Electronic Supplementary Information

A Comprehensive Machine Learning Strategy for Designing High-performance Photoanode Catalysts

Meirong Huang[1,2], Sutong Wang[3], Hongwei Zhu[1*]

[1] *State Key Laboratory of New Ceramics and Fine Processing, School of Materials Science and Engineering, Tsinghua University, Beijing 100084, China*

[2] *State Key Laboratory of Photocatalysis on Energy and Environment, College of Chemistry, Fuzhou University, Fuzhou 350108, China*

[3] *School of Economics and Management, Dalian University of Technology, Dalian 116023, China*

[*]Email: hongweizhu@tsinghua.edu.cn

Methods:

- **Influencing factors**

Supplementary Figures:

- **Figures S1–S25**

Systematic study about the influencing factors:

The screening of cocatalysts for photoanodes is greatly influenced by several factors, including the physical and chemical properties of the cocatalyst (e.g., type, dopant, thickness, morphology, synthesis method), the photoanode (e.g., dopant, microscopic size, morphology, synthesis method), and the electrolyte (e.g., composition, concentration, pH). These factors can cause various cocatalysts to work with different effectiveness and mechanisms[1-3].

Cocatalyst type can affect various properties of the cocatalyst/photoanode catalytic system, such as band alignment, light absorption, characteristics of active sites, sedimentary morphology, and surface wettability. (1) Band alignment is determined by the relative energy band positions of the cocatalyst and photoanode. Semiconductor-type cocatalysts can form heterojunction with photoanodes. Since different band alignments can change hole extraction ability and charge separation efficiency, the alignment should be appropriately regulated to achieve optimal performance[4-6]. (2) The light absorption ability of the catalytic system is affected by the cocatalyst's light absorption properties. While metal-type cocatalysts may reduce light harvesting efficiency due to the shading effect, semiconductor-type cocatalysts can enhance light absorption and produce more photo-generated charge carriers[6, 7]. (3) The active sites of a metal-containing cocatalyst are typically the metal, and the number of active sites depends on their exposure way, which is determined by the cocatalyst type, typical examples are metal-organic frameworks[8-10]. The type of active sites affects the surface reaction kinetics, and more exposed active sites can improve catalytic activity by enhancing the hole injection efficiency[11]. (4) The sedimentary morphology of cocatalysts, such as isolated particles, conformal layer, and isolated layer, depends on the cocatalyst type, typical examples are molecular catalysts. A conformal layer usually provides better stability than isolated particles by passivating surface states and separating photoanode from electrolyte[8, 12, 13]. (5) The hydrophilicity of the catalytic system is mainly determined by the number of hydrophilic groups on the cocatalyst surface. Higher hydrophilicity can improve the full contact between electrolyte and cocatalyst, accelerate the release of gaseous products, and thus improve surface reaction kinetics and catalyst stability[14, 15].

Doping cocatalyst can optimize the interfacial contact between photoanode and cocatalyst to promote charge separation[16], active catalytic sites to provide higher surface OER kinetics[17], expose more defects and edges to act as active sites[18], as well as improve conductivity for efficient charge transfer[17, 18]. Nevertheless, it is important to optimize the doping concentration since high levels may introduce defects that can cause charge recombination[9].

Cocatalyst thickness is also a trade-off parameter that controls the cocatalyst's effectiveness. In the size range of the depletion width introduced by the semiconductor/electrolyte interface, thicker cocatalysts layers promote charge separation by extracting holes more effectively, which results in improved catalytic performance[19]. However, an excessively thick cocatalyst layer can cause shading effects[7], create shunting pathways[20], and hinder charge transport[4], resulting in reduced performance. Thus, optimizing cocatalyst thickness is essential to balance these competing factors and maximize catalytic efficiency[21, 22].

Cocatalyst morphology is determined by the cocatalyst type and its preparation method, usually presented in the form of isolated particles, a conformal layer, or an isolated layer. Conformal layers have the advantage of providing superior stability over isolated particles by passivating surface states and isolating photoanode from electrolyte[5, 12]. However, conformal layers composed of multiple nanoparticles may exhibit limited catalytic activity due to inferior charge transport within the layer caused by numerous defects and granular interfaces.

The preparation method of cocatalysts can impact their specific composition, defect state, crystallinity, morphology, size, and specific surface area, which ultimately affects their catalytic activity and stability[23]. These changes can make a more rational band alignment between the cocatalyst and photoanode, facilitating charge separation and injection[5]. In-situ synthesis techniques, such as electrodeposition, photo-deposition, and photo-assisted electrodeposition, can establish better contact between the photoanode and cocatalyst compared to post-deposition methods like dip coating, spin coating, and drop coating. The enhanced contact can promote charge transport and prolong the carrier lifetime, thereby facilitating the OER process[24].

Photoanode morphology and size can change the effectiveness of cocatalysts. Although several studies have investigated the impact of photoanode morphology and size on its performance, the effect on cocatalysts is often overlooked. Recent research suggests that isolated layered cocatalysts can improve the charge separation efficiency and catalytic activity of denser photoanode layers more efficiently[25]. Denser photoanode layers have superior photon absorption and electron transport but inferior hole transport, whereas porous layers exhibit the opposite characteristics. With the help of cocatalysts that promote hole extraction, denser layers show greater performance enhancement due to more excited charge carriers, good electron transport, and improved hole transport[26].

Doping photoanode can improve the effectiveness of cocatalysts by promoting charge separation. This improvement is likely a result of optimized interfacial bonding strength and band alignment between the photoanode and cocatalyst. Typically, doping the photoanode and

loading the cocatalyst work together in a synergistic manner[27]. However, it is crucial to carefully regulate the doping concentration to prevent excessive dopants or defects from causing serious charge recombination[28].

The preparation method of photoanodes not only affects their morphology and size but also their surface states. Choosing an appropriate preparation method can improve the interaction between the photoanode and cocatalyst, resulting in reduced charge transfer resistance and more efficient charge separation.

Electrolyte composition plays a crucial role in determining the activity and stability of the catalytic system. The interaction between electrolyte anions and metal ions in cocatalysts may cause structural changes in the cocatalysts[29, 30]. In addition, the addition of specific ions to the electrolyte can help to inhibit photoanode dissolution[31] or regenerate the cocatalyst layer[32], leading to improved catalyst stability. Nevertheless, most anions, such as borate, phosphate, and sulfate, are electrochemically inert and do not impact the OER process[33].

Electrolyte pH usually varies with the change of electrolyte composition possessing various pK_A values. Although alkaline conditions are typically favorable for the surface OER on cocatalysts[34], extremely high pH levels may cause cocatalyst dissolution[35]. The influence of electrolyte pH on cocatalyst performance is also decided by the working mechanism of cocatalysts. Therefore, cocatalysts work better in a solution with a proper pH value[36], some cocatalysts can even perform better in acidic solutions[35, 37].

Electrolyte concentration should be high enough to provide effective charge injection at the catalyst/electrolyte interface and reduce solution resistance. When anions are absent or present in low concentrations in the electrolyte, photoanodes and/or cocatalysts show poor activity. However, excessively high electrolyte concentration may lead to the detachment of cocatalysts. Therefore, optimizing the electrolyte concentration is essential for achieving optimal performance of the catalytic system[33].

References:

1. Wang, R.; Kuwahara, Y.; Mori, K.; Qian, X.; Zhao, Y.; Yamashita, H., Modification of Ti-doped Hematite Photoanode with Quasi-molecular Cocatalyst: A Comparison of Improvement Mechanism Between Non-noble and Noble Metals. *ChemSusChem* **2021,** *14* (10), 2180-2187.
2. Singh, A.; Tejasvi, R.; Karmakar, S.; Basu, S., α-Fe2O3 Nanorods Decorated with $NiMnO_3$ Co-catalyst as Photoanode for Enhanced Oxygen Evolution Reaction in Photoelectrochemical Water Splitting. *Materials Today Communications* **2021,** *27*.
3. Wen, P.; Lei, R.; Cao, X.; Ma, Q.; Zhang, G.; Guo, C.; Wang, X.; Qiu, Y., Anchored Ni Nanocrystals Boosting $BiVO_4$ Photoanode for Highly Efficient Water Oxidation via in-situ Generation of Ni@NiOOH Co-catalyst. *Chemical Engineering Journal* **2023,** *454*.

4. Wu, Q.; Liang, X.; Chen, H.; Yang, L.; Xie, T.; Zou, X., Surface-oxidized Titanium Diboride as Cocatalyst on Hematite Photoanode for Solar Water Splitting. *CrystEngComm* **2022,** *24* (12), 2251-2257.
5. Cao, X.; Wang, Y.; Lin, J.; Ding, Y., Ultrathin CoO_x Nanolayers Derived from Polyoxometalate for Enhanced Photoelectrochemical Performance of Hematite Photoanodes. *Journal of Materials Chemistry A* **2019,** *7* (11), 6294-6303.
6. Zhifeng Liu, Q. S., Miao Zhou, Zhengang Guo, Jianhai Kang, Huiyu Yan, Synergistic Enhancement of Charge Management and Surface Reaction Kinetics by Spatially Separated Cocatalysts and p-n Heterojunctions in Pt/$CuWO_4$/Co_3O_4 Photoanode. *Chemical Engineering Journal* **2019,** *2019*, 554-563.
7. Sol A Lee, T. H. L., Changyeon Kim, Mi Gyoung Lee, Min-Ju Choi, Hoonkee Park, Seokhoon Choi, Jihun Oh, and Ho Won Jang, Tailored NiO_x/Ni Cocatalysts on Silicon for Highly Efficient Water Splitting Photoanodes via Pulsed Electrodeposition. *ACS Catalysis* **2018,** *8*, 7261-7269.
8. Vo, T.-G.; Tai, Y.; Chiang, C.-Y., Multifunctional Ternary Hydrotalcite-like Nanosheet Arrays as an Efficient Co-catalyst for Vastly Improved Water Splitting Performance on Bismuth Vanadate Photoanode. *Journal of Catalysis* **2019,** *370*, 1-10.
9. Zhang, Y.; Li, Y.; Ni, D.; Chen, Z.; Wang, X.; Bu, Y.; Ao, J. P., Improvement of $BiVO_4$ Photoanode Performance During Water Photo‐oxidation Using Rh‐doped $SrTiO_3$ Perovskite as a Co‐catalyst. *Advanced Functional Materials* **2019,** *29* (32).
10. Wang, R.; Kuwahara, Y.; Mori, K.; Yamashita, H., Semiconductor‐based Photoanodes Modified with Metal‐organic Frameworks and Molecular Catalysts as Cocatalysts for Enhanced Photoelectrochemical Water Oxidation Reaction. *ChemCatChem* **2021,** *13* (24), 5058-5072.
11. Miao, Y.; Liu, J.; Chen, L.; Sun, H.; Zhang, R.; Guo, J.; Shao, M., Single-atomic-Co Cocatalyst on (040) Facet of $BiVO_4$ Toward Efficient Photoelectrochemical Water Splitting. *Chemical Engineering Journal* **2022,** *427*.
12. Shi, Y.; Yu, Y.; Yu, Y.; Huang, Y.; Zhao, B.; Zhang, B., Boosting Photoelectrochemical Water Oxidation Activity and Stability of Mo-doped $BiVO_4$ Through the Uniform Assembly Coating of NiFe–phenolic Networks. *ACS Energy Letters* **2018,** *3* (7), 1648-1654.
13. Long, X.; Wang, P.; Jin, J.; Zhao, X.; Ma, J., Decorating the Cocatalyst Membrane with Coordinated Tannic Acid and Ternary Metal for Advancing Photoelectrochemical Performance of F-doped Hematite Photoanodes. *ACS Sustainable Chemistry & Engineering* **2021,** *9* (38), 13047-13055.
14. Hu, X.; Wang, Q.; Li, Y.; Meng, Y.; Wang, L.; She, H.; Huang, J., The Hydrophilic Treatment of a Novel Co-catalyst for Greatly Improving the Solar Water Splitting Performance over Mo-doped Bismuth Vanadate. *Journal of Colloid and Interface Science* **2022,** *607* (Pt 1), 219-228.
15. Huang, M.; Lei, W.; Wang, M.; Zhao, S.; Li, C.; Wang, M.; Zhu, H., Large Area High-performance Bismuth Vanadate Photoanode for Efficient Solar Water Splitting. *Journal of Materials Chemistry A* **2020,** *8* (7), 3845-3850.
16. Tsyganok, A.; Klotz, D.; Malviya, K. D.; Rothschild, A.; Grave, D. A., Different Roles of $Fe_{1-x}Ni_xOOH$ Cocatalyst on Hematite (α-Fe_2O_3) Photoanodes with Different Dopants. *ACS Catalysis* **2018,** *8* (4), 2754-2759.

17. Lin, J.; Han, X.; Liu, S.; Lv, Y.; Li, X.; Zhao, Y.; Li, Y.; Wang, L.; Zhu, S., Nitrogen-doped Cobalt-iron Oxide Cocatalyst Boosting Photoelectrochemical Water Splitting of $BiVO_4$ Photoanodes. *Applied Catalysis B: Environmental* **2023,** *320*.
18. Pal, D.; Maity, D.; Sarkar, A.; Sarkar, D.; Khan, G. G., Effect of Defect-rich Co-CeO_x OER Cocatalyst on the Photocarrier Dynamics and Electronic Structure of Sb-doped TiO_2 Nanorods Photoanode. *J Colloid Interface Sci* **2022,** *620*, 209-220.
19. Pokrant, S.; Dilger, S.; Landsmann, S.; Trottmann, M., Size Effects of Cocatalysts in Photoelectrochemical and Photocatalytic Water Splitting. *Materials Today Energy* **2017,** *5*, 158-163.
20. Irani, R.; Plate, P.; Höhn, C.; Bogdanoff, P.; Wollgarten, M.; Höflich, K.; van de Krol, R.; Abdi, F. F., The Role of Ultra-thin MnO_x Co-catalysts on the Photoelectrochemical Properties of $BiVO_4$ Photoanodes. *Journal of Materials Chemistry A* **2020,** *8* (11), 5508-5516.
21. Yu, F.; Li, F.; Yao, T.; Du, J.; Liang, Y.; Wang, Y.; Han, H.; Sun, L., Fabrication and Kinetic Study of a Ferrihydrite-modified $BiVO_4$ Photoanode. *ACS Catalysis* **2017,** *7* (3), 1868-1874.
22. Quiñonero, J.; Gómez, R., Controlling the Amount of Co-catalyst as a Critical Factor in Determining the Efficiency of Photoelectrodes: The Case of Nickel (II) Hydroxide on Vanadate Photoanodes. *Applied Catalysis B: Environmental* **2017,** *217*, 437-447.
23. Zhang, W.; Ma, J.; Xiong, L.; Jiang, H.-Y.; Tang, J., Well-crystallized α-FeOOH Cocatalysts Modified $BiVO_4$ Photoanodes for Efficient and Stable Photoelectrochemical Water Splitting. *ACS Applied Energy Materials* **2020,** *3* (6), 5927-5936.
24. Lu Zhang, M. Y., Zhishan Luo, Jinshui Zhang, Yidong Hou, Photodeposited CoO_x as Highly Active Phases to Boost Water Oxidation on $BiVO_4/WO_3$ Photoanode. *International Journal of Hydrogen Energy* **2019,** *44*, 25652-25661.
25. Eftekharinia, B.; Vayghan, N. S.; Esfandiar, A.; Dabirian, A., Effect of Film Morphology on Water Oxidation Enhancement in NiFeCo Modified Hematite Photoanodes. *Surface and Coatings Technology* **2021,** *421*.
26. Blaise A. Pinaud, P. C. K. V., and Thomas F. Jaramillo, Effect of Film Morphology and Thickness on Charge Transport in Ta_3N_5/Ta Photoanodes for Solar Water Splitting. *The Journal of Physical Chemistry C* **2012,** *116*, 15918-15924.
27. Liu, C.; Luo, H.; Xu, Y.; Zhang, Z.; Liang, Q.; Wang, W.; Chen, Z., Synergistic Cocatalytic Effect of Ultra-thin Metal-organic Framework and Mo-dopant for Efficient Photoelectrochemical Water Oxidation on $BiVO_4$ Photoanode. *Chemical Engineering Journal* **2020,** *384*.
28. Lu, X.; Ye, K.-h.; Zhang, S.; Zhang, J.; Yang, J.; Huang, Y.; Ji, H., Amorphous Type FeOOH Modified Defective $BiVO_4$ Photoanodes for Photoelectrochemical Water Oxidation. *Chemical Engineering Journal* **2022,** *428*.
29. Wang, L.; Mitoraj, D.; Turner, S.; Khavryuchenko, O. V.; Jacob, T.; Hocking, R. K.; Beranek, R., Ultrasmall CoO(OH)x Nanoparticles as a Highly Efficient "True" Cocatalyst in Porous Photoanodes for Water Splitting. *ACS Catalysis* **2017,** *7* (7), 4759-4767.
30. Meirong Huang, Z. H., Hongwei Zhu, Excellent Stability of Molecular Catalyst/$BiVO_4$ Photoanode in Borate Buffer Solution. *Nano Energy* **2020,** *70*, 104487.
31. Sun, Q.; Cheng, T.; Liu, Z.; Qi, L., A Cobalt Silicate Modified $BiVO_4$ Photoanode for Efficient Solar Water Oxidation. *Applied Catalysis B: Environmental* **2020,** *277*.

32. Gao, R.-T.; Wu, L.; Liu, S.; Hu, K.; Liu, X.; Zhang, J.; Wang, L., Boosting the Stability of $BiVO_4$ Photoanodes: In-situ Cocatalyst Passivation and Immobilization by Functional Fluorine Anions. *Journal of Materials Chemistry A* **2021,** *9* (10), 6298-6305.
33. Adeel Mehmood, S. Y. C., Eun Duck Park, $BiVO_4$/Rh–C_i/HCO_3^- Hetero-/homogeneous Dual Co-catalyst-decorated Photoanode System for Photoelectrochemical Water Oxidation. *Journal of Photochemistry & Photobiology, A: Chemistry* **2023,** *436*, 114414.
34. Higashi, M.; Kato, Y.; Iwase, Y.; Tomita, O.; Abe, R., RhO_x Cocatalyst for Efficient Water Oxidation over TaON Photoanodes in Wide pH Range Under Visible-Light Irradiation. *Journal of Photochemistry and Photobiology A: Chemistry* **2021,** *419*.
35. Bledowski, M.; Wang, L.; Neubert, S.; Mitoraj, D.; Beranek, R., Improving the Performance of Hybrid Photoanodes for Water Splitting by Photodeposition of Iridium Oxide Nanoparticles. *The Journal of Physical Chemistry C* **2014,** *118* (33), 18951-18961.
36. Dang, K.; Wang, T.; Li, C.; Zhang, J.; Liu, S.; Gong, J., Improved Oxygen Evolution Kinetics and Surface States Passivation of Ni-B_i Co-Catalyst for a Hematite Photoanode. *Engineering* **2017,** *3* (3), 285-289.
37. Li, H.; Yin, M.; Huang, M.; Xue, X.; Li, X.; Mo, R., Wide-pH-compatible MoS_x Co-catalyst Layer on TiO_2 Nanowire Arrays Photoanode for Simultaneous Acceleration of Charge Carrier Separation and Catalytic Reactions. *Chemical Engineering Journal* **2022,** *450*.

Supplementary Figures

Class	Low	Medium	High
Range (∇y)	0.082 ~1.391	1.391 ~2.695	2.695 ~4.0

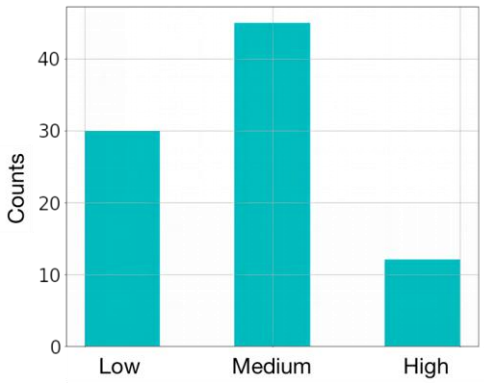


Figure S1. Distribution of the output variable after being categorized. The insert table is the classification basis.

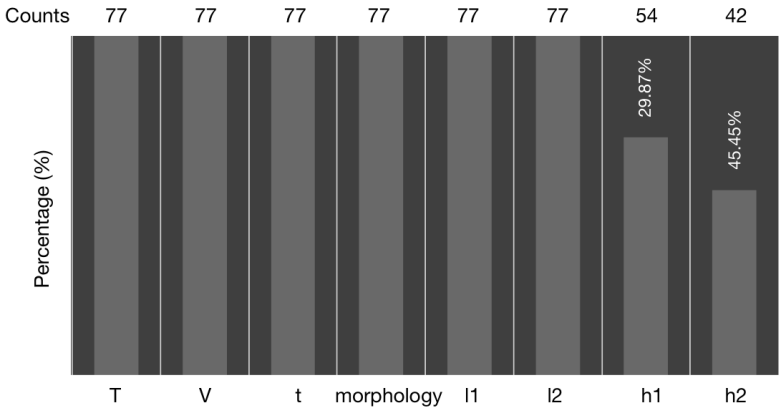


Figure S2. Data integrity of some input variables: anneal temperature (T), deposition voltage (V), deposition time (t), the morphology of photoanode, and the size of photoanode (l1 and l2 are lateral dimensions, h1 is the thickness of a single grain, h2 is the overall thickness). Here, only photoanodes prepared by the electrodeposition-anneal method were considered.

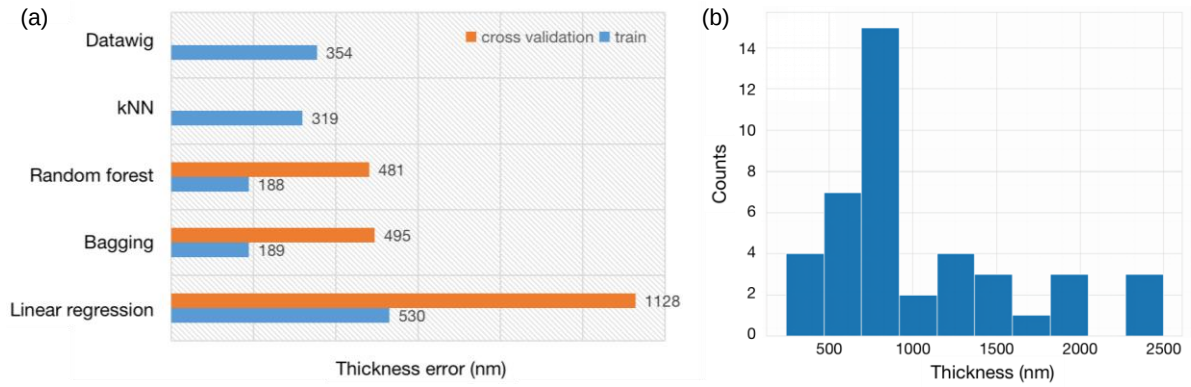


Figure S3. (a) Thickness error of the photoanode thickness filled using single imputation methods: linear regression, bagging, random forest, k-nearest neighbor (kNN), and Datawig (based on a neural network). (b) Thickness distribution of the overall thickness of photoanode.

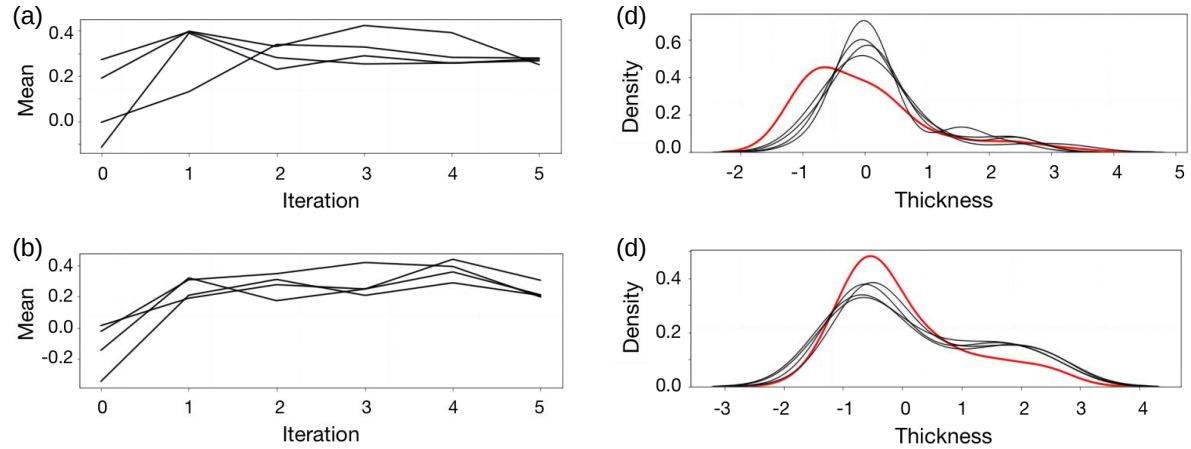


Figure S4. Convergence of the multiple imputations by chained equations (MICE) algorithm used for filling in different input variables: (a) h1, (b) h2. Distributions of the imputed input variables: (c) h1, (d) h2. The red line is the original data, black lines represent the data from each imputation.

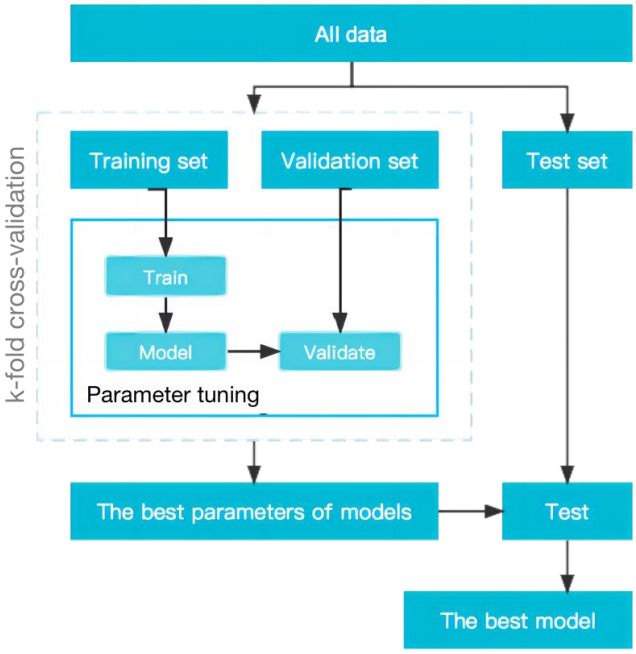


Figure S5. Data splitting for model training, validation, and testing.

(a)

Methods	Equations
Standard (Z-Score)	$x' = \frac{x - mean}{std}$
MinMax	$x' = \frac{x - min}{max - min}$
MaxAbs	$x' = \frac{x}{\lvert max \rvert}$
Robust	$x' = \frac{x - Q(50)}{Q(75) - Q(50)}$
Normalizer	$x' = \frac{x}{\lVert x \rVert_2}$

(b)

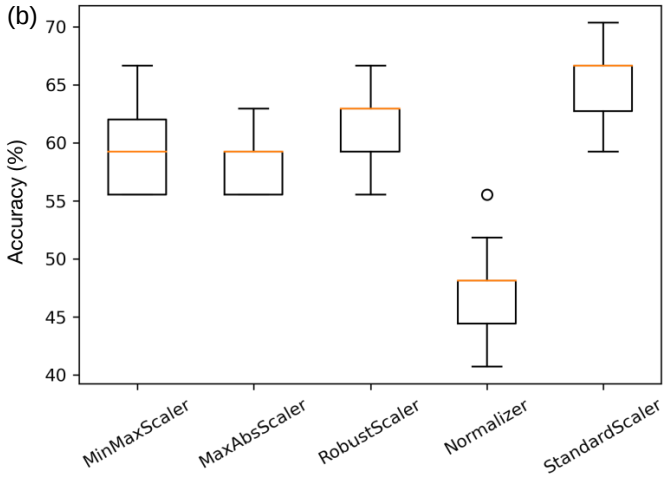


Figure S6. (a) Five data normalization methods and related scaling principles. (b) Test accuracy of neural networks trained on the normalized datasets. The box plot uses boxes and lines to depict the distribution of statistical results, where box limits show the range of the middle 50% of the data with an orange line marking the median value.

(a)

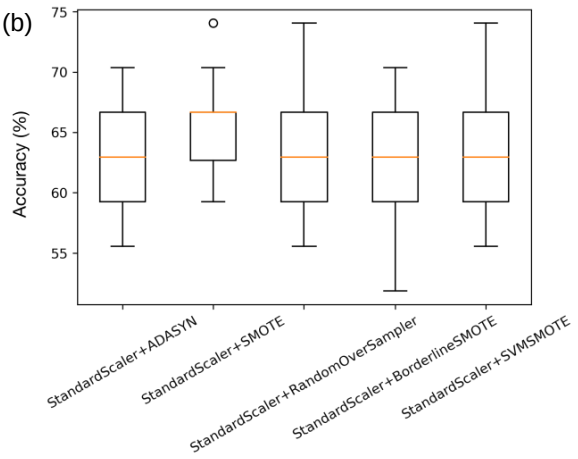
Methods
Collect more data
Convert to anomaly detection problem
Change the class_weight (Cost-sensitive learning)
Move the threshold
Data resampling

Undersampling
- Random
- ENN
- Tomek links
- EasyEnsemble
- BalanceCascade
- NearMiss
- Ensemble

Oversampling
- Random
- SMOTE
- BorderlineSMOTE
- ADASYN
- SVMSMOTE

Combination
- ENN+SMOTE

(b)



(c)

Methods	Maco-F1-score	AUC
ADASYN	0.581	0.739
SMOTE	0.666	0.777
RandomOverSampler	0.597	0.757
BorderlineSMOTE	0.601	0.747
SVMSMOTE	0.59	0.763

Figure S7. (a) Some solutions to data imbalance. Predicted results of neural networks trained on the preprocessed datasets: (b) test accuracy, (c) macro-F1-score and area under the curve (AUC) of the receiver operating characteristic curve (ROC).

(a)

Methods
Proportion of missing values
Low variance filter
High correlation filter
Random forest (Feature importance)
Reverse feature elimination
Forward feature selection
Principle component analysis
Independent component analysis
Local linear embedding
t-distributed stochastic neighbour embedding

(b)

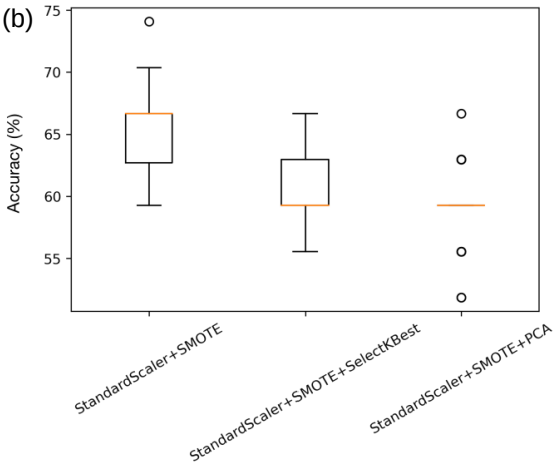
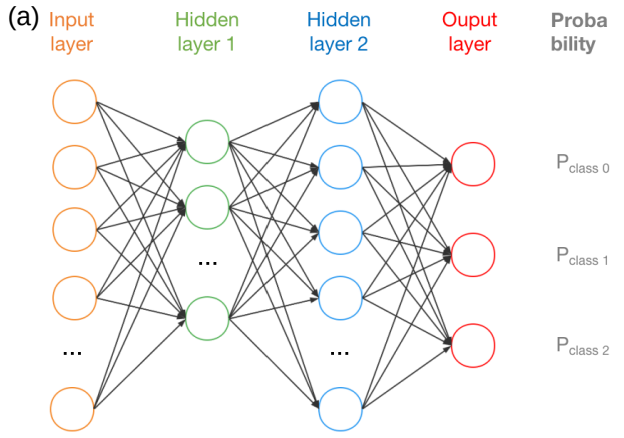


Figure S8. (a) Several dimensionality reduction methods, the first six methods belong to feature selection, while the last four methods belong to feature extraction. (b) Test accuracy of neural networks trained on the preprocessed datasets.



(b)

Hyperparameter	Value
Units1	32
Activation	relu
Initializer	normal
Regularizer	1e-6
Dropout	0.4
Units2	92
Units3	3
	softmax
Learning_rate	0.01
Decay	0.01
Optimizer	adam
Epoch	early stopping

Figure S9. (a) A multi-layer perceptron neural network with two hidden layers. (b) Optimized hyperparameters of the neural network.

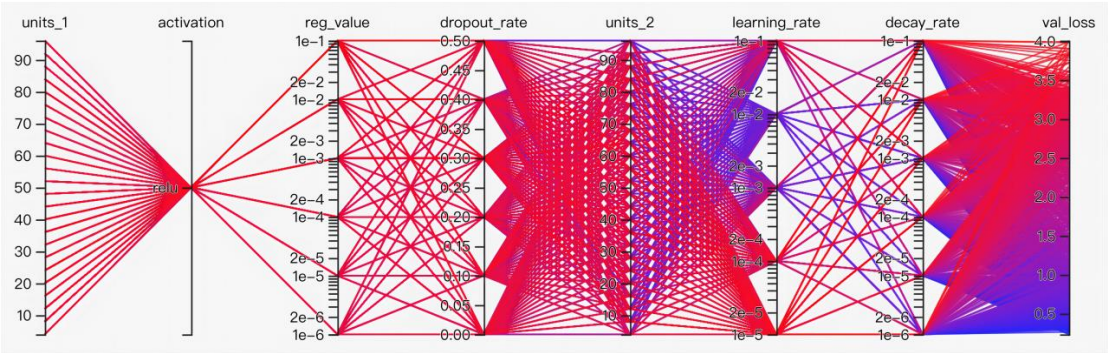


Figure S10. Hyperparameters and their values considered in the neural network.

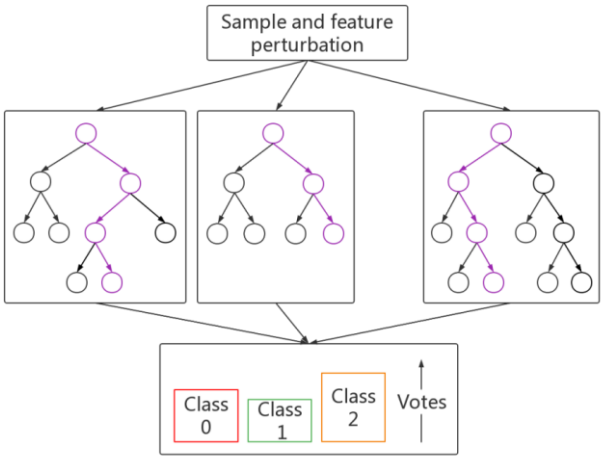


Figure S11. The basic architecture of a random forest model.

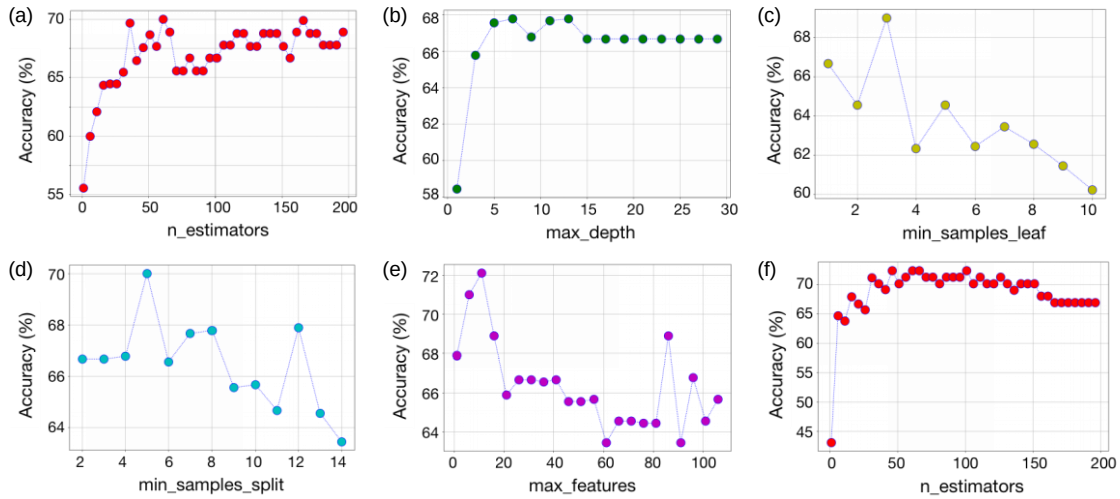


Figure S12. Cross-validation accuracy of random forest models varying with hyperparameters: (a) n_estimators, (b) max_depth, (c) min_samples_leaf, (d) min_samples_split, (e) max_features, (f) n_estimators (second optimization)

Parameter	Value	Accuracy
n_estimators	61	70
max_depth	7	67.78
min_samples_leaf	3	69
min_samples_split	5	70
max_features	11	72.11
'max_depth': **8**, 'max_features': **8**, 'min_samples_leaf': **2**, 'min_samples_split': **5**, 'n_estimators': **46**		

Figure S13. Optimized hyperparameters of the random forest model. The preceding lines show the best values and results (cross-validation accuracy) of each hyperparameter individually optimized, and the last line shows the best parameter combination optimized by the gride search of all parameters.

Parameter	Value	Accuracy
n_estimators	116	68.89
max_samples	0.95	66.67
max_features	0.35	70
n_estimators=80, max_samples=0.8, max_features=0.2		
max_depth	11	74.22
min_samples_split	2	74.22
min_samples_leaf	1	74.22
max_depth=11, min_samples_split=2, min_samples_leaf=1		
max_depth=**11**, min_samples_split=**2**, min_samples_leaf=**1**, max_samples=**0.8**, max_features=**0.2**, n_estimators=**80**		

Figure S14. Optimized hyperparameters of the bagging model.

Parameter	Value	Accuracy
n_estimators	71	61.33
learning_rate	0.7	62.22
n_estimators=80, learning_rate=0.7		
max_depth	2	66.67
min_samples_split	14	68.78
min_samples_leaf	3	68.89
max_depth=5, min_samples_split=10, min_samples_leaf=2		
max_depth=**5**, min_samples_split=**10**, min_samples_leaf=**2**, n_estimators=**80**, learning_rate=**0.6**		

Figure S15. Optimized hyperparameters of the AdaBoost model.

Parameter	Value	Accuracy
n_estimators	51	62.44
max_depth	2	62.67
min_samples_split	2 3	73.11
min_samples_leaf	1	
max_features	8	71.11
subsample	0.8	71.11
learning_rate	0.15	73.22
n_estimators	46	
learning_rate= **0.15**, n_estimators= **46**, max_depth= **2**,min_samples_split= **3**, min_samples_leaf= **1**, max_features= **8**, subsample= **0.8**		

Figure S16. Optimized hyperparameters of the gradient boosting model.

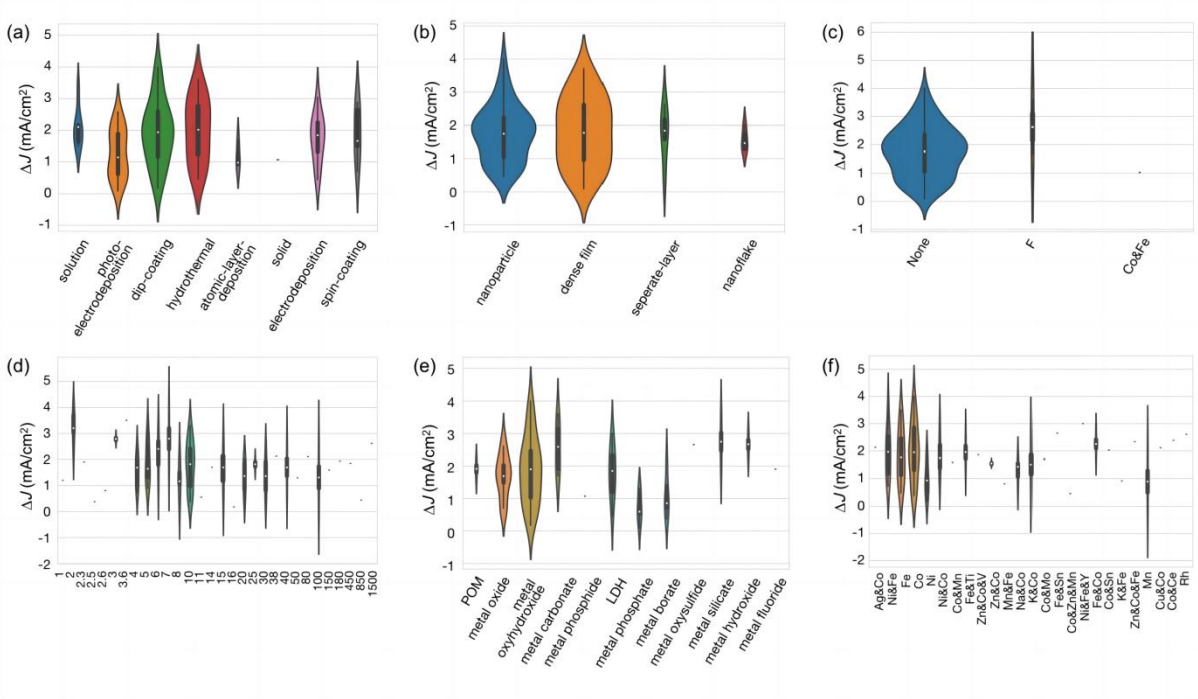


Figure S17. Effect of different cocatalyst parameters on the performance boost: (a) preparation method of cocatalyst, (b) cocatalyst morphology, (c) dopant of cocatalyst, (d) cocatalyst thickness, (e) cocatalyst type, (f) metal type of cocatalyst.

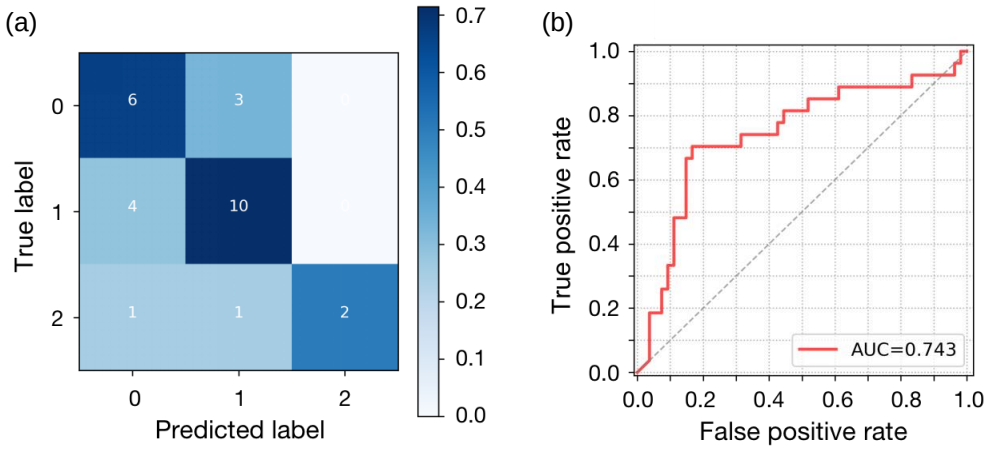


Figure S18. Testing results of the neural network trained on the dataset normalized by the StandardScaler method: (a) confusion matrix, (b) ROC curve.

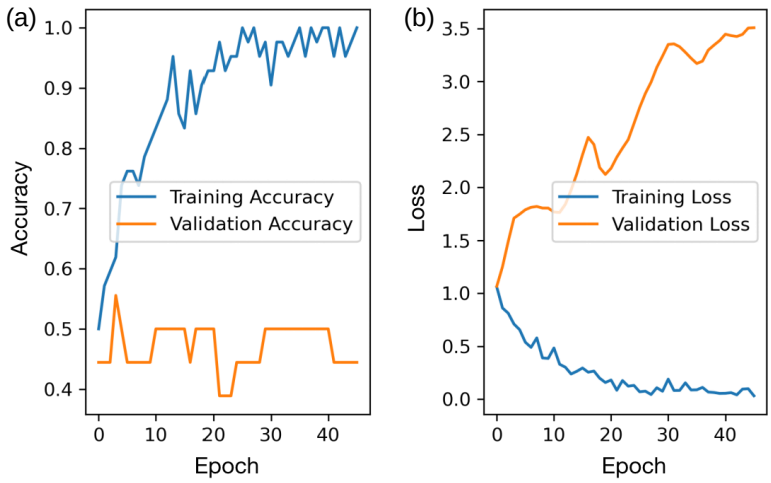


Figure S19. Training curves of the neural network trained on the dataset normalized by the StandardScaler method, showed in the form of (a) accuracy and (b) loss.

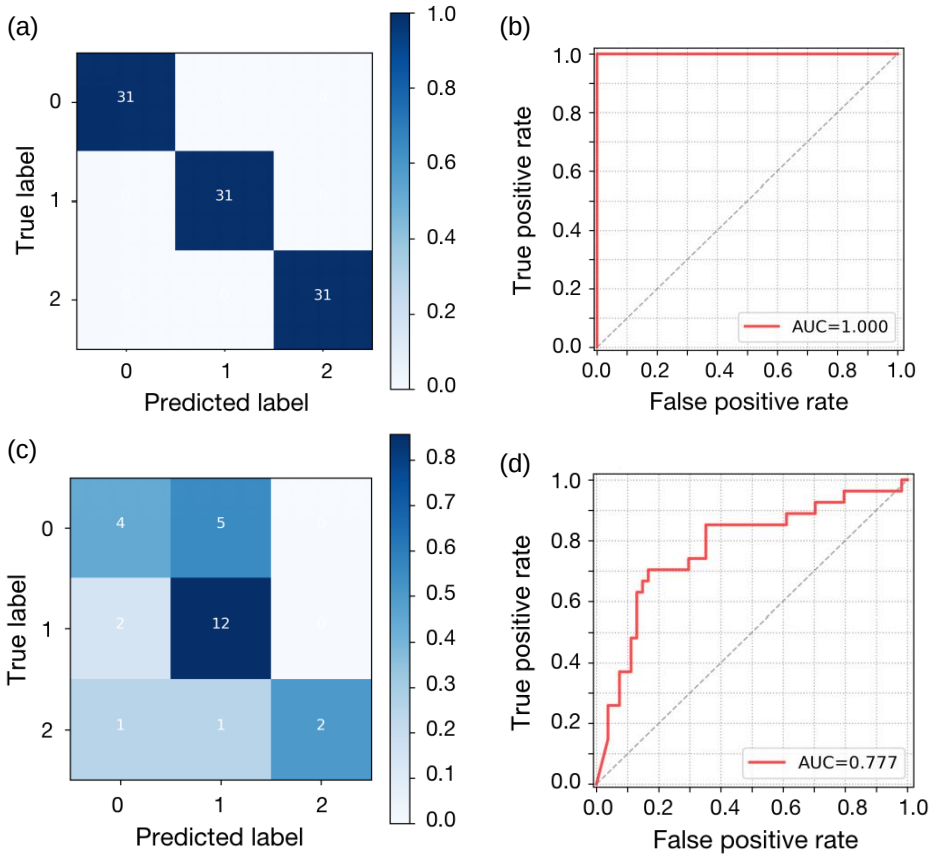


Figure S20. Training (a, b) and testing (c, d) results of the neural network trained on the dataset using the SMOTE oversampling method: (a, c) confusion matrix, (b, d) ROC curve.

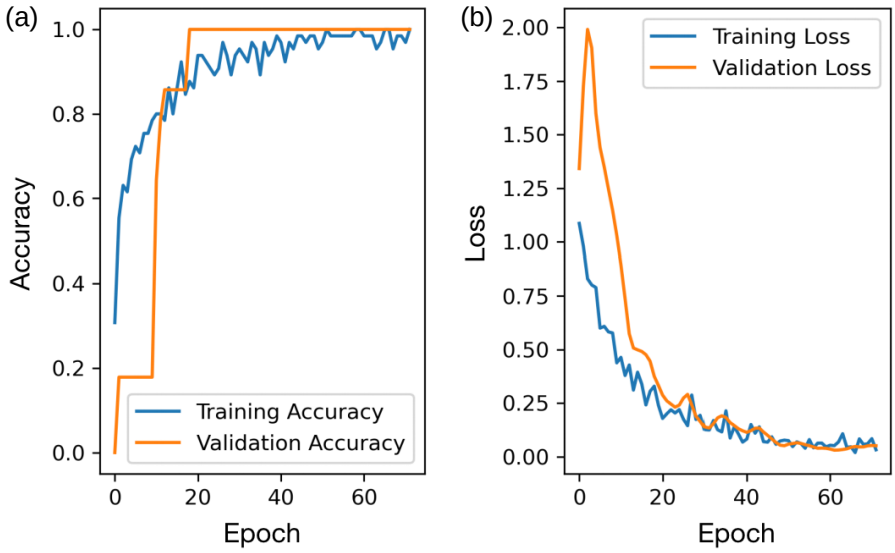


Figure S21. Training curves of the neural network trained on the dataset using the SMOTE oversampling method, showed in the form of (a) accuracy and (b) loss.

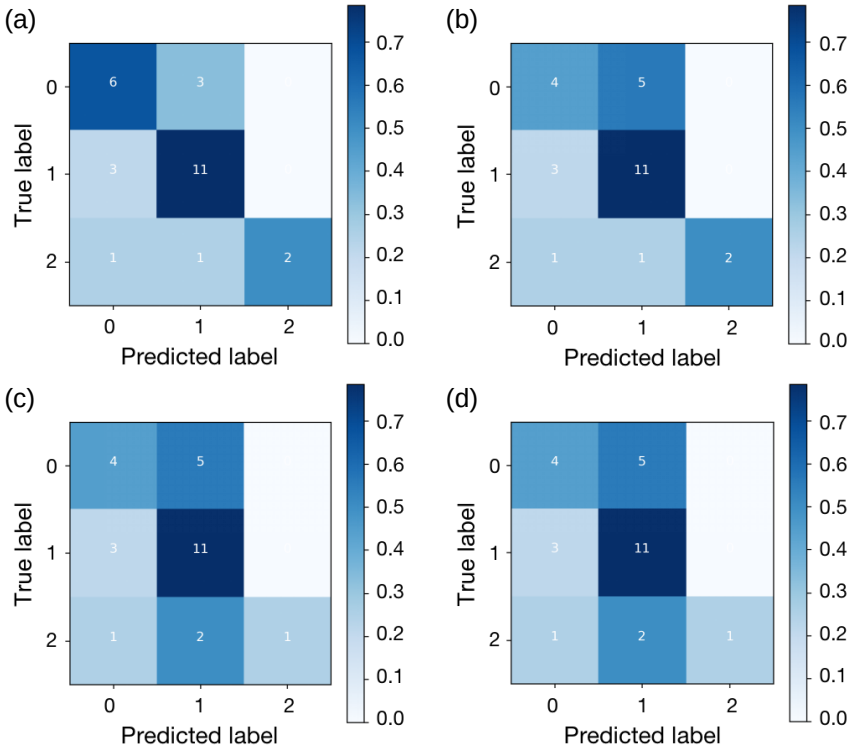


Figure S22. Confusion matrixes of testing results for different ensemble methods: (a) random forest, (b) bagging, (c) AdaBoost, (d) gradient boosting

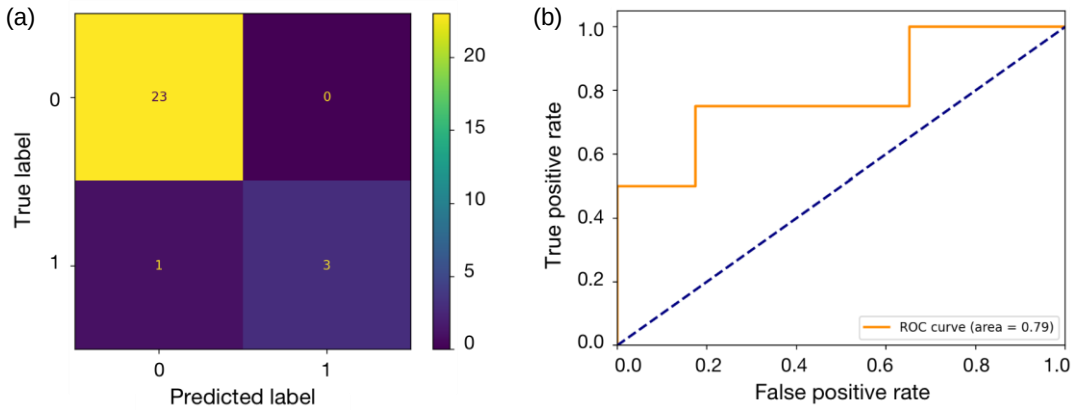


Figure S23. Testing results of the random forest used for the binary classification: (a) confusion matrix, (b) ROC curve.

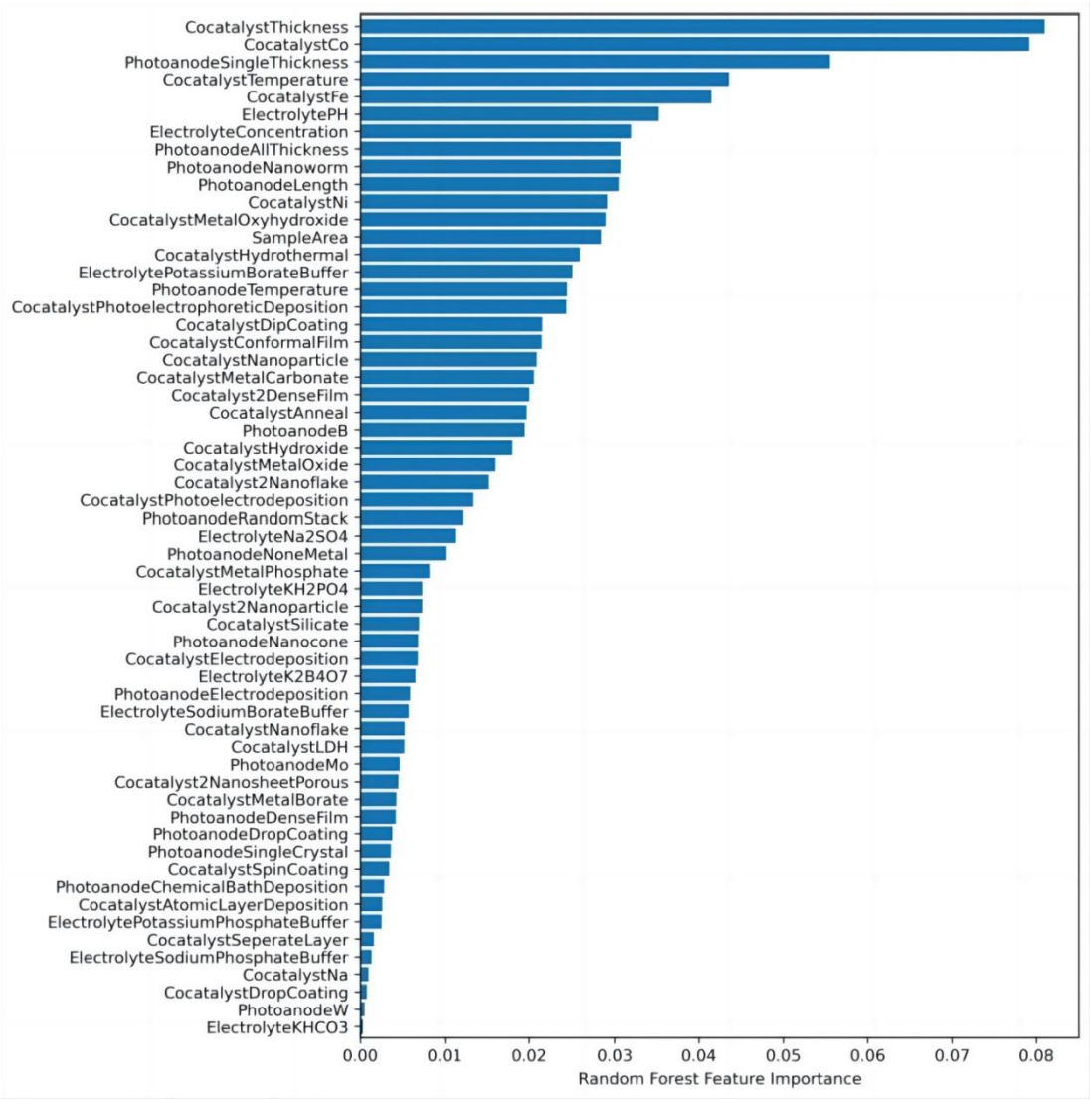


Figure S24. Random forest built-in feature importance of top input variables.

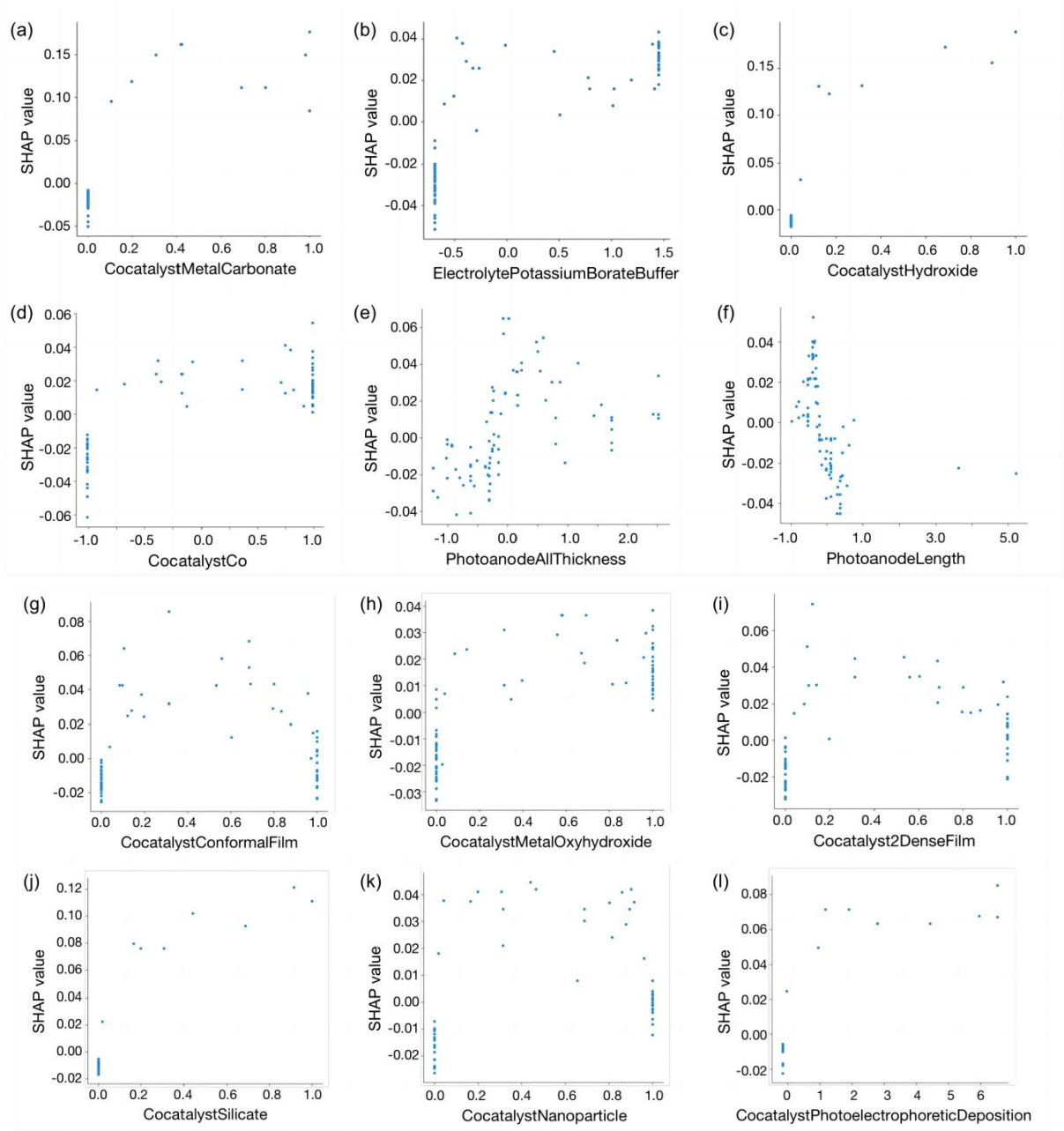


Figure S25. SHAP dependence plots of several input variables: (a) CocatalystMetalCarbonate, (b) ElectrolytePotassiumBorateBuffer, (c) CocatalystHydroxide, (d) CocatalystCo, (e) PhotoanodeAllThickness, (f) PhotoanodeLength, (g) CocatalystConformalFilm, (h) CocatalystMetalOxyhydroxide, (i) Cocatalyst2DenseFilm, (j) CocatalystSilicate, (k) CocatalystNanoparticle, and (l) CocatalystPhotoelectrophoreticDeposition.