Supporting Information for

Machine learning-assisted screening of copper-cerium bimetallic catalysts for tetracycline degradation

Qingyu Huang[a,1], Fei Huang[a,1], Yang Yu[a*], Na Jiang[a], Chuang Yang[a], Yuchen Wang[a], and Yongjun Zhang[a*]

[a] School of Environmental Science and Engineering, Nanjing Tech University, Nanjing 211816, PR China

[b] School of Pharmaceutical Sciences, Nanjing Tech University, Nanjing 211816, PR China

[*]Corresponding author.

E-mail address: yuyang19880421@yeah.net (Yang Yu)

Postal address: School of Environmental Science and Engineering, Nanjing Tech University, Nanjing 211816, PR China.

Supporting Information includes the following contents, as cross-referenced throughout the main article:

Figures

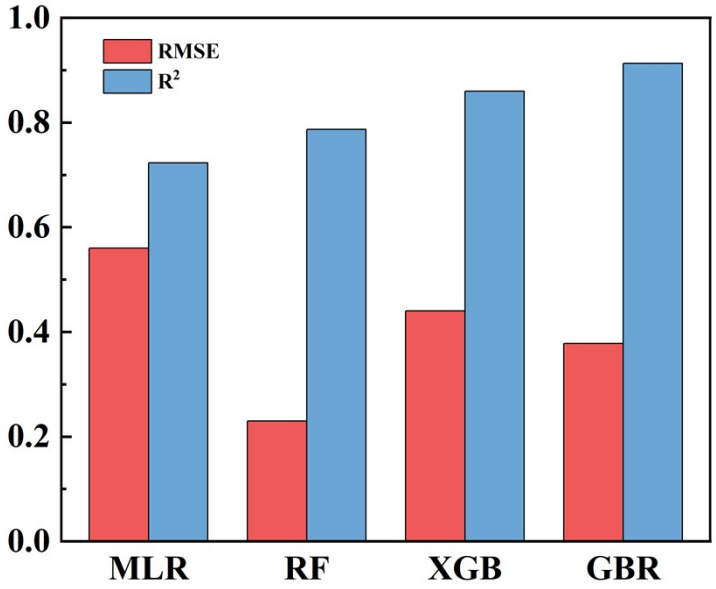

Figure S1. The RMSE and R^2 of different algorithmic models.

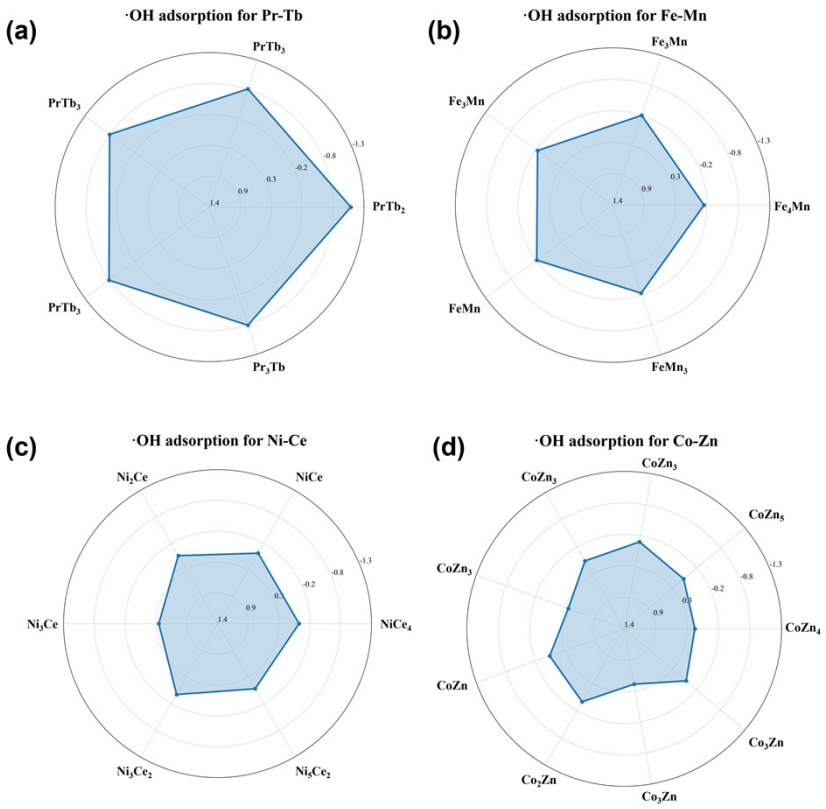

Figure S2. Radar chart of the ·OH adsorption energies predicted by the GBR model for PrTb (a), FeMn (b), NiCe (c), and CoZn (d), respectively.

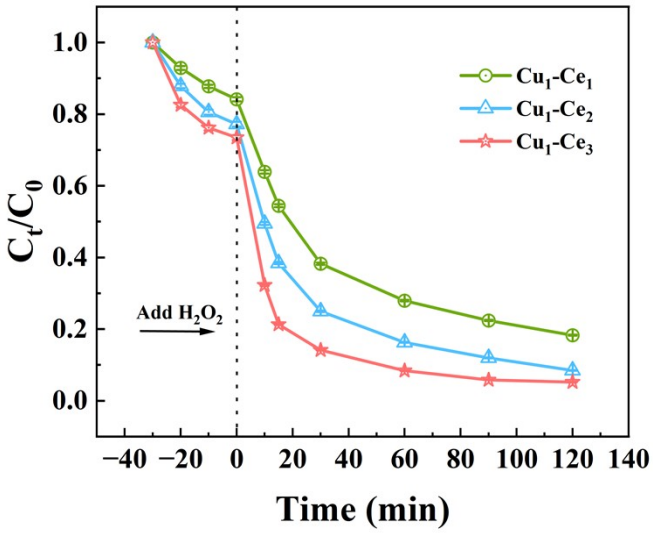

Figure S3. The degradation of TC in CuCe bimetallic systems with varying ratios.

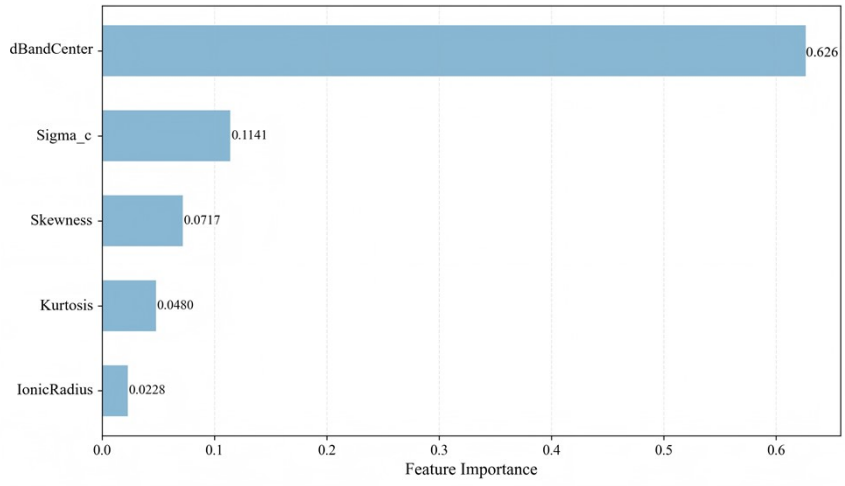

Figure S4. The 5 most important features in the GBR model.

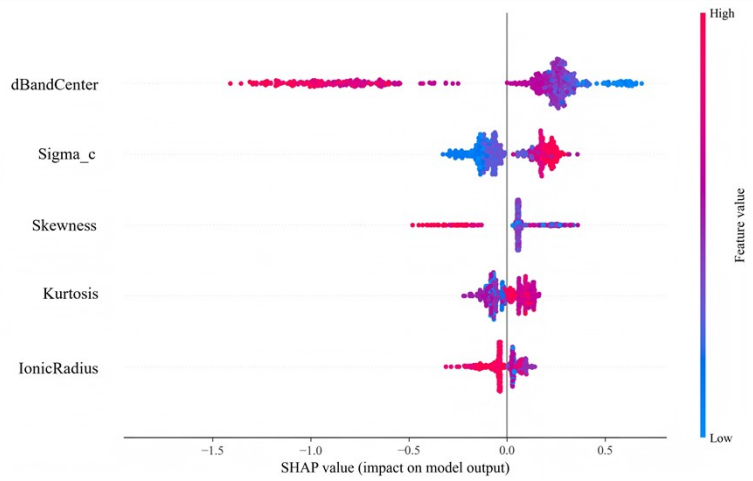

Figure S5. SHAP value distributions for each feature and their impact on bimetallic activity model predictions.

Table S1 Comparative analysis of performance discovery efficiency.

Metric	Traditional Experiment	DFT-only	ML-DFT framework
Search Space	< 20 candidates	< 500 candidates	2,200 combinations
Time Cost	2000 days	400 days	8 hours
Resources	Massive chemical reagents	High-performance computing clusters	Standard workstation and Focused experiment
Physical Insight	Macroscopic observation	Atomic-level mechanisms	Mechanism-driven and Data-guided
Research Paradigm	Linear or Random exploration	Computational Screening	Rapid filtering and Precise validation

Traditional methods are typically limited by a narrow search space and extensive experimental cycles, but ML model completed the screening of 2000 bimetallic combinations within 8 hours. The 90% reduction in time cost, combined with high prediction accuracy ($R^2 = 0.913$), underscores the efficiency of using the d-band center as a key descriptor for the rapid discovery of high-performance catalysts, which is consistent with the advanced screening methodologies reported in recent studies.

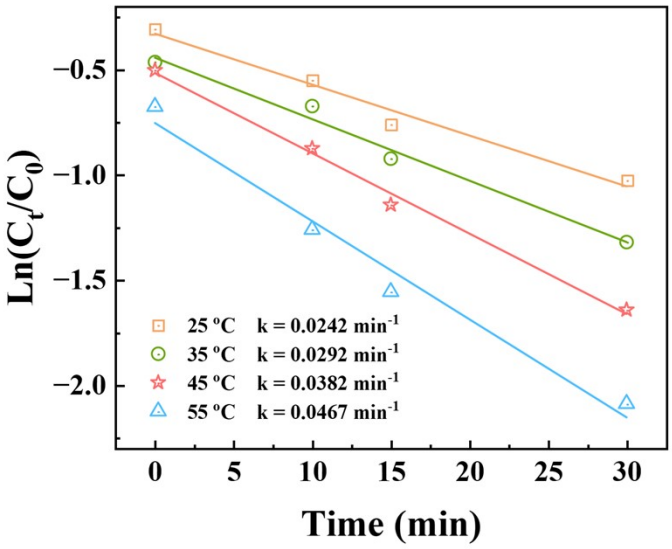

Figure S6. Effects of temperature on the pseudo-first-order rate constant (Kobs).

$$lnk = lnA - \frac{E_a}{R} \times \frac{1}{T} \qquad \text{(Eqs. S1)}$$

$$\ln\left(\frac{k}{T}\right) = -\frac{\Delta H}{R} \times \frac{1}{T} + \ln\left(\frac{K_B}{h}\right) + \frac{\Delta S}{R} \qquad \text{(Eqs. S2)}$$

In Eq. S1, where E_a is the activation energy, A is the pre-exponential factor, and R is the universal gas constant.

In Eq. S2, where k_B is the Boltzmann constant, h is the Planck constant, ΔH is the activation enthalpy, and ΔS is the activation entropy.

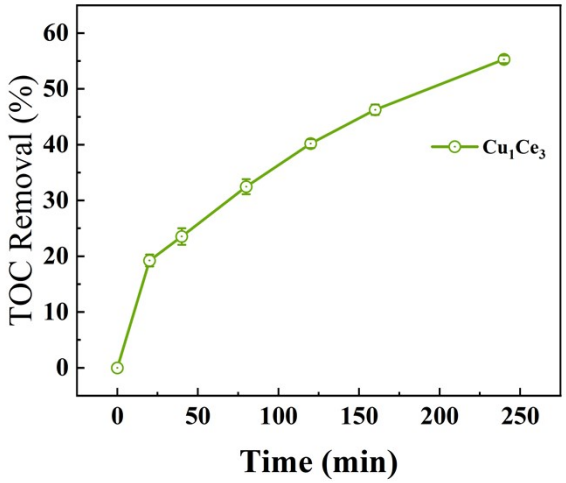

Fig. S7. TOC degradation performance of the Cu_1Ce_3 bimetallic catalyst.

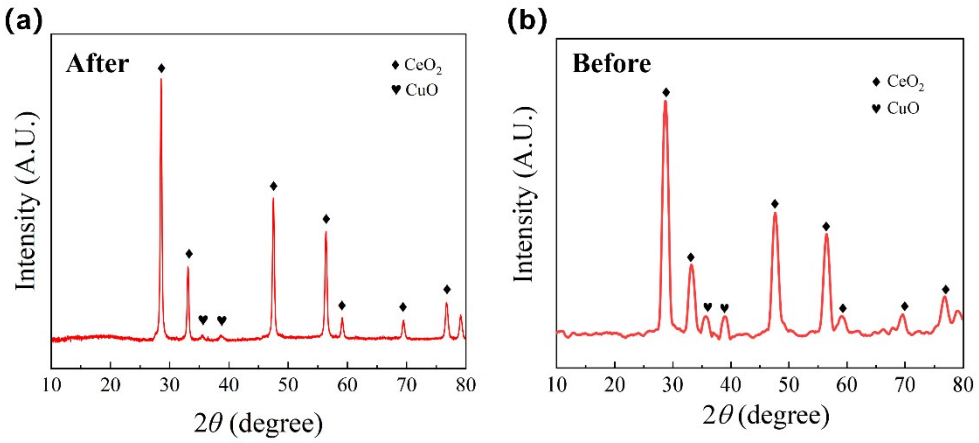

Fig. S8. XRD patterns of Cu_1Ce_3 bimetallic catalyst before and after reaction.

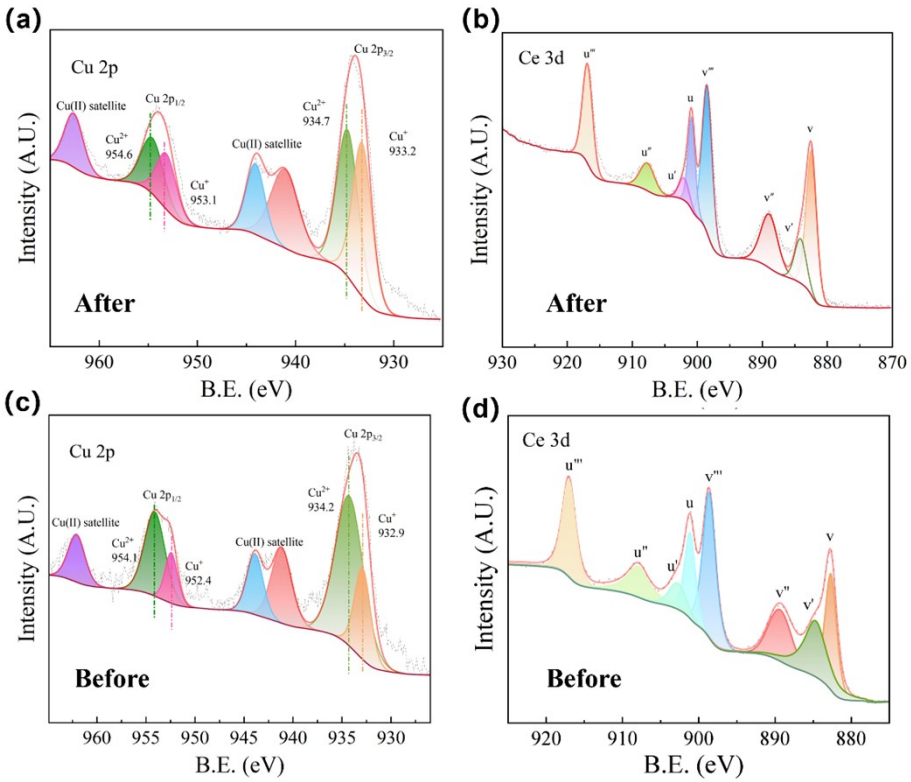

Fig. S9. XPS analysis of Cu_1Ce_3 bimetallic catalyst before and after reaction deconvoluted spectra Cu 2p (a) and Ce 3d (b).

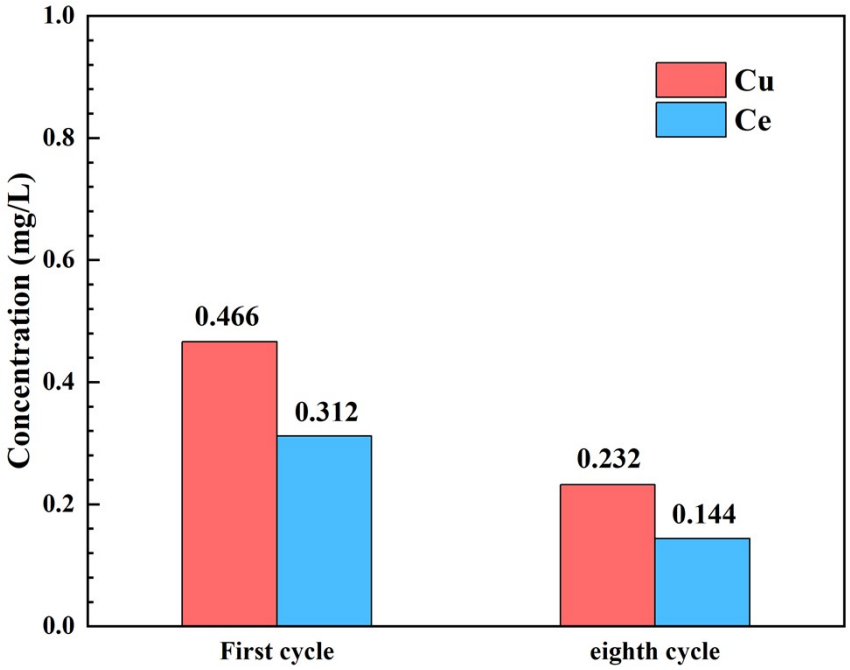

Fig. S10. Metal (Cu and Ce) leaching levels in the first and eighth cycles.

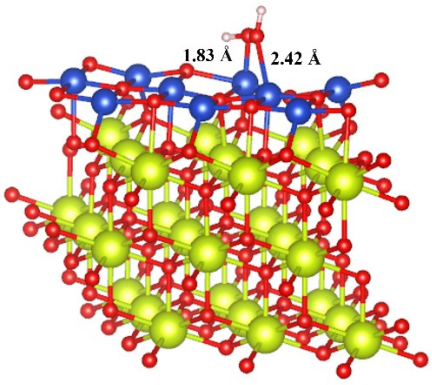

Fig. S11. Optimized adsorption configuration of H_2O_2 on the Cu_1Ce_3 bimetallic catalyst surface.