Progress in Photovoltaics: Research and Applications

WILEY

PROGRESS IN PHOTOVOLTAICS

SPECIAL ISSUE REVIEW

Machine Learning for Stability Enhancement in Perovskite Solar Cells: A Pathway to Commercial Viability

Shen Wang[1,2] | Weiren Zhao[3] | Minjia Zhou[1] | Tanghao Liu[4] | Yi'an Wang[4] | Run Shi[5] | Yunfan Wang[6] | Zhuoqiong Zhang[7]

[1]Faculty of Science, Civil Aviation Flight University of China, Guanghan, China | [2]Leicester International Institute, Dalian University of Technology, Panjin, China | [3]School of Chemical Engineering, Ocean and Life Sciences, Dalian University of Technology, Panjin, China | [4]School of Physical Sciences, Great Bay University, Dongguan, China | [5]School of Materials Science and Engineering, Tsinghua University, Beijing, China | [6]Department of Materials Science and Engineering, City University of Hong Kong, Hong Kong, SAR, China | [7]Department of Physics, Hong Kong Baptist University, Hong Kong, SAR, China

Correspondence: Run Shi (shirun@tsinghua.edu.cn) | Yunfan Wang (yunfawang2-c@my.cityu.edu.hk) | Zhuoqiong Zhang (jonez_zhang@126.com)

Received: 22 April 2025 | **Revised:** 25 July 2025 | **Accepted:** 9 December 2025

Keywords: data-driven | machine learning | perovskite solar cells | stability

ABSTRACT

Organic–inorganic hybrid perovskites hold significant promise for low-cost, high-efficiency, and scalable photovoltaic production. However, stability challenges of perovskite materials hinder their commercial viability. Although significant progress in enhancing stability has been achieved through compositional adjustments, additive engineering, and solvent-based processing strategies, these methods often involve laborious and time-consuming optimization processes. Machine learning (ML) is proving highly effective for accelerating the development and optimization of stable perovskite materials, reducing reliance on trial-and-error methods. This review outlines the fundamental ML workflow and highlights its applications in material screening, mechanism investigation, and characterization analysis in perovskite solar cells (PSCs) research. These key advancements underscore the utility of ML in systematically improving the durability of PSCs. Future integration of ML with high-throughput experimentation is expected to further advance the development of efficient, stable, and commercially viable PSCs, contributing to sustainable energy solutions.

1 | Introduction

Perovskite solar cells (PSCs) have rapidly become a leading candidate among next-generation photovoltaic technologies, owing to their remarkable power conversion efficiency (PCE), cost-effective fabrication processes, and scalability for scalable production [1–5]. Since the first solid-state PSCs were reported in 2009, their PCE has surged from an initial 3.8% to an impressive 27.0%, rivaling commercialized silicon-based solar cells [6]. This extraordinary progress can be attributed to the exceptional optoelectronic properties of perovskite materials, including their high absorption coefficients, tunable bandgaps, long carrier diffusion lengths, and minimal non-radiative recombination [7–9]. However, despite these advantages, the long-term stability of PSCs remains a significant challenge, limiting their path to commercial viability.

The stability issues of PSCs can be broadly divided into two categories: those related to the carrier transport layers and those originating from the functional layer. Among these, the functional layer—the perovskite material itself—lies at the core of stability concerns. Perovskite materials are highly sensitive to environmental factors such as moisture, oxygen, light, and heat, all of which accelerate degradation and lead

Shen Wang, Weiren Zhao, and Minjia Zhou authors contributed equally to the study.

to performance losses [10–14]. For instance, exposure to moisture triggers decomposition of perovskites into their precursor components, causing irreversible damage. In addition to environmental instability, PSCs face intrinsic structural instability. Perovskite materials typically follow the ABX_3 formula, where "A" is a monovalent cation (e.g., methylammonium [MA^+], formamidinium [FA^+], or cesium [Cs^+]), "B" is a divalent metal cation (commonly Pb^{2+} or Sn^{2+}), and "X" is a halide anion (Cl^-, Br^-, or I^-). This structure is composed of a three-dimensional network of corner-sharing BX_6 octahedra, with the A cation occupying the voids. The structural stability of ABX_3 perovskites depends on the ionic composition and tolerance factor, which dictate whether a stable perovskite phase forms. For instance, although $FAPbI_3$ is highly desirable due to its near-optimal bandgap, the relatively large size of FA cations destabilizes the PbI_6 framework, resulting in phase instability and the formation of undesirable non-perovskite phases [5, 15, 16].

To address these challenges, researchers have developed various strategies. Encapsulation techniques, such as polymeric barriers and glass coatings, effectively shield PSCs from moisture and oxygen [17, 18]. Composition engineering, including the incorporation of inorganic cations (e.g., Cs^+) or halides (e.g., Br^-), enhances thermal and photo-stability by reinforcing the perovskite lattice [19–21]. Additive engineering, which involves using passivation agents, reduces defect density, and suppresses ion migration [22, 23]. Similarly, solvent engineering during film formation has been shown to enhance perovskite film quality and stability [24–26]. Despite these advancements, the process of identifying suitable composition, solvents, and additives remains labor-intensive and time-consuming. The vast chemical space of potential candidates requires extensive trial-and-error experimentation, significantly slowing the pace of discovery. The application of machine learning (ML) offers a powerful tool to accelerate this process. Leveraging large datasets and advanced algorithms, ML can identify patterns, predict material properties, and guide experimental design, thereby significantly reducing the time and effort needed for material optimization.

This review details the strategic application of predictive ML techniques to address the critical stability challenges of PSCs. While emphasizing the contributions of ML to enhancing durability, we also consider how advancements in other performance metrics, particularly PCE, can yield insights relevant to stability improvement. Key elements of the predictive ML workflows—including data collection, feature selection, model training, and prediction—are discussed alongside illustrative case studies. These examples showcase the success of ML in improving device longevity and elucidating degradation mechanisms, highlighting its potential to accelerate the development of efficient, stable, and commercially viable PSCs.

2 | The Workflow of the Predictive ML Technique in PSCs Research

In recent years, ML has emerged as a transformative tool in the field of PSCs, driving advancements in material discovery, device performance optimization, and stability enhancement. The growing significance of ML in this domain is evidenced by the rapid increase in published articles and the rising average impact factor of these studies (Figure 1), highlighting the critical role of ML-driven research in advancing PSC technologies. The typical ML workflow in perovskite research consists of four key steps: (1) data preparation, (2) feature selection, (3) model training, and (4) prediction and validation (Figure 2). The following sections will delve deeper into each of these steps, highlighting their significance and illustrating how they contribute to the development of perovskite materials and devices.

2.1 | Data Preparation

In ML-assisted materials study, the input data typically comprises the features (material descriptors) and the labels (stability) [27]. The features represent the chemical information or structural properties of the material, such as cell parameters or chemical composition. The labels, which are the target variables for model learning, stand for the material performances, such as stability that are affected by features.

Datasets for ML in perovskite research can be constructed through various approaches: (1) material databases, such as the Materials Project [28], Inorganic Crystal Structure Database [29], Cambridge Structural Database [30], and Open Quantum Materials [31], provide extensive chemical data across a wide range of materials. When using these

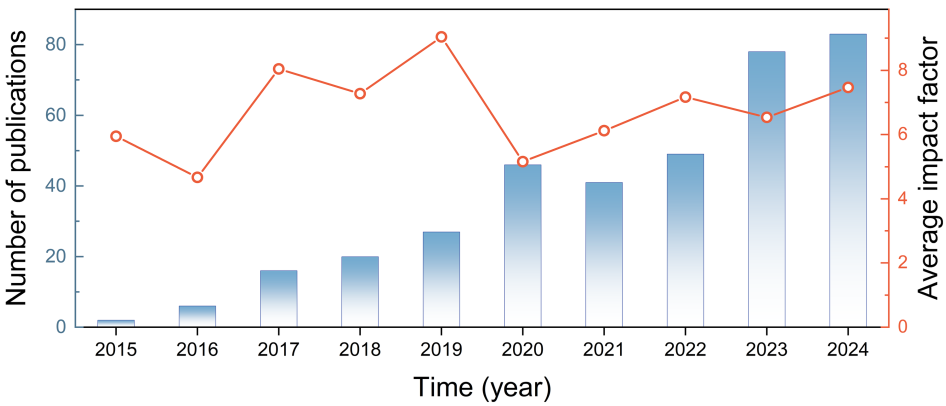

FIGURE 1 | The number of publications (left) on ML applied to PSCs research and the average impact factor (right) of publishing journals over time.

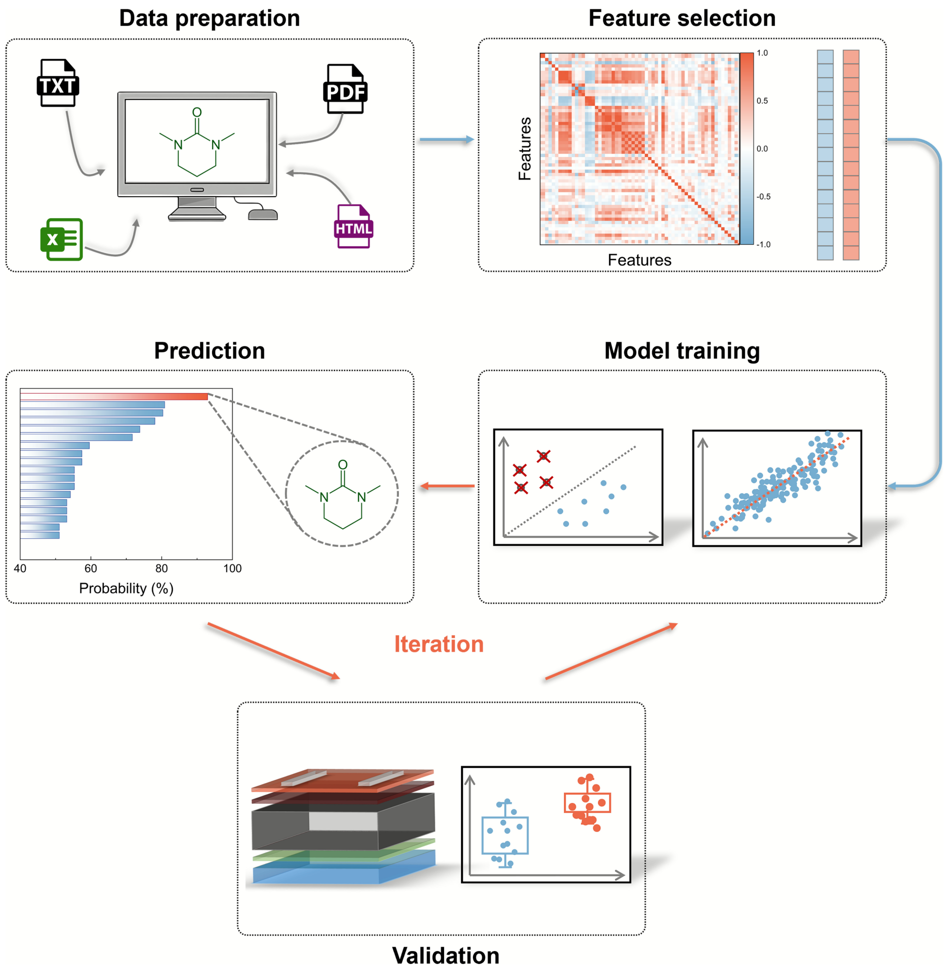

FIGURE 2 | The workflow of ML in PSCs research.

large-scale databases in specific studies, appropriate data screening and cleaning processes are crucial; (2) alternatively, high-throughput experimentation generates high-quality data under controlled conditions but is often costly and resource-intensive [32, 33]; (3) in addition, data extraction from published literature provides another avenue [34–36], though inconsistencies in experimental conditions can introduce bias and data gaps [37]; (4) high-throughput computing, employing quantum chemistry (QC) methods like density functional theory (DFT), efficiently predicts material properties and generates large datasets [38–41], though computational results may not always fully reflect real-world material performance due to the differences in synthesis methods and limitations in DFT calculations, which struggle with weak interactions, point defects, and usually underestimates band gaps in perovskite. By combining these methods, researchers can construct robust and high-quality datasets, advancing ML applications in perovskite research. The reliability of ML techniques in improving PSCs relies on the quality of the data [42], which must be relevant, complete, reliable, and uniformly distributed. Therefore, rigorous data preprocessing is essential [43, 44]. The detailed information on constructing datasets for ML is presented in Table S1. Addressing missing values and outliers is essential to maintain data integrity, while normalization methods, such as *Z*-score normalization and min–max normalization, can improve the accuracy and efficiency of model training [45].

2.2 | Feature Selection

Features or descriptors are attributes used to capture key patterns in data and serve as the fundamental units of input information for ML models. Their selection directly influences the predictive performance of models and the interpretability of material stability. Thus, selected features should be unique, computationally efficient, and sufficient to describe and distinguish the structures of materials in PSCs research, while retaining clear physical and chemical significance [46, 47].

In general, the descriptors in PSCs research can be categorized into those for the organic and inorganic components [48, 49], as illustrated in Figure 3a. Organic descriptors include molecular fingerprints (1D) [49], SMILES strings (1D) [50], physicochemical parameters (1D) [51], and UV/IR spectral images (2D) [52], which are commonly used for organic cations or preparation solvents screening [53, 54]. In contrast, the inorganic component is more challenging to describe due to its extended structures and inherent defects. Local descriptors such as DFT-derived lattice parameters cannot represent the long-range dependencies or point defects of the crystal, complicating modeling efforts. Available descriptors for inorganic components include physicochemical parameters (1D) [55, 56], XRD spectral images (2D) [57, 58], and crystal graphs (2D) [59], where nodes and edges are represented by vectors corresponding to the atoms and bonds, respectively. Most

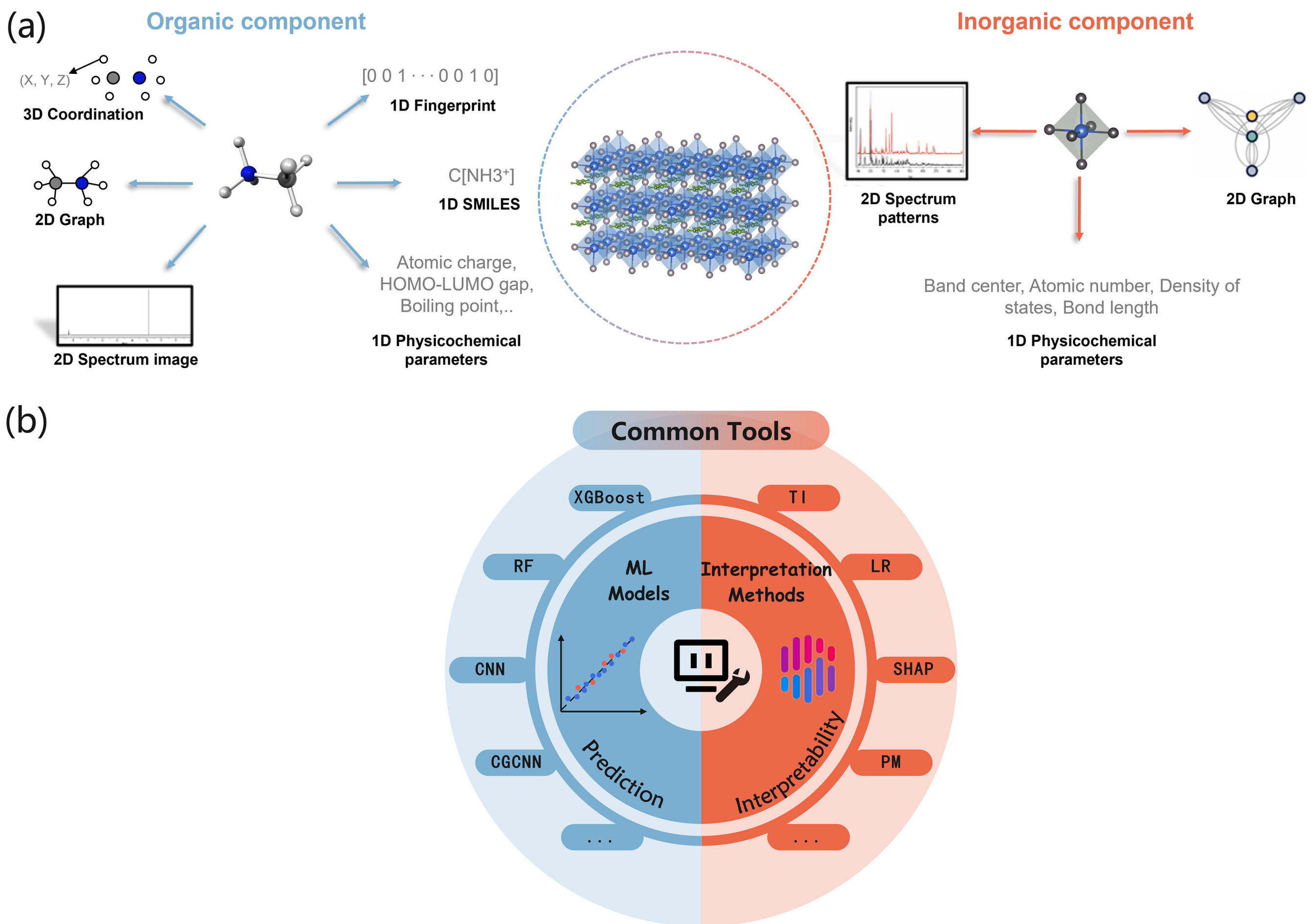

FIGURE 3 | (a) The organic (left) and inorganic (right) components of the descriptors in perovskite research. (b) Commonly used machine learning models (left) and interpretation methods (right).

studies focus on physicochemical parameters derived from QC calculations (the detailed information of descriptors is presented in Table S2). For example, Huang et al. developed a descriptor for perovskite stability using large language models (LLMs) and QC calculations [60]. However, the high computational cost of QC calculations limits its scalability for exploring novel materials. Yang et al. constructed a stability descriptor for A_2BX_6-type perovskites using easily obtainable features to tackle this dilemma, though the predictive performance of the model could be further enhanced [61]. To address these challenges, the development of efficient, structure-based molecular fingerprint descriptors for inorganic materials, enabling faster and more cost-effective exploration of PSC stability, is advocated.

2.3 | Model Training

Supervised learning, where ML models are trained on labeled data, is widely used in perovskite research for regression and classification tasks. In regression, ML models predict decomposition energies to identify high-stability photovoltaic materials [62]. Performance is assessed using metrics such as the coefficient of determination (R^2), mean absolute error (*MAE*), and root mean squared error (*RMSE*), where higher R^2 (closer to 1) and lower *MAE*/*RMSE* indicate better accuracy. In classification, models categorize materials into high or low stability based on thermodynamic properties [63], with accuracy and the area under the curve (*AUC*) serving as common performance indicators. The choice of an ML model is crucial for accurately predicting material properties. Typically, datasets are divided into train and test sets using scaling or cross-validation methods [64]. The model learns data features from the train set and evaluates its performance on the test set. To identify the most suitable model for discovering high-potential materials, multiple models are compared. As illustrated in the left part of Figure 3b, ML models can be broadly categorized into traditional ML models and deep learning models. Traditional ML models, such as XGBoost, rely on statistical methods [38, 65], offering computational efficiency but depending heavily on the quality of descriptors. For instance, Chen et al. demonstrated that a well-trained XGBoost model could accurately predict long-term aging curves exceeding 3000 h [66]. In contrast, deep learning models, including artificial neural networks (ANNs) and crystal graph convolutional neural networks (CGCNNs), excel at processing high-dimensional data such as images and graphs [67]. Kim et al. employed CGCNN to identify novel multi-element metal halide perovskites (MHPs), suggesting two potential high-performance materials [68]. In addition, transfer learning enables deep learning models to leverage large-scale datasets, such as those extracted from literature, which traditional ML models struggle to handle [69, 70]. However, training of deep learning models typically requires higher computational resources and greater technical expertise. Many areas of the broader field experience a surge in the development of predictive models, aiming at achieving higher accuracy in prediction and placing emphasis on the model's generalization ability, that is, to handle data outside the utilized dataset.

Interpretation tools, as illustrated on the right side of Figure 3b, play a crucial role in understanding model predictions. Model-based methods, such as random forests (RF) and decision trees, visualize decision-making processes using tree importance (TI), while linear regression (LR) provides interpretable formulas linking features to outcomes [71]. This makes LR particularly useful for developing perovskite stability descriptors, though its accuracy is often insufficient for complex prediction tasks. For black-box models like deep learning, interpretation tools such as Shapley additive explanations (SHAP) and the perturbation method (PM) are widely applied. These tools analyze trained models to assess the importance of features or functional groups [72, 73]. For instance, Subudhi et al. used SHAP to identify critical factors influencing PSC performance, enabling improvements in stability and efficiency [74]. Such interpretation methods are increasingly valued for their ability to balance model accuracy with explainability. However, explanatory methods based on probability statistics are contingent on the features used in material modeling, resulting in the interpretive results sometimes distorting our ability to judge the chemical essence of the substance. Thus, it is highly recommended to combine QC calculations to gain a deeper insight.

2.4 | Prediction and Validation

Prediction and experimental validation are pivotal steps in the ML workflow for PSCs research. Experimental validation not only assesses the reliability of ML predictions but also provides critical feedback for model refinement. By integrating experimental results into the training process, the model can be iteratively optimized, enhancing its predictive accuracy and robustness. This feedback loop allows the ML model to evolve alongside an expanding experimental dataset, enabling precise property predictions and accelerating the discovery of high-performance materials with increased efficiency.

In summary, the ML workflow in PSCs research offers a systematic framework for accelerating the discovery and optimization of high-performance materials. By following key steps—data preparation, feature selection, model training, experimental validation, and iterative optimization—ML empowers researchers to extract valuable insights from complex datasets and make accurate predictions about material properties. Each step is interconnected, with experimental validation serving as a vital feedback mechanism to improve model performance. This structured approach not only enhances the efficiency of material screening but also establishes a robust foundation for integrating ML into advanced materials science. The workflow's adaptability and versatility make it an indispensable tool for addressing the challenges of modern materials research.

3 | Application of ML Techniques in PSCs Research

After showing the general workflow of ML in materials research, in the subsequent section, a detailed review of case studies where ML methodologies have been effectively implemented in PSCs research is presented.

3.1 | ML Research for Selecting and Designing Materials in PSCs Research

The vast compositional space and complex interplay of parameters in perovskite material design render traditional trial-and-error approaches impractical, highlighting the necessity of ML for efficiently identifying optimal formulations [55, 75]. For instance, in 2023, Zhu et al. developed a comprehensive PSC database using QC calculations. High-throughput screening, guided by trained ML models, was performed to predict material stability, band gaps, and photovoltaic efficiency (Figure 4a). From 177,264 candidates, 434 compounds were selected. Among these, the spectroscopic limited maximum efficiency (SLME) values of the screened PSCs exceeded 20%, with four perovskite cells achieving efficiencies above 23% (Figure 4b) [76]. This study provided critical guidance for developing stable PSCs through data-driven screening. Similarly, in 2024, Zhang et al. systematically screened 60 perovskite compositions. Through iterative ML-guided optimization, the experimental workload was significantly reduced, culminating in a record-breaking PCE of 25.75%, along with outstanding thermal and photochemical stability (Figure 4c) [77]. Notably, the PSCs retained 94% of their initial efficiency after 1920 h of continuous operation in a nitrogen atmosphere at 65°C. This work not only set new performance benchmarks for PSCs but also demonstrated the transformative potential of ML in accelerating photovoltaic innovations. Compared to traditional trial-and-error approaches, ML-assisted virtual screening is both cost- and labor-efficient, making it a promising mainstream method for materials exploration in future laboratories.

In terms of materials design, combining material stability predictors with essential factors enables researchers to design novel materials more effectively [78, 79]. For example, bandgap tuning often results in compositional mixtures that cause severe phase segregation in perovskites, compromising solar cell stability. Luo et al. addressed this challenge by identifying optimal compositions to suppress phase segregation and enhance PSC stability [80]. They constructed a global phase diagram of hybrid perovskites, integrated with bandgap maps, and trained a neural network model to predict the most stable perovskite compositions. The workflow offers a systematic approach to optimizing perovskite materials for improved stability. Additionally, Zhang et al. employed ML for buried-interface engineering in perovskite materials [81]. A deep neural network model was trained using Klekota–Roth fingerprints (KRFPs) and physicochemical descriptors (Figure 5a,b). Correlation analysis identified critical features and key molecular fragments, guiding the design of novel molecules. These newly designed molecules significantly improved the PCE and long-term stability of PSCs, retaining 91.24% of the initial PCE after 1200 h of continuous maximum

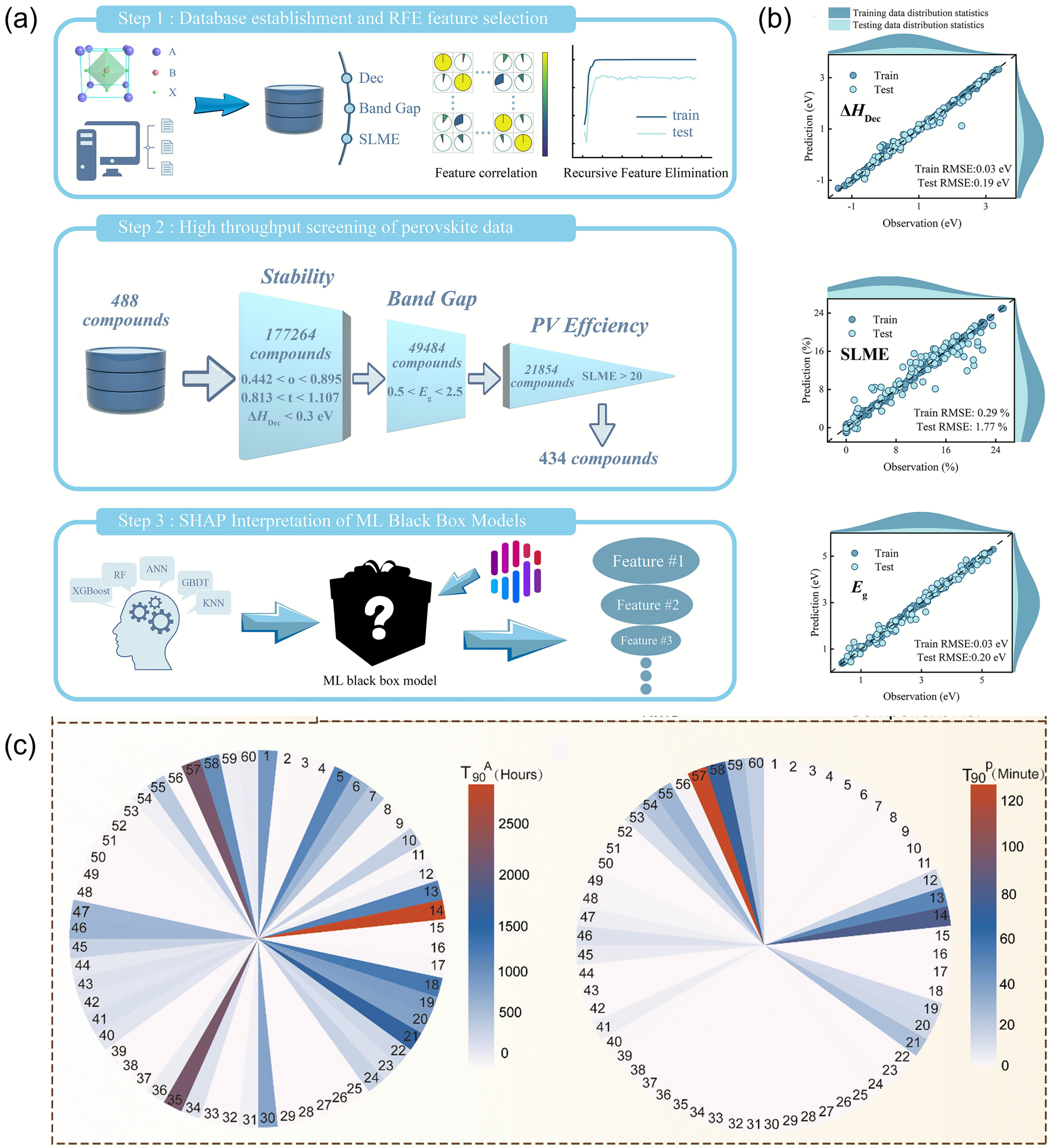

FIGURE 4 | (a) The flowchart of the high-throughput screening guided by trained ML models [76]. (b) Predictive scores of different ML models [76]. Reproduced with permission [76]. Copyright 2024 Elsevier. (c) Colormap representation of the lifetime of 60 different perovskite samples subjected to aging at 50°C under a nitrogen atmosphere (left) and 30°C, 40%–50% RH (right) using MPP tracking [77]. Reproduced with permission [77]. Copyright 2024 Royal Society of Chemistry.

power point tracking (Figure 5c). Furthermore, when integrated with high-throughput laboratories, ML can provide more effective support for materials design. For instance, in 2024, Wu et al. demonstrated a closed-loop workflow to identify new hole-transporting materials tailored for PSCs. By synthesizing 149 molecules, they identified a series of high-performance candidates, achieving PSCs with a certified PCE of 25.9% [82].

3.2 | ML Research for Mechanism Investigation in PSCs Research

ML can be employed to investigate the mechanisms underlying the improved performance of selected or designed materials in PSCs, enabling targeted problem-solving. By leveraging interpretation tools to analyze important features and functional

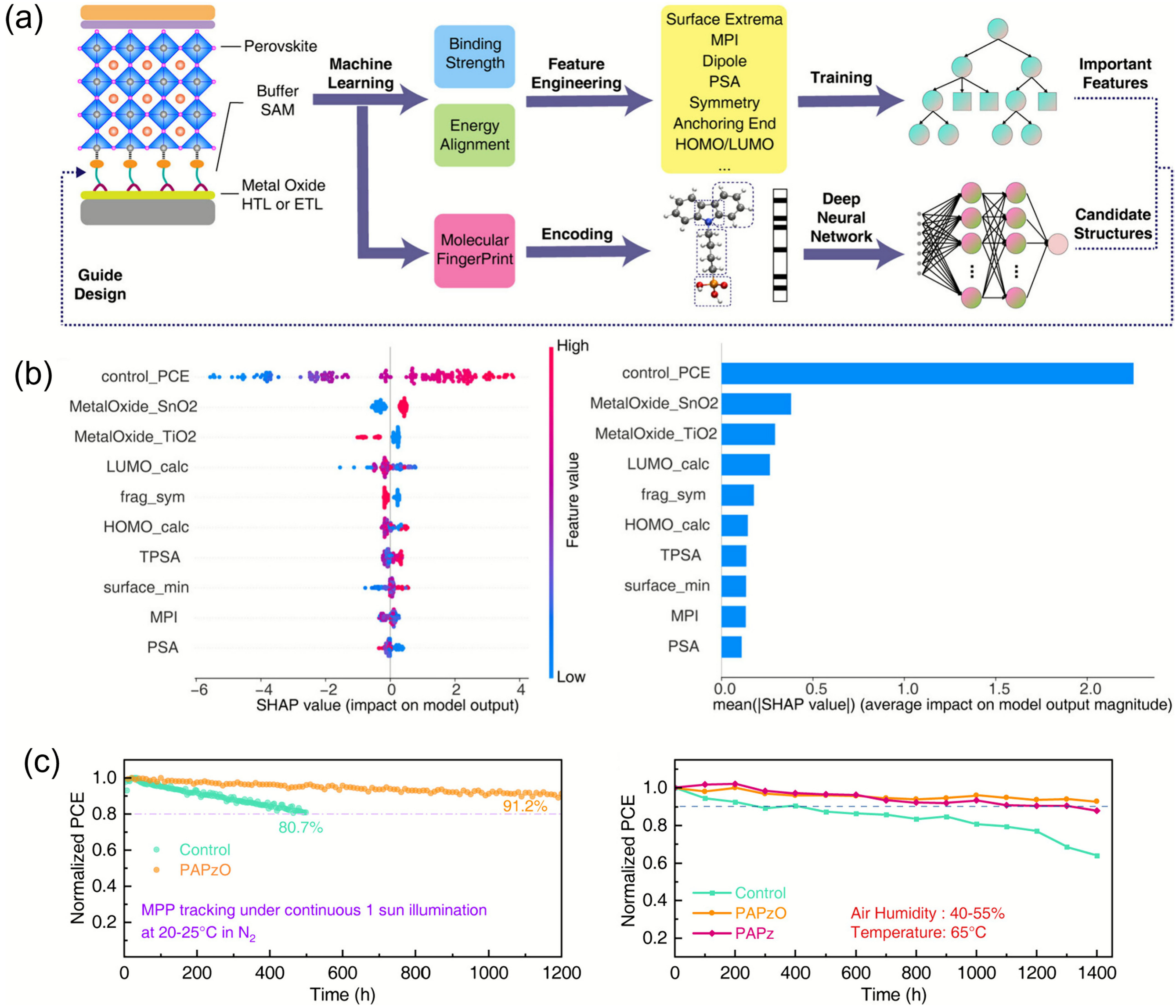

FIGURE 5 | (a) Workflow of the machine-learning process with combinations of features and fingerprints. (b) Important features from the SHAP analysis on the XGBoost model. (c) The stability result of PSCs treated by the ML-designed material [81]. Reproduced with permission [81]. Copyright 2024 American Chemical Society.

groups, researchers can identify the key factors influencing the long-term stability and other performance of PSCs, providing valuable insights for the development of more stable and high-efficiency PSC devices [83, 84].

Interface passivation using an ammonium salt can effectively improve the performance and stability of PSCs. Despite significant PCE improvement achieved in previous studies, the selection criteria for ammonium salts are not fully understood. In 2023, Zhi et al. applied an ML method to investigate the relationship between the molecular features of ammonium salts and the PCE improvement of PSCs. They established an ML model using an experimental dataset of 19 salts to predict the PCE improvement after passivation. Three molecular features (hydrogen bond donor, hydrogen atom, and octane–water partition coefficient) are identified as the most important features of selecting an ammonium salt for passivation (Figure 6a) [85]. This work cleared the mechanism of ammonium salt, which is beneficial to establish the selection criteria of ammonium salts for interface passivation. Additionally, in 2024, Wang et al. utilized genetic programming and extra-trees ML models to identify critical molecular features that decouple multi-molecule contributions in perovskite systems. Combined with DFT calculations, they revealed that stabilized interfacial structures arise from multiple hydrogen bonds and anion··π surface interactions, offering valuable insights into improving the water stability of perovskites [86]. Similarly, they integrated high-throughput experimentation with ML models to investigate the effect of multiple cations on perovskite stability across a wide aging temperature range. Their findings showed that at high aging temperatures, increasing organic cations (e.g., MA), or decreasing inorganic cations (e.g., Cs) in multi-cation perovskites negatively impacts photo- and thermal-stability. However, below 100°C, the trend is reversed (Figure 6b). Consequently, they demonstrated that MA-containing PSCs maintained negligible efficiency loss after 1800 h of operation under illumination at 30°C (Figure 6c) [32].

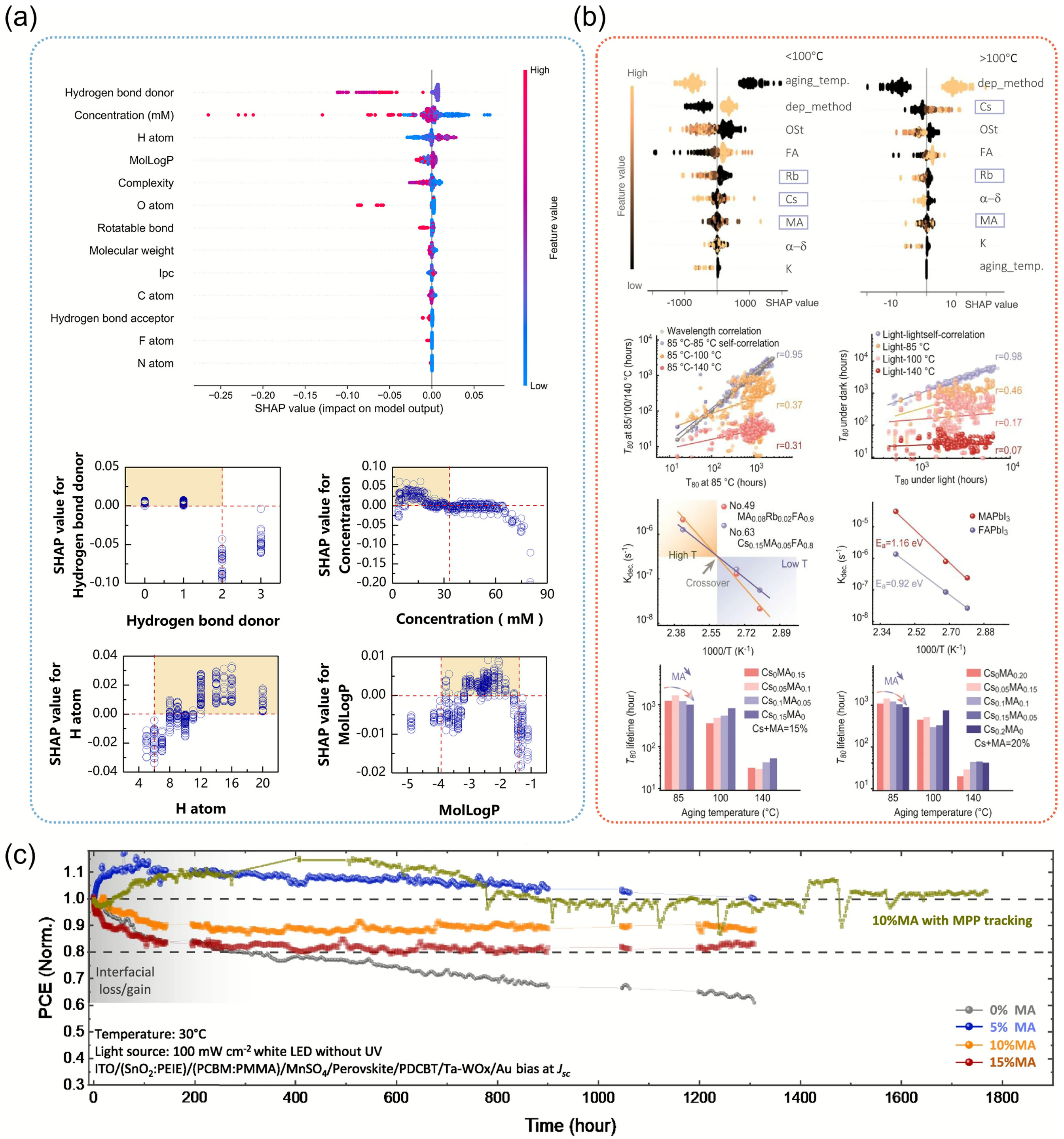

FIGURE 6 | (a) SHAP analysis of the ML ensemble regression model prediction, including the importance ranking of different features and the relationship between the important descriptors of the ensemble model and their SHAP values [85]. Reproduced with permission [85]. Copyright 2023 American Chemical Society. (b) Lifetime analysis of the 64 compositions for mixed-cation lead iodide perovskites [32]. (c) Long-term stability for $Cs_xMA_{0.15-x}FA_{0.85}PbI_3$ (x = 0%, 5%, 10%, and 15%) perovskite solar cells tested under 100 mW cm^{-2} white LED illumination [32]. Reproduced with permission [32]. Copyright 2021 Springer Nature.

3.3 | ML Research Based on Characterization

Characterization techniques are commonly employed to uncover the underlying mechanisms behind device performance improvements [87–89]. However, various subtle and complex correlations remain concealed within the data. ML provides a powerful approach to unravel these intricate relationships, providing deeper insights into the factors that affect material behavior and device efficiency, which are challenging to obtain through traditional analysis methods.

For example, scanning electron microscopy (SEM) is frequently used to study the impact of film morphology on device performance. However, identifying the optimal morphology remains

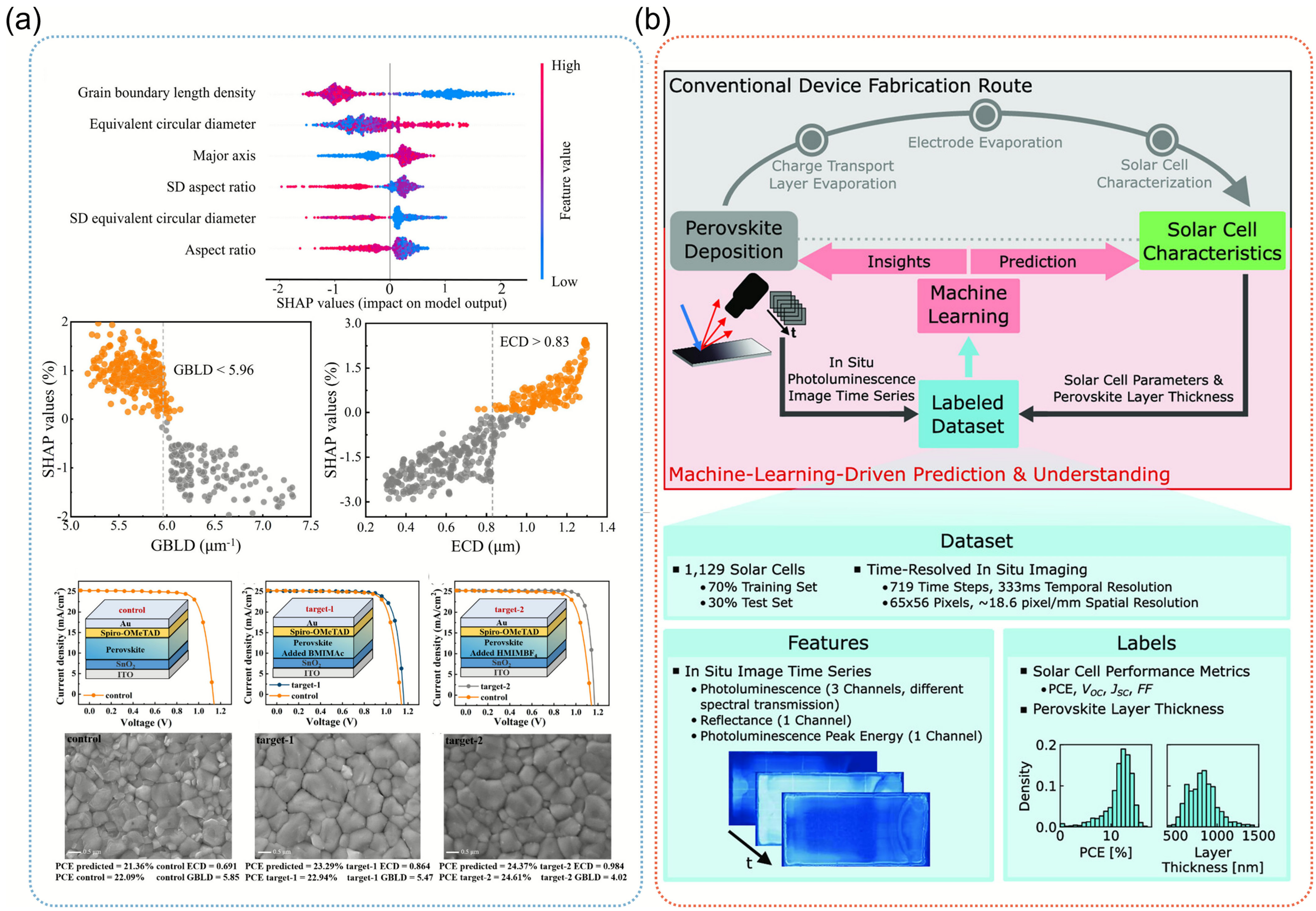

FIGURE 7 | (a) SHAP value analysis plot evaluating the feature importance of features in the SEM image. *J*–*V* curves and SEM images of the devices and films optimized under ML guidance [90]. Reproduced with permission [90]. Copyright 2025 American Chemical Society. (b) Schematic illustration of a feedback loop driven by ML methods applied to experimental in situ data and description of the generated dataset containing 1129 solar cells [91]. Reproduced with permission [91]. Copyright 2023 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim.

challenging. In 2025, Liu et al. developed a multimodal convolutional neural network to extract microstructural features from SEM images of perovskite thin films. The model dynamically adjusted the weights of various modal inputs, including material composition, processing techniques, and microstructure, to improve predictive accuracy. It achieved an impressive R^2 of 0.79 on a dataset of 1583 SEM images. By introducing six SEM-derived features to characterize the grain size of PSCs, the study revealed that a grain boundary length density (GBLD) below 5.96 and an equivalent circular diameter (ECD) above 0.83 significantly enhance PCE (Figure 7a) [90]. Additional experiments validated these findings, demonstrating that optimizing these parameters to control crystallization improved the PCE to 24.61%. The consistency of the results highlights the effectiveness and reliability of the multimodal model. In addition, Laufer et al. introduced a unique in situ photoluminescence (PL) dataset for blade-coated PSCs and analyzed it using unsupervised *k*-means clustering, demonstrating the potential of extracting valuable insights from such data. The correlations between the identified clusters and the measured device performance revealed that the in situ PL signal encodes critical information about the quality of perovskite thin films. By detecting spatial variations in PL patterns, the study identified detrimental mechanisms occurring during thin-film formation, which directly impact device performance. Furthermore, *k*-nearest neighbors (*k*-NN) were employed to predict PSC performance, highlighting the promise of ML-based in-line process monitoring for scalable PSC fabrication [91]. This approach enables the detection, understanding, and mitigation of process variations across fabrication iterations, paving the way for more consistent and efficient device production.

4 | Challenges and Outlook

The integration of artificial intelligence in PSCs research holds transformative potential but is hindered by challenges in data quality, standardization, and accessibility. Literature-extracted data often suffer from performance bias, where reported positive cases can lead models to overestimate material properties [92, 93]. Variations in computational methodologies introduce inherent discrepancies that compromise the reproducibility of DFT calculations, thereby affecting the reliability of the resulting data [94]. Detailed records of experimental equipments and procedures are essential to mitigate human-induced deviations [95]. Ensuring data verifiability is also critical; unverified patent claims and retracted publications must be filtered during database construction. The lack of standardized data formats complicates cross-system research, limiting the efficiency of ML methods. Addressing accessibility requires fostering an open data community based on shared resources, with LLMs offering solutions for faster, more accurate data extraction, processing, and validation [96, 97]. Advances in retrieval-augmented generation [98] and data analysis directly enhance the proficiency of LLMs in targeted literature retrieval and data authenticity verification, which addresses critical data bottlenecks and accelerates PSCs research. Moreover, current material descriptors, mostly based on computationally expensive QC parameters, are insufficient for broader ML applications. The lack of descriptors for disordered structures like amorphous perovskite materials further limits progress. Consequently, low-cost, interpretable, and versatile feature design is critical. Features derived from spatial 3D and topological structures, as well as real experimental data like XRD images, show promise for improving ML performance.

The accuracy of ML models largely depends on the features selected, whereas limited datasets hinder their generalization. Integrating chemical properties into model design allows for a more comprehensive representation of material characteristics. Multi-modal models, which combine diverse data types such as images and crystal graphs, offer significant improvements in prediction accuracy [99, 100]. However, small experimental datasets often lead to overfitting. LLMs, with their strong generalization capabilities, can address this limitation through transfer learning, enabling effective application to small-scale experimental data. On the other hand, the development of smart laboratories is also gaining momentum, with LLM-driven technologies supporting fully automated research workflows, including literature review, experimental design, validation, and data analysis [101, 102]. Despite this progress, accessibility remains a challenge due to the lack of user-friendly platforms and documentation, which limits their use by chemists without computational expertise. Additionally, the high cost of experimental robots slows the transition to automation. Overcoming these barriers by creating cost-effective, customizable hardware and integrated software platforms will revolutionize research methods, making them more efficient, accessible, and innovative.

5 | Conclusions

ML has proven to be a transformative tool in PSCs research, offering solutions to longstanding challenges in material optimization, stability enhancement, and scalable fabrication. By leveraging advanced algorithms and large datasets, ML accelerates the discovery of high-performance materials, identifies critical stability mechanisms, and optimizes fabrication processes. The future of PSCs research will increasingly depend on the integration of ML with experimental and computational approaches. Expanding high-quality datasets, particularly through high-throughput experimentation and in situ characterization, will be essential for improving model accuracy and robustness. Additionally, the development of interpretable ML models will enhance our understanding of the underlying physics and chemistry of PSCs, enabling more effective material design and process optimization. To fully realize the potential of ML, interdisciplinary collaborations across materials science, data science, and computational modeling will be crucial. Addressing challenges such as data standardization, algorithm scalability, and computational cost will further enhance the impact of ML in PSCs research. Ultimately, the continued integration of ML into PSC development has the potential to accelerate the transition of PSCs from laboratory innovation to market-ready technology, contributing significantly to global efforts in achieving sustainable and affordable energy solutions.

Author Contributions

S.W., W.Z., and M.Z. contributed equally to this work. S.W., R.S., Y.W., and Z.Z. conceived and designed this review. S.W., W.Z., and M.Z. wrote the first draft of the paper. T.L. and Y.W. revised the draft. R.S., Y.W., and Z.Z. supervised the project. All co-authors reviewed the paper and provided comments on the paper.

Acknowledgments

M.Z. acknowledges the Youth Foundation of Civil Aviation Flight University of China (25CAFUC05014). S.W. acknowledges the Fundamental Research Funds for the Central Universities (DUT24BK047). T.L. acknowledges the start-up grant from Great Bay University.

Funding

This work was supported by the Youth Foundation of Civil Aviation Flight University of China (25CAFUC05014), Fundamental Research Funds for the Central Universities (DUT24BK047), and the start-up grant from Great Bay University.

Data Availability Statement

The data that support the findings of this study are available in the Supporting Information of this article.

References

1. H. Chen, C. Liu, J. Xu, et al., "Improved Charge Extraction in Inverted Perovskite Solar Cells With Dual-Site-Binding Ligands," *Science* 384, no. 6692 (2024): 189–193.

2. N. 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," *Nature Materials* 13, no. 9 (2014): 897–903.

3. M. A. Green, E. D. Dunlop, M. Yoshita, et al., "Solar Cell Efficiency Tables (Version 64)," *Progress in Photovoltaics: Research and Applications* 32, no. 7 (2024): 425–441.

4. Y. Wang, P. Xiang, A. Ren, et al., "MXene-Modulated Electrode/SnO_2 Interface Boosting Charge Transport in Perovskite Solar Cells," *ACS Applied Materials & Interfaces* 12, no. 48 (2020): 53973–53983.

5. J. W. Lee, S. Tan, S. I. Seok, Y. Yang, and N. G. Park, "Rethinking the A Cation in Halide Perovskites," *Science* 375, no. 6583 (2022): eabj1186.

6. A. Kojima, K. Teshima, Y. Shirai, and T. Miyasaka, "Organometal Halide Perovskites as Visible-Light Sensitizers for Photovoltaic Cells," *Journal of the American Chemical Society* 131, no. 17 (2009): 6050–6051.

7. Z. Li, T. R. Klein, D. H. Kim, et al., "Scalable Fabrication of Perovskite Solar Cells," *Nature Reviews Materials* 3, no. 4 (2018): 1–20.

8. N. G. Park, "Perovskite Solar Cells: An Emerging Photovoltaic Technology," *Materials Today* 18, no. 2 (2015): 65–72.

9. J. Y. Kim, J. W. Lee, H. S. Jung, H. Shin, and N. G. Park, "High-Efficiency Perovskite Solar Cells," *Chemical Reviews* 120, no. 15 (2020): 7867–7918.

10. D. Wang, M. Wright, N. K. Elumalai, and A. Uddin, "Stability of Perovskite Solar Cells," *Solar Energy Materials & Solar Cells* 147 (2016): 255–275.

11. R. Wang, M. Mujahid, Y. Duan, Z. K. Wang, J. Xue, and Y. Yang, "A Review of Perovskites Solar Cell Stability," *Advanced Functional Materials* 29, no. 47 (2019): 1808843.

12. Y. Wang, Z. Zeng, Y. Zhang, et al., "Unlocking the Ambient Temperature Effect on FA-Based Perovskites Crystallization by In Situ Optical Method," *Advanced Materials* 36, no. 17 (2024): 2307635.

13. M. Jung, S. G. Ji, G. Kim, and S. I. Seok, "Perovskite Precursor Solution Chemistry: From Fundamentals to Photovoltaic Applications," *Chemical Society Reviews* 48, no. 7 (2019): 2011–2038.

14. T. Wu, X. Liu, X. Luo, et al., "Lead-Free Tin Perovskite Solar Cells," *Joule* 5, no. 4 (2021): 863–886.

15. J. Prakash, A. Singh, G. Sathiyan, et al., "Progress in Tailoring Perovskite Based Solar Cells Through Compositional Engineering: Materials Properties, Photovoltaic Performance and Critical Issues," *Materials Today Energy* 9 (2018): 440–486.

16. W. Gao, C. Chen, C. Ran, et al., "A-Site Cation Engineering of Metal Halide Perovskites: Version 3.0 of Efficient Tin-Based Lead-Free Perovskite Solar Cells," *Advanced Functional Materials* 30, no. 34 (2020): 2000794.

17. S. Ma, G. Yuan, Y. Zhang, N. Yang, Y. Li, and Q. Chen, "Development of Encapsulation Strategies Towards the Commercialization of Perovskite Solar Cells," *Energy & Environmental Science* 15, no. 1 (2022): 13–55.

18. Q. Lu, Z. Yang, X. Meng, et al., "A Review on Encapsulation Technology From Organic Light Emitting Diodes to Organic and Perovskite Solar Cells," *Advanced Functional Materials* 31, no. 23 (2021): 2100151.

19. Z. Yang, C. C. Chueh, P. W. Liang, et al., "Effects of Formamidinium and Bromide Ion Substitution in Methylammonium Lead Triiodide Toward High-Performance Perovskite Solar Cells," *Nano Energy* 22 (2016): 328–337.

20. M. Lyu and N. G. Park, "Effect of Additives AX (A=FA, MA, Cs, Rb, NH_4; X=Cl, Br, I) in $FAPbI_3$ on Photovoltaic Parameters of Perovskite Solar Cells," *Solar RRL* 4, no. 10 (2020): 2000331.

21. A. He, Y. Wang, P. Tang, et al., "Unraveling the Impact of Bromine Incorporation on the Hybrid Perovskite Film Phase Evolution Mechanism by In Situ Characterization," *Advanced Energy Materials* 13, no. 39 (2023): 2300957.

22. L. Yan, H. Huang, P. Cui, et al., "Fabrication of Perovskite Solar Cells in Ambient Air by Blocking Perovskite Hydration With Guanabenz Acetate Salt," *Nature Energy* 8, no. 10 (2023): 1158–1167.

23. B. Ding, Y. Ding, J. Peng, et al., "Dopant-Additive Synergism Enhances Perovskite Solar Modules," *Nature* 628, no. 8007 (2024): 299–305.

24. Y. Zou, W. Yu, H. Guo, et al., "A Crystal Cap Layer for Formation of Black-Phase $FAPbI_3$ Perovskite in Humid Air," *Science* 385, no. 6705 (2024): 161–167.

25. Y. Wang, Z. Zhang, Z. Zeng, et al., "Intermediate State Formation Extends the Ambient Temperature Processing Window of Solution-Processed Perovskite Solar Cells," *ACS Energy Letters* 10, no. 2 (2025): 647–657.

26. R. Wang, J. Zhu, J. You, et al., "Custom-Tailored Solvent Engineering for Efficient Wide-Bandgap Perovskite Solar Cells With a Wide Processing Window and Low V OC Losses," *Energy & Environmental Science* 17, no. 7 (2024): 2662–2669.

27. W. Lu, R. Xiao, J. Yang, H. Li, and W. Zhang, "Data Mining-Aided Materials Discovery and Optimization," *Journal of Materiomics* 3, no. 3 (2017): 191–201.

28. A. Jain, S. P. Ong, G. Hautier, et al., "Commentary: The Materials Project: A Materials Genome Approach to Accelerating Materials Innovation," *APL Materials* 1, no. 1 (2013), https://pubs.aip.org/aip/apm/article/1/1/011002/119685.

29. G. Bergerhoff, R. Hundt, R. Sievers, and I. D. Brown, "The Inorganic Crystal Structure Data Base," *Journal of Chemical Information and Computer Sciences* 23, no. 2 (1983): 66–69.

30. C. R. Groom, I. J. Bruno, M. P. Lightfoot, and S. C. Ward, "The Cambridge Structural Database," *Acta Crystallographica. Section B: Structural Science, Crystal Engineering and Materials* 72, no. 2 (2016): 171–179.

31. S. Kirklin, J. E. Saal, B. Meredig, et al., "The Open Quantum Materials Database (OQMD): Assessing the Accuracy of DFT Formation Energies," *npj Computational Materials* 1, no. 1 (2015): 1–15.

32. Y. Zhao, J. Zhang, Z. Xu, et al., "Discovery of Temperature-Induced Stability Reversal in Perovskites Using High-Throughput Robotic Learning," *Nature Communications* 12, no. 1 (2021): 2191.

33. N. Bhati, M. K. Nazeeruddin, and F. Maréchal, "Critical Analysis of Decision Variables for High-Throughput Experimentation (HTE) With Perovskite Solar Cells," *Solar Energy* 262 (2023): 111810.

34. Y. Hu, X. Hu, L. Zhang, et al., "Machine-Learning Modeling for Ultra-Stable High-Efficiency Perovskite Solar Cells," *Advanced Energy Materials* 12, no. 41 (2022): 2201463.

35. S. Rath, G. S. Priyanga, N. Nagappan, G. Sudha Priyanga, and T. Thomas, "Discovery of Direct Band Gap Perovskites for Light Harvesting by Using Machine Learning," *Computational Materials Science* 210 (2022): 111476.

36. Y. Liu, W. Yan, S. Han, et al., "How Machine Learning Predicts and Explains the Performance of Perovskite Solar Cells," *Solar RRL* 6, no. 6 (2022): 2101100.

37. T. Liu, S. Wang, Y. Shi, et al., "Machine-Learning Accelerating the Development of Perovskite Photovoltaics," *Solar RRL* 7, no. 23 (2023): 2300650.

38. H. Park, R. Mall, F. H. Alharbi, et al., "Learn-and-Match Molecular Cations for Perovskites," *Journal of Physical Chemistry. A* 123, no. 33 (2019): 7323–7334.

39. J. Yang, J. Fan, and S. Li, "Map the Room-Temperature Dynamic Stabilities of Inorganic Halide Double Perovskites," *Chemistry of Materials* 34, no. 20 (2022): 9072–9085.

40. S. Chakraborty, W. Xie, N. Mathews, et al., "A High-Throughput Computational Screening and Experimental Validation Methodology for Lead-Free and Emergent Hybrid Perovskites," *ACS Energy Letters* 2, no. 4 (2017): 837–845.

41. K. Yang, Y. Li, and J. Zhang, "High-Throughput Screening of Hybrid Quaternary Halide Perovskites for Optoelectronics," *Journal of Materials Chemistry A* 11, no. 12 (2023): 6465–6473.

42. X. Wan, W. Feng, Y. Wang, et al., "Materials Discovery and Properties Prediction in Thermal Transport via Materials Informatics: A Mini Review," *Nano Letters* 19, no. 6 (2019): 3387–3395.

43. S.-A. N. Alexandropoulos, S. B. Kotsiantis, and M. N. Vrahatis, "Data Preprocessing in Predictive Data Mining," *Knowledge Engineering Review* 34 (2019): e1.

44. S. García, J. Luengo, and F. Herrera, "Tutorial on Practical Tips of the Most Influential Data Preprocessing Algorithms in Data Mining," *Knowledge-Based Systems* 98 (2016): 1–29.

45. T. Toyao, Z. Maeno, S. Takakusagi, T. Kamachi, I. Takigawa, and K.-i. Shimizu, "Machine Learning for Catalysis Informatics: Recent Applications and Prospects," *ACS Catalysis* 10, no. 3 (2019): 2260–2297.

46. A. Jain, G. Hautier, S. P. Ong, and K. Persson, "New Opportunities for Materials Informatics: Resources and Data Mining Techniques for

Uncovering Hidden Relationships," *Journal of Materials Research* 31, no. 8 (2016): 977–994.

47. V. Bolón-Canedo and A. Alonso-Betanzos, "Ensembles for Feature Selection: A Review and Future Trends," *Information Fusion* 52 (2019): 1–12.

48. W. Zhang, X. Guo, Z. Cui, et al., "Strategies for Improving Efficiency and Stability of Inverted Perovskite Solar Cells," *Advanced Materials* 36, no. 37 (2024): 2311025.

49. D. Rogers and M. Hahn, "Extended-Connectivity Fingerprints," *Journal of Chemical Information and Modeling* 50, no. 5 (2010): 742–754.

50. D. Weininger, "SMILES, a Chemical Language and Information System. 1. Introduction to Methodology and Encoding Rules," *Journal of Chemical Information and Computer Sciences* 28, no. 1 (1988): 31–36.

51. Y. Mai, J. Tang, H. Meng, et al., "Machine Learning-Based Screening of Two-Dimensional Perovskite Organic Spacers," *Advanced Composites and Hybrid Materials* 7, no. 3 (2024): 104.

52. G. Jung, S. G. Jung, and J. M. Cole, "Automatic Materials Characterization From Infrared Spectra Using Convolutional Neural Networks," *Chemical Science* 14, no. 13 (2023): 3600–3609.

53. W. Liu, Y. Lu, D. Wei, et al., "Screening Interface Passivation Materials Intelligently Through Machine Learning for Highly Efficient Perovskite Solar Cells," *Journal of Materials Chemistry A* 10, no. 34 (2022): 17782–17789.

54. L. J. Diorazio, D. R. J. Hose, and N. K. Adlington, "Toward a More Holistic Framework for Solvent Selection," *Organic Process Research and Development* 20, no. 4 (2016): 760–773.

55. A. Talapatra, B. P. Uberuaga, C. R. Stanek, and G. Pilania, "A Machine Learning Approach for the Prediction of Formability and Thermodynamic Stability of Single and Double Perovskite Oxides," *Chemistry of Materials* 33, no. 3 (2021): 845–858.

56. J. Zhao and X. Wang, "Screening Perovskites From ABO3 Combinations Generated by Constraint Satisfaction Techniques Using Machine Learning," *ACS Omega* 7, no. 12 (2022): 10483–10491.

57. F. Oviedo, Z. Ren, S. Sun, et al., "Fast and Interpretable Classification of Small X-Ray Diffraction Datasets Using Data Augmentation and Deep Neural Networks," *npj Computational Materials* 5, no. 1 (2019): 60.

58. S. Sun, N. T. P. Hartono, Z. D. Ren, et al., "Accelerated Development of Perovskite-Inspired Materials via High-Throughput Synthesis and Machine-Learning Diagnosis," *Joule* 3, no. 6 (2019): 1437–1451.

59. V. A. Aristizabal-Ferreira, J. M. Guevara-Vela, A. Sauza-de la Vega, Á. M. Pendás, G. Fuentes-Pineda, and T. Rocha-Rinza, "Computation of Photovoltaic and Stability Properties of Hybrid Organic–Inorganic Perovskites via Convolutional Neural Networks," *Theoretical Chemistry Accounts* 141, no. 4 (2022): 19.

60. Y. Huang and L. Zhang, "Descriptor Design for Perovskite Material With Compatible Molecules via Language Model and First-Principles," *Journal of Chemical Theory and Computation* 20, no. 15 (2024): 6790–6800.

61. X. Yang, Y. Han, P. Xu, and F. Liu, "Thermodynamic Stability Descriptor of A2BX6-Type Perovskite Materials," *Materials Chemistry and Physics* 333 (2025): 130324.

62. B. Yılmaz and R. Yıldırım, "Critical Review of Machine Learning Applications in Perovskite Solar Research," *Nano Energy* 80 (2021): 105546.

63. M. Chen, Z. Yin, Z. Shan, et al., "Application of Machine Learning in Perovskite Materials and Devices: A Review," *Journal of Energy Chemistry* 94 (2024): 254–272.

64. T. Lu, H. Li, M. Li, S. Wang, and W. Lu, "Predicting Experimental Formability of Hybrid Organic–Inorganic Perovskites via Imbalanced Learning," *Journal of Physical Chemistry Letters* 13, no. 13 (2022): 3032–3038.

65. D. Jain, S. Chaube, P. Khullar, S. Goverapet Srinivasan, and B. Rai, "Bulk and Surface DFT Investigations of Inorganic Halide Perovskites Screened Using Machine Learning and Materials Property Databases," *Physical Chemistry Chemical Physics* 21, no. 35 (2019): 19423–19436.

66. J. Chen, Y. Zhan, Z. Yang, Y. Zang, W. Yan, and X. Li, "Predicting and Analyzing Stability in Perovskite Solar Cells: Insights From Machine Learning Models and SHAP Analysis," *Materials Today Energy* 48 (2025): 101769.

67. D. J. Livingstone, *Artificial Neural Networks: Methods and Applications* (Humana Press, 2008).

68. J.-S. Kim, J. Noh, and J. Im, "Machine Learning-Enabled Chemical Space Exploration of All-Inorganic Perovskites for Photovoltaics," *npj Computational Materials* 10, no. 1 (2024): 97.

69. H. A. H. Mahmoud, "Computerized Prediction of Perovskite Performance Using Deep Learning," *Electronics* 11, no. 22 (2022): 3759.

70. X. Li, Y. Dan, R. Dong, et al., "Computational Screening of New Perovskite Materials Using Transfer Learning and Deep Learning," *Applied Sciences* 9, no. 24 (2019): 5510.

71. R. Lyu, C. E. Moore, T. Liu, Y. Yu, and Y. Wu, "Predictive Design Model for Low-Dimensional Organic–Inorganic Halide Perovskites Assisted by Machine Learning," *Journal of the American Chemical Society* 143, no. 32 (2021): 12766–12776.

72. S. Zhang, T. Lu, P. Xu, Q. Tao, M. Li, and W. Lu, "Predicting the Formability of Hybrid Organic–Inorganic Perovskites via an Interpretable Machine Learning Strategy," *Journal of Physical Chemistry Letters* 12, no. 31 (2021): 7423–7430.

73. Z. Ge, Z. Song, S. X. Ding, and B. Huang, "Data Mining and Analytics in the Process Industry: The Role of Machine Learning," *IEEE Access* 5 (2017): 20590–20616.

74. P. Subudhi, S. Sivapatham, R. Narasimhan, et al., "Enhancing Photovoltaic Performance in Tin-Based Perovskite Solar Cells: A Unified Approach Utilizing Numerical Simulation and Machine Learning Techniques," *Journal of Power Sources* 639 (2025): 236639.

75. T. Wu and J. Wang, "Global Discovery of Stable and Non-Toxic Hybrid Organic-Inorganic Perovskites for Photovoltaic Systems by Combining Machine Learning Method With First Principle Calculations," *Nano Energy* 66 (2019): 104070.

76. C. Zhu, Y. Liu, D. Wang, et al., "Exploration of Highly Stable and Highly Efficient New Lead-Free Halide Perovskite Solar Cells by Machine Learning," *Cell Reports Physical Science* 5, no. 12 (2024): 102321.

77. B. Zhang, H. Zeng, H. Yin, et al., "Combining Component Screening, Machine Learning and Molecular Engineering for the Design of High-Performance Inverted Perovskite Solar Cells," *Energy & Environmental Science* 17, no. 15 (2024): 5532–5541.

78. Z. L. Huang, C. H. Yeh, M. Y. Wang, V. W. H. Lau, H. K. Tian, and C. F. Shih, "Comprehensive Study on Phase Stability of Lead-Free Sn-Based Perovskite $FA_xMA_{1\text{-}x}SnI_3$," *Materials Today Advances* 24 (2024): 100534.

79. Z. Li, Q. Xu, Q. Sun, Z. Hou, and W. J. Yin, "Thermodynamic Stability Landscape of Halide Double Perovskites via High-Throughput Computing and Machine Learning," *Advanced Functional Materials* 29, no. 9 (2019): 1807280.

80. R. Luo, X. Jia, X. Niu, et al., "Machine Learning-Driven Insights for Phase-Stable $FA_xCs_{1-x}Pb\ (I_yBr_{1-y})_3$ Perovskites in Tandem Solar Cells," *JACS Au* 5, no. 4 (2025): 1771–1780.

81. Q. Zhang, H. Wang, Q. Zhao, et al., "Machine-Learning-Assisted Design of Buried-Interface Engineering Materials for High-Efficiency and Stable Perovskite Solar Cells," *ACS Energy Letters* 9, no. 12 (2024): 5924–5934.

82. J. Wu, L. Torresi, M. M. Hu, et al., "Inverse Design Workflow Discovers Hole-Transport Materials Tailored for Perovskite Solar Cells," *Science* 386, no. 6727 (2024): 1256–1264.

83. W. Feng, R. Zhao, X. Wang, et al., "Global Instability Index as a Crystallographic Stability Descriptor of Halide and Chalcogenide Perovskites," *Journal of Energy Chemistry* 70 (2022): 1–8.

84. J. Zhang, S. Langner, J. Wu, et al., "Intercalating-Organic-Cation-Induced Stability Bowing in Quasi-2D Metal-Halide Perovskites," *ACS Energy Letters* 7, no. 1 (2021): 70–77.

85. C. Zhi, S. Wang, S. Sun, et al., "Machine-Learning-Assisted Screening of Interface Passivation Materials for Perovskite Solar Cells," *ACS Energy Letters* 8, no. 3 (2023): 1424–1433.

86. S. Wang, Y. Huang, W. Hu, and L. Zhang, "Data-Driven Optimization and Machine Learning Analysis of Compatible Molecules for Halide Perovskite Material," *npj Computational Materials* 10, no. 1 (2024): 114.

87. Ç. Odabaşı and R. Yıldırım, "Machine Learning Analysis on Stability of Perovskite Solar Cells," *Solar Energy Materials & Solar Cells* 205 (2020): 110284.

88. R. Bhaduri and S. Manasa, "Perovskite Solar Cell Stability Analysis Using Entropy-Based Support Vector Machines Learning," *Progress in Photovoltaics: Research and Applications* 33, no. 9 (2024): 962–979.

89. N. T. P. Hartono, J. Thapa, A. Tiihonen, et al., "How Machine Learning Can Help Select Capping Layers to Suppress Perovskite Degradation," *Nature Communications* 11, no. 1 (2020): 4172.

90. H. Liu, A. Yang, C. Zhong, et al., "Integration of Microstructural Image Data Into Machine Learning Models for Advancing High-Performance Perovskite Solar Cell Design," *ACS Energy Letters* 10, no. 4 (2025): 1884–1891.

91. F. Laufer, S. Ziegler, F. Schackmar, et al., "Process Insights Into Perovskite Thin-Film Photovoltaics From Machine Learning With In Situ Luminescence Data," *Solar RRL* 7, no. 7 (2023): 2201114.

92. M. P. Maloney, C. W. Coley, S. Genheden, et al., "Negative Data in Data Sets for Machine Learning Training," *Organic Letters* 25, no. 17 (2023): 2945–2947.

93. R. Kurczab, S. Smusz, and A. J. Bojarski, "The Influence of Negative Training Set Size on Machine Learning-Based Virtual Screening," *Journal of Cheminformatics* 6 (2014): 32.

94. K. Lejaeghere, G. Bihlmayer, T. Björkman, et al., "Reproducibility in Density Functional Theory Calculations of Solids," *Science* 351, no. 6280 (2016): aad3000.

95. Z. Liu, Y. S. Morozbcd, and O. Isayev, "The Challenge of Balancing Model Sensitivity and Robustness in Predicting Yields: A Benchmarking Study of Amide Coupling Reactions," *Chemical Science* 14, no. 39 (2023): 10835–10846.

96. Z. Zheng, O. Zhang, C. Borgs, J. T. Chayes, and O. M. Yaghi, "ChatGPT Chemistry Assistant for Text Mining and the Prediction of MOF Synthesis," *Journal of the American Chemical Society* 145, no. 32 (2023): 18048–18062.

97. S. Gupta, A. Mahmood, P. Shetty, A. Adeboye, and R. Ramprasad, "Data Extraction From Polymer Literature Using Large Language Models," *Communications Materials* 5, no. 1 (2024): 269.

98. P. Lewis, E. Perez, A. Piktus, et al., "Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks," *Advances in Neural Information Processing Systems* 33, (2020): 9459–9474.

99. W. Rahman, C. Zhong, H. Liu, et al., "Enhancing Perovskite Solar Cell Efficiency and Stability: A Multimodal Prediction Approach Integrating Microstructure, Composition, and Processing Technology," *Nanoscale* 17 (2025): 15935–15949.

100. H. Zhang, F. T. Eickemeyer, Z. Zhou, et al., "Multimodal Host–Guest Complexation for Efficient and Stable Perovskite Photovoltaics," *Nature Communications* 12 (2021): 3383.

101. C. W. Coley, D. A. Thomas, J. A. M. Lummiss, et al., "A Robotic Platform for Flow Synthesis of Organic Compounds Informed by AI Planning," *Science* 365, no. 6453 (2019): eaax1566.

102. T. Song, M. Luo, X. Zhang, et al., "A Multiagent-Driven Robotic AI Chemist Enabling Autonomous Chemical Research on Demand," *Journal of the American Chemical Society* 147, no. 15 (2025): 12534–12545.

103. H. S. Kim, J. Y. Seo, and N. G. Park, "Material and Device Stability in Perovskite Solar Cells," *ChemSusChem* 9, no. 18 (2016): 2528–2540.

Supporting Information

Additional supporting information can be found online in the Supporting Information section. **Table S1:** Comparison of methods for constructing datasets for machine learning in perovskite research. **Table S2:** Classification and characteristics of descriptors used in PSCs research.