

Machine learning-assisted discovery of Lewis base additives for defect passivation in perovskite solar cells

MENG Jiangtao^{1,2}, DING Bin³, LUO Shulin¹, ZHANG Qi¹, LI Yuanliang^{1,2}, WU Shuangjia^{1,2}, ZHAO Zhiyu¹, WANG Gangcheng^{1,2}, LIU Jun¹, WU Guixuan^{1,*}, DING Yong^{4,*}, ZHANG Wenyuan^{1,*}

¹ State Key Laboratory of Coal Conversion, Institute of Coal Chemistry, Chinese Academy of Sciences, Taiyuan 030001, China;

² University of Chinese Academy of Sciences, Beijing 100049, China;

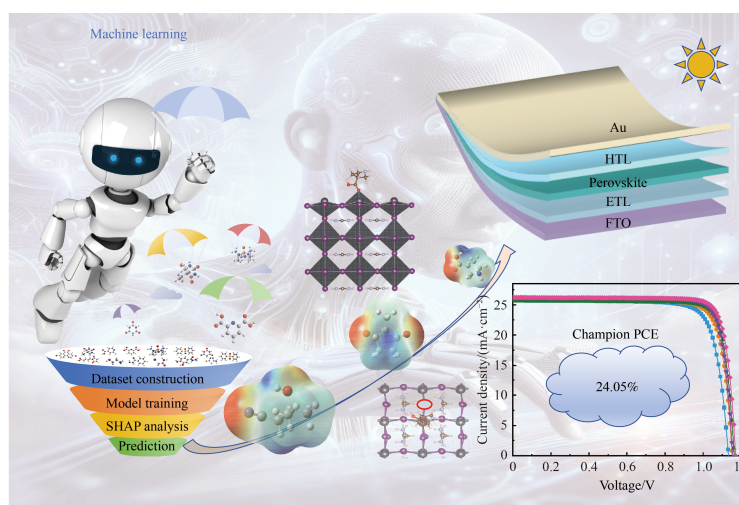
³ College of Chemistry, Chemical Engineering and Materials Science, Soochow University, Suzhou 215123, China;

⁴ College of Renewable Energy, Hohai University, Changzhou 213000, China

Abstract: Defect-induced nonradiative recombination critically restricts the power conversion efficiency (PCE) and stability of perovskite solar cells (PSCs). Lewis base additives show great promise in defect passivation, but current screening methods rely heavily on empirical trial and error and lack clear design principles, making it difficult to efficiently discover high-performance candidate materials. Here, we present a machine learning (ML) framework to intelligently screen Lewis base molecules for defect passivation. We trained six ensemble models on a dataset of 146 experimental data points, with Light Gradient Boosting Machine

(LightGBM) yielding the best classification performance (87% accuracy). Shapley Additive Explanations (SHAP) interpretability analysis subsequently identifies the highest occupied molecular orbital (HOMO) energy (-7.5 to -6.3 eV), additive concentration (2.5 to 6.5 mg/mL), and simplified molecular backbones (O atom ≤ 2 , C atom < 5) as critical design criteria. The ML prediction was experimentally validated: (S)-pyrrolidine-3-carboxylic acid ((S)-PCA) and 2-methyl-1,3-cyclopentanedione (MCPD) (Class II) improved PCE by 2.22% and 2.01%, respectively, while 3-hydroxymethyl-3-methylbutanenitrile (3-HMBN) (Class I) showed minimal gain. Density functional theory (DFT) calculations further confirmed the stronger binding affinities and elevated defect formation energies of Class II additives. Notably, the champion (S)-PCA device achieved a PCE of 24.05%. This work established an ML-accelerated paradigm for the rational design of defect passivators, bridging data science and photovoltaics.

Key words: machine learning; perovskite solar cells; Lewis base; defects passivation; density functional theory



Organic-inorganic hybrid perovskite solar cells (PSCs) have emerged as a leading photovoltaic technology driven by their high power conversion efficiencies (PCEs), low production costs, and facile fabrication processes^[1–4]. Since their initial report in 2009, the PCE of PSCs has surged

from 3.8% to over 26%, now rivaling and, in some cases, surpassing conventional photovoltaic technologies^[5]. This breakthrough is mainly attributed to the excellent optoelectronic properties of halide perovskites, including high absorption coefficient, low exciton binding energy,

Received: 2025-10-30; Revised: 2025-12-16

Foundation items: Supported by the Fundamental Research Program of Shanxi Province (202403021222487) and the Foundation from Chinese Academy of Sciences (YBR2023001, YBR2025003).

* Corresponding author. E-mail: wuguixuan@sxicc.ac.cn, yding@hhu.edu.cn, zhangwenyuan@sxicc.ac.cn.

DOI: 10.1016/S1872-5813(26)60655-X CSTR: 32157.14.2097-213X.2026.JFCT.0010 Associated Data DOI: 10.57760/sciencedb.j00124.00358

本文的英文电子版由 Elsevier 出版社在 ScienceDirect 上出版 (<http://www.sciencedirect.com/science/journal/18725813>)

long carrier diffusion length, tunable bandgap, and high carrier mobility^[6–12]. Despite these advantages, the commercialization of PSCs is hindered by the inherent instability of the materials. The soft, ionic lattice inevitably facilitates the formation of structural defects, particularly at the surface and grain boundaries (GBs) of polycrystalline films. These defect sites act as non-radiative recombination centers, leading to open-circuit voltage losses and accelerating the degradation of device efficiency and long-term operational stability^[13–15]. Consequently, defect passivation has become the crucial strategy for suppressing recombination losses and inhibiting ion migration, which is essential for enhancing PSC efficiency and stability.

To address this challenge, extensive research has been conducted on various chemical passivation materials, including small molecules^[16–18], polymers^[19–20], ionic liquids^[21–22], and inorganic salts. Among these, Lewis base small molecules (e.g., phosphine oxides, thiophenes) have received considerable attention due to their effective passivating properties^[23–25]. As electron donors, Lewis base small molecules can coordinate with undercoordinated Pb²⁺ ions at perovskite grain boundaries or surfaces, forming stable Pb–O/Pb–N bonds. This interaction suppresses deep-level trap states, reduces non-radiative recombination, and enhances film morphology by promoting uniform nucleation during crystallization. Therefore, Lewis base small molecules can significantly improve the overall performance of the device. Their molecular tunability, solution-processability, and applicability across various device architectures further underscore their potential for broad implementation. However, the discovery and optimization of high-performance Lewis base passivators remain largely empirical, relying on intuition-driven screening that lacks a fundamental understanding of the underlying structure-property-performance relationships. Moreover, due to the vast molecular design space, this trial-and-error approach is inherently inefficient^[26]. Therefore, establishing rational design principles is urgent to rapidly and accurately guide the identification and development of novel, efficient passivation materials, thereby accelerating the advancement of perovskite photovoltaic technologies.

To overcome this inherent limitation, machine learning (ML) has recently emerged as a transformative tool in perovskite material design and device performance optimization, leveraging its powerful data mining and predictive capabilities^[27–29]. Previous ML applications in PSCs include identifying key molecular descriptors correlated with efficiency in ammonium salts^[30], systematically evaluating organic halide salts for improved stability^[31], and integrating drift-diffusion physical modeling to select optimal passivators^[32]. However, ML research specifically focused on Lewis base passivators is still in its early stages. Currently, there is a lack of dedicated feature descriptor libraries tailored for Lewis base molecules in PSCs, as well as benchmark datasets that systematically correlate molecular features with device performance metrics such as PCE enhancement. As a result,

existing studies are largely confined to performance validation of individual molecules, leading to fragmented research without generalizable design principles or systematic screening strategies. This severely hinders the rational design and rapid discovery of high-performance Lewis base passivators.

In this work, we proposed an intelligent screening framework based on data-driven ML to rationally identify high-performance Lewis base additives. A dataset for machine learning training was constructed by collecting 146 experimental data points from the literature. Six binary classification prediction models were constructed using six ensemble learning algorithms—including Random Forest (RF), Extreme Gradient Boosting (XGBoost), Gradient Boosting Decision Tree (GBDT), Light Gradient Boosting Machine (LightGBM), Categorical Boosting (CatBoost), and Adaptive Boosting (AdaBoost)—with a threshold of PCE improvement rate $\geq 2\%$. The results show that the LightGBM model exhibited superior predictive accuracy, and subsequent Shapley Additive Explanations (SHAP) analysis revealed key molecular descriptors that governed passivation efficacy, providing in-depth mechanistic analysis and guidance for the screening and design of superior Lewis base passivators. The predictions of the LightGBM model were verified through the binding energy and defect formation energy calculated using the density functional theory (DFT) method. Moreover, the model's predictions were further validated by applying three novel, model-identified additives in device fabrication. The champion additive, (S)-pyrrolidine-3-carboxylic acid ((S)-PCA), yields a certified PCE of 24.05%. This work highlights the immense potential of machine learning in accelerating the identification of excellent Lewis base additives for PSCs, which opens a new avenue for developing more efficient perovskite devices.

1 Results and discussion

1.1 Dataset construction and feature engineering

In this work, we collected 146 data points from the literature where Lewis base small molecules were used as additives in PSCs, representing 44 distinct compounds. These molecules demonstrated effective defect passivation in PSCs. The target variable (Y) for the model is the difference in power conversion efficiency (ΔPCE), which is calculated as the difference between the passivated and initial PCE values ($\Delta\text{PCE} = \text{PCE}_{\text{passivated}} - \text{PCE}_{\text{initial}}$). ΔPCE is divided into two categories: Class I ($\Delta\text{PCE} < 2\%$) and Class II ($\Delta\text{PCE} \geq 2\%$). The ΔPCE values of the Lewis base small molecules are sourced from the literature. Multidimensional features (X) are also included to characterize the Lewis base molecules. We utilized the simplified molecular input line entry system (SMILES) from the PubChem database to obtain molecular structural information^[33]. Subsequently, 7 descriptors—including

atomic composition, physicochemical properties (computed *via* the RDKit library), electronic structure parameters (obtained from DFT calculations, such as the highest occupied molecular orbital (HOMO) energy, the lowest unoccupied molecular orbital (LUMO) energy), and the dipole moment—were incorporated. Furthermore, we also included the molecular concentration (MC) reported in the literature, resulting in a 20-dimensional feature space. The detailed definitions of these features are shown in Table 1.

To address the challenges posed by high-dimensional features and the limited sample data size^[34], a three-tiered feature engineering system was developed. First, we quantified the linear correlation between feature pairs using the Pearson correlation coefficient (r) (Figure S1)^[35]. Features with $|r| > 0.8$ were considered highly correlated, allowing for the selection of a representative feature while excluding the redundant counterpart^[36–37]. For example, HAC was excluded due to its high correlation with the C atom ($|r| = 0.94$), and EM was removed due to its perfect correlation with MW ($|r| = 1.0$). This filtering step substantially reduced feature redundancy (Figure 1). Subsequently, the dataset was discretized into a binary classification sample space, revealing a significant class imbalance (Class I : Class II = 74.66% : 25.34%, a sample ratio of 2.95 : 1) (Figure S2). This imbalance typically leads to model bias, causing the model to favor the majority class^[38].

Table 1 Definition of the 20 features for the organic small molecules in this study

Feature	Definition
C atom, H atom, O atom, N atom, S atom	number of carbon, hydrogen, oxygen, nitrogen and sulfur atoms
HOMO energy	highest occupied molecular orbital energy/eV
LUMO energy	lowest unoccupied molecular orbital energy/eV
MlogP	molecular lipid-water partition coefficient
TPSA	topological polar surface area/ $\text{\AA}^2$
HBD	number of hydrogen bond donors
HBA	number of hydrogen bond receptors
RBC	rotatable bond count
MCP	molecular complexity
HAC	heavy atom count
Gap	HOMO-LUMO gap/eV
MC	molecular concentration/($\text{mg} \cdot \text{mL}^{-1}$)
MW	molecular weight/($\text{g} \cdot \text{mol}^{-1}$)
RS	number of rings
DM	dipole moment
EM	exact mass(Da)

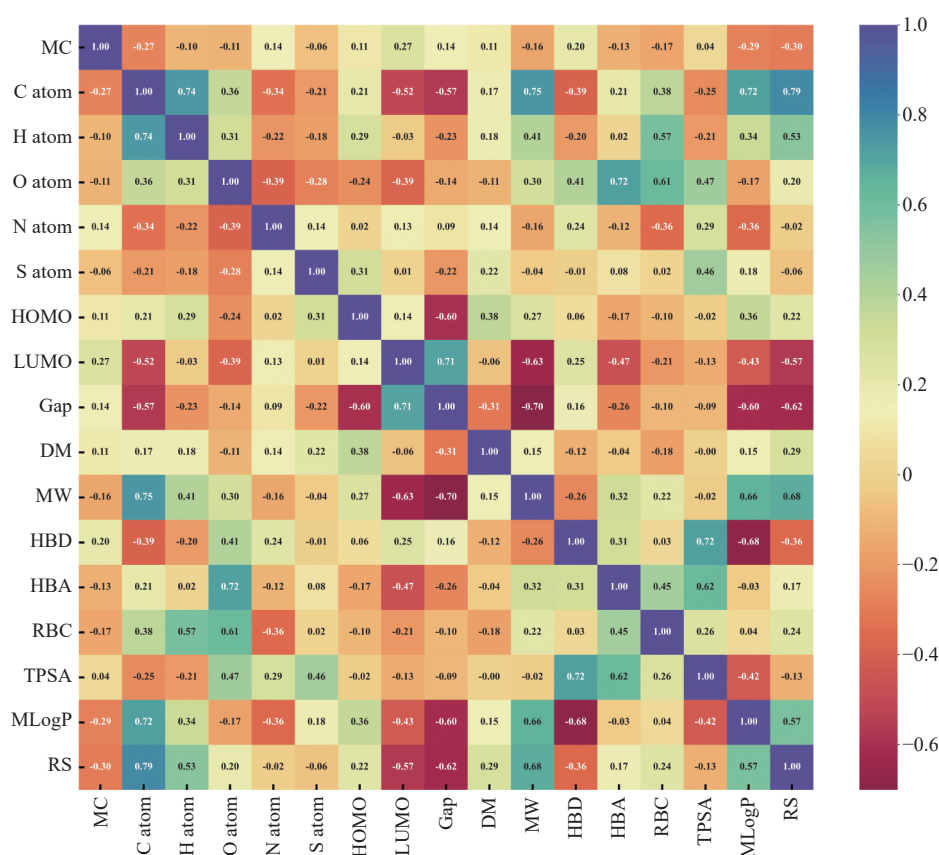


Fig. 1 Pearson correlation coefficient matrix of the 17 selected features

To address this issue, we employed the Synthetic Minority Over-sampling Technique (SMOTE) to correct the class imbalance^[39]. This adjustment effectively balanced the class distribution, increasing the ratio of Class I to Class II to 1.20:1 (Figure S2), thereby improving the classifier's sensitivity to minority class data. Finally, Min-Max normalization was applied to all numerical features, scaling each feature to the [0, 1] range to eliminate the influence of dimensional differences during model training. This progressive data processing framework significantly improved the generalization ability and convergence efficiency of subsequent ML models.

1.2 Model training and comparative evaluation

In a supervised ML framework, the model learns the mapping relationship between input variables X and output variables Y by training on a labeled dataset. ML tasks are conventionally categorized based on prediction objectives into regression (predicting continuous labels) and classification (predicting discrete labels). Although both regression and classification models are applicable, the classification approach demonstrates superior generalization capability when handling small-scale datasets ($n = 146$). Consequently, this work employed classification algorithms to construct ML models, implementing six ensemble learning methods for comparative analysis: RF^[40], XGBoost^[41], GBDT^[42], LightGBM^[43], CatBoost^[44] and AdaBoost^[45].

The dataset was split into training and testing sets in a 4:1 ratio, with the training and testing sets used for model development and performance evaluation, respectively. All models underwent hyperparameter optimization *via* Bayesian optimization with five-fold cross-validation^[46]. Each model's optimal hyperparameter configuration (detailed in Table S1) was determined through 50 iterations of the search process, strictly confined to the training set to prevent data leakage. It is worth noting that we used accuracy as the main metric for evaluating classification performance. As shown in Figure 2(a), the LightGBM model achieved an accuracy of 87% on the test set, indicating its high overall predictive correctness. To quantify the statistical significance of the observed performance differences, McNemar test was conducted. The analytical results showed that no statistically significant difference in test accuracy between LightGBM and RF ($p = 0.48$), GBDT ($p = 0.48$), XGBoost ($p = 1.00$), CatBoost ($p = 1.00$), or AdaBoost ($p = 0.45$). To further evaluate the model's discriminative ability, Receiver Operating Characteristic (ROC) analysis and corresponding Area Under the Curve (AUC) values were examined (Figure 2(b) and Figure S3). The ROC curves illustrate the trade-off between true positive rates and false positive rates across various decision thresholds, while AUC quantifies the model's ability to distinguish between classes. Herein, the AUC values for both Category I and Category II were 0.82, indicating that the LightGBM model has strong inter-class separability.

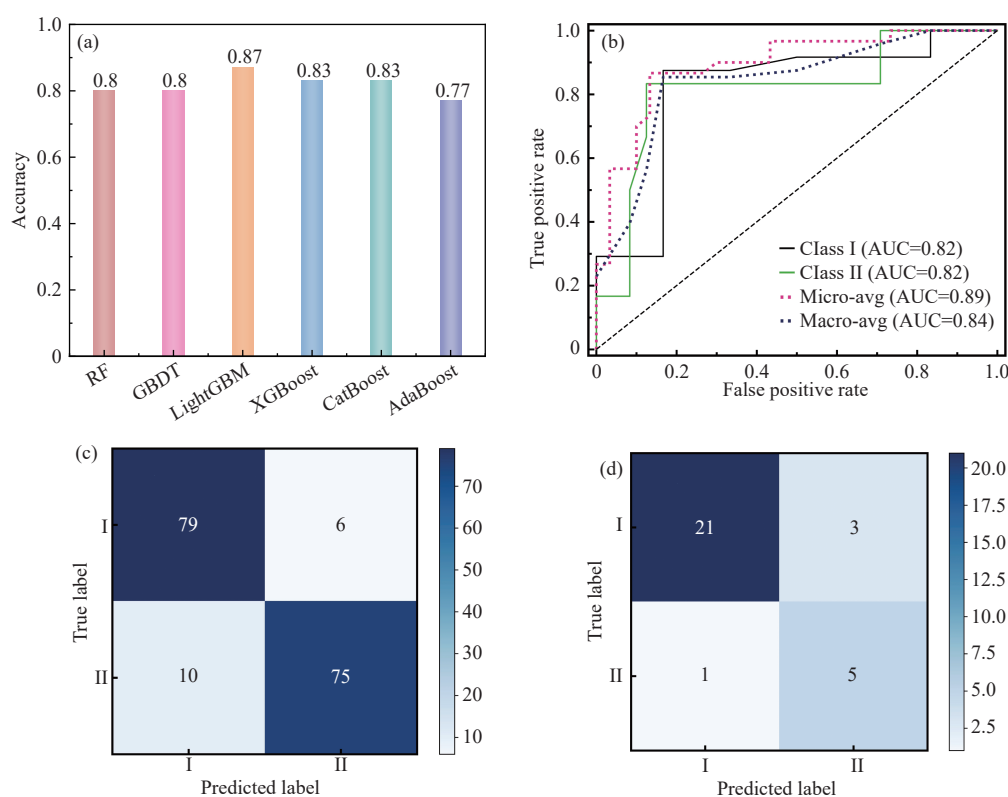


Fig. 2 (a) The accuracy of the ML ensemble model; (b) The ROC curve of the LightGBM model on the test set; (c) The confusion matrix of the LightGBM model on the train set; (d) The confusion matrix of the LightGBM model on the test set

The micro-average AUC was 0.89, and the macro-average AUC was 0.84, further validating the model's performance from the perspectives of sample-level performance and class balance. Finally, the classification accuracy of the model was validated using confusion matrices. The confusion matrices for both the training and testing sets (Figure 2(c), (d)) exhibit high and stable classification accuracy. Results on the test set (Figure 2(d)) demonstrate that among the 24 Class I samples, 3 were misclassified as Class II; conversely, only 1 of 6 Class II samples was incorrectly assigned to Class I. This indicates that the LightGBM model not only demonstrates superior discriminative capability compared to other models but also possesses strong generalization performance (Figure S4). Considering the accuracy metrics, ROC-AUC analysis, and confusion matrix evaluation supports the superior predictive performance and generalizability of the LightGBM model. Therefore, LightGBM was selected as the final predictive model for this work.

1.3 SHAP-driven Interpretability and design criteria

The Shapley Additive Explanations (SHAP) method, proposed by Lundberg and Lee, utilizes game theory to explain the predictions of machine learning models. This work employs this method to deeply analyze the contribution of each feature to the prediction results. Figure 3(a) presents the importance ranking of input features X for the two output classes Y , based on the average absolute SHAP values. The larger the average absolute SHAP value of a feature, the greater its influence on the classification result. The results show that the HOMO energy, MC, O atom and C atom are the four most influential features, having a significant and consistent impact on the molecular classification results for both output classes (Figure S5 and S6).

Figure 3(b) illustrates the SHAP summary for the prediction of class II. A slight vertical jitter was introduced to ensure the visibility of all data points. Positive SHAP values indicate a positive contribution to class II classification, whereas negative SHAP values indicate a negative contribution. Analysis reveals that the feature values of O atom and C atom are negatively correlated with SHAP values, while HOMO energy and MC exhibit a trend of initially increasing and then decreasing SHAP values. This clearly demonstrates the direction and strength of the contributions of these features to the classification of Category II.

Figure 3(c)–(f) further explore the quantitative impact of these important features on the class II prediction:

- (1) HOMO energy: when the HOMO energy is in the range of -7.5 to -6.3 eV, its SHAP value becomes significantly positive (Figure 3(c)).
- (2) MC: an MC in the range of 2.5 to 6.5 mg/mL yields a peak contribution (Figure 3(d)).
- (3) Molecular Structure: the number of oxygen atoms $\leq$

2 (Figure 3(e)) and the number of carbon atoms < 5 (Figure 3(f)) both show a positive effect, indicating that simplifying the molecular backbone enhances passivation performance.

Based on these findings, we proposed a multi-parameter collaborative optimization range: HOMO energy (-7.5 to -6.3 eV), MC (2.5 to 6.5 mg/mL), O atom ≤ 2 and C atom < 5 . This quantitative standard reveals the nonlinear coupling relationships between characteristic parameters, which establishes a measurable theoretical framework for the rational design and performance prediction of Lewis base passivators. This framework effectively bridges the predictive power of ML models with material design principles.

1.4 Molecular screening prediction and verification

To validate the predictions of our model, three previously unexplored Lewis base molecules—3-hydroxymethyl-3-methylbutanenitrile (3-HMBN), 2-methyl-1,3-cyclopentanedione (MCPD) and (S)-PCA—were selected for theoretical and experimental investigation. We performed density functional theory (DFT) calculations to evaluate the electrostatic potential (MEP) distribution of the molecules^[47–48] (Figure 4(a)–(c)). The results confirmed that these molecules possess Lewis basic properties, as significant negative electrostatic potentials were observed on the N and O atoms.

These atoms can coordinate with the under-coordinated Pb^{2+} ions, promoting electrostatic interaction with the perovskite surface^[49]. To further investigate these interactions, the binding energies of the three molecules with the most stable PbI_2 -terminated (001) perovskite surface were calculated (Figure S7)^[16]. The results show that the binding energies of 3-HMBN, MCPD and (S)-PCA are -0.71 , -0.82 and -0.92 eV, respectively. This indicates a favorable interaction, with (S)-PCA exhibiting the strongest and most stable binding affinity.

Furthermore, we evaluated the defect formation energies of the pristine perovskite (0.91 eV) and the passivated systems. The defect formation energy increased significantly to 2.25 eV (3-HMBN), 2.27 eV (MCPD) and 2.38 eV ((S)-PCA), indicating that these candidate passivating agents can effectively suppress the formation of intrinsic defects. The calculated binding energies and defect formation energies (Figure 4(d)) exhibit excellent correlation with the LightGBM model predictions: higher binding energies and defect formation energies correspond to more effective defect passivation. Consequently, Class II molecules ((S)-PCA and MCPD) exhibit higher ΔPCE values than the Class I molecule (3-HMBN), predicted with the highest classification probability (95.83%), demonstrated superior theoretical performance compared to MCPD (92.28%). In contrast, 3-HMBN was predicted as a Class I molecule (75.19% probability), which is consistent with its lower ΔPCE and weaker passivation ability. These findings provide strong theoretical support for the model's ability to distinguish effective passivators.

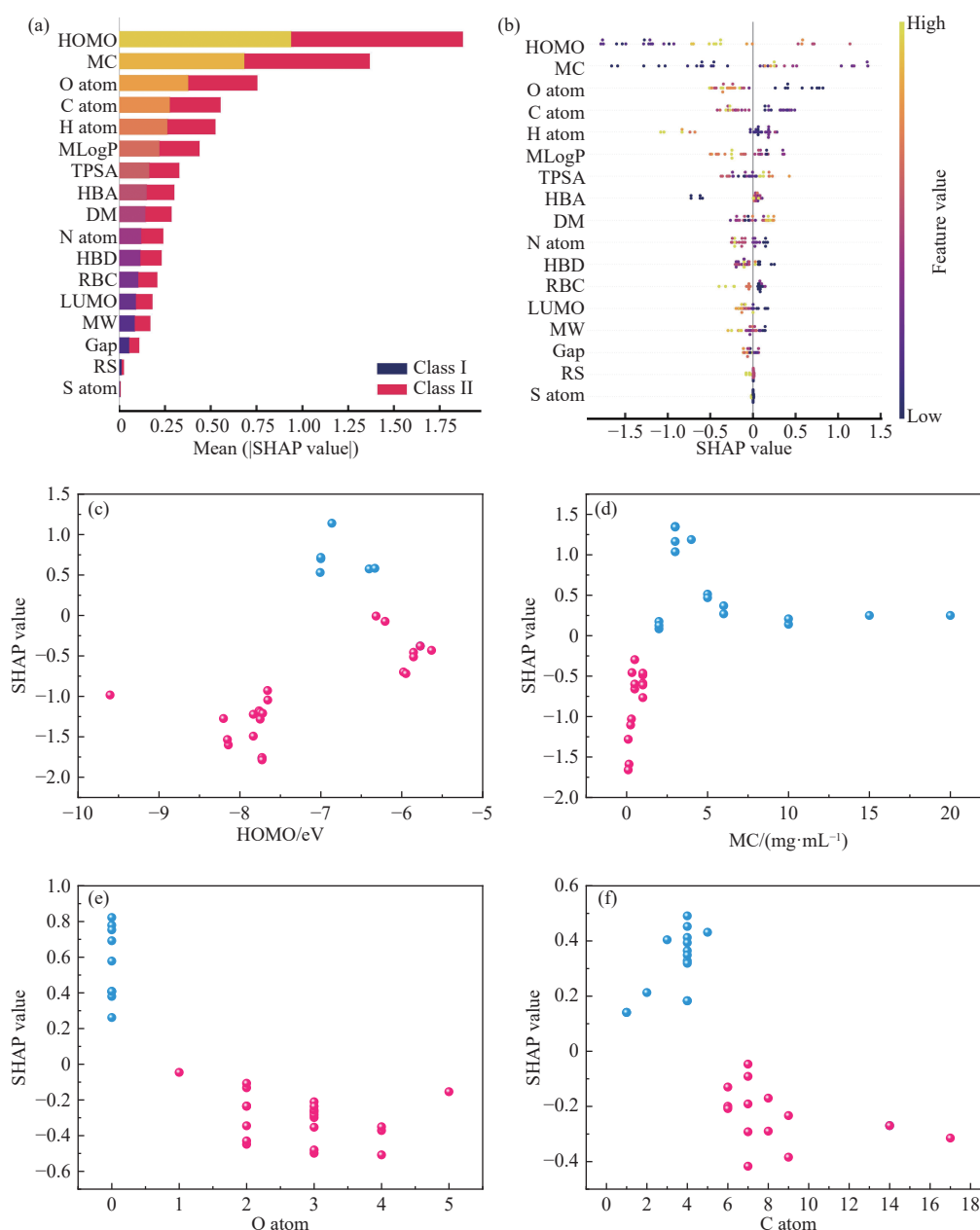


Fig. 3 (a) The ranking of feature X importance for both Class I and Class II obtained from the LightGBM model; (b) The SHAP values of various features X with different values for Class II, where the color bar on the right indicates the relative value size of features X_i ; The relationships between the values of important features and their SHAP values: (c) HOMO; (d) MC; (e) O atom; (f) C atom

To experimentally validate these theoretical and computational findings, we fabricated FTO/SnO₂@TiO₂/perovskite/PEAI/Spiro-OMeTAD/Au devices. The additives were introduced into the perovskite precursor solutions at their optimal probability MC predicted by LightGBM: 3-HMBN (1.0 mg/mL), MCPD (4.5 mg/mL) and (S)-PCA (4.5 mg/mL). Under ambient conditions, we measured the current density-voltage (J - V) characteristics of the control device and the target devices under AM1.5G illumination (100 mW/cm²), as shown in Figure 4(e), with the corresponding photovoltaic parameters summarized in

Table S2. The results show that the power conversion efficiency (PCE) of all three modified devices was significantly improved compared to the control device (21.85%). Furthermore, the PCE statistics, as shown in the histogram of 14 devices in Figure 4(f), indicate excellent reproducibility. Notably, the (S)-PCA device achieved the highest certified PCE of 24.05%, marking a substantial improvement of 2.22%. The close agreement between experimental data and theoretical modeling further validates the reliability of our ML-based predictive model.

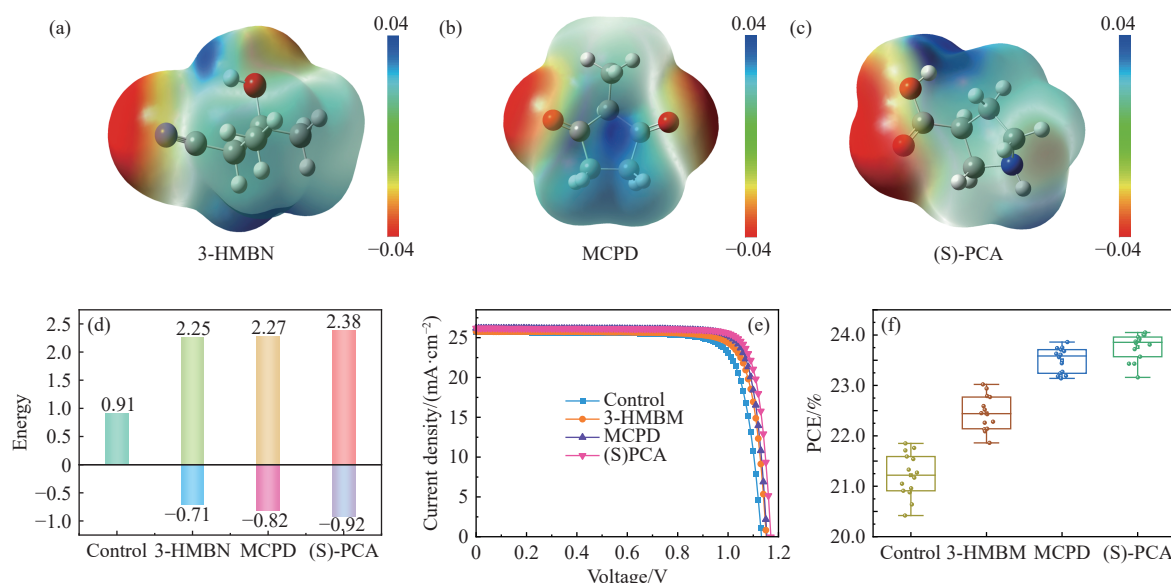


Fig. 4 The molecular electrostatic potential of (a) 3-HMBN, (b) MCPD, (c) (S)-PCA; (d) Binding energy and V_T -related defect formation energy on the surface; (e) $J-V$ curves of the control device and (S)-PCA device; (f) PCE distribution from 14 photovoltaic devices

To gain deeper insights into the passivation effect of the champion additive, (S)-PCA, we conducted the steady-state photoluminescence (PL) and the time-resolved photoluminescence (TRPL) analyses to analyze the charge generation and recombination kinetics for different perovskite films. The (S)-PCA film exhibited enhanced PL intensity and a significantly longer decay time (621.32 ns) compared to the control film (445.36 ns) (Figure S8, Figure 5

(a) and Table S3), indicating a drastic suppression of non-radiative recombination losses^[50–51]. We further performed quantitative evaluation of the electron transport device (FTO/SnO₂/perovskite/PCBM/Au) using space-charge-limited current (SCLC) measurements, which determined the trap state density (N_t) based on the trap-filling limit voltage (V_{TFL})^[52] (Figure 5(b)). The V_{TFL} values for the control and films were 0.552 and 0.356 V, respectively.

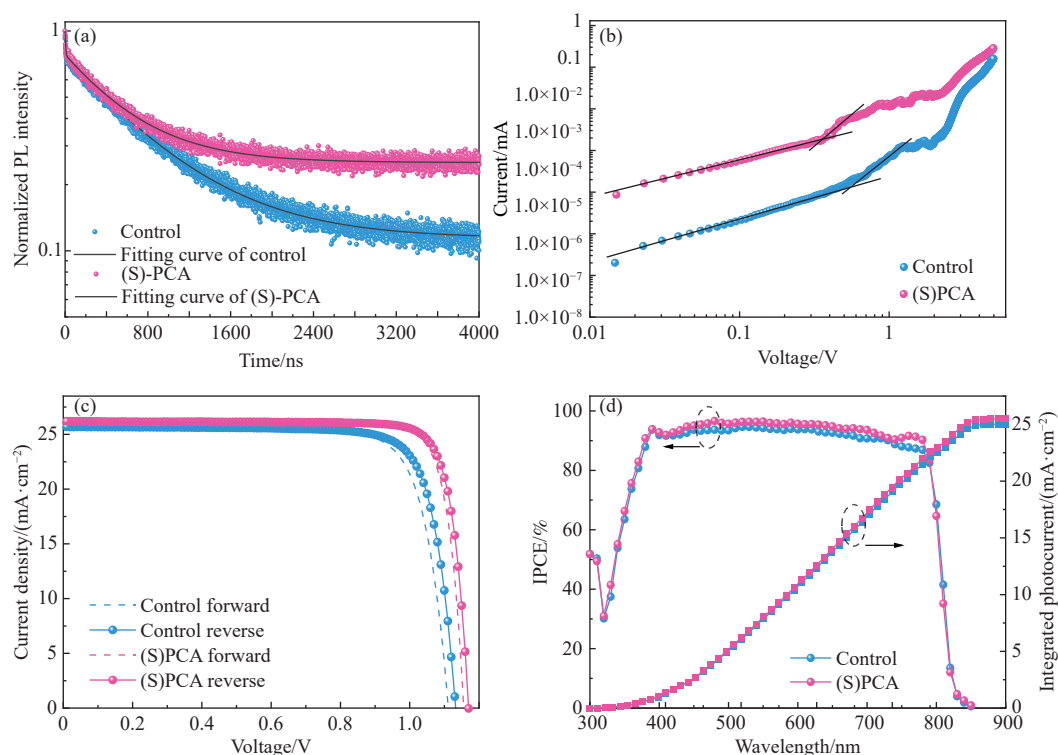


Fig. 5 (a) TRPL spectra of control and (S)-PCA films on glass substrates; (b) Dark current-voltage characteristics of electron only devices with and without (S)-PCA passivation; (c) $J-V$ curves of control and (S)-PCA champion devices under reverse and forward scanning; (d) IPCE spectrum and corresponding integrated J_{sc}

This indicates a significant reduction in N_t from $5.94 \times 10^{15} \text{ cm}^{-3}$ (control) to $3.83 \times 10^{15} \text{ cm}^{-3}$ ((S)-PCA), definitively showing that (S)-PCA effectively represses defect formation.

Electrochemical impedance spectroscopy (EIS) measurements (Figure S9 and Table S4) showed a decrease in the charge transfer resistance (R_{ct}) of the (S)-PCA device, indicating higher charge extraction and transport efficiency, which is consistent with a lower density of defect state^[53]. To systematically evaluate the impact of (S)-PCA modification on device performance and hysteresis effects, current density-voltage (J - V) characteristics were measured under both forward and reverse scans. As shown in Figure 5(c), the champion (S)-PCA device demonstrated superior performance across all metrics: PCE of 24.05%, V_{oc} of 1.112 V, J_{sc} of 26.11 mA/cm² and FF of 82.84%, with significantly reduced hysteresis compared to the control device (Table S5). Furthermore, the (S)-PCA-based device exhibited significantly enhanced incident photon-to-electron conversion efficiency (IPCE) across nearly the entire spectral range, as shown in Figure 5(d). The integrated J_{sc} values reached 25.03 and 25.52 mA/cm², for the control and (S)-PCA devices, respectively, which are very close to the J_{sc} values measured from J - V curves (deviation < 5%). This consistency further validated the superior optoelectronic performance of the (S)-PCA device.

2 Conclusion

In summary, we developed a ML-driven framework for the rapid and rational identification of high-performance Lewis base passivators in PSCs. By combining experimental datasets with multidimensional molecular descriptors (electronic, structural, and compositional features), the LightGBM model achieved excellent predictive accuracy (87%) in predicting additives capable of increasing PCE by $\geq 2\%$. SHAP-based interpretability analysis identified key molecular design rules governing the passivation performance: a lower HOMO energy (-7.5 to -6.3 eV), an optimal MC window (2.5 to 6.5 mg/mL), and reduced numbers of O and C atoms (≤ 2 and < 5 , respectively). The accuracy of the ML model was validated through both experimental assessments and DFT calculations. Ultimately, a device modified with (S)-PCA achieved a record-high certified PCE of 24.05% ($\Delta\text{PCE} = +2.22\%$). This work not only delivers a high-efficiency passivator but also establishes a generalizable, ML-accelerated paradigm. This effectively reveals the complex structure-performance relationships in PSCs materials, paving the way for data-driven discovery of next-generation photovoltaic materials.

Declaration of Competing Interest

The authors declare no conflict of interest.

References

- [1] ZHANG S, YE F Y, WANG X Y, et al. Minimizing buried interfacial defects for efficient inverted perovskite solar cells[J]. Science, 2023, 380(6643): 404–409.
- [2] PENG W, MAO K T, CAI F C, et al. Reducing nonradiative recombination in perovskite solar cells with a porous insulator contact[J]. Science, 2023, 379(6633): 683–690.
- [3] ZHANG T K, WANG F, KIM H B, et al. Ion-modulated radical doping of spiro-OMeTAD for more efficient and stable perovskite solar cells[J]. Science, 2022, 377(6605): 495–501.
- [4] LI Y J, HE Z Y, CHEN X J, et al. Dual-functional interfacial engineering with Portulaca oleracea phytochemicals enables efficient and eco-safe perovskite solar cells[J]. Appl Surf Sci, 2025, 711: 164045.
- [5] WANG Y Y, LU C X, LIU M C, et al. Solvent-assisted reaction for spontaneous defect passivation in perovskite solar cells[J]. Nat Photonics, 2025.
- [6] RAVI M, ZHAI M D, CHEN C, et al. Tuning Pyrrolo[3, 2-b] pyrrole core-based hole transport materials properties via addition of fluorine for highly efficient and stable planar perovskite solar cells[J]. Appl Surf Sci, 2025, 680: 161312.
- [7] WANG K, JIN Z W, LIANG L, et al. All-inorganic cesium lead iodide perovskite solar cells with stabilized efficiency beyond 15%[J]. Nat Commun, 2018, 9(1): 4544.
- [8] SONG Z N, LI C W, CHEN L, et al. Perovskite solar cells go bifacial—mutual benefits for efficiency and durability[J]. Adv Mater, 2022, 34(4): 2106805.
- [9] LI X C, WU X, LI B, et al. Modulating the deep-level defects and charge extraction for efficient perovskite solar cells with high fill factor over 86%[J]. Energy Environ Sci, 2022, 15(11): 4813–4822.
- [10] CHEN Y, XIAO Y, HU J T, et al. Multifunctional ytterbium oxide buffer for perovskite solar cells[J]. Nature, 2024, 625(7995): 516–522.
- [11] YU S Q, XIONG Z, ZHOU H T, et al. Homogenized NiO_x nanoparticles for improved hole transport in inverted perovskite solar cells[J]. Science, 2023, 382(6677): 1399–1404.
- [12] LIU M C, WANG Y Y, LU C X, et al. Localized oxidation embellishing strategy enables high-performance perovskite solar cells[J]. Angew Chem, 2024, 136.
- [13] YANG S, DAI J, YU H, et al. Tailoring passivation molecular structures for extremely small open-circuit voltage loss in perovskite solar cells[J]. J Am Chem Soc, 2019, 141(14): 5781–5787.
- [14] DENG J D, ZHANG H F, WEI K, et al. Molecular bridge assisted bifacial defect healing enables low energy loss for efficient and stable perovskite solar cells[J]. Adv Funct Mater, 2022, 32(52): 2209516.
- [15] FU L, LI H, WANG L, et al. Defect passivation strategies in perovskites for an enhanced photovoltaic performance[J]. Energy Environ Sci, 2020, 13(11): 4017–4056.
- [16] GUO J J, SUN J G, HU L, et al. Indigo: A natural molecular passivator for efficient perovskite solar cells[J]. Adv Energy Mater, 2022, 12(22): 2200537.
- [17] WANG Y, YANG H, ZHANG K, et al. Tautomeric

- passivation strategy-assisted photostable perovskite solar modules[J]. *ACS Energy Lett*, 2022, 7(10): 3646–3652.
- [18] ZHANG W Y, HE L, MENG Y, et al. Dual-site synergistic passivation for highly efficient and stable perovskite solar cells[J]. *Adv Energy Mater*, 2022, 12(46): 2202189.
- [19] WANG M H, ZHAO Y P, JIANG X Q, et al. Rational selection of the polymeric structure for interface engineering of perovskite solar cells[J]. *Joule*, 2022, 6(5): 1032–1048.
- [20] HAN T H, LEE J W, CHOI C, et al. Perovskite-polymer composite cross-linker approach for highly-stable and efficient perovskite solar cells[J]. *Nat Commun*, 2019, 10(1): 520.
- [21] WU J P, LI M H, FAN J T, et al. Regioselective multisite atomic-chlorine passivation enables efficient and stable perovskite solar cells[J]. *J Am Chem Soc*, 2023, 145(10): 5872–5879.
- [22] ZHU Y Q, LV P, HU M, et al. Synergetic passivation of metal-halide perovskite with fluorinated phenmethylammonium toward efficient solar cells and modules[J]. *Adv Energy Mater*, 2023, 13(8): 2203681.
- [23] LEE J W, KIM H S, PARK N G. Lewis acid–base adduct approach for high efficiency perovskite solar cells[J]. *Acc Chem Res*, 2016, 49(2): 311–319.
- [24] LEE J W, BAE S H, HSIEH Y T, et al. A Bifunctional lewis base additive for microscopic homogeneity in perovskite solar cells[J]. *Chem*, 2017, 3(2): 290–302.
- [25] LI C W, WANG X M, BI E B, et al. Rational design of Lewis base molecules for stable and efficient inverted perovskite solar cells[J]. *Science*, 2023, 379(6633): 690–694.
- [26] XU J, CHEN H, GRATER L, et al. Anion optimization for bifunctional surface passivation in perovskite solar cells[J]. *Nat Mater*, 2023, 22(12): 1507–1514.
- [27] ZHAO Y C, ZHANG J Y, XU Z W, et al. Discovery of temperature-induced stability reversal in perovskites using high-throughput robotic learning[J]. *Nat Commun*, 2021, 12(1): 2191.
- [28] LIN B, JIANG J, ZENG X C, et al. Temperature-pressure phase diagram of confined monolayer water/ice at first-principles accuracy with a machine-learning force field[J]. *Nat Commun*, 2023, 14(1): 4110.
- [29] HU Y J, HU X B, ZHANG L, et al. Machine-learning modeling for ultra-stable high-efficiency perovskite solar cells[J]. *Adv Energy Mater*, 2022, 12(41): 2201463.
- [30] ZHI C Y, WANG S, SUN S J, et al. Machine-learning-assisted screening of interface passivation materials for perovskite solar cells[J]. *ACS Energy Lett*, 2023, 8(3): 1424–1433.
- [31] HARTONO N T P, THAPA J, TIIHONEN A, et al. How machine learning can help select capping layers to suppress perovskite degradation[J]. *Nat Commun*, 2020, 11(1): 4172.
- [32] LIU W, MENG N, HUO X M, et al. Machine learning enables intelligent screening of interface materials towards minimizing voltage losses for p-i-n type perovskite solar cells[J]. *J Energy Chem*, 2023, 83: 128–137.
- [33] WEININGER D. SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules[J]. *J Chem Inf Comput Sci*, 1988, 28(1): 31–36.
- [34] LIU R, GILLIES D F. Overfitting in linear feature extraction for classification of high-dimensional image data[J]. *Pattern Recognit*, 2016, 53: 73–86.
- [35] ZHOU H M, DENG Z H, XIA Y Q, et al. A new sampling method in particle filter based on Pearson correlation coefficient[J]. *Neurocomputing*, 2016, 216: 208–215.
- [36] PUJOL A, BROKHATTINGEN N, MATAMBISSO G, et al. Detecting temporal and spatial malaria patterns from first antenatal care visits[J]. *Nat Commun*, 2023, 14(1): 4004.
- [37] DURAND S, LIAN Q C, JING J, et al. Joint control of meiotic crossover patterning by the synaptonemal complex and HEI10 dosage[J]. *Nat Commun*, 2022, 13(1): 5999.
- [38] JAISWAL A, DWIVEDI P, DEWANG R k. Handling imbalance dataset issue in insider threat detection using machine learning methods[J]. *Comput Electr Eng*, 2024, 120: 109726.
- [39] SONG C, ZHAO T Y, XU L, et al. Probabilistic prediction of uniaxial compressive strength for rocks from sparse data using Bayesian Gaussian process regression with Synthetic Minority Oversampling Technique (SMOTE)[J]. *Comput Geotech*, 2024, 165: 105850.
- [40] HOUBORG R, MCCABE M F. hybrid training approach for leaf area index estimation via Cubist and random forests machine-learning[J]. *ISPRS J Photogramm Remote Sens*, 2018, 135: 173–188.
- [41] HONG Z X, TAO M, LIU L L, et al. An intelligent approach for predicting overbreak in underground blasting operation based on an optimized XGBoost model. *Eng[J]. Appl Artif Intell*, 2023, 126: 107097.
- [42] ZHOU F, ZHANG Q, SORNETTE D, et al. Cascading logistic regression onto gradient boosted decision trees for forecasting and trading stock indices[J]. *Appl Soft Comput*, 2019, 84: 105747.
- [43] FAN J L, MA X, WU L F, et al. Light Gradient Boosting Machine: An efficient soft computing model for estimating daily reference evapotranspiration with local and external meteorological data[J]. *Agric Water Manage*, 2019, 225: 105758.
- [44] JOO C, PARK H, LIM J, et al. Machine learning-based heat deflection temperature prediction and effect analysis in polypropylene composites using catboost and shapley additive explanations[J]. *Eng Appl Artif Intell*, 2023, 126: 106873.
- [45] CAO Y, MIAO Q G, LIU J C, et al. Advance and Prospects of AdaBoost Algorithm[J]. *Acta Autom Sin*, 2013, 39(6): 745–758.
- [46] WONG T T. Performance evaluation of classification algorithms by k-fold and leave-one-out cross validation[J]. *Pattern Recognit*, 2015, 48(9): 2839–2846.
- [47] LU T, CHEN F W. Multiwfn: A multifunctional wavefunction analyzer[J]. *J Comput Chem*, 2012, 33(5): 580–592.
- [48] LU T. A comprehensive electron wavefunction analysis toolbox for chemists, Multiwfn[J]. *J Chem Phys*, 2024, 161(8).
- [49] LI Z, LI B, WU X, et al. Organometallic-functionalized interfaces for highly efficient inverted perovskite solar

- cells[J]. Science, 2022, 376(6591): 416–420.
- [50] GUO Y, HE X, LIU X, et al. One- step implementation of plasmon enhancement and solvent annealing effects for airprocessed high- efficiency perovskite solar cells[J]. J Mater Chem A, 2018, 6(47): 24036–24044.
- [51] ZHENG H Y, WU W W, XU H F, et al. Self-additive low-dimensional ruddlesden–popper perovskite by the incorporation of glycine hydrochloride for high-performance and stable solar cells[J]. Adv Funct Mater, 2020, 30(15): 2000034.
- [52] DUIJNSTEE E A, BALL J M, LE Corre V M, et al. Toward understanding space-charge limited current measurements on metal halide perovskites[J]. ACS Energy Lett, 2020, 5(2): 376–384.
- [53] WANG R, XUE J J, WANG K L, et al. Constructive molecular configurations for surface-defect passivation of perovskite photovoltaics[J]. Science, 2019, 366(6472): 1509–1513.

机器学习辅助发现用于钙钛矿太阳能电池缺陷钝化的路易斯碱添加剂

孟江涛^{1,2}, 丁 斌³, 罗树林¹, 张 祺¹, 李元亮^{1,2}, 武双甲^{1,2}, 赵治宇¹, 王刚成^{1,2},
刘 军¹, 吴桂选^{1*}, 丁 勇^{4*}, 张文媛^{1*}

(1. 中国科学院山西煤炭化学研究所 煤炭高效低碳利用全国重点实验室, 山西 太原 030001; 2. 中国科学院大学, 北京 100049;
3. 苏州大学 化学化工与材料科学学院, 江苏 苏州 215123; 4. 河海大学 新能源学院, 江苏 常州 213000)

摘 要: 缺陷诱导的非辐射复合严重制约着钙钛矿太阳能电池(PSCs)的功率转换效率(PCE)与稳定性。尽管路易斯碱添加剂在缺陷钝化方面前景广阔, 但当前的筛选方法严重依赖经验性试错, 缺乏高效发现高性能候选分子所需明确的设计原则。本研究提出一种机器学习(ML)框架, 用于智能筛选适用于缺陷钝化的路易斯碱分子。基于包含 146 个实验数据点的数据集, 训练了六种集成模型, 其中, LightGBM 取得了最佳分类性能(准确率 87%)。沙普利加和解释(SHAP)可解释性分析确定了最高占据分子轨道(HOMO)能级(−7.5–(−6.3) eV)、添加剂质量浓度(2.5–6.5 mg/mL)以及简化分子骨架(O 原子数≤2, C 原子数<5)为关键设计准则。机器学习预测结果得到了实验验证, (S)-吡咯烷-3-羧酸((S)-PCA)和 2-甲基-1,3-环戊二酮(MCPD)(II 类添加剂)分别将 PCE 提升了 2.22% 和 2.01%, 而 3-羟甲基-3-甲基丁腈(3-HMBN)(I 类添加剂)则改善甚微。密度泛函理论(DFT)计算进一步证实了 II 类添加剂具有更强的结合亲和力与更高的缺陷形成能。最优的 (S)-PCA 器件实现了 24.05% 的 PCE。本工作建立了一种机器学习加速的缺陷钝化剂理性设计新范式, 为数据科学与光伏研究搭建了桥梁。

关键词: 机器学习; 钙钛矿太阳能电池; 路易斯碱; 缺陷钝化; 密度泛函理论

中图分类号: TM914.4; TP181

文献标识码: A

作者简介:



孟江涛(1997), 硕士研究生, 研究方向为第一性原理计算模拟, 数据驱动机器学习钙钛矿太阳能电池材料理性设计的研究, 18235041845@163.com。



丁勇(1987), 博士, 教授, 研究方向为光伏钙钛矿太阳能电池, 大面积组件, 叠层电池的研究, yding@hhu.edu.cn。



吴桂选(1984), 博士, 研究员, 研究方向为化工与能源材料热力学性质与数据库开发, 基于数据库和人工智能的多目标优化与设计, 复杂反应器的过程模拟与优化及废弃物资源化综合利用的研究, wuguixuan@sxicc.ac.cn。



张文媛(1994), 博士, 助理研究员, 研究方向为钙钛矿太阳能电池, 第一性原理计算, 全固态电池电解质的研究, zhangwenyuan@sxicc.ac.cn。