

Bayesian reverse design of high-efficiency perovskite solar cells based on experimental knowledge constraints

Cite as: Appl. Phys. Lett. **125**, 063901 (2024); doi: [10.1063/5.0216447](https://doi.org/10.1063/5.0216447)

Submitted: 29 April 2024 · Accepted: 22 July 2024 ·

Published Online: 5 August 2024



Hongyu Liu,¹ Zhengxin Chen,¹ Yaping Zhang,¹ Jiang Wu,² Lin Peng,¹ Yanan Wang,¹ Xiaolin Liu,^{1,a)} Xianfeng Chen,^{3,4,a)} and Jia Lin^{1,a)}

AFFILIATIONS

¹College of Mathematics and Physics, Shanghai University of Electric Power, Shanghai 200090, China

²College of Energy and Mechanical Engineering, Shanghai University of Electric Power, Shanghai 200090, China

³State Key Laboratory of Advanced Optical Communication Systems and Networks, School of Physics and Astronomy, Shanghai Jiao Tong University, Shanghai 200240, China

⁴Collaborative Innovation Center of Light Manipulation and Applications, Shandong Normal University, Jinan 250358, China

^{a)}Authors to whom correspondence should be addressed: xliu@shiep.edu.cn; xfchen@sjtu.edu.cn; and jlin@shiep.edu.cn

ABSTRACT

To alleviate high costs and lengthy trial-and-error periods associated with traditional optimization methods for perovskite solar cells (PSCs), we developed a data-driven reverse design framework for high-efficiency PSCs. This framework integrates machine learning and Bayesian optimization (BO) to accelerate the optimization process of PSCs by intelligently recommending the most promising parameter configurations for PSCs, such as device structure and fabrication processes. To improve the robustness of the framework, we first designed a two-stage sampling strategy to alleviate the issue of imbalanced dataset classes. Subsequently, by integrating “experimental knowledge constraints” into the BO process, we achieved precise parameter configurations, thus avoiding discrepancies between predicted and actual results due to parameter mismatches. Finally, using SHapley Additive exPlanations, we unveiled key factors influencing the power conversion efficiency (PCE), such as the composition of perovskite solvents. Our framework not only precisely predicted the PCE of PSCs with an area under the curve of 0.861 but also identified the optimal parameter configurations, achieving a high probability of 0.981. This framework offers substantial support for minimizing redundant experiments and characterizations, effectively accelerating the optimization process of PSCs.

Published under an exclusive license by AIP Publishing. <https://doi.org/10.1063/5.0216447>

During the development of perovskite solar cells (PSCs), their excellent power conversion efficiency (PCE), low-cost manufacturing process, and sensitive response to a broad spectrum have collectively established their leading position in the future development of the photovoltaic field.^{1,2} Over the past decade, the PCE of single-junction PSCs has surged from 3.8% to 26.1%^{3–5} through continuous optimization in device structure and fabrication processes.^{6–18} As research progressively deepens, an undeniable fact gradually emerges: the PCE of PSCs is highly sensitive to the configuration and fabrication process parameters of each functional layer. Whether it is the strategic choice of charge transport layer (CTL) materials or the precise input of additives, every step of adjustment is directly related to the final PCE. Therefore, the rational design of device structure and fabrication processes is crucial for achieving high-efficiency PSCs.

Faced with the increasing number of potential structural configurations and fabrication process parameter combinations, as well as the

complex interactions between device performance and parameters, finding the optimal configuration for each layer through experiments to effectively optimize PSC often means a long trial-and-error process and substantial consumable costs. This method not only heavily relies on the extensive experience and intuitive judgment of researchers but also struggles to systematically cover enough parameter combinations and their corresponding impacts on performance due to the complexity and variability of experimental conditions. An efficient, accurate, and easy-to-operate design method for high-efficiency PSCs is urgently needed. Fortunately, a substantial amount of experimental data have been generated during the research process, particularly for MAPbI₃ devices, providing an opportunity to use artificial intelligence (AI) technology to promote the design of high-efficiency PSCs through a data-driven approach.

In recent years, machine learning (ML), an AI technology with powerful pattern recognition capabilities, has been primarily applied in

the form of forward prediction in the field of PSCs, where the input is the parameters of PSCs and the output is the performance of PSCs.^{19–27} ML can learn hidden patterns between various influencing factors and PSC performance from vast data, enabling precise predictions. For example, Li *et al.*,²⁵ Gok *et al.*,²⁶ and Lu *et al.*²⁷ predicted the performance of PSCs based on different input parameters such as perovskite bandgap, perovskite composition, and fabrication process. However, in the design of PSCs, what is more crucial is designing the appropriate device structure and fabrication processes based on specific requirements, such as high PCE, which embodies a concept of reverse design. It is worth noting that Bayesian optimization (BO), an efficient global optimization technique (as shown in Fig. S1), is a crucial technology to achieve performance-oriented reverse design process, and has made significant progress in accelerating material performance improvement and experimental process optimization.^{28–30} For example, MacLeod *et al.* utilized BO to enhance the optoelectronic performance of thin films by optimizing the cobalt doping ratio and annealing time of the spiro-OMeTAD.²⁸ The powerful ability of BO in handling complex parameter spaces and achieving precise objective optimization is very suitable for the reverse design of PSCs. Nevertheless, resolving the issue of parameter mismatch is an imminent challenge within this process.

Different from the research focused on improving specific materials, our work shifts the optimization focus to the device level, including all functional layers. While this strategy broadens the optimization horizon, it inevitably introduces a series of hard constraints, such as the compatibility between materials and additives, resulting in a high interdependency among certain parameters. BO typically does not automatically consider these hard constraints, which may lead to parameter mismatch issues in optimization results, thereby affecting the consistency between predicted results and actual applications. This explains why the application of BO to optimize systems with highly interdependent parameters is more intricate. To address the issue, we

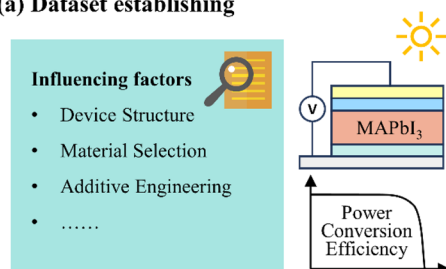
introduce “experimental knowledge constraints.” This approach is based on existing experimental parameter configurations, constructing multiple dictionaries containing relevant feature combinations to guide and constrain the BO process. Through this method, we can achieve precise and reasonable parameter configurations, thereby significantly enhancing the accuracy and practicality of reverse design.

In this study, we combine the powerful pattern recognition capability of ML with the efficient global optimum identification ability of BO to construct a reverse design framework for designing high-efficiency PSCs. Initially, we established the MAPbI₃ device dataset based on an experimental database and designed a two-stage sampling strategy to more effectively alleviate class imbalance issues in the dataset. Subsequently, the voting model achieved accurate predictions of device PCE and assessed the potential for the devices to become high-efficiency PSCs. In addition, by employing BO combined with “experimental knowledge constraints,” we intelligently recommend the device parameter configuration with the most potential to achieve high-efficiency PSCs. Finally, utilizing SHAP (SHapley Additive exPlanations), we revealed the key factors affecting the PCE.

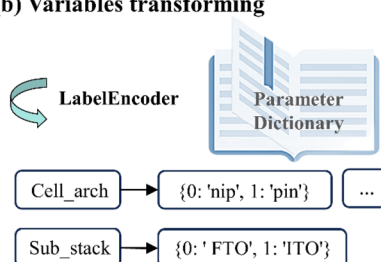
Figure 1 shows the workflow diagram of our framework. The MAPbI₃ dataset we constructed, containing 3526 device data points, is derived from the open database of PSCs created by Jacobsson *et al.*, which compiles peer-reviewed experimental data up until February 2020.³¹ As shown in Table S1, we used the device structure and fabrication process parameters in the data as the input features of the ML model.

Based on recent literature and considering the proportion of PSCs with high PCE in the dataset [as shown in Fig. 2(a)], we classify the devices into two categories, high efficiency (positive category) and low efficiency (negative category), based on whether the PCE value reaches 17%, and use them as labels for the model output.² Figures 2(b)–2(d) show the distribution of device architecture (p-i-n or n-i-p) and CTL materials in the MAPbI₃ dataset. It is worth noting that in Figs. 2(c) and 2(d), we provide lists of electron transport layer (ETL)

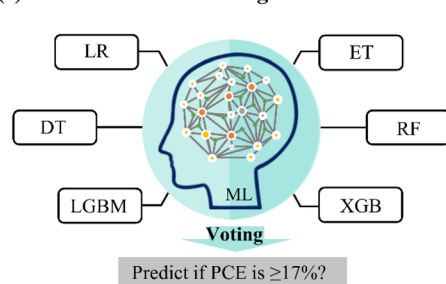
(a) Dataset establishing



(b) Variables transforming



(c) ML model constructing



(d) Optimal parameters finding

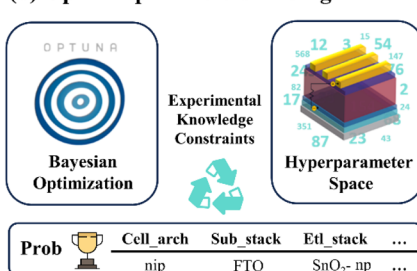


FIG. 1. The reverse design framework for PSCs, where the procedure includes (a) dataset establishment, (b) variable transformation, (c) ML model construction, and (d) optimal parameters identification.

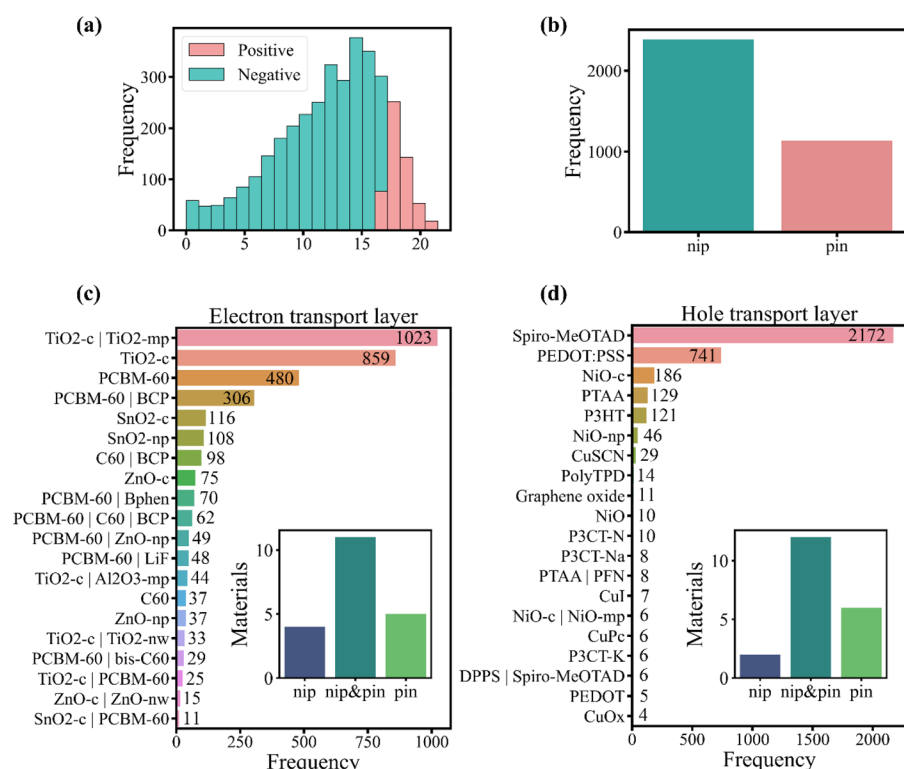


FIG. 2. The distribution of PCE, device architecture, and CTL materials in each functional layer in the MAPbI₃ dataset: (a) PCE of the device ("Jv_PCE"), (b) device architecture ("Cell_arch"), (c) electron transport layer ("Etl_stack"), and (d) hole transport layer ("Htl_stack"). In the subplots for (c) and (d), the counts of materials used in the two device architectures are tabulated: "nip" represents exclusive use in nip devices, "nip+pin" indicates usage in both structures, and "pin" signifies exclusive use in inverted devices.

and hole transport layer (HTL) materials by frequency, revealing the differences in ETL and HTL materials used in different device architectures. Additionally, due to the significant difference in thickness between carbon electrodes and metal electrodes, this dataset only includes devices employing metal electrodes.

Given the MAPbI₃ dataset encompasses a large number of textual features related to materials and fabrication processes, we employed the LabelEncoder method to convert these textual features into numerical labels, thereby transforming device information into ML recognizable low-dimensional matrices. This approach not only avoids the challenges and limitations associated with obtaining descriptors of material physical properties and their lack of universality but also avoids the dimensionality explosion caused by using one-hot encoding.

During the ML model construction process, we employed six algorithms to construct classification models. These included logistic regression (LR), decision tree (DT), and four tree-based ensemble learning models, namely random forest (RF), Light Gradient-Boosting Machine (LGBM), Extremely Randomized Trees (ET), and eXtreme Gradient Boosting (XGB).^{32–36} Furthermore, three classification evaluation indicators were selected to comprehensively assess the performance of the models: accuracy, F1 score, and area under the curve (AUC).³⁷ Additionally, we employed the tenfold cross-validation results as model scores to reduce performance fluctuations caused by data segmentation and to enhance the models' generalization capability.³⁸ Figure 3(a) compares the performance of six models on the dataset, and we choose the more robust ensemble models RF, LGBM, ET, and XGB for subsequent work.

Observing Fig. 2(a), we discovered a severe class imbalance issue in the dataset, manifested as a ratio of positive to negative classes of

1:6. This imbalance led models to tend to predict the majority classes (negative samples), resulting in high accuracy but low F1 scores, as shown in Fig. 3(a). We devised a two-stage sampling strategy to mitigate class imbalance, which first involves RandomUnderSampler (RUS) sampling followed by Synthetic Minority Over-sampling Technique (SMOTE). As illustrated in Fig. 3(b) and Fig. S2, we utilized BO to determine the optimal undersampling rate $\alpha_{us} = 0.38$ and over-sampling rate $\alpha_{os} = 0.93$, resulting in a peak F1 score of 0.633, although at the cost of a slight accuracy reduction of about 0.07. Figure 3(c) displays the distribution of classes before and after the two-stage sampling. The use of this method not only effectively alleviates the class imbalance issue within the dataset but also reduces the potential for information loss or noise introduced by employing a single sampling method (the comparison of sampling strategies can be found in Table S2).

To ensure outstanding model performance while minimizing computational resource consumption, as shown in Fig. 3(d) and Fig. S3, we selected 16 key features to construct the optimal feature subset (refer to Table S3) for subsequent modeling processes. Additionally, to further improve the robustness of the model, we used BO to fine-tune the hyperparameters of four models, as shown in Table S4. In the process of identification of optimal parameters, we focused more on the model's ability to rank output probability scores rather than solely on classification accuracy. Voting algorithm is an ensemble learning algorithm that follows the "majority rules" principle.³⁹ This method combines predictions from multiple models according to specific rules to reach a final collective decision, thereby enhancing the overall robustness of the model. For this purpose, we constructed an additional voting model based on high-performance RF and XGB models (with

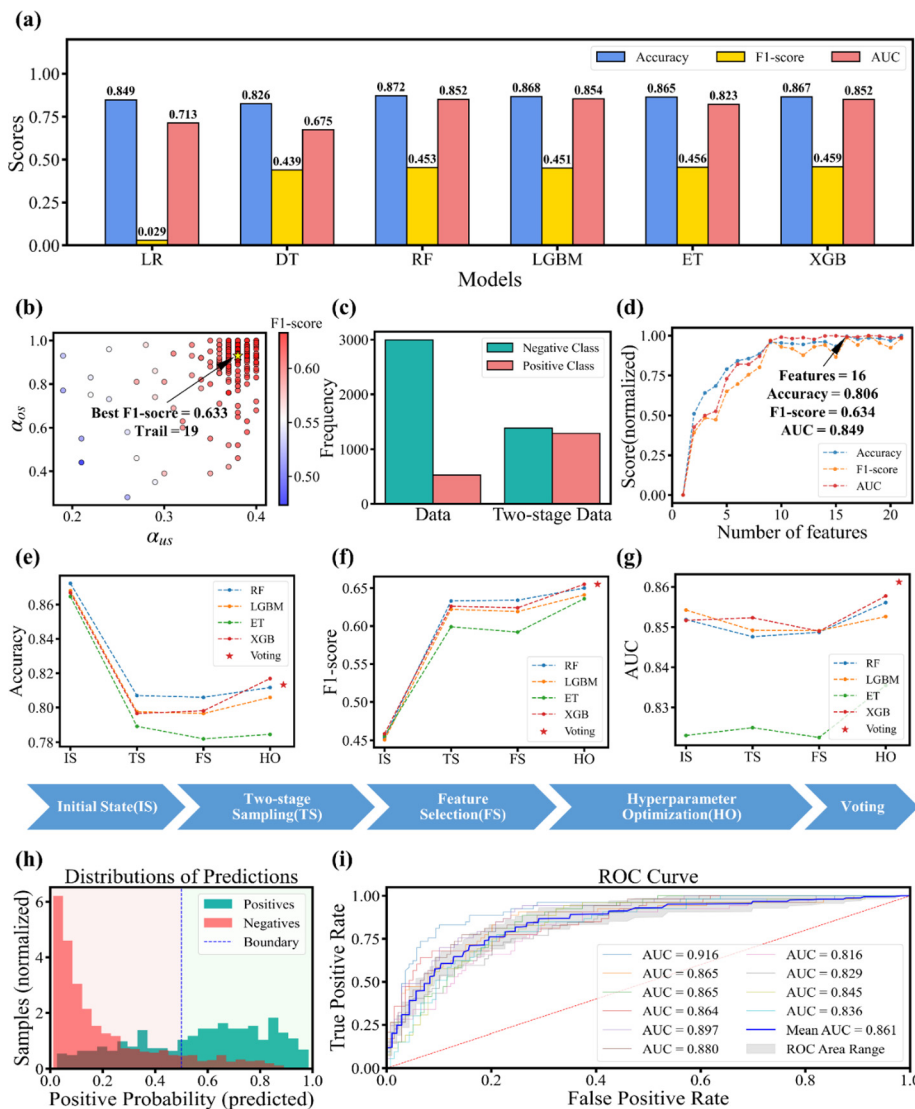


FIG. 3. The performance visualization during ML model modeling process. (a) The performance comparison of six classification models on the dataset. (b) The F1 score chart of the two-stage sampling strategy under various sampling ratios. (c) Comparison of sample distribution before and after implementing the two-stage sampling strategy. (d) Influence of the number of features on the performance of the RF model, with scores normalized. Select corresponding features from high to low based on the importance determined by the model. The evaluation metric scores of models in the following stages: initial state, two-stage sampling, feature selection, hyperparameter optimization, and voting. (e) Accuracy, (f) F1 score, and (g) AUC. (h) Visualization of the voting model output probability distribution. The legend colors reflect the true class of the samples. (i) The ROC curve of the voting model.

weights set to 0.47 and 0.53, respectively), achieving a peak AUC of 0.861. Figures 3(e)–3(g) delineate the influence of each stage strategy throughout the modeling process on the model's performance, confirming the effectiveness of our optimization efforts. The performance of the optimized models is shown in Table S5.

In order to more intuitively present the excellent performance of the voting model, we visualized it and presented the results in Fig. 3. The distribution of model prediction probabilities and its confusion matrix are clearly presented in Fig. 3(h) and Fig. S4, respectively. The proportion of false positive (FP) samples is minimal and significantly decreases as the positive probability increases. Notably, all samples with a probability exceeding 0.927 were accurately classified, indicating a high level of reliability for samples predicted as positive class with probabilities exceeding this threshold. The model exhibits excellent performance in each round of tenfold cross-validation, with a maximum AUC value of 0.916 and an average value of 0.861, as shown in Fig. 3(i).

In the process of optimal parameter identification, the Optuna, a hyperparameter optimization framework, was used to perform the BO process.⁴⁰ During this process, we limit the parameter search space to a range defined by the features in the optimal feature subset, as shown in Fig. 4(a), ensuring that the optimization process focuses on the factors most critical to PCE. We selected the indicator of the device's potential for high-efficiency (the probability of the device being classified as positive by the voting model) as the objective function. This intended to continuously update the probability surrogate model through iteration to fit the objective function in order to identify the most promising combination of device structures and fabrication process parameters.

The proper configuration of parameters is crucial for achieving high PCE, yet BO often overlooks this aspect. For example, the nip and pin architectures exhibit differences in charge transfer and collection mechanisms, directly impacting the selection of materials for the CTL [as illustrated in Figs. 2(c) and 2(d)]. Failure to consider and

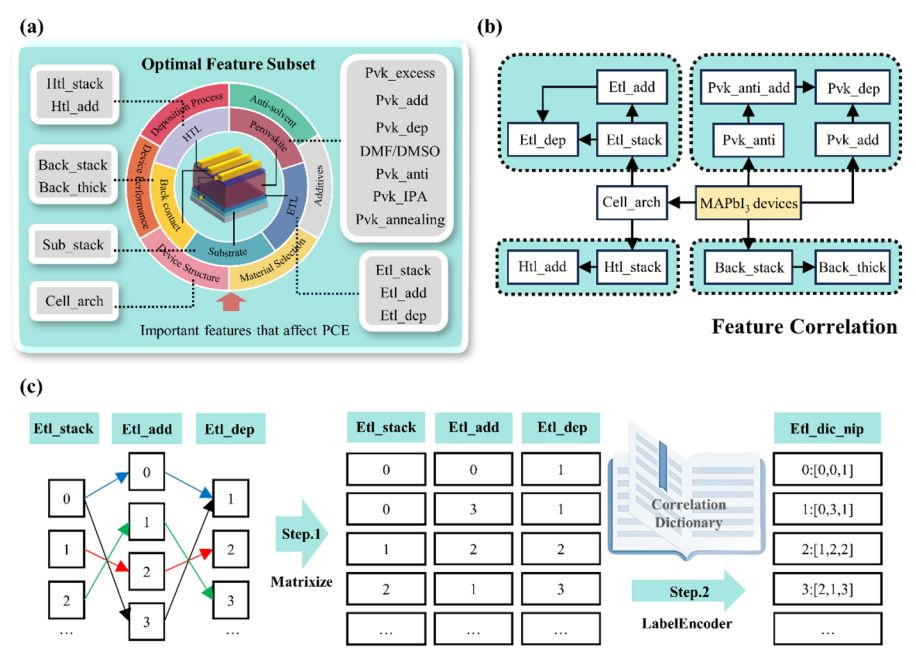


FIG. 4. The process of establishing “experimental knowledge constraints.” (a) The optimal subset of features. (b) The correlation between features. (c) The process of establishing “Etl_dic_nip” feature association dictionary.

constrain these differences may lead to mismatches between the device architecture and CTL materials, resulting in significant discrepancies between predicted and actual results. Similar situations include the matching of materials and additives. Faced with parameter configuration issues, we innovatively introduced “experimental knowledge constraints” into the BO process. This strategy utilizes prior knowledge from specific experimental configurations to avoid undesirable parameter settings, ensuring that the reverse design results align with physical principles and experimental facts while efficiently approaching target performance metrics. Figures 4(b) and 4(c) show the introduction process of “experimental knowledge constraints.” First, it establishes the correlation between features based on experimental knowledge and the functional layer to which they belong. Then, it utilizes matrix transformation and the LabelEncoder method to convert all parameter configurations of relevant features into multiple feature association dictionaries. The parameter spaces of these dictionaries were used to

replace the original parameter spaces of the related features, thereby imposing constraints on parameter selection during the BO process. The adoption of this strategy enables precise parameter configuration during the BO process, significantly enhancing the accuracy and practicality of reverse design. For detailed information, please refer to Tables S6 and S7, and Fig. S5.

Figure 5(a) displays the optimal parameter configuration results for the MAPbI₃ device, derived through BO. The device structure is FTO|SnO₂-np|MAPbI₃|PTAA|AgAl, with a remarkably high positive probability of 0.981. To gain a deeper understanding of the positive impact of these parameters on the device PCE, we reviewed the relevant literature. Previous studies have indicated that Zr-doping can enhance the electron transport properties and adjust the energy-level structure of the SnO₂-np layer, thereby improving the performance of PSCs.⁴¹ Excessive introduction of PbI₂ and PbBr₂ into the precursor of MAPbI₃ perovskite can improve the surface quality of the film, making

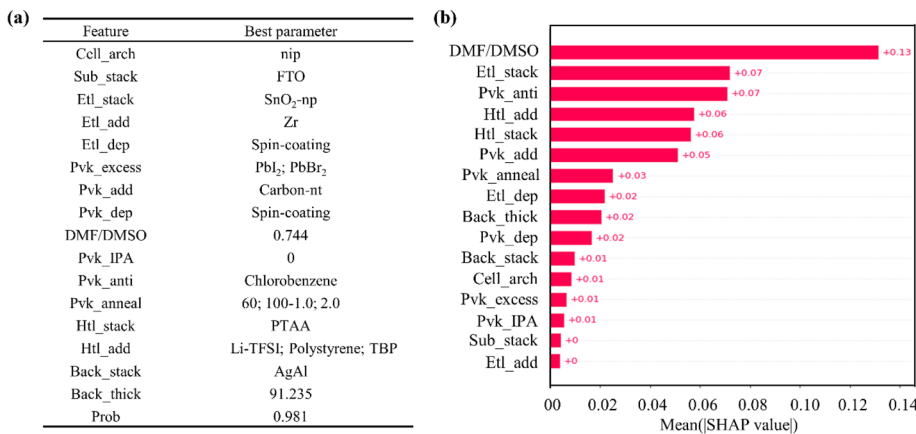


FIG. 5. The results of reverse design and model analysis. (a) The optimal parameter configuration for MAPbI₃ devices. (b) The feature importance of the voting model based on SHAP values.

it smoother with fewer defects, thus improving the device performance.⁴² Additionally, carbon nanotubes (Carbon-nt) have been employed to improve the charge extraction capability of PSCs.^{43,44} There is research suggesting that the optimal DMF:DMSO ratio for nip devices is 0.7:0.3, which is a mere 0.044 variance from our suggested ratio.⁴⁵ This minor discrepancy is primarily attributed to differences in other parameter settings. Furthermore, the use of an appropriate amount of chlorobenzene as anti-solvent can improve the morphology and crystallinity of the MAPbI₃ film, promoting the accumulation of photogenerated carriers at the interface, thereby accelerating the electron transport rate and enhancing the PCE.⁴⁶ Conventional two-step annealing processes are also widely employed to improve film quality. Although PTAA is a commonly used HTL material in the pin device, it is also applied in the nip device. Doping PTAA with Li-TFSI, polystyrene, and TBP can improve the electrical properties of PTAA, thereby enhancing the PCE of PSCs.⁴⁷ Finally, AgAl-based electrode demonstrates higher PCE and stability compared to single Ag or Al electrodes.⁴⁸ In summary, through the coordinated optimization and precise control of the parameters for the device structure and fabrication processes, the potential for this device to become a high-efficiency PSC has been significantly enhanced.

SHAP is a sophisticated tool for model interpretability, which helps researchers understand the decision-making process of models and reveal the complex mechanisms between device parameters and their performance.⁴⁹ To determine the crucial factors affecting the PCEs of PSCs, we employed SHAP values to quantify the contribution of each feature to model predictions. In this section, we focus on analyzing the top three features that have the greatest impact on PCE. As shown in Fig. 5(b), the ratio of DMF to DMSO in perovskite solvents has a significant impact on the performance of PSCs (with a SHAP value of 0.13), consistent with recent research.^{45,50,51} The mixture of DMF and DMSO solvents, with their stronger ability to form intermediate phases, can promote the formation of larger grain perovskite films, thereby enhancing light absorption and charge transfer efficiency. However, an inappropriate ratio of DMF to DMSO can severely affect the crystallization rate and PCEs. Additionally, the ETL is located between the perovskite layer and the electrode, responsible for the effective transfer of electrons and preventing the migration of holes toward the cathode. The matching energy levels between the ETL and the perovskite layer facilitate the extraction of electrons, while mismatched energy levels can create an interfacial barrier, reducing the efficiency of electron transport and the overall PCEs. Therefore, the selection of ETL materials is crucial for enhancing the PCEs of PSCs.^{52,53} Furthermore, anti-solvent treatment significantly improves film quality and interface characteristics by promoting crystallization of the perovskite film, thereby enhancing the PCEs of PSCs.^{54,55} Improper selection of anti-solvent can lead to a decrease in film quality and even the formation of cracks, significantly impacting device performance. Additionally, the selection of materials and additives for the HTL can also notably impact the final PCEs of PSCs.^{47,56}

In this work, we have developed a data-driven framework for reverse designing PSCs, aiming to overcome the inefficiency and high cost issues associated with traditional experimental optimization methods. By applying the two-stage sampling strategy and the “experimental knowledge constraint” method, our framework has not only accurately predicted the PCE of PSCs with an AUC of 0.861 but also identified the optimal parameter configuration, achieving a

high probability of 0.981. The corresponding device structure is FTO/SnO₂-np/MAPbI₃/PTAA/AgAl. Additionally, we have identified key factors that significantly impact PCE, such as the composition of the perovskite solvents. Although our work is primarily focused on single-junction PSCs, the method we propose is also applicable to the design of other photovoltaic devices with similar structures, such as perovskite/silicon tandem solar cells (TSCs).^{57–59} This approach allows for the various components of TSCs (such as nanotextures) to be considered as a unified system for the synergistic optimization of device configuration and manufacturing processes, rather than being optimized in isolation. The framework provides substantial support for reducing redundant experiments, thereby effectively accelerating the optimization process of photovoltaic devices.

See the [supplementary material](#) for modeling details, including a comparison of effectiveness of sampling strategies, ranking of 16 features and their importance in feature selection, the impact of feature selection on the performance of four models, parameter space and optimal hyperparameters for model hyperparameter optimization, detailed establishment process of the feature correlation dictionary and the included feature information, and the parameter space for MAPbI₃ devices in the process of finding optimal parameters.

This work was partially sponsored by the National Natural Science Foundation of China (Grant Nos. 12174246 and 52076126), the National Key Research and Development Program (Grant No. 2021YFE0109800), the Key Laboratory of Clean Power Generation and Environmental Protection Technology in Mechanical Industry, Science and Technology Commission of Shanghai Municipality (Grant No. 21010501300), and the Shuguang Program of Shanghai Education Development Foundation and Shanghai Municipal Education Commission (Grant No. 22SG51).

AUTHOR DECLARATIONS

Conflict of Interest

The authors have no conflicts to disclose.

Author Contributions

Hongyu Liu and Zhengxin Chen contributed equally to this work.

Hongyu Liu: Data curation (equal); Formal analysis (equal); Validation (equal); Writing – original draft (equal). **Zhengxin Chen:** Formal analysis (equal); Investigation (equal); Methodology (equal); Validation (equal). **Yaping Zhang:** Investigation (equal); Methodology (equal). **Jiang Wu:** Resources (equal); Validation (equal). **Lin Peng:** Formal analysis (equal); Methodology (equal). **Yanan Wang:** Data curation (equal); Writing – review & editing (equal). **Xiaolin Liu:** Formal analysis (equal); Validation (equal); Visualization (equal). **Xianfeng Chen:** Funding acquisition (equal); Resources (equal). **Jia Lin:** Conceptualization (equal); Supervision (equal); Writing – review & editing (equal).

DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding authors upon reasonable request.

REFERENCES

- ¹N.-G. Park, "Perovskite solar cells: An emerging photovoltaic technology," *Mater. Today* **18**(2), 65 (2015).
- ²J. Y. Kim, J. W. Lee, H. S. Jung, H. Shin, and N. G. Park, "High-efficiency perovskite solar cells," *Chem. Rev.* **120**(15), 7867 (2020).
- ³A. Kojima, K. Teshima, Y. Shirai, and N. G. Park, "Organometal halide perovskites as visible-light sensitizers for photovoltaic cells," *J. Am. Chem. Soc.* **131**(17), 6050 (2009).
- ⁴H. Chen, C. Liu, J. Xu, A. Maxwell, W. Zhou, Y. Yang, Q. Zhou, A. S. R. Bati, H. Wan, Z. Wang *et al.*, "Improved charge extraction in inverted perovskite solar cells with dual-site-binding ligands," *Science* **384**(6692), 189 (2024).
- ⁵M. A. Green, E. D. Dunlop, M. Yoshita, N. Kopidakis, K. Bothe, G. Siefer, and X. Hao, "Solar cell efficiency tables (Version 63)," *Prog. Photovoltaics* **32**(1), 3 (2023).
- ⁶J. W. Lee, D. J. Seol, A. N. Cho, and N. G. Park, "High-efficiency perovskite solar cells based on the black polymorph of $\text{HC}(\text{NH}_2)_2\text{PbI}_3$," *Adv. Mater.* **26**(29), 4991 (2014).
- ⁷S. Liu, W. Huang, P. Liao, N. Pootrakulchote, H. Li, J. Lu, J. Li, F. Huang, X. Shai, X. Zhao *et al.*, "Correction: 17% efficient printable mesoscopic PIN metal oxide framework perovskite solar cells using cesium-containing triple cation perovskite," *J. Mater. Chem. A* **6**(9), 4220 (2018).
- ⁸P. W. Liang, C. Y. Liao, C. C. Chueh, F. Zuo, S. T. Williams, X. K. Xin, J. Lin, and A. K. Jen, "Additive enhanced crystallization of solution-processed perovskite for highly efficient planar-heterojunction solar cells," *Adv. Mater.* **26**(22), 3748 (2014).
- ⁹H. Min, M. Kim, S. U. Lee, H. Kim, G. Kim, K. Choi, J. H. Lee, and S. I. Seok, "Efficient, stable solar cells by using inherent bandgap of alpha-phase formamidinium lead iodide," *Science* **366**(6466), 749 (2019).
- ¹⁰Y.-J. Kang, S.-N. Kwon, S.-P. Cho, Y.-H. Seo, M.-J. Choi, S.-S. Kim, and S.-I. Na, "Antisolvent additive engineering containing dual-function additive for triple-cation p-i-n perovskite solar cells with over 20% PCE," *ACS Energy Lett.* **5**(8), 2535 (2020).
- ¹¹M. Hou, H. Zhang, Z. Wang, Y. Xia, Y. Chen, and W. Huang, "Enhancing efficiency and stability of perovskite solar cells via a self-assembled dopamine interfacial layer," *ACS Appl. Mater. Interfaces* **10**(36), 30607 (2018).
- ¹²W. Kong, W. Li, C. Liu, H. Liu, J. Miao, W. Wang, S. Chen, M. Hu, D. Li, A. Amini *et al.*, "Organic monomolecular layers enable energy-level matching for efficient hole transporting layer free inverted perovskite solar cells," *ACS Nano* **13**(2), 1625 (2019).
- ¹³C. Wang, Y. Zhao, T. Ma, Y. An, R. He, J. Zhu, C. Chen, S. Ren, F. Fu, D. Zhao *et al.*, "A universal close-space annealing strategy towards high-quality perovskite absorbers enabling efficient all-perovskite tandem solar cells," *Nat. Energy* **7**(8), 744 (2022).
- ¹⁴L. Huang, Y. H. Lou, and Z. K. Wang, "Buried interface passivation: A key strategy to breakthrough the efficiency of perovskite photovoltaics," *Small* **19**(38), e2302585 (2023).
- ¹⁵B. A. Al-Asbahi, S. M. H. Qaid, M. Hezam, I. Bedja, H. M. Ghaithan, and A. S. Aldwayyan, "Effect of deposition method on the structural and optical properties of $\text{CH}_3\text{NH}_3\text{PbI}_3$ perovskite thin films," *Opt. Mater.* **103**, 109836 (2020).
- ¹⁶B. Li, S. Li, J. Gong, X. Wu, Z. Li, D. Gao, D. Zhao, C. Zhang, Y. Wang, and Z. Zhu, "Fundamental understanding of stability for halide perovskite photovoltaics: The importance of interfaces," *Chem* **10**(1), 35 (2024).
- ¹⁷Z. Huang, Y. Bai, X. Huang, J. Li, Y. Wu, Y. Chen, K. Li, X. Niu, N. Li, G. Liu *et al.*, "Anion- π interactions suppress phase impurities in FAPbI_3 solar cells," *Nature* **623**(7987), 531 (2023).
- ¹⁸Z. Liang, Y. Zhang, H. Xu, W. Chen, B. Liu, J. Zhang, H. Zhang, Z. Wang, D. H. Kang, J. Zeng *et al.*, "Homogenizing out-of-plane cation composition in perovskite solar cells," *Nature* **624**(7992), 557 (2023).
- ¹⁹Z. Chen, H. Wang, B. Liu, H. Zhou, Y. Cao, Y. Wang, L. Peng, X. Liu, J. Lin, X. Chen *et al.*, "Cationic perturbation strategy to solve the information drought in material explainable machine learning," *Phys. Rev. B* **109**(8), 085306 (2024).
- ²⁰S. Lu, Q. Zhou, Y. Ouyang, Y. Guo, Q. Li, and J. Wang, "Accelerated discovery of stable lead-free hybrid organic-inorganic perovskites via machine learning," *Nat. Commun.* **9**(1), 3405 (2018).
- ²¹C. Zhi, S. Wang, S. Sun, C. Li, Z. Li, Z. Wan, H. Wang, Z. Li, and Z. Liu, "Machine-learning-assisted screening of interface passivation materials for perovskite solar cells," *ACS Energy Lett.* **8**(3), 1424 (2023).
- ²²N. T. P. Hartono, J. Thapa, A. Tiihonen, F. Oviedo, C. Batali, J. J. Yoo, Z. Liu, R. Li, D. F. Marron, M. G. Bawendi *et al.*, "How machine learning can help select capping layers to suppress perovskite degradation," *Nat. Commun.* **11**(1), 4172 (2020).
- ²³Y. Yu, X. Tan, S. Ning, and Y. Wu, "Machine learning for understanding compatibility of organic-inorganic hybrid perovskites with post-treatment amines," *ACS Energy Lett.* **4**(2), 397 (2019).
- ²⁴J. Hu, Z. Chen, Y. Chen, H. Liu, W. Li, Y. Wang, L. Peng, X. Liu, J. Lin, X. Chen *et al.*, "Interpretable machine learning predictions for efficient perovskite solar cell development," *Sol. Energy Mater. Sol. Cells* **271**, 112826 (2024).
- ²⁵J. Li, B. Pradhan, S. Gaur, and J. Thomas, "Predictions and strategies learned from machine learning to develop high-performing perovskite solar cells," *Adv. Energy Mater.* **9**(46), 1901891 (2019).
- ²⁶E. C. Gok, M. O. Yildirim, M. P. U. Haris, E. Eren, M. Pegu, N. H. Hemasiri, P. Huang, S. Kazim, A. U. Oksuz, and S. Ahmad, "Predicting perovskite bandgap and solar cell performance with machine learning," *Sol. RRL* **6**(2), 2100927 (2022).
- ²⁷Y. Lu, D. Wei, W. Liu, J. Meng, X. Huo, Y. Zhang, Z. Liang, B. Qiao, S. Zhao, D. Song *et al.*, "Predicting the device performance of the perovskite solar cells from the experimental parameters through machine learning of existing experimental results," *J. Energy Chem.* **77**, 200 (2023).
- ²⁸B. P. MacLeod, F. G. L. Parlane, T. D. Morrissey, F. Hase, L. M. Roch, K. E. Dettelbach, R. Moreira, L. P. E. Yunker, M. B. Rooney, J. R. Deeth *et al.*, "Self-driving laboratory for accelerated discovery of thin-film materials," *Sci. Adv.* **6**(20), eaaz8867 (2020).
- ²⁹Z. Liu, N. Rolston, A. C. Flick, T. W. Colburn, Z. Ren, R. H. Dauskardt, and T. Buonassisi, "Machine learning with knowledge constraints for process optimization of open-air perovskite solar cell manufacturing," *Joule* **6**(4), 834 (2022).
- ³⁰W. Li, J. Hu, Z. Chen, H. Jiang, J. Wu, X. Meng, X. Fang, J. Lin, X. Ma, T. Yang *et al.*, "Performance prediction and optimization of perovskite solar cells based on the Bayesian approach," *Sol. Energy* **262**, 111853 (2023).
- ³¹T. J. Jacobsson, A. Hultqvist, A. García-Fernández, A. Anand, A. Al-Ashouri, A. Hagfeldt, A. Crovetto, A. Abate, A. G. Ricciardulli, A. Vijayan *et al.*, "An open-access database and analysis tool for perovskite solar cells based on the FAIR data principles," *Nat. Energy* **7**(1), 107 (2021).
- ³²D. W. Hosmer, Jr., S. Lemeshow, and R. X. Sturdivant, *Applied Logistic Regression* (John Wiley & Sons, 2013).
- ³³W. Loh, "Fifty years of classification and regression trees," *Int. Stat. Rev.* **82**(3), 329 (2014).
- ³⁴G. Biau and E. Scornet, "A random forest guided tour," *Test* **25**(2), 197 (2016).
- ³⁵C. Bentéjac, A. Csörgő, and G. Martínez-Muñoz, "A comparative analysis of gradient boosting algorithms," *Artif. Intell. Rev.* **54**(3), 1937 (2020).
- ³⁶P. Geurts, D. Ernst, and L. Wehenkel, "Extremely randomized trees," *Mach. Learn.* **63**, 3 (2006).
- ³⁷Z. Vujović, "Classification model evaluation metrics," *Int. J. Adv. Comput. Sci. Appl.* **12**(6), 599 (2021).
- ³⁸T. Fushiki, "Estimation of prediction error by using K-fold cross-validation," *Stat. Comput.* **21**, 137 (2011).
- ³⁹R. Polikar, "Ensemble based systems in decision making," *IEEE Circuits Syst. Mag.* **6**(3), 21 (2006).
- ⁴⁰T. Akiba, S. Sano, T. Yanase, T. Ohta, and M. Koyama, paper presented at the Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, 2019.
- ⁴¹Y. W. Noh, J. H. Lee, I. S. Jin, S. H. Park, and J. W. Jung, "Tailored electronic properties of Zr-doped SnO nanoparticles for efficient planar perovskite solar cells with marginal hysteresis," *Nano Energy* **65**, 104014 (2019).
- ⁴²J. Zhang, X. Li, L. Wang, J. Yu, S. Wageh, and A. A. Al-Ghamdi, "Enhanced performance of $\text{CH}_3\text{NH}_3\text{PbI}_3$ perovskite solar cells by excess halide modification," *Appl. Surf. Sci.* **564**, 150464 (2021).
- ⁴³P. Schulz, A. M. Dowgiallo, M. Yang, K. Zhu, J. L. Blackburn, and J. J. Berry, "Charge transfer dynamics between carbon nanotubes and hybrid organic metal halide perovskite films," *J. Phys. Chem. Lett.* **7**(3), 418 (2016).
- ⁴⁴S. Agbolaghi, "Efficacy beyond 17% via engineering the length and quality of grafts in organic halide perovskite/CNT photovoltaics," *New J. Chem.* **43**(26), 10567 (2019).
- ⁴⁵Y.-H. Seo, E.-C. Kim, S.-P. Cho, S.-S. Kim, and S.-I. Na, "High-performance planar perovskite solar cells: Influence of solvent upon performance," *Appl. Mater. Today* **9**, 598 (2017).

- ⁴⁶L. Yang, Y. Gao, Y. Wu, X. Xue, F. Wang, Y. Sui, Y. Sun, M. Wei, X. Liu, and H. Liu, "Novel insight into the role of chlorobenzene antisolvent engineering for highly efficient perovskite solar cells: Gradient diluted chlorine doping," *ACS Appl. Mater. Interfaces* **11**(1), 792 (2019).
- ⁴⁷E. J. Juarez-Perez, M. R. Leyden, S. Wang, L. K. Ono, Z. Hawash, and Y. Qi, "Role of the dopants on the morphological and transport properties of spiro-MeOTAD hole transport layer," *Chem. Mater.* **28**(16), 5702 (2016).
- ⁴⁸X. Jia, Z. Jiang, X. Chen, J. Zhou, L. Pan, F. Zhu, Z. Sun, and S. Huang, "Highly efficient and air stable inverted polymer solar cells using LiF-modified ITO cathode and MoO₃/AgAl alloy anode," *ACS Appl. Mater. Interfaces* **8**(6), 3792 (2016).
- ⁴⁹S. M. Lundberg, G. Erion, H. Chen, A. DeGrave, J. M. Prutkin, B. Nair, R. Katz, J. Himmelfarb, N. Bansal, and S. I. Lee, "From local explanations to global understanding with explainable AI for Trees," *Nat. Mach. Intell.* **2**(1), 56 (2020).
- ⁵⁰N. J. Jeon, J. H. Noh, Y. C. Kim, W. S. Yang, S. Ryu, and S. I. Seok, "Solvent engineering for high-performance inorganic-organic hybrid perovskite solar cells," *Nat. Mater.* **13**(9), 897 (2014).
- ⁵¹N. Ahn, D.-Y. Son, I.-H. Jang, S. M. Kang, M. Choi, and N.-G. Park, "Highly reproducible perovskite solar cells with average efficiency of 18.3% and best efficiency of 19.7% fabricated via Lewis base adduct of lead(II) iodide," *J. Am. Chem. Soc.* **137**(27), 8696 (2015).
- ⁵²G. Yang, H. Tao, P. Qin, W. Ke, and G. Fang, "Recent progress in electron transport layers for efficient perovskite solar cells," *J. Mater. Chem. A* **4**(11), 3970 (2016).
- ⁵³W. Ke, G. Fang, Q. Liu, L. Xiong, P. Qin, H. Tao, J. Wang, H. Lei, B. Li, J. Wan *et al.*, "Low-temperature solution-processed tin oxide as an alternative electron transporting layer for efficient perovskite solar cells," *J. Am. Chem. Soc.* **137**(21), 6730 (2015).
- ⁵⁴K. Wang, T. You, R. Yin, B. Fan, J. Liu, S. Cui, H. Chen, and P. Yin, "Precise nucleation regulation and defect passivation for highly efficient and stable carbon-based CsPbI₂Br perovskite solar cells," *ACS Appl. Energy Mater.* **4**(4), 3508 (2021).
- ⁵⁵A. D. Taylor, Q. Sun, K. P. Goetz, Q. An, T. Schramm, Y. Hofstetter, M. Litterst, F. Paulus, and Y. Vaynzof, "A general approach to high-efficiency perovskite solar cells by any antisolvent," *Nat. Commun.* **12**(1), 1878 (2021).
- ⁵⁶L. Calió, S. Kazim, M. Grätzel, and S. Ahmad, "Hole-transport materials for perovskite solar cells," *Angew. Chem. Int. Ed.* **55**(47), 14522 (2016).
- ⁵⁷G. Ayvazyan, F. Gasparyan, and V. Gasparian, "Optical simulation and experimental investigation of the crystalline silicon/black silicon/perovskite tandem structures," *Opt. Mater.* **140**, 113879 (2023).
- ⁵⁸F. Sahli, J. Werner, B. A. Kamino, M. Bräuninger, R. Monnard, B. Paviet-Salomon, L. Barraud, L. Ding, J. J. Diaz Leon, and D. Sacchetto, "Fully textured monolithic perovskite/silicon tandem solar cells with 25.2% power conversion efficiency," *Nat. Mater.* **17**(9), 820 (2018).
- ⁵⁹K. A. Bush, A. F. Palmstrom, Z. J. Yu, M. Boccard, R. Cheacharoen, J. P. Mailoa, D. P. McMeekin, R. L. Z. Hoyer, C. D. Bailie, and T. Leijtens, "23.6%-efficient monolithic perovskite/silicon tandem solar cells with improved stability," *Nat. Energy* **2**(4), 1 (2017).