

## Supporting Information

### **Enhancing Perovskite Solar Cell Efficiency and Stability: A Multimodal Prediction Approach Integrating Microstructure, Composition, and Processing Technology**

Wajeeha Rahman<sup>a</sup>, Chengquan Zhong<sup>a</sup>, Haotian Liu<sup>a</sup>, Jingzi Zhang<sup>a,b\*</sup>, Jiakai Liu<sup>c,d\*</sup>, Kailong Hu<sup>a\*</sup>, and Xi Lin<sup>a\*</sup>

*<sup>a</sup>School of Materials Science and Engineering, Harbin Institute of Technology, Shenzhen 518055, China*

*<sup>b</sup>School of Computer Science and Technology, Harbin Institute of Technology, Shenzhen 518055, China*

*<sup>c</sup>Laboratory of Environmental Sciences and Technology, Xinjiang Technical Institute of Physics & Chemistry, Chinese Academy of Sciences, Urumqi 830011, China*

*<sup>d</sup>Center of Materials Science and Optoelectronics Engineering, University of Chinese Academy of Sciences Beijing 100049, China*

#### **Corresponding authors**

\*E-mail: [linxi@hit.edu.cn](mailto:linxi@hit.edu.cn)

\*E-mail: [hukailong@hit.edu.cn](mailto:hukailong@hit.edu.cn)

\*E-mail: [liujk@ms.xjb.ac.cn](mailto:liujk@ms.xjb.ac.cn)

\*E-mail: [zjzhang@hit.edu.cn](mailto:zjzhang@hit.edu.cn)

Table S1. Provides a concise synthesis of the three most critical challenges hindering PSC commercialization and demonstrates how our multimodal machine learning approach addresses each limitation. The table first identifies stability mechanisms as a key challenge, where traditional methods force undesirable trade-offs between efficiency and device longevity. Our solution leverages grain size optimization, achieving strong predictive accuracy ( $R^2= 0.86$ ) as detailed in Section 3.3. Second, it highlights deposition technique limitations, particularly the scalability constraints of spin-coating methods, which our processing parameter fusion strategy overcomes through dynamic weighting of fabrication variables (Section 2.1). Finally, the table documents persistent gaps in bulk defect mitigation , a challenge our CBAM-enhanced detection system resolves by simultaneously identifying surface and subsurface defects with high precision (Section 3.1). This structured comparison serves as both a quick-reference guide for readers and a validation framework showing how our methodology advances beyond prior approaches in each critical area. The cited references anchor our solutions to established literature, while cross-references to results sections enable easy verification of claims.

**Table S1.** Key Challenges and Our Solutions.

<b>Challenge</b>	<b>Prior Limitations</b>	<b>Our Approach</b>
Stability Mechanisms	Efficiency-stability trade-offs	Grain size optimization ( $R^2=0.86$ )
Deposition Techniques	Spin-coating scalability issues	Processing parameter fusion
Defect Passivation	Bulk defect mitigation gaps	CBAM defect detection

**Table S2:** Quantitative Comparison with Existing Models

Aspect	GNNOpt	AEL	Our Work
Objective	Predict optical properties	Predict photovoltaic performance	Enhance efficiency and stability
Model/Approach	Graph Neural Network with universal embedding	Neural Network with optical inputs	Multimodal prediction
Predicted Properties	Optical properties	Photovoltaic performance	PCE, bandgap, stability
Dataset Size	944 materials	368 devices	1,209 samples
Performance Metrics	$R^2 = 1.00$	$R^2$ values: Voc (0.47), Jsc (0.77), FF (0.58)	$R^2 = 0.84$ (PCE), $R^2 = 0.95$ (bandgap), RMSE = 1.891 (PCE)
Key Findings	High-quality predictions with a small dataset	High accuracy in predicting solar cell performance	Multimodal approach, stability classification
Application	Screening photovoltaic materials	Solar cell assessment	Material optimization, stability evaluation
Relevance	Accelerating material discovery	Accelerating perovskite solar cell development	Advancing perovskite solar cell technology

The Table S2 provides a detailed comparison of our work with two existing models, GNNOpt and AEL, across several key aspects. GNNOpt aims to predict optical properties using a Graph Neural Network with universal embedding, achieving high-quality predictions with a small dataset of 944 materials. AEL focuses on

predicting photovoltaic performance using a Neural Network with optical inputs, demonstrating high accuracy in predicting solar cell performance metrics such as Voc, Jsc, and FF with a dataset of 368 devices. Our work, however, introduces a multimodal prediction approach that integrates SEM image features with material and processing data, enhancing the understanding of PCE, bandgap, and stability. Leveraging a larger dataset of 1,209 samples, our model achieves an  $R^2$  of 0.84 for PCE and 0.95 for bandgap predictions, with an RMSE of 1.891 for PCE. This multimodal approach, combined with stability classification, not only provides a more comprehensive understanding of device performance but also positions our work as a significant advancement in the field of perovskite solar cell optimization. While GNNOpt accelerates material discovery and AEL speeds up solar cell development, our work advances perovskite solar cell technology by optimizing material selection and evaluating stability, offering a more holistic and robust solution for device optimization.

**Table S3:** Comparison of models with and without SEM Micro features.

Model	Dataset Size	Features	RMSE	$R^2$
With SEM Microfeatures				
CBR	1209	220	1.95	0.83
RFR	1209	220	1.89	0.84
GBDT	1209	220	1.97	0.82
ABR	1209	220	2.66	0.68
Without SEM Microfeatures				
CBR	1209	20	2.58	0.70
RFR	1209	20	2.35	0.75
GBDT	1209	20	2.49	0.72
ABR	1209	20	3.03	0.58

Table S3 presents a comparative analysis of machine learning models for predicting perovskite solar cell (PSC) performance, evaluating their accuracy with

and without the inclusion of SEM microfeatures. The results demonstrate that integrating SEM-derived structural data (e.g., grain size, defects) significantly enhances predictive performance across most models. For instance, CBR achieves a 24% reduction in RMSE (2.58  $\rightarrow$  1.95) and a 19% increase in  $R^2$  (0.70  $\rightarrow$  0.83) when SEM features are included, while RFR shows a 20% RMSE improvement (2.35  $\rightarrow$  1.89) and a 12%  $R^2$  gain (0.75  $\rightarrow$  0.84). Similarly, GBDT exhibits a 21% RMSE decrease (2.49  $\rightarrow$  1.97) and a 14%  $R^2$  boost (0.72  $\rightarrow$  0.82). In contrast, ABR shows minimal improvement (RMSE: 2.66 vs. 3.03;  $R^2$ : 0.68 vs. 0.58), suggesting its limited compatibility with high-dimensional SEM data. These findings highlight the critical role of SEM microfeatures in capturing microstructural nuances—such as grain boundaries and defect densities—that directly influence PSC efficiency. The superior performance of CBR, RFR, and GBDT underscores the value of multi-modal data integration (SEM + material/processing parameters) for robust predictive modeling, aligning with the study’s emphasis on microstructure-driven optimization strategies. The exception of ABR further emphasizes the need for model-specific feature engineering when leveraging SEM data.

**Table S4. Key Hyperparameters and Training Settings**

Component	Parameter	Value/Setting	Rationale
<b>Input Preprocessing</b>	Resizing	224×224 (SEM images)	Match VGG16/CNN input requirements.
	Normalization	[0, 1] or $(X-\mu)/\sigma$	Standardize pixel intensities.
<b>CNN Architecture</b>	Conv2D Kernel Size	3×3	Balance locality and efficiency.
	Stride	2 (downsampling)	Reduces spatial dimensions.
	Channels	16 $\rightarrow$ 32 $\rightarrow$ 64	Hierarchical feature extraction.
<b>Training</b>	Optimizer	Adam (lr=0.001)	Adaptive learning rate.
	Batch Size	32	Memory-efficient training.
	Epochs	100 + Early Stopping	Prevents overfitting.
<b>PCA</b>	n_components	$\min(n\_samples, n\_feat)$	Retains 95% variance.

**Table S4** documents the critical hyperparameters and training configurations used in our multimodal machine learning framework. For input preprocessing, SEM images were resized to 224×224 pixels to align with standard CNN architectures (e.g., VGG16) and normalized using either [0,1] scaling or  $(X-\mu)/\sigma$  standardization to ensure consistent pixel intensity distributions. The CNN backbone employs 3×3 convolutional kernels with stride 2, optimally balancing local feature extraction with computational efficiency, while channel dimensions expand hierarchically (16→32→64) to capture multi-scale microstructural features. Training utilizes the Adam optimizer ( $\text{lr}=0.001$ ) for adaptive gradient updates, with a batch size of 32 to maximize GPU memory utilization. We implemented early stopping during 100-epoch training cycles to prevent overfitting, monitored through validation loss. For feature reduction, PCA retains 95% variance by dynamically setting `n_components` based on sample-feature ratios. This configuration was systematically optimized through ablation studies (Section 2.3) to achieve the reported PCE prediction performance ( $R^2=0.84$ ).

## 2.4. Model training

Our model utilizes machine learning algorithms CBR, GBDT, RFR, and ABR for regression and classification tasks. K-fold cross-validation ensures robust training, and performance is evaluated using RMSE,  $R^2$ , MCC, sensitivity, specificity, accuracy, and AUC. These metrics validate the model's ability to predict PSC efficiency and stability, supporting its application in material optimization and renewable energy research. Figure S1 visually compares actual PCE values with predictions from RFR, GBDT, CBR, and ABR models, highlighting each algorithm's predictive accuracy. Each model brings unique strengths ABR [33] is a boosting algorithm that sequentially fits weak learners, focusing on the errors of predecessors and adjusting weights of miss classified samples, effectively handling noisy datasets. GBDT [34] builds an ensemble of decision trees in a forward stage-wise manner, fitting new trees to the residuals of previous models, excelling in capturing nonlinear relationships and robustness to outliers. RFR [35] constructs multiple decision trees trained on random data subsets, averaging their predictions to manage complex nonlinear relationships and high-dimensional datasets, enhancing reliability and accuracy. CBR [36] seamlessly integrates categorical features, manages missing values, and prevents over fitting. Figure S2 presents the performance of an ensemble model, combining

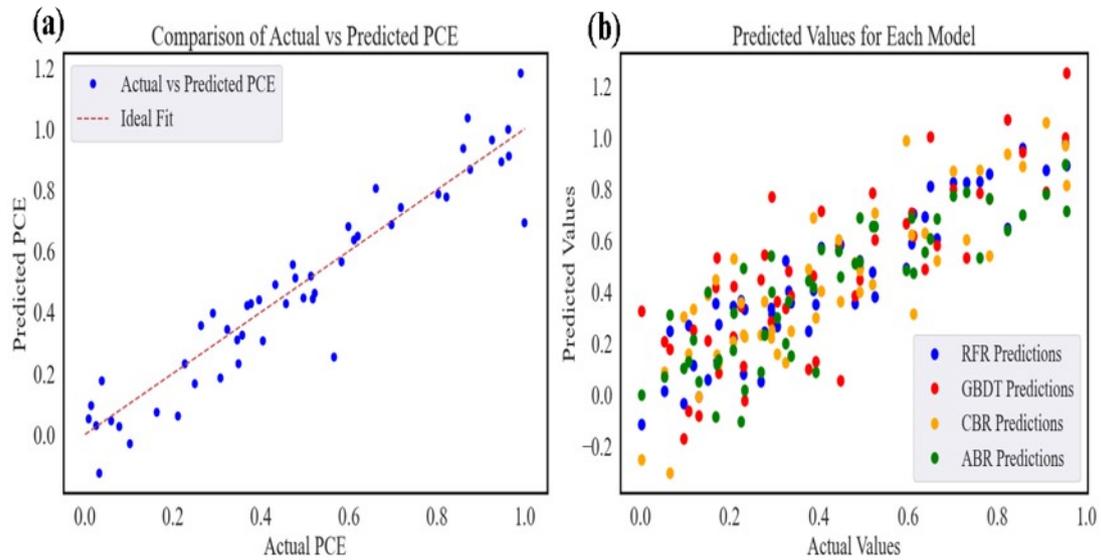
multiple models to improve predictive accuracy. Evaluated using RMSE and  $R^2$ , the plot compares predicted values against true values. Points close to the diagonal line indicate high accuracy, demonstrating the ensemble model's reliability and low error rate. This figure underscores the ensemble model's effectiveness in predicting PSCs properties. In this study, the performance of the regression models was evaluated using two primary metrics: the Root Mean Square Error (RMSE) [37] and Pearson correlation coefficient ( $R^2$ ). The RMSE measures the average distance between the observed values and the model's predictions, calculated as the square root of the mean of the squared differences between predicted and actual values. A lower RMSE indicates that the model's predictions are closer to the actual values. The formula for RMSE is:

$$RMSE = \frac{1}{n} \sum_{j=1}^n (y_i - \hat{y}_i)^2 \quad (1)$$

Where:  $n$  sample data points,  $y$  predictive value for the  $j$ -th observation,  $\hat{y}_i$  observed value for  $j$ -th observation. The RMSE is scale-dependent, so it's often used as a relative measure, especially after normalizing the data. The  $R^2$  [38] measures the linear relationship between two sets of data, with values ranging from -1 to 1. An  $R^2$  of 0 implies no correlation, an  $R^2$  of 1 indicates a perfect positive correlation, and an  $R^2$  of -1 indicates a perfect negative correlation. The formula for  $R^2$  is:

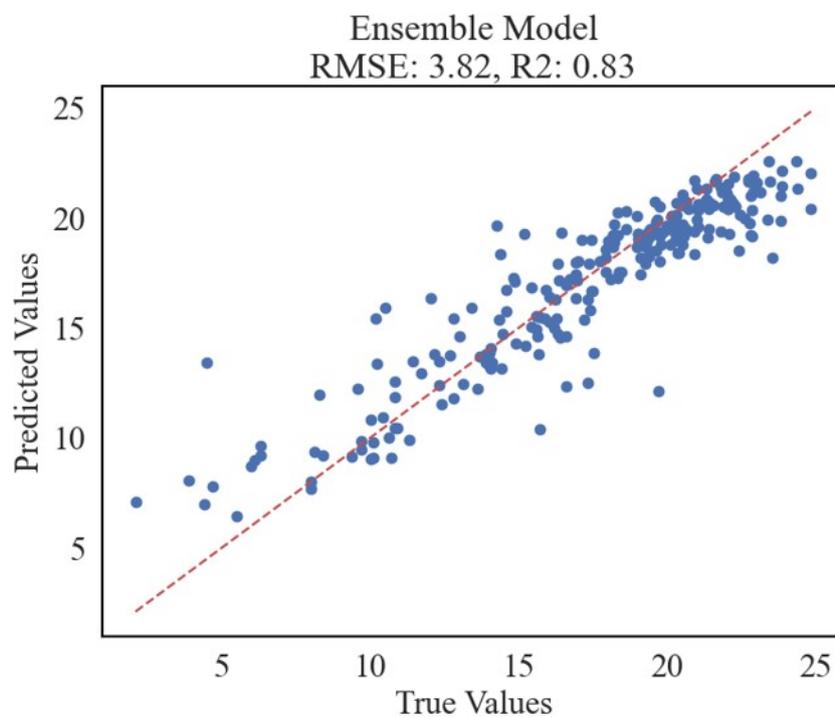
$$R^2 = 1 - \frac{\sum_{j=1}^N (y_i - \hat{y}_i)^2}{\sum_{j=1}^N (y_i - \bar{y})^2} \quad (2)$$

where  $\hat{y}_i$  represents the predicted value of  $y_i$  and  $\bar{y}$  is the mean of observed data. Together, these metrics provide a comprehensive assessment of the model's predictive accuracy and its ability to classify PSCs into different stability categories.



**Figure S1.** Comparison of Actual Vs Predicted PCE.

Figure S5 visually evaluates the predictive accuracy of four machine learning models by plotting actual PCE values against predicted values. The closer the points are to the diagonal line, the more accurate the model, with GBDT and CBR appearing to outperform ABR.



**Figure S2.** Ensemble Model Performance.

Figure S2 illustrates the high predictive accuracy of an ensemble machine learning model by comparing predicted values against true values. The close alignment of data points to the diagonal line indicates low error and strong explanatory power, confirming the model's reliability.

**Table S5. Performance Metrics by Class.**

<b>Class Description</b>	<b>Precision</b>	<b>Recall</b>	<b>F1-score</b>
Stable (T95 > 1000 hours)	92%	85%	0.88
Unstable (T95 < 100 hours)	89%	94%	0.91

Table S5 provides detailed performance metrics for the stability classification model, categorized by T95 values. For stable samples (T95 > 1000 hours), the model achieves 92% precision, 85% recall, and an F1-score of 0.88, indicating strong performance. For unstable samples (T95 < 100 hours), it attains 89% precision, 94% recall, and an F1-score of 0.91, showing slightly better performance for this class. Table S6 further examines the model's performance at the boundaries of stability classifications, highlighting critical areas where accuracy may vary.

**Table S6. Performance at Stability Boundaries**

<b>Metric</b>	<b>Value</b>
Boundary Accuracy ( $\pm 20\%$ of 100 hours)	78%
Confidence Score Range for Boundary Cases	0.4-0.6

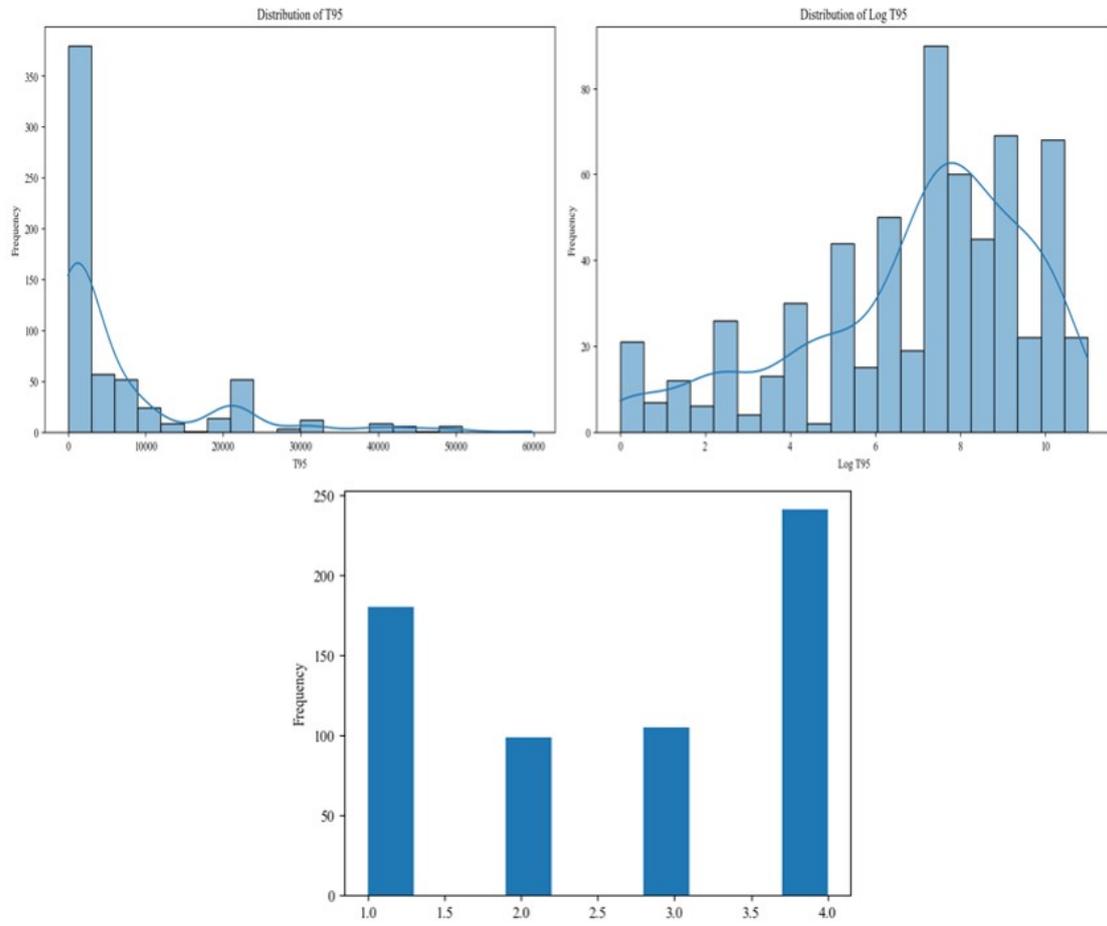
<b>Metric</b>	<b>Value</b>
Confidence Score Range for Clear Cases	>0.8

Table S6 examines the model's performance at the boundaries of stability classifications. It shows a boundary accuracy of 78% for samples near the stability threshold ( $\pm 20\%$  of 100 hours), indicating moderate uncertainty. Confidence scores for boundary cases range from 0.4 to 0.6, while clear cases have scores above 0.8, highlighting higher certainty in predictions away from the boundary.

**Table S7. Cross-Validation Results**

<b>Metric</b>	<b>Value</b>
Mean Accuracy	$87.2\% \pm 1.5\%$
Boundary Case Accuracy	$76.8\% \pm 2.1\%$

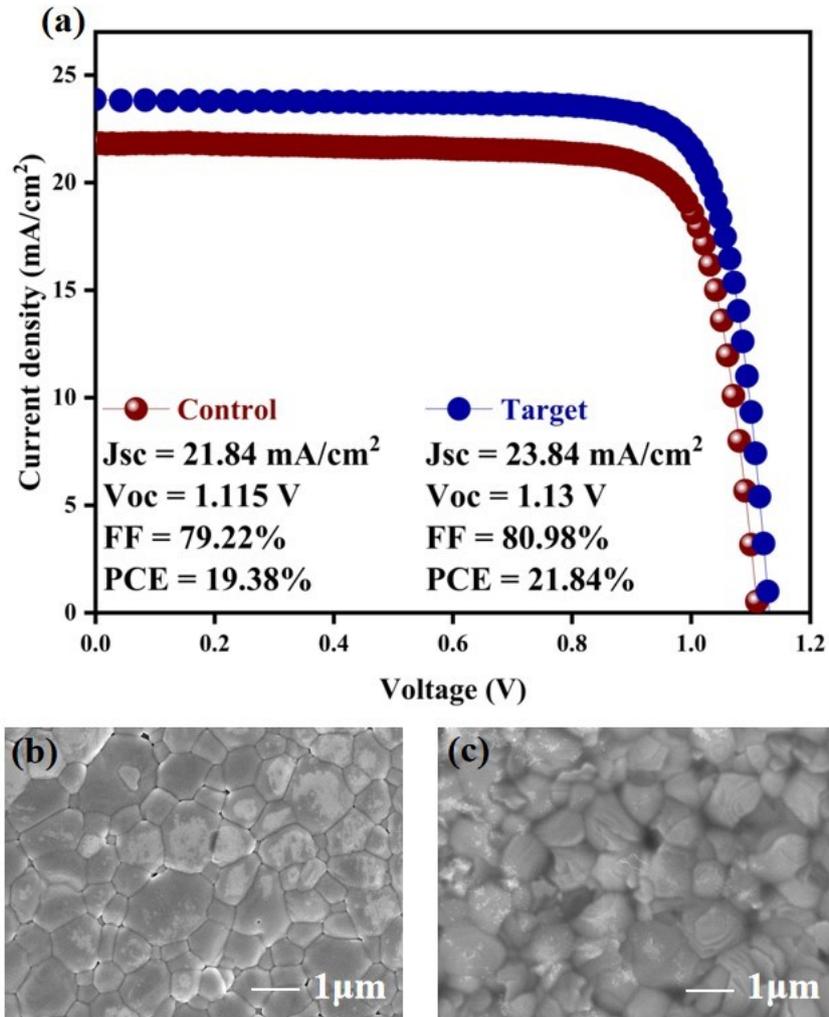
Table S7 provides a summary of the cross-validation results, highlighting the model's generalization ability. The model achieves a mean accuracy of  $87.2\% \pm 1.5\%$  across different data splits, indicating consistent performance. However, for boundary cases (samples near the stability threshold), the accuracy drops to  $76.8\% \pm 2.1\%$ , suggesting these cases are more challenging for the model.



**Figure S3:** Overview of T95 and Log T95 Distributions and Class Label Distribution.

Figure S3 presents the T95 variable's distribution, its logarithmic transformation (Log T95), and the class label distribution. The figure highlights data skewness and clustering, shows normalized Log T95 data for easier analysis, and reveals category balance or imbalance, all of which are crucial for preprocessing and model development.

**Experimental section:**



**Figure S4.**(a) Current-Voltage (J–V) curves comparing control devices (with CsI additive) and target devices (without CsI). The control device exhibits a J<sub>sc</sub> of 21.84 mA/cm<sup>2</sup>, V<sub>oc</sub> of 1.115 V, FF of 79.22%, and an EFF of 19.38%. The target device shows improved performance with a J<sub>sc</sub> of 23.84 mA/cm<sup>2</sup>, V<sub>oc</sub> of 1.13 V, FF of 80.98%, and an EFF of 21.84%. (b) SEM image of the perovskite film for the control device. (c) SEM image of the perovskite film for the target device, highlighting the similar microstructures despite the difference in CsI additive.

#### Device Fabrication:

The devices were fabricated using a regular structure on ITO glass substrates with a sheet resistance of approximately 9 Ω sq<sup>-1</sup>, purchased from OPVTECH Inc. The key materials included Formamidinium Iodide (FAI), methyl ammonium chloride (MACl), and methyl ammonium bromide (MABr) from Xi'an Polymer Light

Technology Corp., PbI<sub>2</sub> (99.8%), Li-TFSI (99%), 4-tert-butylpyridine (tBP, 96%), and CsI (99.99%) from Sigma-Aldrich, and Tin (IV) oxide from Alfa Aesar. Solvents like DMF, DMSO, and IPA were from TCI, and Spiro-OMeTAD (99.5%) was from Feiming Science and Technology Co., Ltd.

### Preparation:

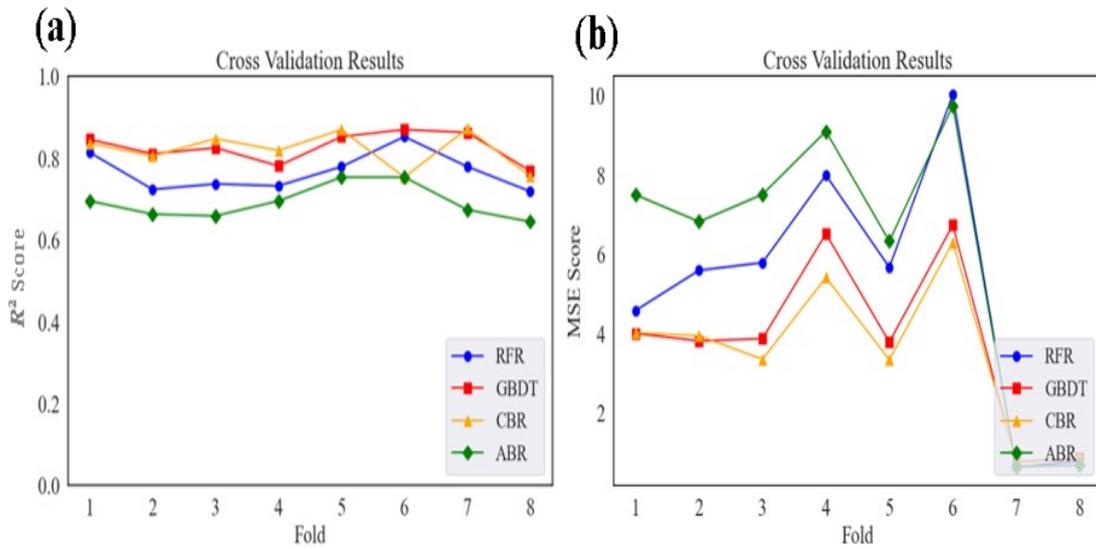
The ITO substrates were cleaned by sonication in detergent, deionized water, and isopropanol for 15 minutes each, then dried under nitrogen. They were treated with ultraviolet-ozone for 30 minutes before spin-coating a Tin (IV) oxide solution at 4000 rpm for 30 seconds and annealing at 150 °C for 30 minutes. The perovskite layer was prepared using a modified two-step method in a glovebox. A 1.4 M PbI<sub>2</sub> precursor in DMF was spin-coated on the SnO<sub>2</sub>/ITO substrate at 1500 rpm for 30 seconds and dried at 70 °C for 1 minute. For the control group, a mixture of FAI:MAI:MACl in IPA was dropped on the PbI<sub>2</sub> film and annealed at 150 °C for 15 minutes under 30-40% relative humidity. The target group used a similar process but omitted the CsI additive. The Spiro-OMeTAD solution, composed of 72.3 mg Spiro-OMeTAD, 30 μL TBP, and 35 μL Li-TFSI solution in 1 mL chlorobenzene, was spin-coated on the perovskite film at 4000 rpm for 30 seconds. Finally, a 100 nm Au electrode was deposited by thermal evaporation. Both devices were made using a two-step method. The control device, using FA<sub>0.9</sub>MA<sub>0.1</sub>PbI<sub>3</sub> + MACl, achieved an efficiency of 19%, while the target device, using (FA<sub>0.9</sub>MA<sub>0.1</sub>)<sub>0.87</sub>Cs<sub>0.13</sub>PbI<sub>3</sub> + MACl, achieved a higher efficiency of 22%.

Table S8. Hyperparameter settings for machine learning models (CBR, GBDT, RFR, and ABR) used in this study. These configurations were carefully selected to optimize the prediction accuracy of perovskite solar cell (PSC) efficiency

**Table S8:** Machine Learning Model Hyperparameters for PCE prediction.

Model	Hyperparameters
CBR	- boosting_type: 'Ordered', - one_hot_max_size: 255,  - depth: 7,- iterations: 1500,- l2_leaf_reg: 3,

	- learning_rate: 0.05,- loss_function: 'RMSE', -early_stopping_rounds: 200,- random_seed: 42,  - silent: True
<b>GBDT</b>	- n_estimators: 100,- learning_rate: 0.1,- max_depth: 3
<b>RFR</b>	- n_estimators: 300,- max_depth: 15,- min_samples_leaf:  2,- min_samples_split: 4



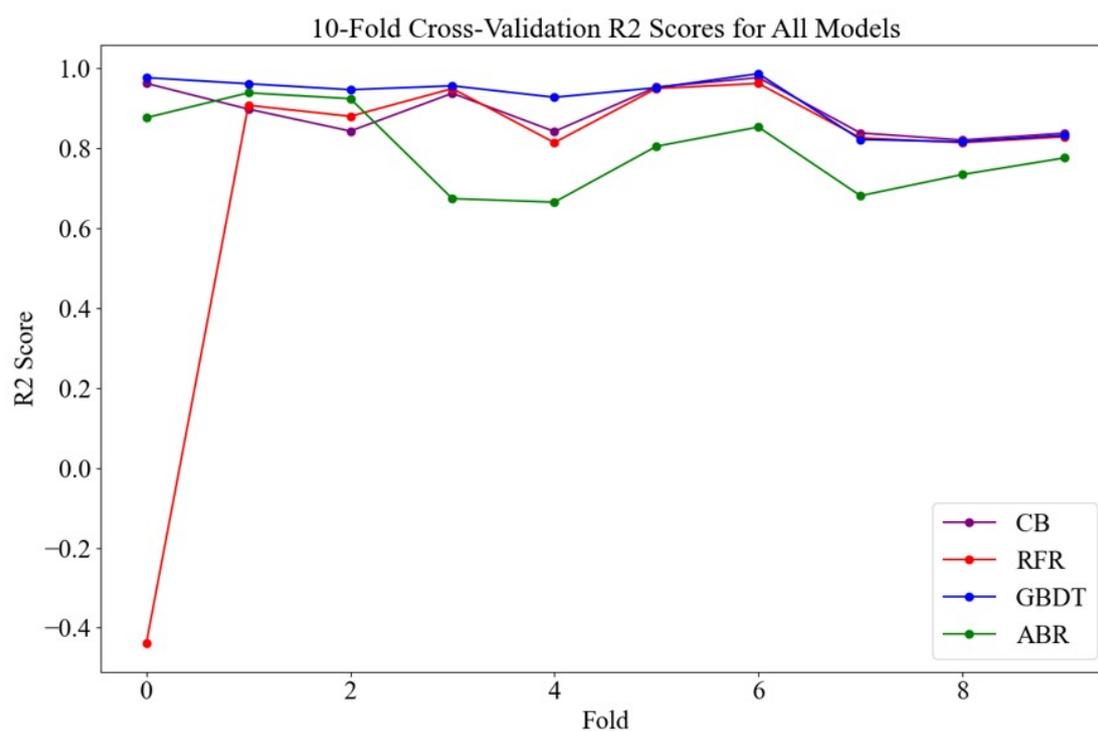
**Figure S5.** Comparative Cross-Validation Performance Assessment of Regression Models.

Figure S6, titled "10-Fold Cross-Validation R<sup>2</sup> Scores for All Models," compares the R<sup>2</sup> scores of four machine learning models—CatBoost (CB), Random Forest (RFR), Gradient Boosting (GBDT), and AdaBoost (ABR)—across 10 folds of cross-validation. Higher R<sup>2</sup> scores on the Y-axis indicate better performance in explaining the variance in the data.

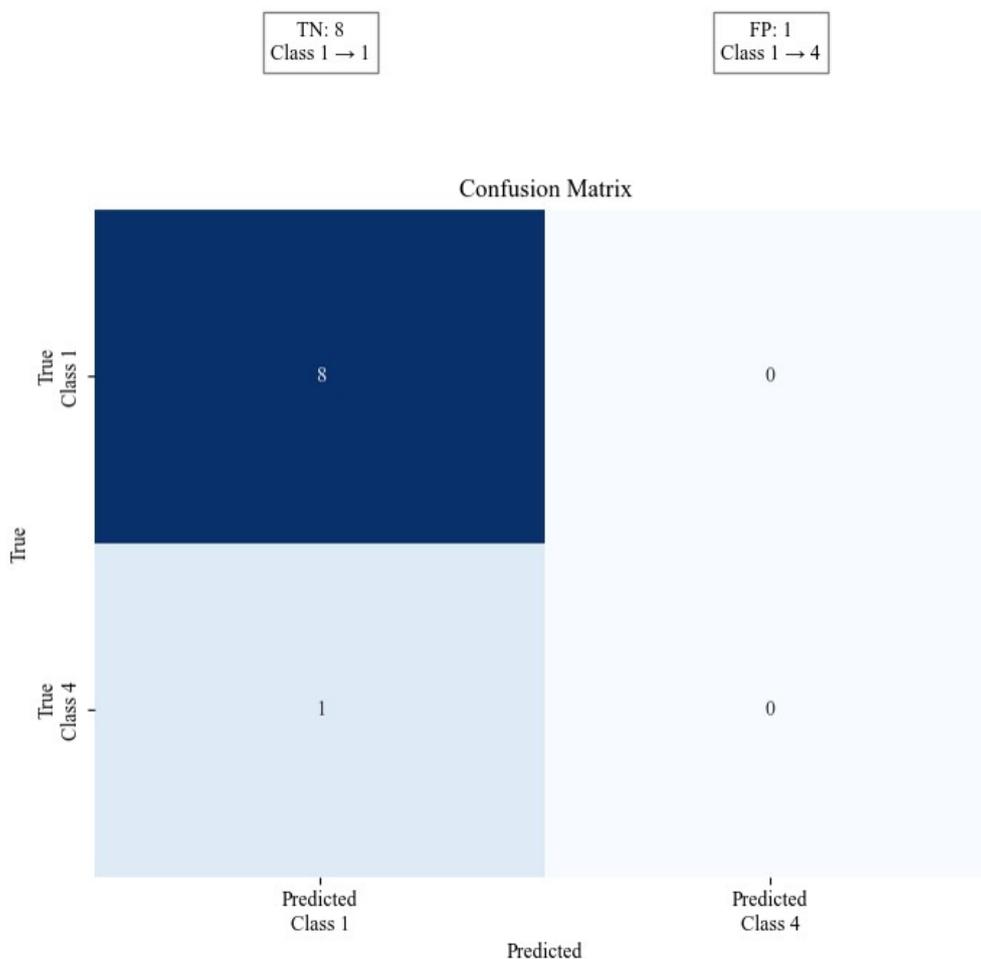
Each row in Table S9 corresponds to a different machine learning model, and the settings provided in the second column are the hyperparameters used to configure each model in the script. These hyperparameters are the key settings that define the behavior and performance of the models during training.

**Table S9:** Machine Learning Model Hyperparameters for bandgap prediction.

Models	Hyperparameters
<b>ABR</b>	- n_estimators: 100,- learning_rate: 0.5, - random_state: 42
<b>GBDT</b>	- n_estimators: 100, - learning_rate: 0.1,- max_depth: 3, - random_state: 42
<b>RFR</b>	- n_estimators: 100, - max_depth: 5, - random_state: 42
<b>CBR</b>	- iterations: 100, - learning_rate: 0.1, - depth: 6,- random_seed: 42, - verbose: 0



**Figure S6:** Cross-Validation  $R^2$  Scores Overview.



**Figure S7.** Confusion Matrix for Stability Classification of PSC.

Figure S7 presents a confusion matrix evaluating the model's performance in classifying perovskite solar cell stability. It shows perfect accuracy for tested samples, with no misclassifications between Class 1 (unstable) and Class 4 (very stable). In independent validation, the model achieved 88.9% accuracy, correctly identifying 8 out of 9 unstable samples, with only one misclassified as very stable. This highlights the model's reliability in predicting unstable PSCs.