

Supplementary Materials

Multi-activity derivative automated labeling and margin-sampling active learning for efficient determination of phase boundaries

Dingding Chen^{1,2}, Guanglong Xu^{1,*}, Bincan Deng^{1,2}, Fuwen Chen¹, Jiheng Fang^{3,4}, Zhuo Wang⁵, Chen Xuan², Yuwen Cui^{1,*}, Aimin Zhang^{3,4}

¹Sino-Spain Joint Laboratory on Biomedical Materials (S2LBM) & College of Materials Science and Engineering, Nanjing Tech University, Nanjing 211816, Jiangsu, China.

²School of Mathematics and Physics, Xi'an Jiaotong-Liverpool University, Suzhou 215123, Jiangsu, China.

³State Key Laboratory of Precious Metal Functional Materials, Kunming Institute of Precious Metals, Kunming 650106, Yunnan, China.

⁴Yunnan Precious Metals Laboratory Co., Ltd., Kunming 650106, Yunnan, China.

⁵MatAi Co. Ltd., Chengdu 610213, Sichuan, China.

***Correspondence to:** Dr. Guanglong Xu, Prof. Yuwen Cui, Sino-Spain Joint Laboratory on Biomedical Materials (S2LBM) & College of Materials Science and Engineering, Nanjing Tech University, Nanjing 211816, Jiangsu, China. E-mail: guanglongxu@njtech.edu.cn; ycui@njtech.edu.cn

Supplementary Materials 1. ML model details

The main manuscript describes the thermodynamic basis, data-pool construction, automated labeling procedure, acquisition functions, and stopping rule. This section therefore reports only the model hyperparameters and computational settings needed to reproduce the numerical experiments.

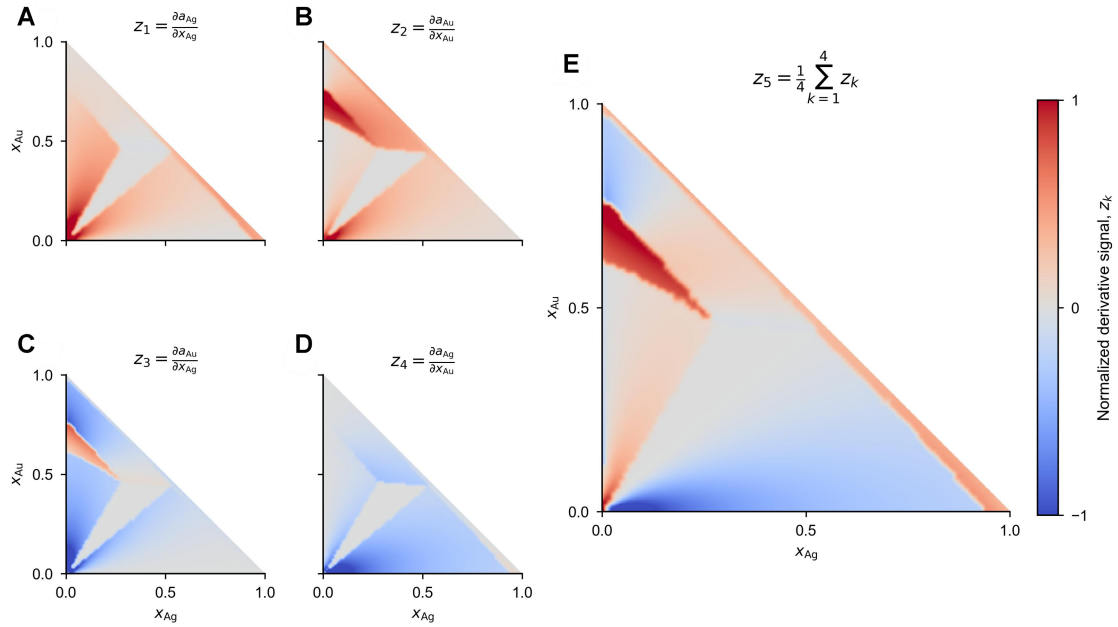
Supplementary Table 1. Machine-learning hyperparameters used for the two systems.

Model	Pd-Pt	Au-Ag-Ge
MLP	Hidden layers=(100, 50); ReLU; Adam; max_iter=2000; batch_size=auto; learning_rate_init=0.001; alpha=0.0001; random_state=42	Same as Pd-Pt
GPC	1.0 $\times$ RBF(length_scale=1.0); one-vs-rest; random_state=42	C(1.0,[10 ⁻³ ,10 ³]) $\times$ RBF(1.0,[10 ⁻² ,10 ²]) + WhiteKernel(1.0,[10 ⁻³ ,10 ³]); one-vs-rest; random_state=42
SVC	RBF; C=10; gamma=scale; probability=True; random_state=42	Same as Pd-Pt
RF	n_estimators=100; criterion=Gini; max_depth=None; min_samples_split=2; min_samples_leaf=1; max_features=sqrt; random_state=42	Same as Pd-Pt
XGB	n_estimators=100; max_depth=6; learning_rate=0.3; subsample=1.0; colsample_bytree=1.0; random_state=42	Same as Pd-Pt

The analyses were implemented in Python 3.10 using NumPy 2.2.6, pandas 2.3.3, SciPy 1.15.3, scikit-learn 1.6.1, Matplotlib 3.10.8, and XGBoost 3.1.3.

Supplementary Materials 2. Additional evidence for the composite channel

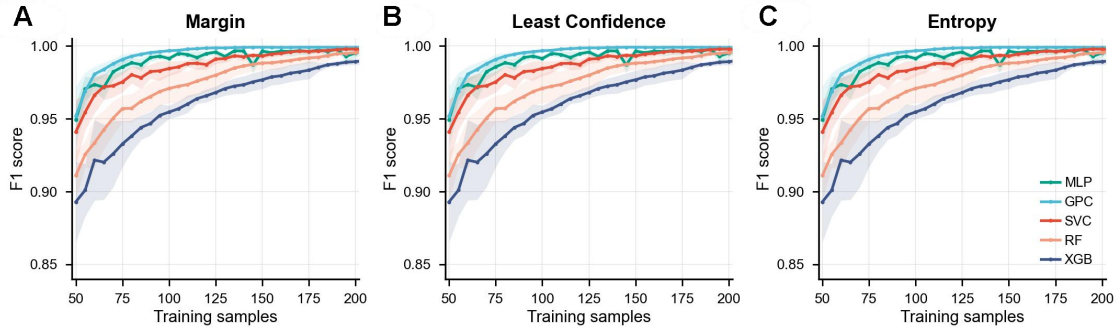
The following figures provide the visual evidence requested during peer review. They are not repeated in the main manuscript.



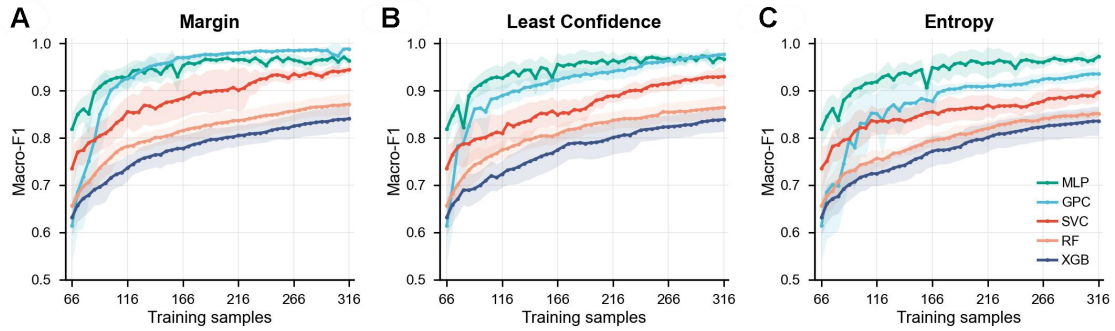
Supplementary Figure 1. Spatial distributions of the four individual activity-derivative channels and their equal-weight auxiliary composite channel for Au-Ag-Ge. Each channel is independently normalized for visualization.

Supplementary Materials 3. Ten-seed repeated experiments

For the randomized stability analysis, every model-acquisition combination was repeated using ten independently generated class-stratified initial sets with seeds 11, 22, 33, 44, 55, 66, 77, 88, 99, and 110. Initial set size, candidate pool, query batch size, model hyperparameters, and active learning budget were held fixed. Five points were uniformly sampled without replacement from the remaining candidate pool in each iteration, using the same initialization and sampling budget as in uncertainty sampling.



Supplementary Figure 2. Active-learning results for Pd-Pt across ten randomized class-stratified initial training sets. Curves show mean Macro-F1 with standard-deviation bands; bars summarize the mean labeled-sample requirement with standard deviations.



Supplementary Figure 3. Active-learning results for Au-Ag-Ge across ten randomized class-stratified initial training sets. Curves show mean Macro-F1 with standard-deviation bands; bars summarize the mean labeled-sample requirement with standard deviations.

Supplementary Figures 2 and 3 show the mean learning curves and standard-deviation bands. Supplementary Tables 2 and 3 report convergence probability, labeled-sample requirement, final Macro-F1, and boundary errors across the ten initial sets.

Supplementary Table 2. Ten-seed repeated-run statistics for Pd-Pt. The sample requirement denotes the first attainment of the reference Macro-F1 level of 0.98.

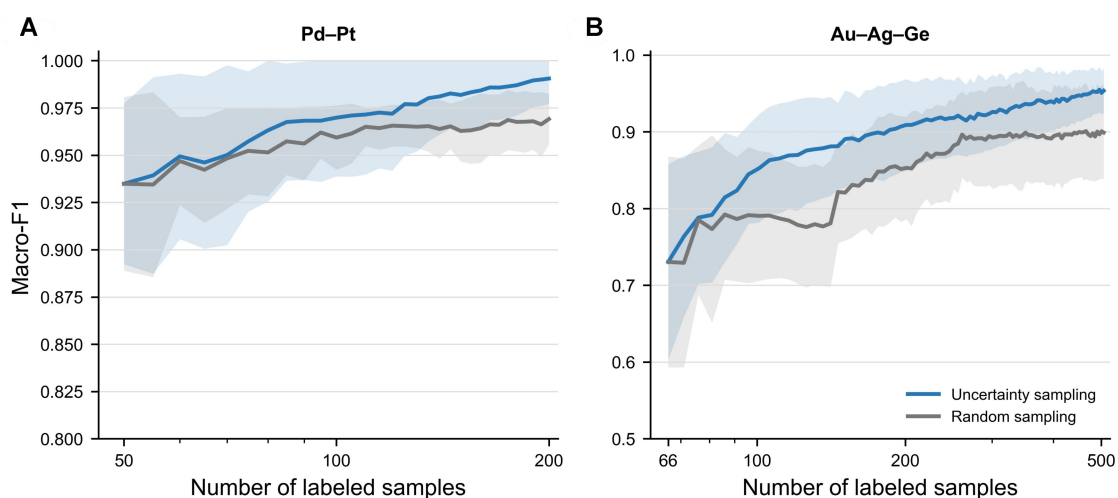
Model	Strategy	Target reached	Samples to target (mean $\pm$ SD)	Final Macro-F1 (mean $\pm$ SD)	Boundary displacement (mean $\pm$ SD)	Triple-junction error
MLP	MS	10/10	70.0000 $\pm$ 6.2361	0.9960 $\pm$ 0.0022	0.0399 $\pm$ 0.0239	N/A
MLP	LC	10/10	70.0000 $\pm$ 6.2361	0.9960 $\pm$ 0.0022	0.0399 $\pm$ 0.0239	N/A
MLP	EA	10/10	70.0000 $\pm$ 6.2361	0.9960 $\pm$ 0.0022	0.0399 $\pm$ 0.0239	N/A
GPC	MS	10/10	61.5000 $\pm$ 4.1164	0.9988 $\pm$ 0.0000	0.0059 $\pm$ 0.0000	N/A
GPC	LC	10/10	61.5000 $\pm$ 4.1164	0.9988 $\pm$ 0.0000	0.0059 $\pm$ 0.0000	N/A
GPC	EA	10/10	61.5000 $\pm$ 4.1164	0.9988 $\pm$ 0.0000	0.0059 $\pm$ 0.0000	N/A
SVC	MS	10/10	76.5000 $\pm$ 8.8349	0.9980 $\pm$ 0.0004	0.0142 $\pm$ 0.0041	N/A
SVC	LC	10/10	76.5000 $\pm$ 8.8349	0.9980 $\pm$ 0.0004	0.0142 $\pm$ 0.0041	N/A
SVC	EA	10/10	76.5000 $\pm$ 8.8349	0.9980 $\pm$ 0.0004	0.0142 $\pm$ 0.0041	N/A
RF	MS	10/10	123.0000 $\pm$ 16.7000	0.9995 $\pm$ 0.0003	0.0053 $\pm$ 0.0033	N/A
RF	LC	10/10	123.0000 $\pm$ 16.7000	0.9995 $\pm$ 0.0003	0.0053 $\pm$ 0.0033	N/A
RF	EA	10/10	123.0000 $\pm$ 16.7000	0.9995 $\pm$ 0.0003	0.0053 $\pm$ 0.0033	N/A
XGB	MS	10/10	157.0000 $\pm$ 13.1656	0.9985 $\pm$ 0.0004	0.0152 $\pm$ 0.0040	N/A
XGB	LC	10/10	157.0000 $\pm$ 13.1656	0.9985 $\pm$ 0.0004	0.0152 $\pm$ 0.0040	N/A
XGB	EA	10/10	157.0000 $\pm$ 13.1656	0.9985 $\pm$ 0.0004	0.0152 $\pm$ 0.0040	N/A

Supplementary Table 3. Ten-seed repeated-run statistics for Au-Ag-Ge. The sample requirement denotes the first attainment of the reference Macro-F1 level of 0.95.

Model	Strategy	Target reached	Samples to target (mean $\pm$ SD)	Final Macro-F1 (mean $\pm$ SD)	Boundary displacement (mean $\pm$ SD)	Triple-junction error (mean $\pm$ SD)
MLP	MS	10/10	131.5000 $\pm$ 22.2923	0.9632 $\pm$ 0.0114	0.1668 $\pm$ 0.0656	1.8831 $\pm$ 3.0024
MLP	LC	10/10	149.0000 $\pm$ 35.4495	0.9674 $\pm$ 0.0061	0.1485 $\pm$ 0.0261	0.6327 $\pm$ 0.1760
MLP	EA	10/10	161.0000 $\pm$ 42.0317	0.9725 $\pm$ 0.0089	0.1235 $\pm$ 0.0419	0.6217 $\pm$ 0.5579
GPC	MS	10/10	134.5000 $\pm$ 11.0680	0.9884 $\pm$ 0.0011	0.0446 $\pm$ 0.0046	0.6370 $\pm$ 0.2495
GPC	LC	10/10	223.0000 $\pm$ 42.3084	0.9770 $\pm$ 0.0052	0.1039 $\pm$ 0.0280	0.3646 $\pm$ 0.1531
GPC	EA	2/10	306.0000 $\pm$ 7.0711	0.9352 $\pm$ 0.0116	0.3646 $\pm$ 0.0923	1.0868 $\pm$ 0.6865
SVC	MS	5/10	276.0000 $\pm$ 27.6134	0.9448 $\pm$ 0.0080	0.3258 $\pm$ 0.2142	1.6968 $\pm$ 0.6859
SVC	LC	3/10	307.6667 $\pm$ 5.7735	0.9303 $\pm$ 0.0172	0.6276 $\pm$ 0.4628	4.0047 $\pm$ 4.7441
SVC	EA	0/10	Not reached	0.8971 $\pm$ 0.0161	0.7973 $\pm$ 0.2617	3.9980 $\pm$ 4.7514
RF	MS	0/10	Not reached	0.8714 $\pm$ 0.0219	0.8792 $\pm$ 0.2833	1.9412 $\pm$ 0.4426
RF	LC	0/10	Not reached	0.8648 $\pm$ 0.0242	0.9469 $\pm$ 0.2832	1.1756 $\pm$ 0.3242
RF	EA	0/10	Not reached	0.8514 $\pm$ 0.0088	0.9889 $\pm$ 0.2555	0.5234 $\pm$ 0.4018
XGB	MS	0/10	Not reached	0.8406 $\pm$ 0.0265	1.1657 $\pm$ 0.2868	2.2523 $\pm$ 0.8166
XGB	LC	0/10	Not reached	0.8389 $\pm$ 0.0255	1.1583 $\pm$ 0.2946	2.1242 $\pm$ 1.2619
XGB	EA	0/10	Not reached	0.8357 $\pm$ 0.0306	1.2281 $\pm$ 0.3371	1.7157 $\pm$ 0.7465

Supplementary Materials 4. Baseline comparisons and ablation analysis

This section evaluates the individual contributions of the sampling components under controlled conditions. Unless otherwise stated, all comparisons used the same data pools, uniformly distributed initial samples, MLP settings, and a batch size of five, with reference Macro-F1 levels of 0.98 for Pd-Pt and 0.95 for Au-Ag-Ge. The comparison between uncertainty sampling and random sampling establishes the sampling baseline; the comparison between Top-25 + Diversity and Direct Top-5 isolates the effect of k-center diversity selection; and the comparison between fixed and hierarchically refined grids evaluates the effect of adaptive candidate-space construction.



Supplementary Figure 4. Comparison of uncertainty-guided sampling (US) and random sampling (RS) under uniform initialization for Pd-Pt and Au-Ag-Ge. Curves terminate before the first abrupt non-representative decline, consistently for both US and RS.

Supplementary Table 4. Mean labeled-sample requirement for reaching the common reference Macro-F1 level.

System	Sampling	Reference Macro-F1	Mean samples	SD	Reduction relative to RS
Pd-Pt	US	0.98	124.7	84.1	50.9%
Pd-Pt	RS	0.98	254.0	87.6	Reference
Au-Ag-Ge	US	0.95	390.0	237.9	65.9%
Au-Ag-Ge	RS	0.95	1,143.0	890.4	Reference

Supplementary Table 5. Effect of k-center diversity selection for MLP in the Au-Ag-Ge system. Top-25 combined Diversity selection. “Top-25+ Diversity” denotes selecting five queries from the 25 most uncertain candidates using k-center diversity selection, whereas “Direct Top-5” denotes directly selecting the five candidates with the highest uncertainty scores.

Strategy	Selection method	Iterations	Labeled samples	Macro-F1	Accuracy	Sample reduction
MS	Top-25 + Diversity	15	141	0.9592	0.9717	19.9%
MS	Direct Top-5	22	176	0.9580	0.9766	Reference
LC	Top-25 + Diversity	22	176	0.9740	0.9838	5.4%
LC	Direct Top-5	24	186	0.9778	0.9857	Reference
EA	Top-25 + Diversity	26	196	0.9633	0.9774	29.0%
EA	Direct Top-5	42	276	0.9575	0.9731	Reference

Supplementary Table 6. Effect of the hierarchical adaptive grid under uniform initialization

Syste m	Grid mode	Target Macro -F1	Iteration s	Labele d sample s	Accessibl e grid points	Grid reductio n	Macro- F1 achieve d
Pd-Pt	Fixed	0.98	4	70	1,717	Referenc e	0.9860
Pd-Pt	Adaptiv e	0.98	2	60	257	85.0%	0.9825
Au- Ag-Ge	Fixed	0.95	7	101	5,049	Referenc e	0.9547
Au- Ag-Ge	Adaptiv e	0.95	15	141	761	84.9%	0.9538

Supplementary Materials 5. Probability calibration methods

Probability calibration was assessed before active querying using ten equal-width bins over the interval $[0, 1]$. Each prediction was assigned to a bin according to its maximum class probability. A non-empty bin contains at least one evaluated point. The multiclass Brier score and expected calibration error (ECE) were calculated as

$$\text{Brier} = \frac{1}{N} \sum_{n=1}^N \sum_{c=1}^C [p_{n,c} - \mathbb{1}(y_n = c)]^2 \quad (0.2)$$

$$\text{ECE} = \sum_{m=1}^M \frac{|B_m|}{N} |\text{acc}(B_m) - \text{conf}(B_m)| \quad (0.2)$$

where $p_{n,c}$ is the predicted probability of class c for point n ; B_m is the set of points in confidence bin m ; and $\text{acc}(B_m)$ and $\text{conf}(B_m)$ are the empirical accuracy and mean confidence within that bin, respectively. Lower values indicate better probability calibration.

Supplementary Materials 6. Animations of the adaptive mesh grid in active learning

In separated files.