Supporting Information

Exploring the Chemical Space of Metal Clusters via Machine Learning

Ning-Zheng Li,[1,3] Zi-Yue Wang,[2,3] Zi-Yu Li,[1,3,4*] Qiang Shi,[1,3,4] Qing-Yu Liu,[1,3,4] Sheng-Gui He[1,3,4,5*]

[1]State Key Laboratory for Structural Chemistry of Unstable and Stable Species, Institute of Chemistry, Chinese Academy of Sciences, Beijing 100190, People's Republic of China.

[2]Institute of Software, Chinese Academy of Sciences, Beijing 100190, People's Republic of China

[3]University of Chinese Academy of Sciences, Beijing 100049, People's Republic of China.

[4]Beijing National Laboratory for Molecular Sciences and Center for Carbon Neutral Chemistry, Institute of Chemistry, Chinese Academy of Sciences, Beijing 100190, People's Republic of China.

[5]Nanjing Institute of Atomic Scale Manufacturing, Nanjing 211899, People's Republic of China.

***Corresponding Authors.** E-mail: liziyu2010@iccas.ac.cn; shengguihe@iccas.ac.cn

Contents

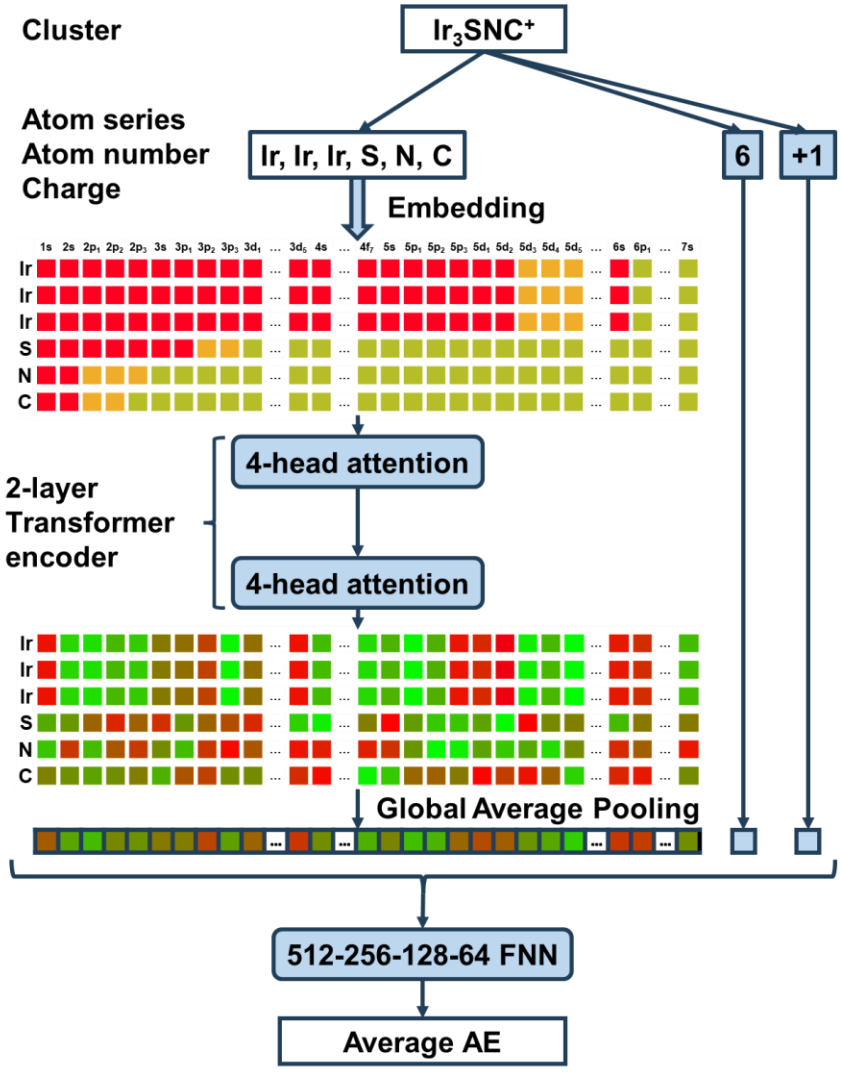

Fig. S1. Illustrative Example of the CTEN Model Architecture.

In the CTEN model, a metal cluster (e.g., Ir_3SNC^+), is first decomposed into its constituent atoms–Ir, Ir, Ir, S, N, and C–together with two global attributes: the total number of atoms and the overall electric charge. Each atom is represented by a 56-dimensional atomic vector that encodes the electron population of individual atomic orbitals, including 1s, 2s, $2p_{1\text{-}3}$, 3s, $3p_{1\text{-}3}$, $3d_{1\text{-}5}$, 4s, $4p_{1\text{-}3}$, $4d_{1\text{-}5}$, $4f_{1\text{-}7}$, 5s, $5p_{1\text{-}3}$, $5d_{1\text{-}5}$, $5f_{1\text{-}7}$, 6s, $6p_{1\text{-}3}$, $6d_{1\text{-}5}$, and 7s orbitals. The resulting sequence of atomic vectors is processed by a transformer encoder consisting of two layers of a four-head self-attention mechanism, enabling the model to capture interactions among different atomic species. To convert the variable-length atomic sequence into a fixed-length representation, global average pooling is applied to the encoder outputs. Positional encoding, which is commonly employed in natural language processing, is deliberately omitted so that different permutations of the same atomic set yield identical feature vectors. As a consequence, clusters with identical elemental compositions but different sizes (e.g., Ir_3SNC^+ and $Ir_6S_2N_2C_2^+$), as well as clusters with the same composition but different charge states (e.g., Ir_3SNC^+ and Ir_3SNC), produce identical pooled feature vectors from the transformer encoder. To distinguish such cases, the total number of atoms and the net charge are explicitly provided as additional inputs to the model. Finally, the pooled feature vector, concatenated with the atom count

and charge, is passed through a fully connected neural network comprising four hidden layers with 512, 256, 128, and 64 neurons, respectively, to generate the final predicted properties.

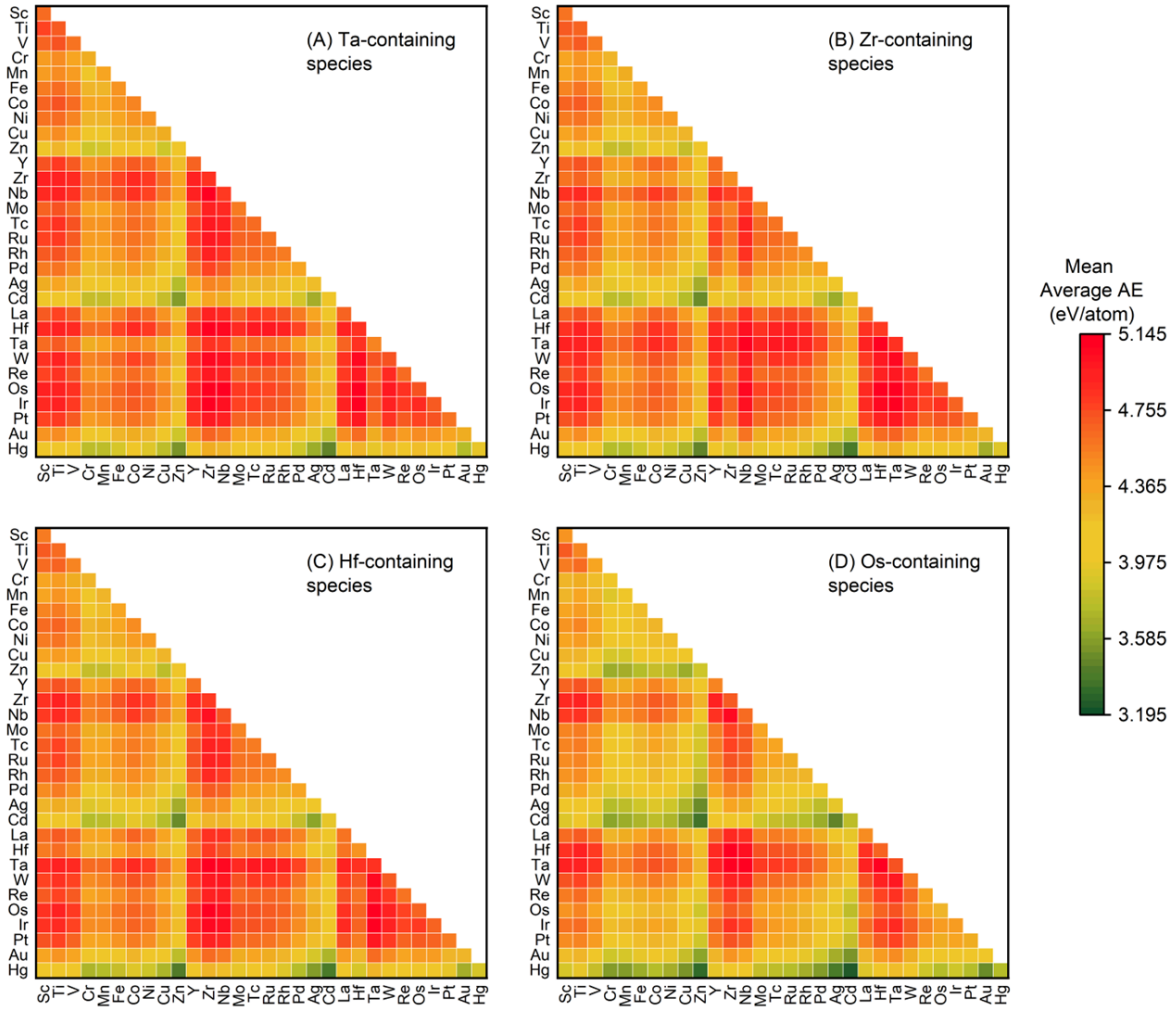

Fig. S2. Heatmap distributions of mean average AE of metal clusters containing (A) Ta, (B) Zr, (C) Hf, and (D) Os.

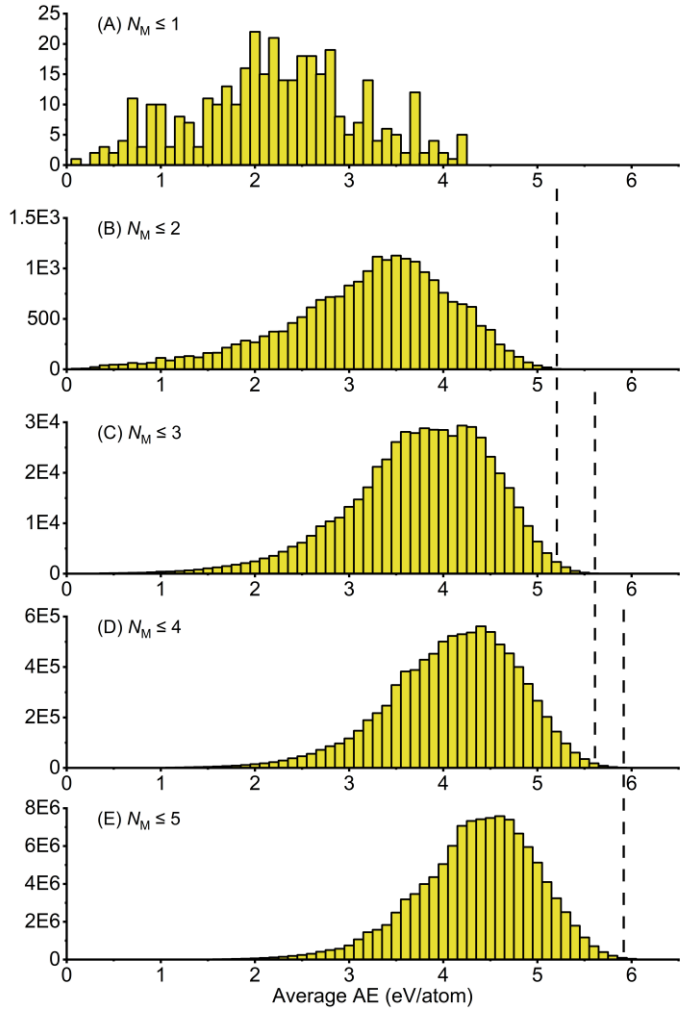

Fig. S3. Distribution of predicted average AE of all clusters in different sizes of chemical spaces. (A) $N_M \leq 1$, (B) $N_M \leq 2$, (C) $N_M \leq 3$, (D) $N_M \leq 4$, and (E) $N_M \leq 5$.

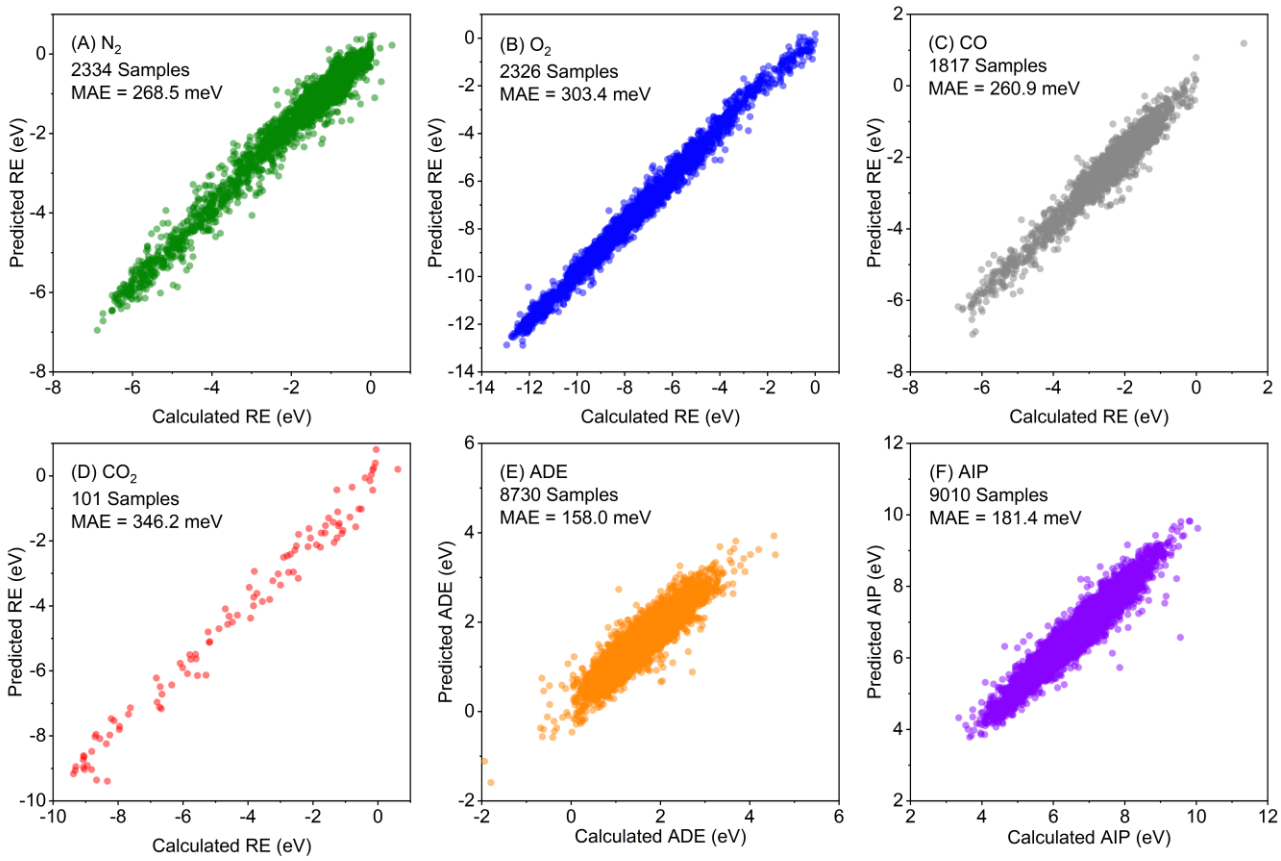

Fig. S4. Calculated and predicted REs, ADEs and AIPs of metal clusters in the database. (A) RE with N_2, (B) RE with O_2, (C) RE with CO, (D) RE with CO_2, (E) ADEs, (F) AIPs.

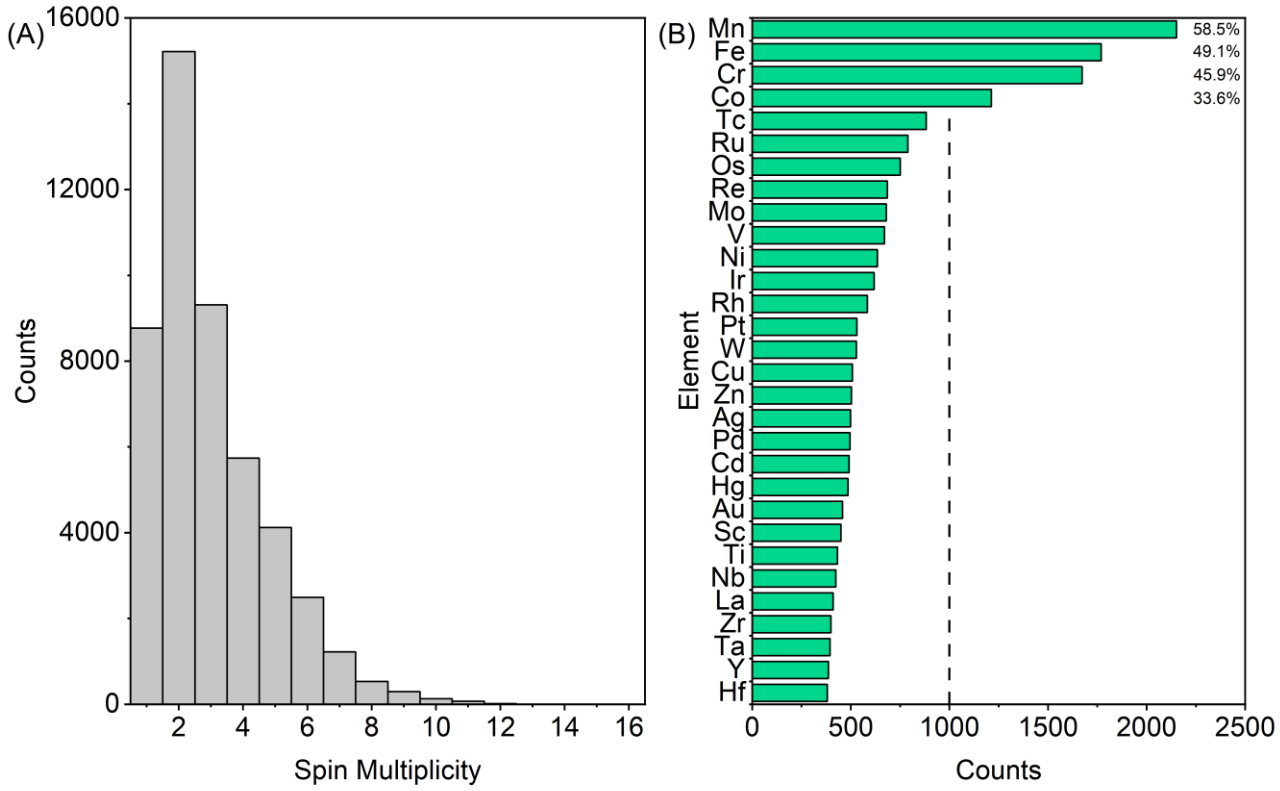

Fig. S5. (A) Distribution of spin multiplicities of the species in the DFT dataset and (B) distribution of the species with high spin multiplicities ($S \geq 5$) species, the percentages of high-spin ($S \geq 5$) species containing Mn, Fe, Cr and Co are listed.

Table S1. Optimized hyperparameters of RF, GBRT, SVR and MLP using two types of descriptors.

Descriptor	RF	GBRT	SVR	MLP
Composition	Estimators = 2745 Max depth = 26 Min sample leaf = 1 Min sample split = 13	Estimators = 4502 Max depth = Min sample leaf = 9, Min sample split = 10 Learning rate = 0.060	C = 23.4 Epsilon = 0.033 Gamma = 0.027 Kernel = RBF	Hidden layers are the same as in CTEN. 166 epochs Learning rate = 0.0014
Orbital-level Atomic Vectors	Estimators = 2899 Max depth = 23 Min sample leaf = 1 Min sample split = 2	Estimators = 3463 Max depth = 8 Min samples leaf = 3 Min samples split = 17 Learning rate = 0.069	C = 23.4 Epsilon = 0.033 Gamma = 0.027 Kernel = RBF	Hidden layers are the same as in CTEN. 647 epochs Learning rate = 0.00065