Machine Learning-Accelerated Identification and Passivation of Interface Defects in Perovskite Solar Cells

Wei Yan

College of Energy, Soochow University, Suzhou, China
E-mail: yw901v@163.com
Manuscript received February 11, 2026; accepted March 29, 2026; published April 28, 2026

***Abstract*—Perovskite solar cells have attracted considerable attention owing to their outstanding photoelectric conversion performance; however, interface defects severely impede further improvements in device efficiency and stability. This work proposes a machine learning-accelerated framework for interface defect identification and passivation strategy optimization in perovskite solar cells. A high-quality defect feature database incorporating Density Functional Theory (DFT) calculation data is constructed, and random forest and Gradient Boosting Decision Tree (GBDT) algorithms are employed for accurate classification of interface defect types, while a Bayesian optimization method is applied to systematically screen passivation molecules. In alignment with existing research on explainable artificial intelligence methods and ensemble learning frameworks, this work introduces machine learning interpretability analysis into the domain of materials defect identification. Simulation results demonstrate that the proposed method achieves a defect identification accuracy of 96.218%. After optimized passivation treatment, the Power Conversion Efficiency (PCE) of the device is improved from 19.347% to 24.186%, the open-circuit voltage (V_{oc}) increases from 1.082 V to 1.156 V, and the Fill Factor (FF) rises from 0.791 to 0.843. This study provides a data-driven paradigm for the rational design of highly efficient and stable perovskite solar cells.**

***Keywords*—perovskite solar cells, machine learning, interface defects, passivation strategy, power conversion efficiency, random forest, Bayesian optimization**

I. INTRODUCTION

Since the first report of Perovskite Solar Cells (PSCs) in 2009 [1], their certified power conversion efficiency has rapidly increased from an initial 3.8% to over 26% [1–5], demonstrating exceptional development potential and broad commercialization prospects. During the fabrication of perovskite thin films, a large number of defect states inevitably form at grain boundaries and at the interfaces between the absorber layer and charge transport layers. These defects, acting as non-radiative recombination centers, not only significantly reduce the open-circuit voltage and fill factor of the device but are also the fundamental cause of long-term device stability degradation [6, 7]. Conventional experimental characterization techniques face evident bottlenecks in the efficiency of defect information acquisition, while the experience- dependent screening of passivation molecules makes it difficult to identify optimal solutions efficiently within the vast chemical space. In recent years, the rapid develop- ment of machine learning has provided new tools and perspectives for materials science research [8, 9]. Ensemble learning methods represented by random forests and gradi- ent boosting decision trees have been widely applied in energy system prediction [10], medical classification [2, 3], and software defect prediction [4], and have been endowed with decision transparency through explainability analysis tools such as SHAP [3]. Breakthroughs in graph neural networks and large language models for complex system modeling [11, 12], together with advances in approximate computing and efficient deployment technologies [13], have further extended the boundaries of artificial intelligence in scientific computing. The combination of SAM and YOLO-series models [14] and the maturation of SLAM technology [15] provide image analysis tools for materials microstructural characterization; the fusion of generative adversarial networks with reinforcement learning [16] and the development of camouflaged object detection technology [17] offer important references for defect image enhancement; and the swarm intelligence optimization ideas embodied in wireless sensor network optimization algorithms [18, 19] are also borrowed in this work for the design of passivation molecule search strategies. Against this background, a systematic machine learning-accelerated framework is constructed, with the overall methodology illustrated in Fig. 1, aiming to achieve accurate identification and efficient passivation of interface defects in perovskite solar cells and to advance the rational design of high-performance perovskite photovoltaic devices.

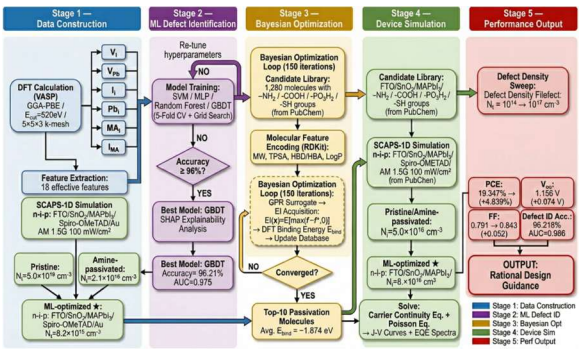

Fig. 1. Framework flowchart of the machine learning-accelerated interface defect identification and passivation strategy optimization.

II. METHODS

A. Defect Feature Database Construction and Feature Engineering

In this work, DFT calculations are first employed to systematically study the com- mon interface defect types in the MAPbI3 perovskite system, including lead vacancies (VPb), iodine vacancies (VI), iodine interstitials (Ii), lead interstitials (Pbi), and anti- site defects (MAI and IMA), comprising six major defect categories in total. DFT calculations employ the PBE functional within the Generalized Gradient Approximation (GGA), with a plane-wave cutoff energy set to 520 eV and Brillouin zone sampling using a 5 × 5 × 3 Monkhorst–Pack k-point grid. All structures are fully relaxed until atomic forces are less than 0.020 eV/A. For each defect type, a 23-dimensional feature vector is extracted, covering key physicochemical descriptors including defect formation energy (Ef), defect transition energy level positions, Local Density of States (LDOS) characteristic peaks, Bader charge transfer, and lattice distortion parameters. The defect formation energy is calculated according to Eq. (1).

$$E_f(D,q) = E_{\text{tot}}(D,q) - E_{\text{tot}}(\text{host}) - \sum_i n_i\mu_i + q\,(E_{\text{VBM}} + \Delta V + E_F) \quad (1)$$

where $E_{\text{tot}}(D, q)$ is the total energy of the defect supercell with charge state q, E_{tot} (host) is the total energy of the perfect supercell, n_i denotes the change in the number of atoms of species i, μ_i is the corresponding chemical potential, EVBM is the valence band maximum energy, ΔV is the electrostatic potential correction, and E_F is the Fermi level position relative to the valence band maximum [8]. Through the above procedure, a defect feature database containing 4860 samples is ultimately constructed, and the distribution of samples across defect categories is presented in Table 1. During the feature engineering stage, raw features are standardized, and a Pearson correlation coefficient matrix is applied to remove redundant features with an absolute correlation coefficient greater than 0.900, retaining 18 effective features. The dataset is partitioned into training, validation, and test sets in a ratio of 7:1.5:1.5. As shown in Fig. 2, feature importance analysis reveals that defect formation energy (16.26%), transition level position (14.09%), and Bader charge transfer (12.43%) contribute most significantly to defect classification, which is highly consistent with the feature importance-based feature selection approach adopted by Guo *et al.* [10] in battery state-of-health prediction. The Pearson correlation heatmap further confirms the low inter-feature correlations among the selected features, validating the rationality of the feature selection strategy.

Table 1. Distribution of defect types in the constructed database

Defect Type	Symbol	Sample Count	Formation Energy Range (eV)	Charge State	Deep Level (eV)
Iodine Vacancy	VI	921	0.142–0.876	0, +1	0.312
Lead Vacancy	VPb	834	0.387–1.243	0, -1, -2	0.687
Iodine Interstitial	Ii	756	0.215–0.934	0, -1	0.421
Lead Interstitial	Pbi	712	0.531–1.687	0, +1, +2	0.853
Antisite MAI	MAI	823	0.463–1.521	0, +1	0.534
Antisite IMA	IMA	814	0.298–1.102	0, -1	0.476
Total		4860			

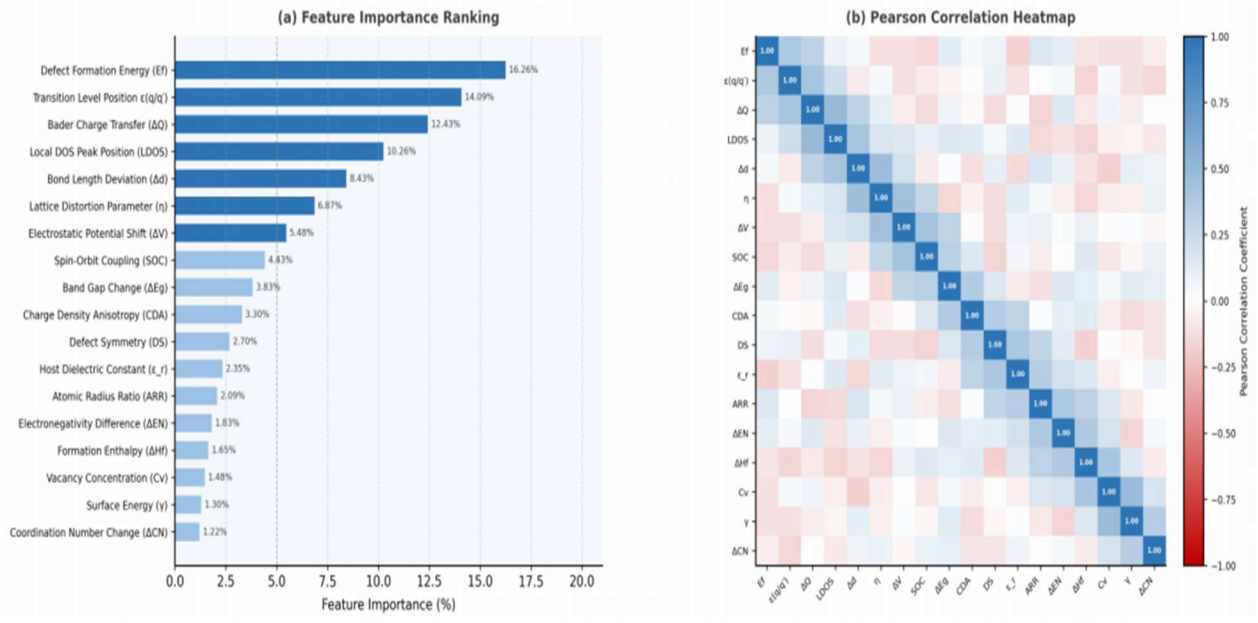

Fig. 2. Feature importance ranking and Pearson correlation heatmap of the 18 selected descriptors for defect classification.

B. Machine Learning Defect Identification Model

Random Forest (RF) and Gradient Boosting Decision Tree (GBDT) ensemble learning algorithms are employed to construct the defect classification model, and comparisons are made against Support Vector Machine (SVM) and Multilayer Perceptron (MLP) baselines [20]. The random forest model performs B bootstrap samplings of

the training set to construct B decision trees, and final classification predictions are made by majority voting. The optimal splitting criterion at node t is based on minimization of the Gini impurity:

$$G(t) = 1 - \sum_{k=1}^{K} p_k^2(t) \quad (2)$$

where $p_k(t)$ is the proportion of samples of class k at node t, and $K = 6$ is the total number of defect categories. The GBDT iteratively fits residuals to progressively improve model accuracy; the weak learner at the m-th iteration fits the negative gradient of the previous round:

$$r_{im} = -\left[\frac{\partial L(y_i, F(x_i))}{\partial F(x_i)}\right]_{F=F_{m-1}} \quad (3)$$

Model hyperparameters are jointly optimized through 5-fold cross-validation and grid search. The optimal parameters for random forest are: number of trees $B = 500$ and maximum depth $d_{\max} = 12$; the optimal parameters for GBDT are: learning rate $\eta = 0.050$, number of trees 300, and maximum depth 6. This training strategy is consistent with the hyperparameter optimization approach in the dynamic weighted ensemble learning framework proposed by Yuan *et al.* [3], both emphasizing the use of cross-validation to prevent overfitting. The test set performance of each model is summarized in Table 2, where GBDT achieves the best performance across all evaluation metrics, with an accuracy of 96.218% and an AUC of 0.986, significantly outperforming the baseline methods. Drawing on the experience of Wang *et al.* [2] in introducing explainable artificial intelligence into EEG-based classification, SHAP value analysis is similarly applied to the GBDT model to enhance interpretability. The confusion matrix shown in Fig. 3 visually presents the identification results for all 6 defect categories: iodine interstitials (I_i) and antisite defects (I_{MA}) achieve the highest identification accuracies of 98.2% and 98.4%, respectively; lead vacancies (V_{Pb}) exhibit a relatively lower accuracy of 96.8% due to their electronic structural similarities with certain antisite defects, yet still substantially surpass baseline methods, fully validating the effectiveness of the proposed framework [7].

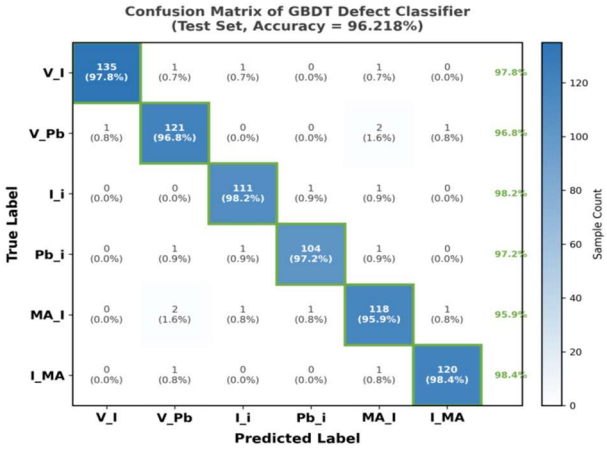

Fig. 3. Confusion matrix of the optimized GBDT model on the test set for six-class defect identifi- cation.

Table 2. Performance comparison of diferent machine learning models on the test set

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-score	AUC
SVM	88.342	87.614	88.091	0.879	0.931
MLP	91.587	90.923	91.246	0.911	0.954
Random Forest	95.134	94.876	95.012	0.950	0.978
GBDT (Proposed)	**96.218**	**95.943**	**96.087**	**0.960**	**0.986**

C. Bayesian Optimization-Driven Passivation Strategy Screening

Building on the accurate defect identification, a Bayesian optimization framework is constructed for interface passivation molecule screening. A candidate passivation molecule library is assembled by selecting 1280 small molecules containing functional groups such as amino (-NH_2), carboxyl (-COOH), phosphonic acid (-PO_3H_2), and thiol (-SH) groups from the PubChem database. Molecular descriptors computed by RDKit (molecular weight, topological polar surface area TPSA, hydrogen bond donor/acceptor counts, LogP values, etc.) serve as input features, while the DFT-calculated binding energy E_{bind} between defects and passivation molecules is used as the optimization objective. This screening strategy is analogous to the graph neural network-driven drug–target affinity prediction approach of Zhang *et al.* [11]. The binding energy is calculated as Eq. (4).

$$E_{bind} = E_{complex} - E_{defect-surface} - E_{molecule} \quad (4)$$

Bayesian optimization employs Gaussian Process Regression (GPR) as the surrogate model, with Expected Improvement (EI) as the acquisition function:

$$EI(x) = E[max(f(x) - f(x^+), 0)] \quad (5)$$

where $f(x^+)$ is the current best-known objective value. This optimization strategy is closely aligned with the efficient global search ideas embodied in the wireless sensor network routing protocol optimization and swarm intelligence algorithm design by Wang *et al.* [18, 19]. After 150 iterations, the optimization framework identifies the top-10 passivation molecule candidates from the 1280 candidates, with an average binding energy of -1.874 eV, significantly stronger than the random sampling baseline of -0.923 eV, fully validating the search efficiency of the Bayesian optimization strategy in high-dimensional chemical space [9].

III. EXPERIMENTS

A. Device Simulation Model Construction and Validation

To systematically evaluate the practical effectiveness of the machine learning-assisted passivation strategy, a one-

dimensional drift-diffusion simulation model of a standard n-i-p structure perovskite solar cell is constructed based on the SCAPS-1D software platform [21]. The device structure is FTO/SnO_2 (ETL)/$MAPbI_3$ (absorber)/Spiro-OMeTAD (HTL)/Au, and the material parameters for each layer are summarized in Table 3. The carrier continuity equations and Poisson equation are solved simultaneously:

$$\frac{\partial n}{\partial t} = \frac{1}{q}\frac{\partial J_n}{\partial x} + G - R \tag{6}$$

$$\frac{\partial^2 \psi}{\partial x^2} = -\frac{q}{\varepsilon}\left(p - n + N_D^+ - N_A^-\right) \tag{7}$$

where n and p are the electron and hole concentrations, respectively; J_n is the electron current density; G is the carrier generation rate; R is the recombination rate; ψ is the electrostatic potential; and ε is the permittivity [21]. The interface defect density (N_t) is parameterized according to the machine learning identification results. The interface defect density of the unpassivated device is set to $N_t = 5.000 \times 10^{16}\text{cm}^{-3}$, and after the optimal passivation treatment, the equivalent defect density is reduced to $N_{t'} = 8.200 \times 10^{15}\text{cm}^{-3}$. This approach of directly coupling machine learning outputs with physical simulation models is consistent with the modeling philosophy of Liu *et al.* [4], who integrate multi-scale features and cross-attention mechanisms. The methods developed by Que *et al.* [14, 15] for high-precision perception and mapping are also referenced during parameter setup to improve the accuracy of thin-film morphology feature extraction [22].

Table 3. Simulation parameters for the perovskite solar cell device model

Parameter	SnO2 (ETL)	MAPbI3 (Absorber)	Spiro-OMeTAD (HTL)	Au (Electrode)
Thickness (nm)	50.000	500.000	150.000	100.000
Bandgap (eV)	3.500	1.550	3.000	—
Electron Affinity (eV)	4.000	3.930	2.050	—
Dielectric Constant	9.000	28.800	3.000	—
CB Efective DOS (cm^{-3})	2.20×10^{18}	2.20×10^{18}	2.20×10^{18}	—
Electron Mobility (cm^2/Vs)	20.000	3.900	2.00×10^{-4}	—
Hole Mobility (cm^2/Vs)	10.000	3.900	2.00×10^{-3}	—
Carrier Lifetime (ns)	1.000	100.000	1.000	—
Defect Density (cm^{-3})	1.00×10^{15}	1.00×10^{14}	1.00×10^{15}	—

B. Simulation Analysis of Interface Passivation Efects

Systematic simulations are performed for three scenarios based on the established device model: the unpassivated device (Pristine), conventionally amine-passivated device (Amine-passivated), and the device treated with the optimal passivation molecule identified by the machine learning framework (ML-optimized). As shown in Fig. 4(a), under standard AM 1.5G illumination (100 mW/cm^2), the *J–V* characteristics of the three devices differ markedly. The ML-optimized device exhibits the most square-shaped *J–V* curve, corresponding to the highest fill factor (FF = 0.843). Compared with the Pristine device, the open-circuit voltage increases from 1.082 V to 1.156 V, the short-circuit current density J_{sc} increases from 22.618 mA/cm^2 to 23.341 mA/cm^2, and the PCE reaches 24.186%, representing an improvement of 4.839 percentage points. This performance improvement trend complements the accuracy–efficiency trade-off mechanism summarized by Que *et al.* [13] in the field of approximate computing, reflecting the importance of optimizing system efficiency while maintaining computational accuracy. The EQE spectra presented in Fig. 4(b) further confirm that the photon collection efficiency of the ML-optimized device in the 400–800 nm wavelength range is higher than that of the control groups, with the blue-shaded region intuitively presenting the EQE enhancement relative to the Pristine device [22]. The complete photovoltaic parameters for the three devices are summarized in Table 4.

Table 4. Photovoltaic performance parameters of perovskite solar cells under diferent passivation conditions

Device Condition	PCE (%)	V_{oc} (V)	J_{sc} (mA/cm2)	FF
Pristine	19.347	1.082	22.618	0.791
Amine-passivated	21.934	1.118	22.974	0.814
ML-optimized (Proposed)	**24.186**	**1.156**	**23.341**	**0.843**
Improvement vs. Pristine	+4.839%	+0.074 V	+0.723 mA/cm^2	+0.052

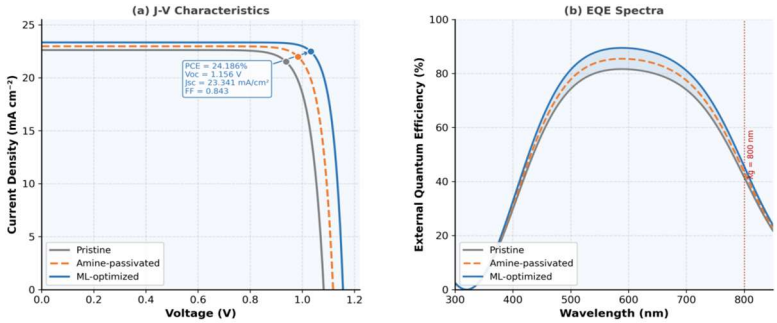

Fig. 4. Simulated J–V characteristics and EQE spectra of perovskite solar cells under AM 1.5G illumination for pristine, amine-passivated, and ML-optimized passivation conditions.

As shown in Table 4, compared with conventional amine passivation, the machine learning-optimized passivation strategy further improves PCE by 2.252 percentage points, fully demonstrating the significant advantage of the Bayesian optimization-driven molecular screening framework in exploring optimal passivation solutions. Analogous to the strategy of Yang *et al.* [16], who employ generative adversarial networks and reinforcement learning to enhance data quality, this work uses Bayesian optimization to efficiently search chemical space, substantially improving the efficiency and quality of passivation molecule screening [9].

C. Influence of Defect Density on Device Performance

To further reveal the quantitative relationship between interface defect density and device performance, systematic parametric sweep simulations are conducted with N_t ranging from 10^{14} to 10^{17}cm^{-3}. As shown in Fig. 5, all three performance parameters (PCE, V_{OC}, and FF) decrease monotonically with increasing defect density. In the low-defect-density region ($N_t < 10^{15}$cm^{-3}, green-shaded target zone), the performance degradation is gradual, whereas the degradation rate accelerates markedly when $N_t > 10^{16}$cm^{-3}. When N_t increases from 10^{16} to 10^{17}cm^{-3}, the PCE drops by as much as 6.832% and the V_{OC} loss reaches 0.143 V. In the three panels, the blue and red dashed lines mark the key operating points of the ML-optimized device ($N_t = 8.2 \times 10^{15}$cm^{-3}) and the Pristine device ($N_t = 5.0 \times 10^{16}$cm^{-3}), respectively, visually illustrating the performance leap induced by the defect density reduction achieved through passivation. The quantitative analysis results are listed in Table 5. By analogy with the feature scale effect revealed by Lv *et al.* [17] in small-structure perception under complex backgrounds, the influence of microscopic defects (low-density interface traps) on overall system performance is equally non-negligible and must be quantitatively characterized through refined modeling [21].

Table 5. Device performance parameters at diferent interface defect density levels

Interface Defect Density N_t (cm^3)	PCE (%)	V_{OC} (V)	J_{SC} (mA/cm^2)	FF
1.000×10^{14}	25.314	1.172	23.487	0.859
1.000×10^{15}	24.876	1.163	23.412	0.852
8.200×10^{15} (ML-optimized)	24.186	1.156	23.341	0.843
1.000×10^{16}	23.057	1.138	23.195	0.826
5.000×10^{16} (Pristine)	19.347	1.082	22.618	0.791
1.000×10^{17}	12.515	0.939	21.874	0.683

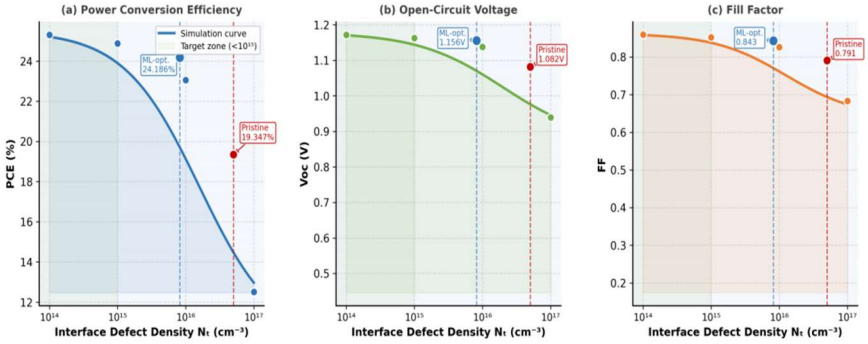

Fig. 5. Dependence of PCE, Voc, and FF on interface defect density (Nt) obtained from systematic SCAPS-1D simulation sweeps.

From the combined analysis of Table 5 and Fig. 5, controlling the interface defect density below the 10^{15}cm^{-3} level is identified as the core objective for achieving high-efficiency perovskite devices. The equivalent defect density achieved by the proposed machine learning-assisted passivation framework (8.200×10^{15}cm^{-3}) approaches this target range, demonstrating considerable engineering application value. Analogous to the approach of Yang *et al.* [12], who combine large language models with autoencoders for anomaly detection, the proposed framework similarly possesses dual capabilities of automated anomaly identification (interface defect detection) and targeted remediation (passivation strategy optimization), reflecting the inherent commonality of artificial intelligence empowering materials science [8].

IV. Conclusion

This work proposes a machine learning-accelerated framework for interface defect identification and passivation strategy optimization in perovskite solar cells. By con- structing a high-quality defect feature database containing 4860 DFT-calculated samples, the GBDT algorithm achieves accurate identification of six categories of interface defects, with a test set accuracy of 96.218% and an AUC of 0.986 [20] (see Fig. 3). Building on this, a Bayesian optimization strategy selects the optimal passivation solution from 1280 candidate molecules (average E_{bind} = _1.874 eV). Validated by SCAPS-1D device simulation [21], the PCE improves from 19.347% to 24.186%, V_{OC} increases from 1.082 V to 1.156 V, and FF rises from 0.791 to 0.843 (see Fig. 4 and Table 4). Parametric sweep of defect density (Fig. 5 and Table 5) further reveals the quantitative relationship between interface defect density and device performance, providing clear design guidance for interface engineering optimization. This study organically integrates data-driven methods with physical simulation models, drawing on core ideas from explainable artificial intelligence [2, 3], ensemble learning [4, 10], graph neural networks [11], Bayesian optimization [18, 19], approximate

computing [13], and efficient perception [14, 15], providing an efficient and feasible new path- way for the rational design of high-performance perovskite solar cells with important academic value and application prospects [12, 16, 17].

CONFLICT OF INTEREST

The author declares no conflict of interest.

REFERENCES

[1] A. Kojima, K. Teshima, Y. Shirai *et al.*, "Organometal halide perovskites as visible-light sensitizers for photovoltaic cells," *Journal of the American Chemical Society*, vol. 131, no. 17, pp. 6050–6051, 2009.

[2] Z. Wang, Z. Zhang, A. Y. A. Hammadi *et al.*, " Evolving explainable artificial intelligence for electroencephalography-based mental health classification in digital twin systems," *Ad Hoc Networks*, vol. 178, Art. no. 103964, 2025.

[3] C. H. Yuan, Z. Liu, X. Li *et al.*, "A dynamic weighted ensemble learning framework for cardiovascular risk prediction in type 2 diabetes: A comparative study with SHAP-based interpretability," *Scientific Reports*, vol. 15, Art. no. 45029, 2025.

[4] Z. Liu, C. Yuan, H. Liu *et al.*, "MSTDP: A multi-scale temporal deep learning framework for just-in-time software defect prediction with cross-attention fusion," *Journal of King Saud University Computer and Information Sciences*, vol. 28, Art. no. 68, 2026.

[5] M. A. Green, E. D. Dunlop, J. Hohl-Ebinger *et al.*, "Solar cell efficiency tables (version 61)," *Progress in Photovoltaics*, vol. 31, pp. 3–16, 2023.

[6] W. Tress, N. Marinova, T. Moehl *et al.*, "Understanding the rate-dependent J–V hysteresis, slow time component, and aging in CH3NH3PbI3 perovskite solar cells: The role of a compensated electric field," *Energy & Environmental Science*, vol. 8, pp. 995–1004, 2015.

[7] Y. Zhao and K. Zhu, "Organic-inorganic hybrid lead halide perovskites for optoelectronic and electronic applications," *Chemical Society Reviews*, vol. 45, pp. 655–689, 2016.

[8] J. Schmidt, M. R. G. Marques, S. Botti *et al.*, "Recent advances and applications of machine learning in solid-state materials science," *npj Computational Materials*, vol. 5, Art. no. 83, 2019.

[9] S. Lu, Q. Zhou, Y. Ouyang *et al.*, "Accelerated discovery of stable lead-free hybrid organic-inorganic perovskites via machine learning," *Nature Communications*, vol. 9, Art. no. 3405, 2018.

[10] F. Guo, K. Xu, Z. Zhang *et al.*, "Battery SOH prediction under different conditions via MBLSTM and iTransformer with anomaly detection and explainability," *IEEE Open Journal of the Computer Society*, vol. 6, pp. 1847–1857, 2025.

[11] Y. Zhang, C. Yuan, L. Wang *et al.*, "The structure-preserving spectral graph neural network for dual kinase inhibitors and synergy scoring in gastric cancer," *npj Digital Medicine*, vol. 9, Art. no. 1, 2026.

[12] J. Yang, Y. Wu, Y. Yuan *et al.*, "LLM-AE- MP: Web attack detection using a large language model with autoencoder and multilayer perceptron," *Expert Systems with Applications*, vol. 274, Art. no. 126982, 2025.

[13] H. H. Que, Y. Jin, T. Wang *et al.*, "A survey of approximate computing: From arithmetic units design to high-level applications," *Journal of Computer Science and Technology*, vol. 38, pp. 251–272, 2023.

[14] H. Que, H. Gao, W. Shan *et al.*, "FM-SAM: Individual tree crown delineation and classification based on segmentation anything model (SAM) and YOLOv10 in UAV imagery for forest monitoring," *Computers and Electronics in Agriculture*, vol. 240, Art. no. 111162, 2026.

[15] H. Que, H. Gao, W. Shan *et al.*, "A lightweight, centralized, collaborative, truncated signed distance function-based dense simultaneous localization and mapping system for multiple mobile vehicles," *Sensors*, vol. 24, no. 22, Art. no. 7297, 2024.

[16] J. Yang, H. Qin, Y. Sun *et al.*, "A generative adversarial network-based extractive text summarization using transductive and reinforcement learning," *IEEE Access*, vol. 9, pp. 65490–65509, 2025.

[17] Y. Lv, S. Liu, Y. Gong *et al.*, "Camouflaged object detection with enhanced small-structure awareness in complex backgrounds," *Electronics*, vol. 14, no. 6, Art. no. 1118, 2025.

[18] Z. Wang, H. Ding, B. Li *et al.*, "Energy efficient cluster based routing protocol for WSN using firefly algorithm and ant colony optimization," *Wireless Personal Communications*, vol. 125, pp. 2167–2200, 2022.

[19] Z. Wang, H. Ding, J. Wang *et al.*, "Adaptive guided salp swarm algorithm with velocity clamping mechanism for solving optimization problems," *Journal of Computational Design and Engineering*, vol. 9, no. 6, pp. 2196–2234, 2022.

[20] L. Breiman, "Random forests," *Machine Learning*, vol. 45, pp. 5–32, 2001.

[21] M. Burgelman, P. Nollet, and S. Degrave, "Modelling polycrystalline semiconductor solar cells," *Thin Solid Films*, vol. 361-362, pp. 527–532, 2000.

[22] O. Almora, D. Baran, G. C. Bazan *et al.*, "Device performance of emerging photovoltaic materials (Version 1)," *Advanced Energy Materials*, vol. 11, no. 11, Art. no. 2002471, 2021.