outstanding predictive performance, but its mathematical formulation also directly reveals the key physicochemical factors influencing hydrogen adsorption free energy.

2. Supporting Fig. S1-S13

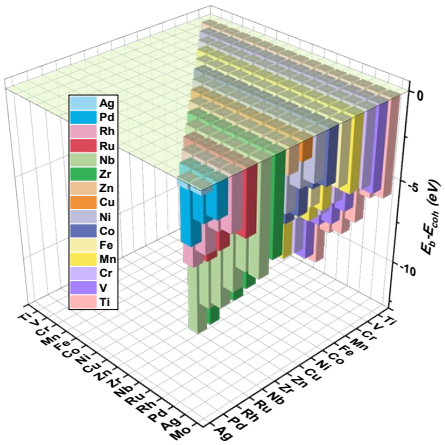


Fig. S1. The difference between binding and cohesive energy ($E_b - E_{coh}$) for TM@S_V-MoS_2.

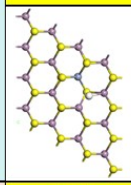
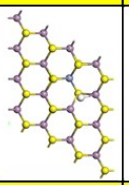
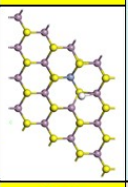
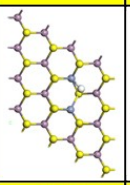
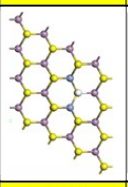
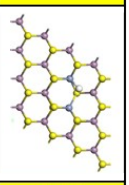
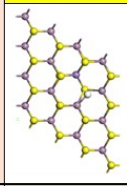
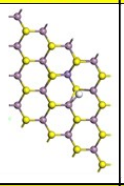
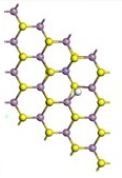
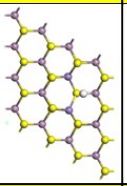
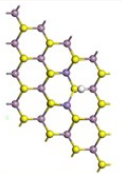
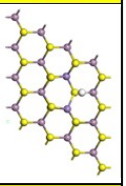
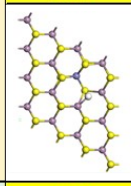
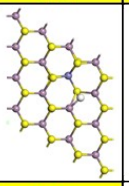
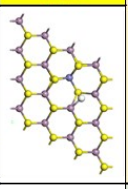
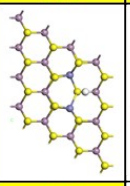
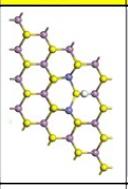
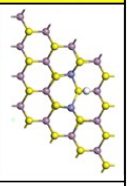
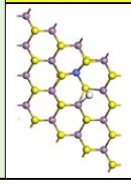
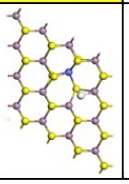
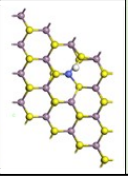
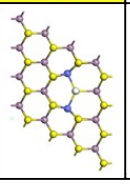
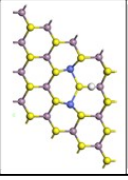
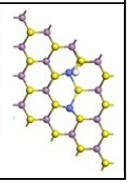
Systems	S_V	Mo	TM	Systems	S_V	Mo	TM
$1Cr@S_V\text{-}MoS_2$	-0.49	-0.49	-0.49	$2Cr@S_V\text{-}MoS_2$	-0.38	-0.38	-0.38
							
$1Mn@S_V\text{-}MoS_2$	-0.23	-0.22	-0.20	$2Mn@S_V\text{-}MoS_2$	-0.10	-0.10	-0.10
							
$1Fe@S_V\text{-}MoS_2$	-0.26	-0.26	-0.26	$2Fe@S_V\text{-}MoS_2$	0.19	0.20	0.19
							
$1Co@S_V\text{-}MoS_2$	-0.20	-0.24	0.65	$2Co@S_V\text{-}MoS_2$	0.86	0.96	1.04
							

Fig. S2. Investigation of H adsorption sites on $TM@S_V\text{-}MoS_2$ systems.

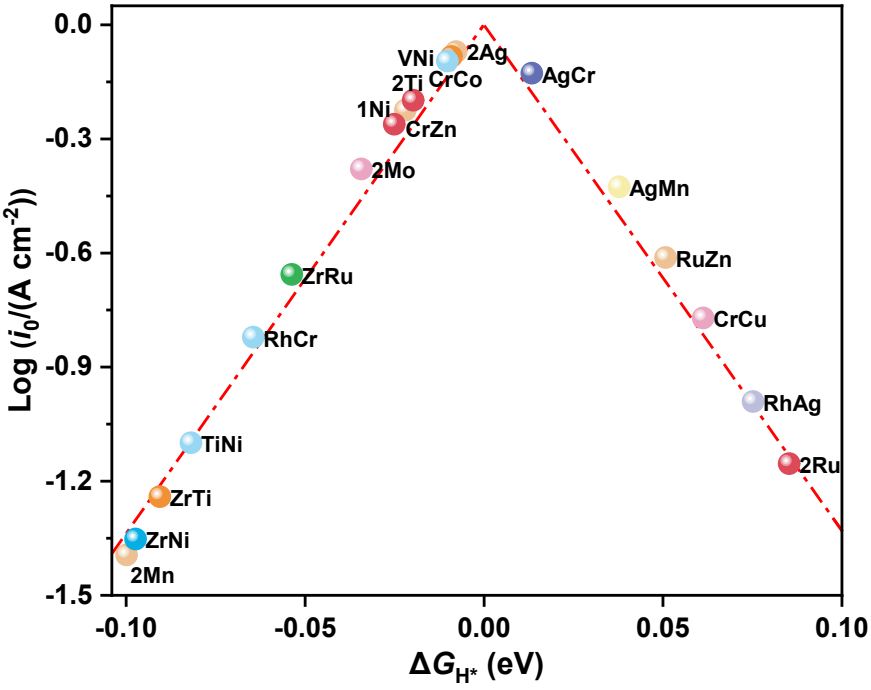


Fig. S3. Volcano plot correlating the exchange current density (i_0) with ΔG_{H^*}, where ΔG_{H^*} ranges from -0.1 to 0.1 eV.

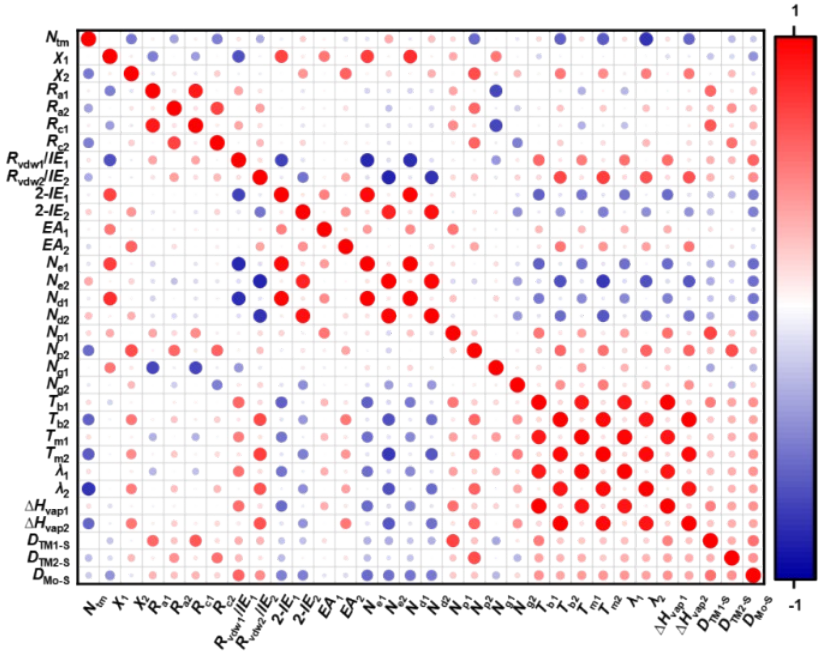


Fig. S4. A heat map of the Pearson correlation coefficient matrix between 32 initial features.

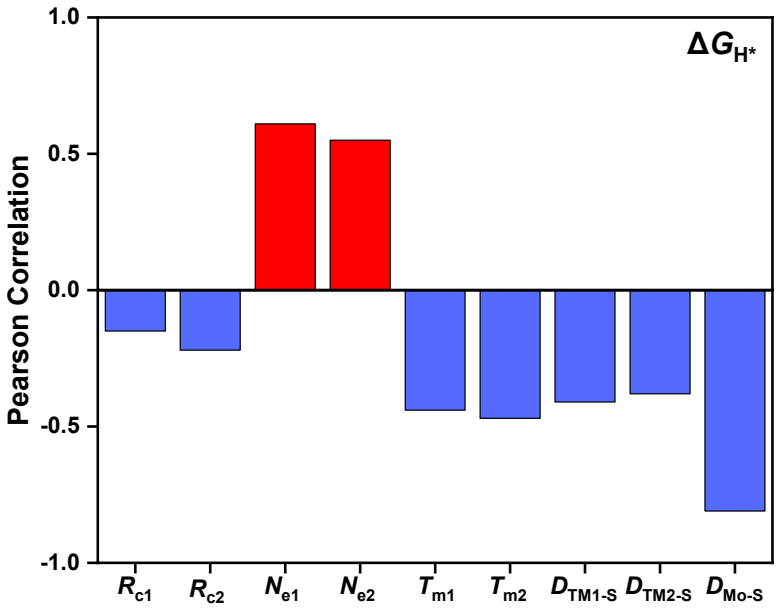


Fig. S5. Pearson correlation plot between the optimized features and the target value.

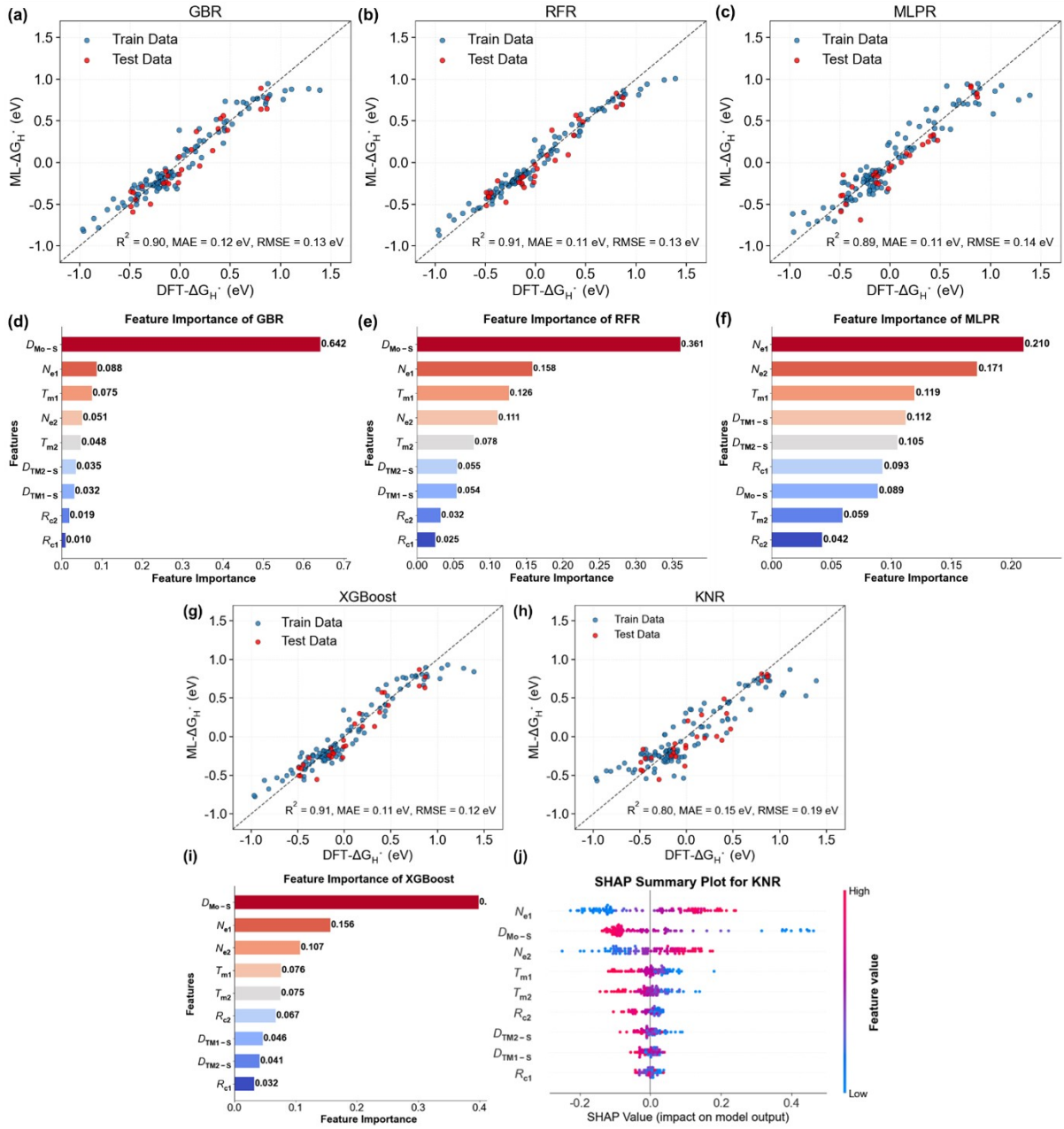


Fig. S6. Comparison of DFT-calculated and SVR-predicted ΔG_{H*} values for GBR, RFR, MLPR, XGBoost and KNR models, and the analysis of feature importance.

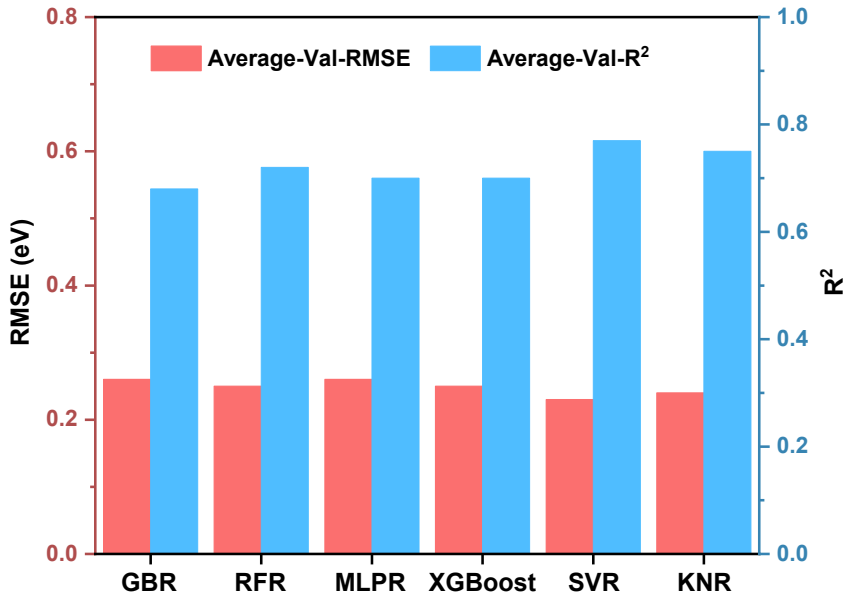


Fig. S7. The average RMSE and R^2 of GBR, RFR, MLPR, XGBoost and KNR models after 5-fold cross-validation.

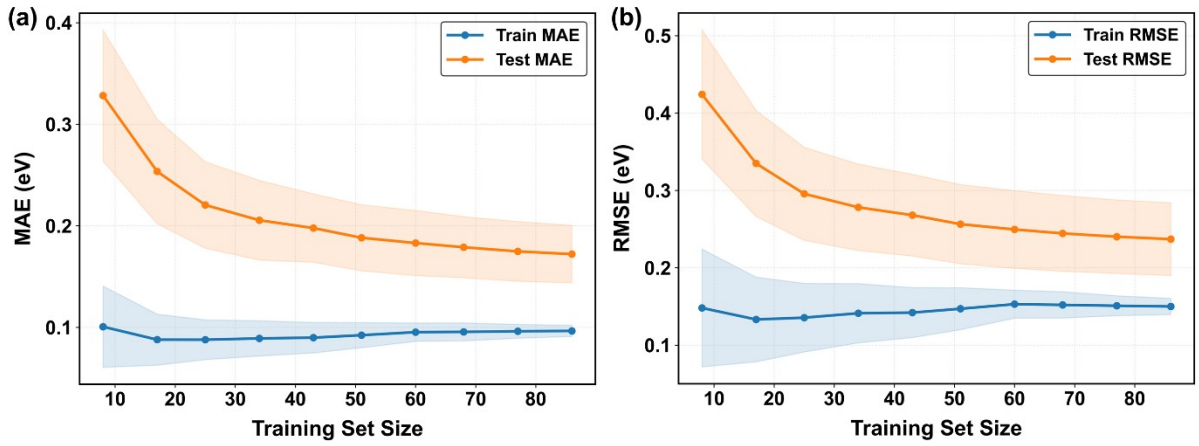


Fig. S8. Learning curve analysis of the SVR model. Evolution of (a) MAE and (b) RMSE with increasing training set size.

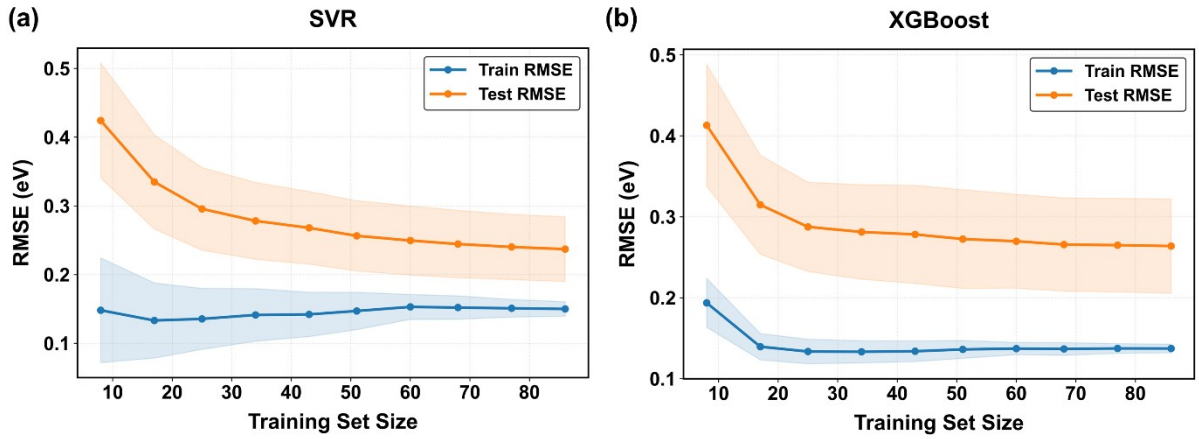


Fig. S9. Learning curve analysis of the (a) SVR model and (b) XGBoost model.

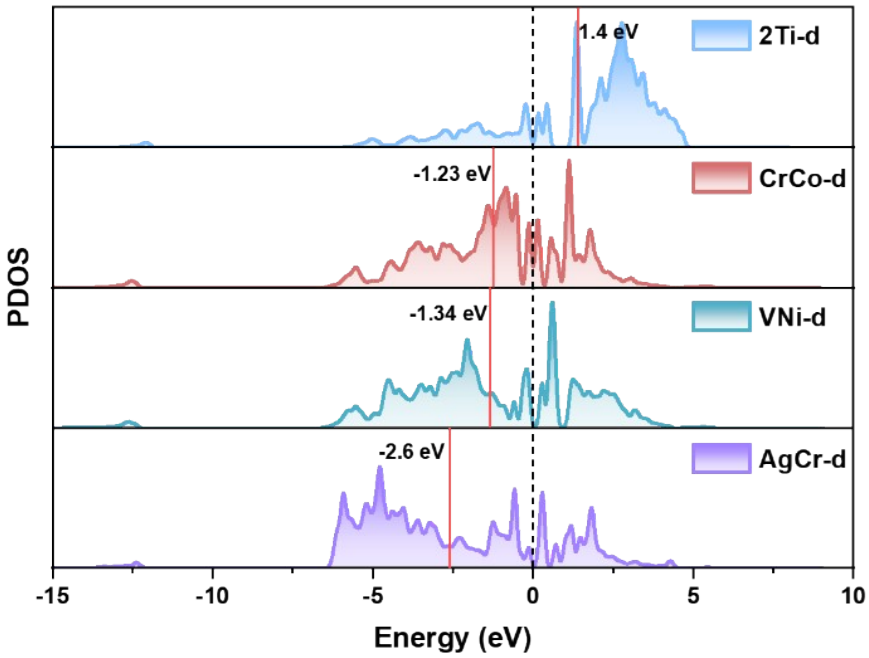


Fig. S10. PDOS of 2Ti@S_V-MoS_2, CrCo@S_V-MoS_2, VNi@S_V-MoS_2, and AgCr@S_V-MoS_2. The corresponding d-band center values are indicated.

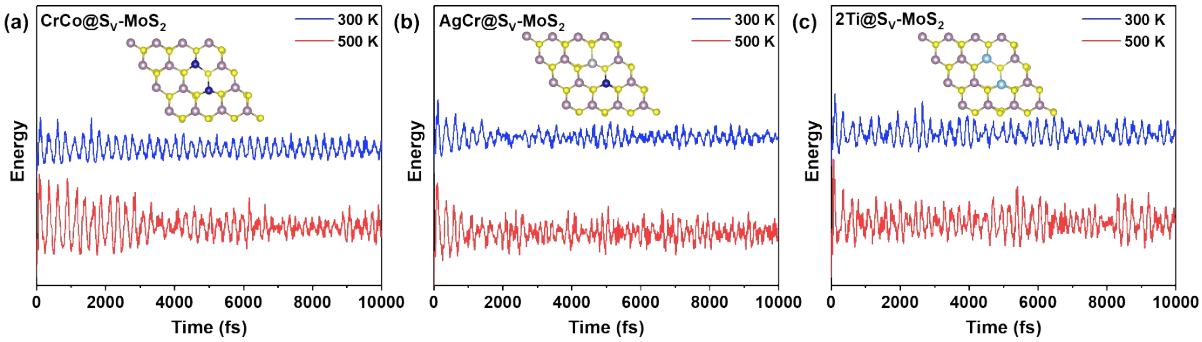


Fig. S11. AIMD of CrCo@S_V-MoS_2, AgCr@S_V-MoS_2 and 2Ti@S_V-MoS_2 at both 300 K and 500 K.

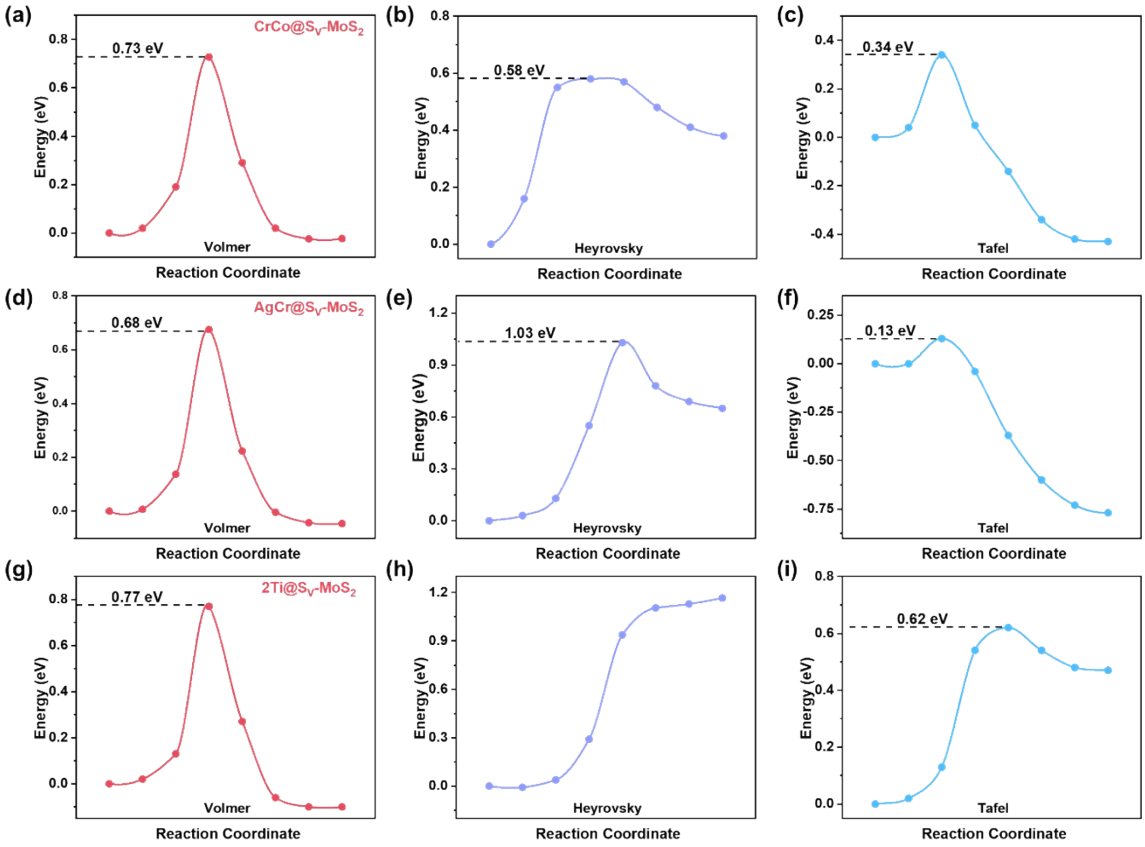


Fig. S12. CINEB of (a-c) CrCo@S_V-MoS_2, (d-f) AgCr@S_V-MoS_2, and (g-i) 2Ti@S_V-MoS_2.

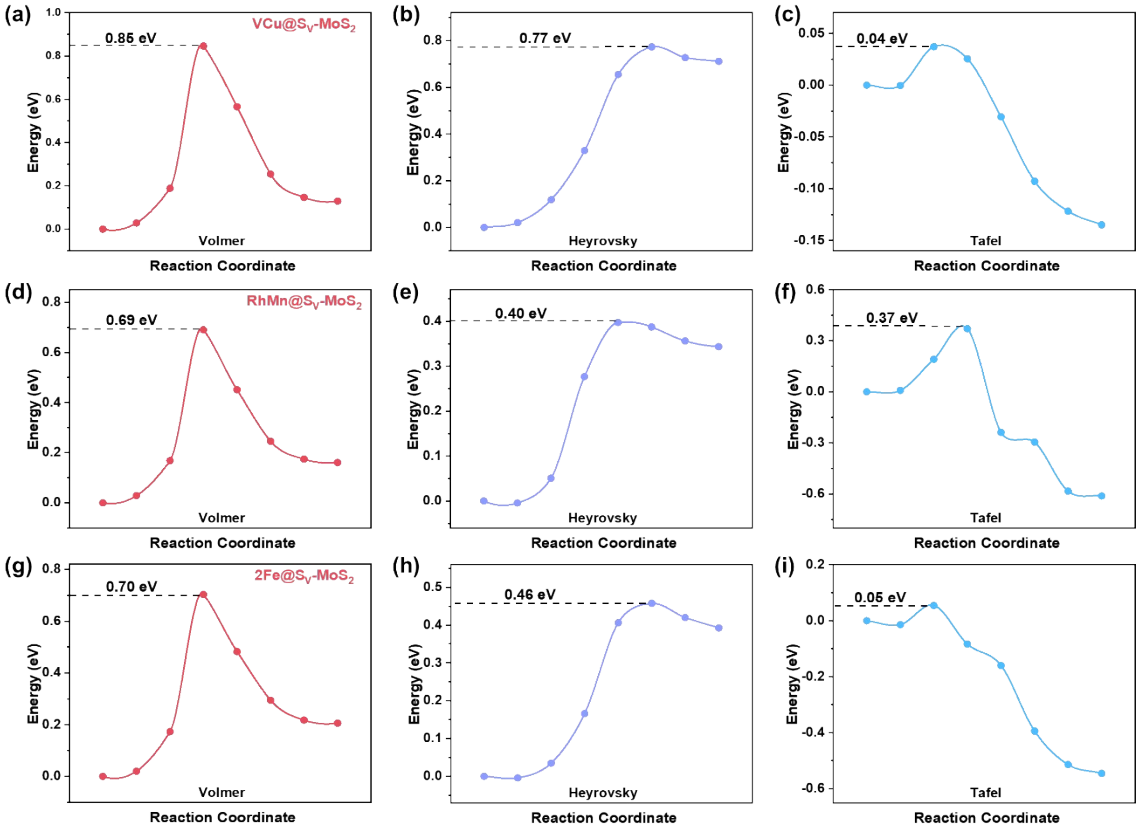


Fig. S13. CI-NEB of (a-c) VCu@S_V-MoS_2, (d-f) RhMn@S_V-MoS_2, and (g-i) 2Fe@S_V-MoS_2.

3. Supporting Table S1-S8

Table S1. Comparison of $|\Delta G_{H^*}|$ for TM@S_V-MoS_2 in 4×4×1 and 5×5×1 supercells. The supercell size has a limited impact on adsorption energy.

Materials	$\|\Delta G_{H^*}\|$ (4×4×1)/eV	$\|\Delta G_{H^*}\|$ (5×5×1)/eV
2Ti@S_V-MoS_2	0.0198	0.0067
AgCr@S_V-MoS_2	0.0134	0.0316
CrCo@S_V-MoS_2	0.0102	0.0166
VNi@S_V-MoS_2	0.0091	0.0094

Table S2. The initial features used in training model and intrinsic descriptor of atoms.

Features	
Doped transition metal number	N_{tm}
Pauling electronegativity	χ
Metal atomic radius	R_{a}
Covalent radius	R_{c}
Van der Waals radius	R_{vdW}
The first ionization energy	IE
The second ionization energy	2-IE
Combined feature	R_{vdW}/IE
Electronic Affinity	EA
Valence electron number	N_{e}
d electron number	N_{d}
Period number	N_{p}
Group number	N_{g}
Boiling point	T_{b}
Melting point	T_{m}
Fusion heat	λ
Vaporization heat	ΔH_{vap}
Distance between TM_1 and S	$D_{\mathrm{TM1\text{-}S}}$
Distance between TM_2 and S	$D_{\mathrm{TM2\text{-}S}}$
Distance between Mo and S	$D_{\mathrm{Mo\text{-}S}}$

Table S3. The hyperparameters of all models used in this work.

Algorithm	Hyperparameters
GBR	random_state=0, learning_rate=0.03, max_depth=3, n_estimators=100, subsample=0.8, min_samples_leaf=5, min_samples_split=20
RFR	random_state=0, n_estimators=120, max_depth=5,min_samples_leaf=1, min_samples_split=4,max_features='sqrt'
MLPR	random_state=0, activation='logistic', hidden_layer_sizes=(4,), learning_rate_init=0.1, max_iter=500, alpha=0.01
XGBoost	random_state=0, learning_rate=0.1, n_estimators=200, max_depth=3, gamma=0.1, min_child_weight=3, subsample=0.7, colsample_bytree=0.8
SVR	kernel='rbf', C=0.5, epsilon=0.05, gamma=0.1
KNR	n_neighbors=8

Table S4. Descriptors identified in the SISSO model.

	d_1	d_2	d_3	c_1	c_2	c_3	c_0	R^2
1D-1F	$N_{e1} \times N_{e2}$			0.01344			-0.81394	0.69
1D-2F	$\frac{Ne1 \times Ne2}{Rc2}$			0.01819			-0.82931	0.74
1D-3F	$\frac{Ne1 \times Ne2}{Rc2 \times DMo}$			0.04020			-0.80016	0.76
2D-1F	$N_{e1} \times N_{e2}$	$\lvert T_{m1}\text{-}T_{m2}\rvert$		0.01385	-0.00026		-0.68977	0.84
2D-2F	$\frac{Ne1 \times Ne2}{Rc2}$	$\frac{\lvert Tm1 - Tm2\rvert}{Rc2}$		0.01856	-0.00031		-0.71179	0.86
2D-3F	$\frac{Ne1 \times Ne2}{\sqrt{Tm1}}$	$T_{m1} + \lvert T_{m1}\text{-}T_{m2}\rvert$		0.00033	-0.00029		-0.10455	0.85
3D-1F	$N_{e1} \times N_{e2}$	$DMo - S^6$	$\lvert T_{m1}\text{-}T_{m2}\rvert$	0.01001	-0.00710	-0.00022	0.61880	0.87
3D-2F	$\frac{DMo - S^3}{Ne2}$	$\frac{\exp(-Ne2)}{Ne1}$	$(N_{e1}\text{-}N_{e2}) \times T_{m1}$	-1.2289	438.19984	0.00006	2.05812	0.92

Table S5. Changes in energy E, work function Φ, charge q, the extrapolated reaction energy ΔE and activation energy E_a. The energy difference between E_{TS} and E_{IS} (E_{TS}–E_{IS}) represents the barriers obtained from the CI-NEB calculations.

Materials	Reaction	E_{FS}–E_{IS}	Φ_{FS}–Φ_{IS}	E_{TS}–E_{IS}	Φ_{TS}–Φ_{IS}	q_{FS}–q_{IS}	q_{TS}–q_{IS}	ΔE	E_a/eV
2Ti@ S_V-MoS_2	Vlomer	-0.10	1.04	-1.23	-0.74	0.77	0.66	-0.57	0.58
	Heyrovsky	/	/	/	/	/	/	/	/
AgCr@ S_V-MoS_2	Vlomer	-0.05	0.79	-1.09	-0.47	0.68	0.45	-0.50	0.56
	Heyrovsky	0.65	0.76	-0.50	0.46	1.03	0.28	-0.18	1.01
CrCo@ S_V-MoS_2	Vlomer	-0.02	0.60	-1.10	-0.35	0.73	0.30	-0.54	0.65
	Heyrovsky	0.38	1.04	-0.50	0.12	0.58	1.29	-0.51	0.25
VNi@ S_V-MoS_2	Vlomer	-0.16	0.76	-1.14	-0.59	0.66	0.43	-0.54	0.55
	Heyrovsky	0.56	0.87	-0.45	0.37	0.68	0.70	-0.39	0.54

Table S6. Changes in energy E, work function Φ, charge q, the extrapolated reaction energy ΔE and activation energy E_a. The energy difference between E_{TS} and E_{IS} (E_{TS}–E_{IS}) represents the barriers obtained from the CI-NEB calculations.

Materials	Reaction	E_{FS}–E_{IS}	Φ_{FS}–Φ_{IS}	E_{TS}–E_{IS}	Φ_{TS}–Φ_{IS}	q_{FS}–q_{IS}	q_{TS}–q_{IS}	ΔE	E_a/eV
VCu@ S_V-MoS_2	Vlomer	0.13	0.79	-1.14	-0.32	0.85	0.45	-0.56	0.72
	Heyrovsky	0.71	0.71	-0.42	0.56	0.77	0.54	-0.32	0.69
RhMn@ S_V-MoS_2	Vlomer	0.16	0.80	-1.01	-0.24	0.69	0.41	-0.47	0.59
	Heyrovsky	0.34	0.73	-0.47	0.17	0.40	0.54	-0.35	0.30
2Fe@ S_V-MoS_2	Vlomer	0.21	0.80	-0.99	-0.19	0.70	0.40	-0.47	0.61
	Heyrovsky	0.39	0.69	-0.49	0.22	0.46	0.53	-0.38	0.36

Table S7. The initial features used in training model and intrinsic descriptor of atoms.

Metal	χ	R_a (Å)	R_c (Å)	R_{vdw} (Å)	N_e	N_d	N_p	N_g
Ti	1.54	1.40	1.36	2.46	4	2	4	4
V	1.63	1.35	1.25	2.42	5	3	4	5
Cr	1.66	1.40	1.27	2.45	6	5	4	6
Mn	1.55	1.40	1.39	2.45	7	5	4	7
Fe	1.83	1.40	1.25	2.44	8	6	4	8
Co	1.88	1.35	1.26	2.40	9	7	4	8
Ni	1.91	1.35	1.21	2.40	10	8	4	8
Cu	1.90	1.35	1.38	2.38	11	10	4	1
Zn	1.65	1.35	1.31	2.39	12	10	4	2
Zr	1.33	1.55	1.48	2.52	4	2	5	4
Nb	1.60	1.45	1.37	2.56	5	4	5	5
Mo	2.16	1.45	1.45	2.45	6	5	5	6
Ru	2.20	1.30	1.26	2.46	8	7	5	8
Rh	2.28	1.35	1.35	2.44	9	8	5	8
Pd	2.20	1.40	1.31	2.15	10	10	5	8
Ag	1.93	1.60	1.53	2.53	11	10	5	1

Metal	*IE* (eV)	*IE*-2(eV)	*EA*(eV)	T_b(K)	T_m(K)	λ(kJ/mol)	ΔH_{vap}(kJ/mol)
Ti	6.8281	13.5750	0.0755	3560	1941	18.7	425
V	6.7462	14.6600	0.5276	3680	2183	22.8	453
Cr	6.7665	16.4850	0.6759	2944	2180	20.5	339
Mn	7.4340	15.6400	-0.5182	2334	1519	13.2	220
Fe	7.9030	16.1880	0.1532	3134	1811	13.8	347
Co	7.8810	17.0800	0.6622	3200	1768	16.2	375
Ni	7.6399	18.1690	1.1572	3186	1728	17.2	378
Cu	7.7270	20.2920	1.2357	2835	1357.77	13.1	300
Zn	9.3942	17.9640	-0.6011	1180	692.68	7.35	119
Zr	6.6341	13.1600	0.4332	4682	2128	21	580
Nb	6.7589	14.3000	0.9173	5017	2750	26.8	690
Mo	7.0924	16.1700	0.7473	4912	2896	36	600
Ru	7.3605	16.7900	1.0392	4423	2607	25.7	580
Rh	7.4589	18.0300	1.1429	3968	2237	21.7	495
Pd	8.3368	19.3800	0.5622	3236	1828.05	16.7	380
Ag	7.5762	21.4500	1.3044	2435	1234.93	11.3	255

Table S8. The E_{TM} of each metal by DFT calculation.

Metal	E_{TM} (eV)
Ti	-2.1803
V	-3.7885
Cr	-5.4402
Mn	-5.1430
Fe	-3.2045
Co	-1.7100
Ni	-0.3162
Cu	-0.2261
Zn	0.0103
Zr	-2.0258
Nb	-3.1730
Mo	-4.6345
Ru	-2.3315
Rh	-1.1608
Pd	-1.4510
Ag	-0.1814

References

1. K. Chan, J. K. Nørskov, *J. Phys. Chem. Lett.*, 2015, **6**, 2663-2668.

2. R. Ouyang, S. Curtarolo, E. Ahmetcik, M. Scheffler, L. M. Ghiringhelli, *Phys. Rev. Mater.* 2018, **2**, 083802.