

Machine-Learning Accelerating the Development of Perovskite Photovoltaics

Tiantian Liu,* Sen Wang, Yinguang Shi, Lei Wu, Ruiyu Zhu, Yong Wang, Jun Zhou,* and Wallace C. H. Choy*

Perovskite solar cells (PSC) are a potential candidate for next-generation photovoltaic technology. Despite the significant advancements in the field of PSCs, the ongoing development of stable and efficient metal halide perovskite materials, along with their successful integration into photovoltaic applications, remains challenges. These challenges originate from the diverse range of device structures and perovskite compositions, requiring meticulous consideration and optimization. Traditional trial-and-error methods are characterized by their sluggishness and labor-intensive nature. Recently, the emergence of extensive datasets and advancements in computer hardware have facilitated the utilization of machine learning (ML) across multiple domains, including in various fields for material discovery and experimental optimization. Herein, the fundamental procedure of ML is briefly introduced, and latest progress of ML in the materials development and solar cell fabrication is comprehensively reviewed. The utilization of ML in PSCs at all stages of design can be categorized into four main areas: screening perovskite material, fabrication process optimization, device structure optimization, and understanding mechanism. The challenges and outlooks on the future development of ML are finally discussed. It is highly expected that this review can offer valuable guidance for the design and development of highly efficient and stable PSCs.

1. Introduction

Metal halide perovskites (MHPs) have become revolutionary semiconductors because of their outstanding physical properties such as high charge-carrier mobility, long carrier-diffusion length, and strong optical absorption.^[1–4] They have been successfully applied into various optoelectronics devices including solar cells, light-emitting diodes, photodetectors, etc.^[5–7] Among these, perovskite solar cell (PSC) is the most prominent because of its sharply increased power conversion efficiency (PCE) from 3.8% to a certified 26.1%.^[8,9] The excellent efficiency progress of PSC makes it a potential alternative for the current solar technologies. Despite their high efficiency, the majority of PSCs are currently confined to small-area devices and encounter significant long-term stability challenges, which make PSCs fall short of the commercial photovoltaic requirements.^[10–14]

Various strategies including composition engineering,^[15–17] passivation approaches,^[18–22] interface engineering,^[23–25] etc., have been developed to improve the efficiency and stability of small/large area PSCs. Nevertheless, traditional trial-and-error methods of developing new strategies to enhance efficiency and stability of PSCs are time-consuming due to multiple parameters that must be taken into consideration.^[26–29] Moreover, the data of experiment become increasingly complex with the development of synthesis and characterization techniques. It is difficult to quickly figure out the relationship between materials descriptors and properties.^[30–35] Theoretical calculations based on first-principle and density-functional theory (DFT) are widely applied to develop new perovskites with excellent optoelectronic properties through predicting the chemical structures and optoelectronic properties of materials. These methods compute material properties directly from fundamental physical quantities (mass, charge, and the Coulomb force of electrons, etc.) based on the principles of quantum mechanics. However, these computational methods are very costly and time-consuming for complex compounds.^[36–38] Therefore, it is necessary to adopt new technological means to improve and accelerate the systematic research of PSCs, to further improve their stability and efficiency, as well as promote the commercialization of PSCs.

T. Liu, S. Wang, Y. Shi, L. Wu, R. Zhu, J. Zhou
School of Chemistry and Chemical Engineering
Xi'an University of Architecture and Technology
Xi'an 710055, China
E-mail: liutiantian@xauat.edu.cn; zhounjun@xauat.edu.cn

Y. Wang
State Key Laboratory of Silicon Materials and School of Materials Science and Engineering
Zhejiang University
Hangzhou, Zhejiang 310027, China

Y. Wang
ZJU-Hangzhou Global Scientific and Technological Innovation Center
Zhejiang University
Hangzhou 310014, China

W. C. H. Choy
Department of Electrical and Electronic Engineering
The University of Hong Kong
Hong Kong 999077, China
E-mail: chchoy@eee.hku.hk

 The ORCID identification number(s) for the author(s) of this article can be found under <https://doi.org/10.1002/solr.202300650>.

DOI: 10.1002/solr.202300650

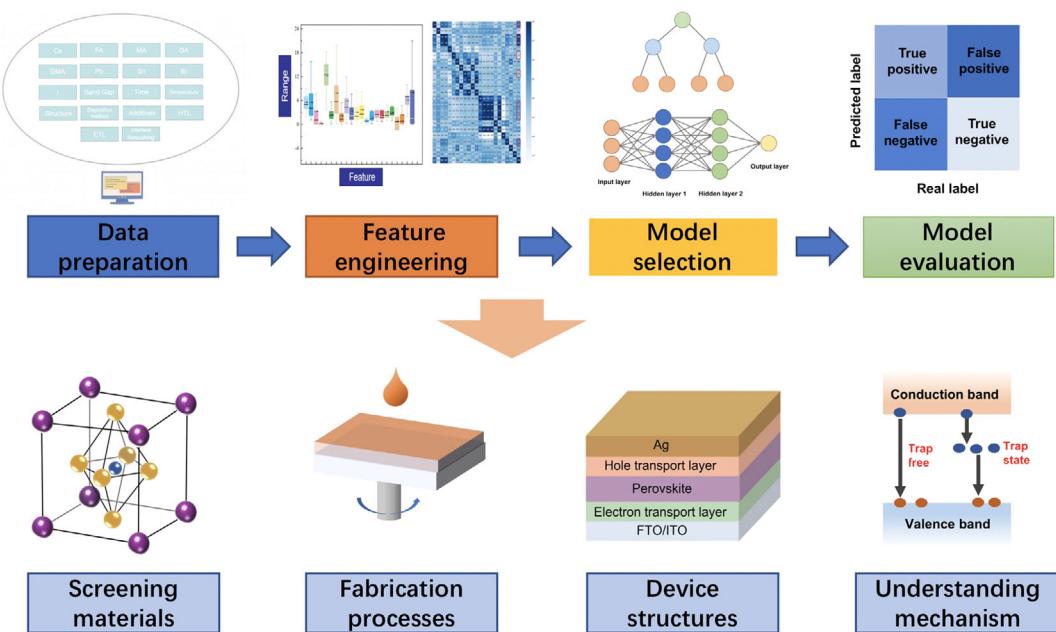


Figure 1. The general workflow of machine learning (ML).

Recently, the application of artificial intelligence in scientific research has significantly changed the way and speed by accelerating data processing, optimizing experimental design, and so on.^[39,40] As an important branch of artificial intelligence, machine learning (ML) combines the knowledge of computer science, statistics, mathematics, and other disciplines, which have been making great influence on various fields because of its strong abilities for fast execution, large dataset, predictability of future, and easy integration with the experiments.^[41–46] The purpose of ML is constructing a statistical algorithm, and then evaluating and predicting the objects with the guidance of the algorithm.^[47] Additionally, ML is also employed for mapping the relationships and quickly predicting the photovoltaic performance of materials.^[48–52] A well-trained ML model can predict properties for thousands of materials in mere seconds, while it may take DFT calculations hours of computational effort to calculate a single material property. At the same time, the accuracy of ML models is comparable to that of DFT. Therefore, ML has gradually emerged as a prominent tool in the development of new stable perovskites, as well as new strategies to improve PCEs and stability of PSCs.^[53]

Considering the significant potential of ML in the scientific research and application in the development of highly efficient and stable PSCs, we comprehensively review the implementation of ML in the field of PSCs. We first introduce the fundamental procedure of ML, followed by systematic discussions on latest progress in applying ML to materials development and solar cell fabrication. Finally, we highlight the challenges in this field and provide future research directions toward ML-assisted material development and solar cell fabrication.

2. Integrating ML Techniques with the Design of PSCs

Figure 1 shows the general workflow of ML, which includes steps like data preparation, feature engineering, model selection, and model evaluation.

2.1. Data Preparation

Constructing high-quality datasets is crucial for ML application.^[54] High-quality data can be collected from authoritative databases. For example, the commonly used experimental databases containing chemical structures of molecules and crystals include Cambridge Structural Database,^[55] Crystallography Open Database,^[56] Inorganic Crystal Structure Database,^[57] and PubChem,^[58] etc. The computational databases are Materials Project (MP),^[59] Open Quantum Materials Database,^[60] and Novel Materials Discovery,^[61] etc.

In addition to databases, data for ML studies can also be directly collected from experiment. However, the generation of huge experimental datasets required by ML is quite challenging, leading to a less widespread application of ML in PSCs research based on experimental data. One solution to address this challenge is employing the computationally generated data to augment the experimental results. For example, Yu et al.^[62] combined the experimental and computational results to predict the compatibility of 50 amines for the post treatment of MAPbI₃. JChem for Office software was used to access the chemical databases and calculate the 53 descriptors of the studied amines. After cross correlations, 22 features are eliminated and the rest 31 features are used for training SVM model to classify amine reactivity. In addition, the experimental data can also be collected

from published research papers. Liu and coworkers analyzed ≈ 100 peer-reviewed publications on PSCs to explore the influences of five factors, including grain size, defect density, bandgap, fluorescence lifetime, and surface roughness, on the efficiency and stability of PSCs.^[63] However, these dataset suffer from large deviations and incompleteness, which inevitably exist because these data are usually measured under different experimental conditions.

The quantity and quality of data is of vital importance for the ML-predicted results. Insufficient datasets can impact the accuracy of the models, whereas excessively large datasets can be time-consuming. Hence, it is important to balance the accuracy and time efficiency in the datasets. To ensure the accuracy, the amount of data should be at least larger than three times of descriptors.^[64] The data collected from experimental or computational databases might exhibit heterogeneity and contain missing values, which can directly impact the performance of the ML model. Therefore, it is essential to regularize the data to ensure consistency and accuracy of origin data. The data regularization includes the removal of redundant values, dealing with missing data, and modification of data outliers. **Data normalization is used in the process to transform the values of a dataset into a common scale, which can increase the accuracy of the model.** The commonly used data normalization methods include min–max normalization, Z-score normalization, and decimal scaling, etc.^[65]

2.2. Feature Engineering

Features, also known as predictors, are independent variables of the dataset. A specific set of features represent the key information associated with the target properties of a material, e.g., atomic parameters (atomic radius, atom types) molecular parameters (electronic coupling, electronegativity, etc.), and structural parameters (atomic radial distribution, Goldschmidt tolerance factor, etc.), electronic properties (bandgap, mobility, electron density, etc.), and the technological conditions of experimental process. Selecting suitable feature is a critical step, as it is fundamental to ensure the applicability and feasibility of ML models.

The reasonably good material feature should be universal, meaningful, reversible, and readily available.^[66,67] Moreover, it is important to avoid redundant features during training, as larger dataset is required when dealing with a higher number of features. Therefore, it is crucial to eliminate less important features to prevent overfitting, especially when the dataset is relatively small. Dimensionality reduction methods, such as feature selection, linear algebra techniques, and projection methods, can be employed to reduce the number of features.^[66]

The traditional feature selection is done manually, which is difficult to identify the most representative features and labor intensive. However, with the advancement of high-performance computing, deep learning enables computers to automatically extract features from the data and integrate them into the model-building process.^[68] This not only diminishes the necessity for manual feature engineering but also enhances prediction accuracy.

2.3. Model Selection

ML models can be divided into supervised and unsupervised learning according to the labeled dataset. Supervised learning models learn from a set of labeled input and output data to find the relationship between them and achieve the required prediction. Supervised learning models can be further divided into regression and classification. The target property of a classification model in ML is a discrete value, whereas target property is a continuous value in a regression model. The unsupervised learning models do not use the labeled datasets and can discover the intrinsic regularity of data on their own. Currently, supervised learning is the most widely used approaches in materials science. Therefore, we mainly focus on supervised learning models later.^[42,69–71]

K-Nearest Neighbors (kNN): It identifies a sample by considering the majority of its nearest k neighbors in feature space. The kNN is a lazy learning technique since generalization of training data is delayed until a system query is made. Consequently, the prediction process in kNN is time-consuming when dealing with a large training dataset.

Naive Bayes: The primary mathematical foundation of Naïve Bayes is Bayes' theorem and the assumption of conditional independence among features. In the Naïve Bayes' classification method, samples are categorized by assigning them to the category with the highest estimated probability.

Decision Tree and Random Forest: Decision tree (DT) consists of internal nodes, which represents distinguishing conditions for features, and leaf nodes that indicate different classes. DT method can potentially generate overly complex trees, resulting in overfitting. Therefore, unreliable branches are deleted to enhance both the speed and accuracy. Random forest (RF) is an ensemble learning method for classification and regression containing a multitude of DTs during training time. In classification tasks, RF outputs the class that the majority of its trees predict, while in regression, the prediction is obtained by averaging the individual tree predictions. RF often yields higher accuracy compared to a single DT. However, it can be sensitive to various factors, including parameters, noise, environmental variations, and other influences.

Artificial neural network: Artificial neural network (ANN) composes a large quantity of connected nodes, also called neurons, which simulates the intricate data processing mechanisms in nervous systems. ANN requires a large amount of training data and is susceptible to overfitting.

Ridge Regression: Ridge regression (RR) is used to identify the multiple-regression data that are highly correlated. RR effectively addresses the problem of overfitting. It is to be noted that the coefficient estimates generated by RR exhibit a bias. Kernel ridge regression (KRR) integrates RR with the kernel trick, which are nonparametric regression that are capable of modeling linear and nonlinear relationships.

Gradient Boosting Regression: Gradient boosting regression (GBR) is an ensemble method combining the weak predictive models to get better performance as a whole. GBR typically offers better accuracy and higher flexibility. Notably, GBR handles the missing data on its own without the need for imputation. Nevertheless, GBR can be computationally intensive, demanding

a substantial number of trees (≈ 1000), which is time and memory exhaustive.

2.4. Model Evaluation

ML model evaluation is required to ensure the ML model with high accurate predictions. The commonly used model evaluation methods include cross validation (CV) and bootstrapping sampling.

K-fold CV and leave-one-out CV are two commonly used CV methods. In k-fold CV method, the data are divided into k separate folds. One fold is used as initial test set and the others as the initial training set. Then, the process is iterated until each fold is utilized as the test set once. After k times of circulation, all test results are averaged to exhibit final ML output. Leave one out cross-validation represents a special case where k equals the number of input data points, and it is employed when the data size is very small. Bootstrapping sampling is based on random sampling with replacement, which can be employed to evaluate statistical characteristics such as variance, confidence intervals, bias, prediction error, etc.^[72,73]

Model performance can be evaluated by computing ML metrics, and the selection of appropriate metrics depends on the type of algorithms being used, such as, classification and regression. For classification algorithms, the typical metrics such as confusion matrix, receiver-operating characteristic curve and area under curve are employed to evaluate model performance. The typical metrics for regression include mean squared error (MSE), root MSE (RMSE), mean absolute error (MAE), and coefficient of determination (R^2).

3. Application of ML in Perovskites Field

3.1. Achievements of ML in PSCs

Various studies have been conducted in the field of perovskites, employing a variety of ML techniques and investigating various subjects. A comprehensive summary of ML-based works has been provided in Table 1 for researchers to quickly obtain the detailed studies. The research on ML has been focused on the exploration and discovery of perovskites that exhibit desirable characteristics, such as suitable bandgap, high efficiency, and exceptional stability, as well as the investigation of highly efficient and stable PSCs.

In majority of these papers, ML was used to predict properties of perovskite materials and identify potential candidates for solar cells. The dataset is generally extracted from traditional computationally (mostly density-functional theory [DFT]) created data. DFT is used to obtain the structural parameters and important properties of perovskites including thermodynamic stability, bandgaps, formability, impurity energy states, electron mobilities, and other electronic properties.^[74,75] The ML methods are then trained on these datasets to predict the properties of the perovskite material, such as the bandgaps, stability, and electron transport properties.^[62,63,76] As ML can considerably enhance the efficiency of material screening and decrease the computational cost of DFT, combining ML and DFT in the field of perovskite materials demonstrate the great potential. Lu et al.^[77] presented a

set of research methods of screening potential perovskite materials by combining ML and DFT, which has a significantly effect on the subsequent studies (Figure 2a). The training dataset comprised large amounts of data for 212 perovskites, which was computed by DFT. The relative significance of descriptors was then analyzed using the GBR method to reduce the redundancy of features. The top 14 most important features were retained such as tolerance factor, octahedral factor etc. (as shown in Figure 2a). Then, the ML methods were tested and optimized using three model evaluation metrics (R^2 , r, MSE). Finally, the optimal model based on the GBM algorithm was obtained, which was further used to predict the bandgaps of 5158 potential candidate materials. After setting the primary limits of material filtering (stability, suitability of the bandgap, accessibility for experiment, and toxicity), six lead-free perovskite materials were filtered. The ML-predicted bandgaps are very close to the bandgaps obtained from DFT calculations. Wu et al.^[78] further speeded up the discovery of hidden hybrid organic-inorganic perovskites (HOIPs) for photovoltaics based on multiple criterions including charge neutrality condition, stability condition, and bandgaps (Figure 2b,c). They used charge neutrality as the first screening criterion and reduced the database from 230 808 candidates to 77 748 candidates. Goldschmidt tolerance factor and octahedral factor were used as stability criteria to further screen and left 38 086 potential candidates. Finally, MLs (GBR, support vector regression (SVR), and KRR) were employed to predict the bandgaps of perovskites, which were trained from the DFT generated dataset of 1346 theoretically known HOIPs. After further screening based on suitable bandgaps, 686 orthorhombic-like HOIPs candidates were selected for light-harvesting applications. DFT calculations were also performed to verify the results. In addition, 132 stable and nontoxic (Cd-, Pb- and Hg-free) HOIPs with suitable bandgap are found to be promising in solar cell applications.

Although DFT data have been widely used in ML research works, there are still some issues. For example, the assumptions and models used in the computational work may be unrealistic under experimental conditions.^[79] In addition, not all the screened perovskites can be used in PSCs. Therefore, the combination of ML with the experiments needs to be strengthened. One effective way is to create a loop between ML and experiments. For example, the most promising perovskite candidates can be selected using ML, and then experiments should be conducted to verify the ML results. The database can be updated with the data obtained from the experiments.

For ML works based on experimental data, ML is primarily used for feature analysis to identify the critical features that affecting the device performance and to guide the experimental process, such as the fabrication process and device structure.^[80] The underlying factors affecting the device performance and long-term stability are also investigated using ML. Therefore, complete feature construction logic and high interpretability of the model are required. Shapley additive explanation (SHAP) is an effective and widely used method for explaining the output of ML models. For example, Zhang et al.^[81] proposed an interpretable strategy combining ML with an SHAP to predict the formability of hybrid perovskites, which paid more attention to the evaluation of features and explanation of models. They used SHAP explanation and statistical analysis on the model and candidate data (Figure 2d). They discovered a remarkable

Table 1. Overview of ML application in perovskites field.

Reference	Predicted properties	ML algorithms	Data source	Summary	Prediction accuracy
Lu et al. 2018 ^[77]	Bandgaps	GBR, KRR, SVM, DT, GPR, MLP	Computation	The nonlinear relationship between the input features and bandgaps was identified with ML to screen perovskites for photovoltaics from 5158 unexplored perovskites.	GBR is the best with R^2 , r , and MSE to be 0.97, 0.985, and 0.086.
Takahashi et al. 2018 ^[185]	Bandgaps	RF	Computation	The Li- and Na-based perovskites are evaluated with DFT and RF where 11 undiscovered Li(Na)-based perovskites possess ideal bandgap and formation energy for solar cell applications.	Accuracy is 98%.
Wu et al. 2019 ^[78]	Bandgaps	GBR, SVM, KRR	Computation	Using a target-driven method combining ML with DFT to screen hidden hybrid perovskites based on bandgap and stability for photovoltaics.	R^2 , MAE, and MSE of GBR are 0.827, 0.377, and 0.201.
Gladikh et al. 2020 ^[186]	Bandgaps	ACE, DT, KRR, ERT, AdaBoost, GBR and SVM	Computation	MLs were trained to study relation between bandgaps of ABX_3 perovskites and the properties of the constituent elements (e.g., electron affinities, electronegativities, and atomic radii).	ACE is better on the small dataset.
Jao et al. 2020 ^[187]	Bandgaps	neural network	Computation	Four sets of different variational related loss functions, named as Element Codes were generated, which was used as the sole descriptor to predict the bandgap of double perovskites.	Accuracy of 95.1% and MAE of 0.266 eV are achieved.
Park et al. 2020 ^[118]	Bandgaps	XGBoost, deep learning (DL)	Computation	The impediment of constructing a highly accurate predictor for perovskite bandgap when the inorganic network undergoes the deformation was addressed.	XGBoost is the best with R^2 of 0.99, r of 0.99, and MAE of 0.07 eV.
Li et al. 2020 ^[188]	Bandgaps	LR, RF, ANN	Experiment	The compositions of the perovskites with desirable bandgaps and high iodide ratio for suppressing halide segregation were predicted by NN.	ANN is the best with RMSE of 0.05 eV and r of 0.993.
Cai et al. 2022 ^[88]	Bandgaps PCE	SVR, RF, GBM, ANN, etc.	Experiment	The relationship between key parameters (bandgap and Sn composition) and photovoltaic performance in $MASn_xPb_{1-x}I_3$ perovskites was established. An optimized Sn: Pb composition ratio near 0.6 is finally obtained for high-performance PSCs, and then verified by experiments.	GBM is the best for bandgap with RMSE of 0.0386 eV. ANN is the best for PCE with RMSE of 1.9626%.
Yang et al. 2021 ^[134]	Bandgaps	RR, SVM, RF, KRR, GBDT	Computation	Selection of functional materials with suitable photoelectric properties and thermal stability from 16 400 candidates. Finally, 10 excellent double perovskites were selected to guide subsequent experiment.	GBDT has the best performance with R^2 of 0.9863, r of 0.9940, and MSE of 0.0567 eV.
Rath et al. 2022 ^[86]	Bandgaps	XGBoost, RF, LR, SVM, ANN, DT, GBM	Computation	XGBoost was used to classify bandgap types of ABX_3 inorganic perovskites and SHAP was used to determine the factors affecting bandgap types. The absence of transition metals and elements	XGBoost predicts direct bandgap perovskites with 81% precision.

Table 1. Continued.

Reference	Predicted properties	ML algorithms	Data source	Summary	Prediction accuracy
Agiorgousis et al. 2019 ^[189]	Bandgaps and stability	RF	Computation	(groups IIIA to VIIA with atomic number greater than 20) increases the probability of direct bandgap. ML was used to classify potential photovoltaic absorbers using data from the periodic table. Five chalcogenide double perovskites were identified as promising alternative considering thermodynamic stability, kinetic stability, and optical absorption.	An RF algorithm achieves a cross-validation accuracy of 86.4%.
Lu et al. 2019 ^[130]	Bandgaps, spontaneous polarization	GBC and GBR	Computation	Eight potential ferroelectric photovoltaic perovskites with excellent thermal stability, suitable bandgap, and considerable spontaneous polarization are presented.	The accuracy of energy difference and bandgap exceeds 90%.
Ma et al. 2019 ^[190]	Bandgaps, Stability, and mobility	SVR, RFR, GBR	Computation	ML was used to screen potential optoelectronic materials of 5000 2D octahedral oxyhalides based on nontoxicity, low cost, dynamical stability, high mobility, and enhanced optical absorption.	GBR model is the best model with the lowest MSE of 0.086 and largest R^2 of 0.835.
Kanno et al. 2019 ^[99]	Bandgaps and carrier effective mass	Multiple linear regression, RR, Lasso regression, SVR, GPR	Computation	A high-throughput material search scheme combining ML and DFT was developed and applied to screen 28, 125, 225 AA'BB'X ₃ X' ₃ double perovskites based on properties of the candidates (bandgaps and carrier-effective mass), the feasibility of their synthesis, toxicity, and cost.	SVR with a Gaussian kernel is the best with R^2_{train} , R^2_{test} , and RMSE of 0.89, 0.65, and 0.81 eV, respectively.
Im et al. 2019 ^[191]	Bandgaps and heat of formation	GBRT	Computation	Applied the GBRT model to a dataset of electronic structures for candidate halide double perovskites to predict heat of formation and bandgap.	The RMSE for heat of formation and bandgap is 0.036 and 0.322 eV.
Saidi et al. 2020 ^[117]	Bandgaps and structure	Convolutional neural network (CNN)	Computation	A hierarchical ML model was designed where each neural network element had a designated role in the prediction process from predicting complex features of the perovskites (e.g., lattice constant and octahedral tilt angle).	RMSE for lattice constants, octahedral angle, and bandgap are 0.01 Å, 5°, and 0.02 eV.
Stanley et al. 2020 ^[97]	Bandgaps, formation energy, and convex hull distance	KRR, convex hull	Computation	An ML approach with a generalized element-agnostic fingerprint was employed to predict key properties (bandgap, formation energy, and convex hull distance)	RMSE for bandgap, formation energy, and convex hull distance are 0.146, 0.015, and 0.011 eV atom ⁻¹ , respectively.
Gao et al. 2021 ^[135]	Bandgaps and stability	XGBoost, ANN, SVR	Computation	A strategy combining ML and DFT calculation was used to screen 5,796 inorganic double perovskites. Two novel lead-free inorganic double perovskites with suitable bandgaps, high thermal stability and good optical properties were obtained.	XGBoost has the best performance with R^2 of 0.956, r of 0.977, and MSE of 0.102.
Wu et al. 2021 ^[192]	Bandgaps and formability	GBM, SVM, RF, DT, KRR, KNN, etc.	Computation	204 stable perovskites with optimal bandgaps were selected from 11 370 mixed double-halide perovskites.	GBM is the best. For bandgap, R^2 is 0.97; and MSE is 0.06 eV.

Table 1. Continued.

Reference	Predicted properties	ML algorithms	Data source	Summary	Prediction accuracy
Mannodi-Kanakkithodi et al. 2022 ^[89]	Bandgaps and electronic properties	ANN, RF, GPR	Computation	The ML was used to predict the structural, energetic, electronic, and optical properties of an extensive dataset of 17 955 compounds, and perform high-throughput screening in terms of stability, bandgap, and defect tolerance.	ANN has the best performance.
Aristizabal-Ferreira et al. 2022 ^[96]	Bandgaps, refraction index, atomization energy	Crystal graph neural networks	Computation	Perovskites with suitable optoelectronic properties (bandgaps, refractive indexes, volumes of unit cells and volumetric densities, and atomization energies) were screened from 3840 organic–inorganic perovskites for photovoltaics.	The <i>r</i> of bandgap, refraction Index, atomization energy is 0.95, 0.97, and 0.99, respectively.
Zhao et al. 2022 ^[98]	Bandgaps and formation energy	eXtreme gradient boosting decision tree (XGDT), GBDT KNN, SVM, RF, LASSO, ANN, etc.	Computation	The nonlinear–ensemble model effectively describes the nonlinear relationship between material features and target property.	XGDT is the best. For formation energy and bandgap, RMSE is 0.269 and 0.106 eV.
Yang et al. 2021 ^[136]	Bandgaps and formability	SVM, LR, RF, BPANN	Experiment	The SVC classifier filters out perovskite candidates from enormous virtual samples. Then, the bandgaps of candidate perovskites are predicted by the SVR. Finally, 60 promising oxide double perovskites for PSCs were screened out from 6529 virtual samples.	SVM is the best. For formability, the accuracy is 96.8%. For bandgap, <i>r</i> , MSE, and MAE are 0.919, 0.173 eV, 0.286 eV.
Xu et al. 2018 ^[93]	Formability	SVM, DT, and kNN	Computation	MLs were used to predict the correlation between perovskite formability and intrinsic properties such ionic radii, electronegativity, atomic number, tolerance factor, and octahedral factor.	Nearly 90% prediction accuracy for 16 applied ML models.
Jain et al. 2019 ^[194]	Formability	SVM	Computation	MLs were used to predict the formability of 454 inorganic perovskites and integrated it with DFT to determine bulk structural optimizations and electronic structure predictions.	Accuracy is 95%.
Zhang et al. 2021 ^[81]	Formability	XGBoost, KNN, SVM, GBM	Computation	ML was combined with a shapley additive explanations (SHAP) approach to screen 18 560 virtual samples based on formability. The SHAP analysis reveals the relation between features and formation probability.	XGBoost has stronger predictive performance with a higher accuracy of 95%.
Pilania et al. ^[195]	Formability	SVM	Experiment	After exploring a wide range of features, ionic radii, tolerance factor, and octahedral factor were identified to be the most important factors.	The accuracy is 92.1%.
Talapatra et al. 2021 ^[92]	Formability and thermodynamic stability	RF	Computation	ML models were used to predict the formability and thermodynamic stability of perovskite oxides. From the intersection of large chemical space of formable and stable oxide perovskites, 414 compositions were identified.	For formability and stability, the accuracy is 94% and 94.1%; precision is 93.4% and 93.3%.
Schmidt et al. 2017 ^[90]	Thermodynamic stability	Extremely randomized tree (ERT), ANN, AdaBoost	Computation	Thermodynamic stability was obtained through predicting the	ERT is the best. MAE is 121 meV atom ⁻¹ .

Table 1. Continued.

Reference	Predicted properties	ML algorithms	Data source	Summary	Prediction accuracy
Park et al. 2019 ^[122]	Stability	RF, generalized linear model, DL, GBM	Computation	energy above the convex hull of 20 000 randomly selected perovskites.	The R^2 and MAE for enthalpy change are 0.99 and 0.0011 eV ion ⁻¹ , respectively.
Li et al. 2019 ^[91]	Decomposition energies	KRR	Computation	ML mapping between the perovskite stability and compositional ionic radii was established, which was validated by the experimental formability.	R^2 is 0.92 and RMSE is 34.38 meV atom ⁻¹
Mannodei-Kanakkithodi et al. 2019 ^[101]	Defect formation energies	Correlation analysis	Computation	They studied the density of states and energy levels related to the substitutional defect of partial substitution of Pb in MAPbBr ₃ . The substitution by Zr, Hf, Nb, and Sc, and Sb can shift the Fermi level and change the perovskite conductivity.	MAE between 0.2 and 0.4 eV
Hu et al. 2022 ^[102]	Ion adsorption energy	KNN, Kriging, RF, Rpart, SVM, XGBoost	Computation	The ion adsorption of 11 976 2D A ₂ BX ₄ perovskites was predicted by ML and high-throughput calculations. The ML-predicted virtual design space was used to select stable lead-free ion/perovskite systems with suitable bandgaps and halogen mixing features.	XGBoost achieved the best accuracy with r of 0.97, and R^2 of 0.93.
Sun et al. 2019 ^[84]	Perovskite structures	Fully connected deep neural network	Experiment	They characterized 75 perovskite compositions with bandgaps between 1.2 and 2.4 eV and then utilized an NN to classify compounds into 0D, 2D, and 3D structures.	Cross-validated accuracy of 90% is achieved.
Oviedo et al. 2019 ^[196]	Perovskite structures	All convolutional neural network	Experiment	The crystallographic dimensionality and space group were predicted from limited number of thin-film X-Ray diffraction (XRD) patterns. The scarce data problem is overcome by coupling ML with a model-agnostic, physics-informed data augmentation strategy.	Accuracies for dimensionality and space group classification are 93% and 89%.
Xie et al. 2020 ^[197]	Perovskite structures	SVM, DT, RF, ANN, SR-SISSO	Experiment	By relating physical parameters (electronegativity, atomic radii) with mathematical operations, they identified an analytical equation predicting the octahedral tilting for 49 perovskite oxides.	SR-SISSO has the best performance with accuracy of 86.7%.
Lyu et al. 2021 ^[139]	Perovskite structures	LR, SVM, KNN, DT, Gaussian Naïve Bayes	Experiment	Four structural features, including steric effect index, eccentricity, largest ring size, and hydrogen-bond donor, had been identified as the key factors. A quantified equation was created to calculate the probability of forming 2D perovskite for a selected amine on the basis of the four features.	SVM has the best performance with accuracy of 82%.
Massuyeau et al. 2022 ^[198]	Perovskite structure	RF, CNN	Experiment	The structure type was identified from powder XRD of new hybrid lead halides. The distances between Pb	CNN has the best performance with accuracy of 92%.

Table 1. Continued.

Reference	Predicted properties	ML algorithms	Data source	Summary	Prediction accuracy
Li et al. 2019 ^[44]	PCE	LR, KNN, SVM, RF, ANN	Experiment	ions and the volume of unit cells were identified as key factors for ML. ML was used to optimize material composition, develop design strategies, and predict the performance of PSCs. New perovskite compositions were experimentally synthesized to validate the practicability of the model.	ANN is the best. For PCE, RMSE and r are 3.23% and 0.8, respectively.
She et al. 2021 ^[159]	PCE	DT, ET, RF, XGBoost, AdaBoost, GBM	Experiment	ML approach was used to explore strategies for PCE improvement. Efficient dopants of electron-transport layer and features of upgrading PCE were studied through prediction of PCEs.	RF is the best. The accuracy, RMSE, and R^2 are 66.48%, 3.81%, and 0.43, respectively.
Liu et al. 2022 ^[94]	PCE	LR, KNN, SVR, RF, MLP, GBDT, XGBoost	Experiment	The factors affecting the PCE were explained and analyzed through the ML and SHAP. Among the 13 features, the content of FA^+ plays the most important role in improving the PCE of PSCs.	XGBoost is the best for bandgap with r of 0.96. RF is the best for PCE with r of 0.86.
Nakanishi et al. 2022 ^[95]	PCE	LR	Experiment	They proposed a multivariate analysis for experimental screening using 133 types of stable $\text{A}_2\text{Sn}(\text{IV})_6$ OD pseudoperovskites to predict the PCE of $\text{ASn}(\text{II})\text{I}_3$. A new $(\text{FA}_{0.92}\text{IM}_{0.08})_{0.9}\text{PEA}_{0.1}\text{SnI}_3$ Sn-PSC with a PCE of 7.22% was identified.	Experimental validation of 0.982 is achieved.
Liu et al. 2022 ^[166]	PCE	LR, RF, XGBoost ANN	Experiment	ML was employed to predict the relation between the PCE and interface passivation material at the atomic level, as well as rapidly screen interface materials for PSCs. The predictions are further validated using DFT and experiments.	RF has the best performance with RMSE of 0.64% and r of 0.9.
Cueto et al. 2022 ^[160]	PCE	KRR	Experiment	The effect of HTM on the PCE of the PSCs was predicted. Features describing the properties of the HTM (fingerprints, structural properties, electronic properties, and additives), as well as the perovskite type and cell architecture, were used.	For homogeneous database and heterogeneous database, RMSE is 2.7% and 3%.
Jiang et al. 2021 ^[176]	Current density–voltage (J – V) curves	GPR	Experiment	ML was trained with J – V curves of PSCs with varying alkali metal doping concentrations. Then the model was used to predict the J – V curves of PSCs with different doping concentrations.	GPR shows R^2 of 0.998, RMSE of 0.305 and MAE of 0.209.
Odabaşı et al. 2020 ^[103]	Hysteresis and reproducibility	Association rule mining	Experiment	Association rule mining was used to evaluate the impact of materials and deposition techniques on the reproducibility and hysteresis of PSCs.	
Priya et al. 2021 ^[100]	Conductivity and the type of charge carriers	LR, SVM, ANN, RF, XGBoost, etc.	Experiment	Average ionic radius, minimum electronegativity, minimum atomic mass, minimum formation energy of oxides for all B-site, and B-site dopant ions of the perovskite were the crucial	XGBoost is the best. For conductivity, R^2 and RMSE are 0.99 and 0.24 Scm^{-1} .

Table 1. Continued.

Reference	Predicted properties	ML algorithms	Data source	Summary	Prediction accuracy
Zhai et al. 2022 ^[104]	Area-specific resistance	SVR, GPR, RF, ANN, etc.	Experiment	predictors for conductivity and the type of charge carriers.	ANN has the best performance with MSE of $0.0131 \Omega\text{cm}^2$.
Zhang et al. 2022 ^[177]	External quantum efficiency	CNN, DNN, XGBoost, SVM, etc.	Experiment	The ionic Lewis acid strength was introduced as an effective physical descriptor for the prediction of area-specific resistance of perovskite oxides.	DNN has the best performance with average accuracy up to 96%.
Hartono et al. 2020 ^[180]	Stability	LR, KNN, SVM, RF, GBDT, MLP	Experiment	Multiple fingerprints as input were used to train the enhanced molecular information model to maximize the information of the molecules and significantly mitigate the duplicated problem.	RF has the best performance with RMSE of 70.8 min.

consistency between the trend of feature influence obtained from the model explanation using SHAP and the results of data analysis. The SHAP analysis shows that the ionic radii and lattice constant of the B site are positively related to formability, while the ionic radius of the A site, the tolerant factor, and the first ionization energy of the B site have negative relations.

3.2. Properties Predicted by ML

The properties of perovskite materials and solar cells can be predicted using ML, which facilitates the design of novel materials and the optimization of PSCs based on these predicted properties. Experiments can also be tailored and executed based on these predictions. Table 1 provides a summary of the properties forecasted through ML.

3.2.1. Perovskite Structure

The perovskites adopts a network of corner sharing BX_6 octahedra, encapsulating A-site cations in the ABX_3 formula, where A is a monovalent organic or inorganic cation such as $\text{HC}(\text{NH}_2)_2^+(\text{FA}^+)$, $\text{CH}_3\text{NH}_3^+(\text{MA}^+)$, and Cs^+ ; B represents a divalent metal cation (e.g., Pb^{2+} , Sn^{2+}); and X represents a monovalent halide anion (Br^- , I^-).^[82] The Goldschmidt tolerance factor and the octahedral factor are widely used to identify the formability of perovskite structure. The Goldschmidt tolerance is calculated using the equation $t = (r_A + r_X)/[\sqrt{2}(r_B + r_X)]$, where r_A , r_B , r_X are the ionic radii of the A, B and X species. The octahedral factor is determined by $\mu = r_B/r_X$. For 3D halide perovskites, empirical observations indicate that the t typically range from 0.813 to 1.107, while μ falls within the range of 0.442 to 0.895. When t falls within the range

0.9–1, perovskite exhibits a cubic phase. It adopts an orthorhombic phase when t is between 0.813 and 0.9, and a tetragonal phase when t is between 0.9 and 1.07.^[83] Since the A, B, and X sites can accommodate ions of varying sizes, resulting in the flexibility of composition. Therefore, perovskites can exhibit 3D structure of continuous corner sharing BX_6 octahedra, 2D-layered structure with corner sharing octahedra separated by cations layers, 1D structure with octahedra connected in a chain, and 0D structure with individual octahedra. Sun et al.^[84] utilized a fully connected deep NN to classify perovskite materials into 0D, 2D, and 3D structures based on experimental X-Ray diffraction data, which is 90% accuracy and more than ten times faster than human analysis.

3.2.2. Optical Properties

It is widely recognized that the bandgap is a critical determinant of PCE. Hence, the bandgap serves as an important evaluation criterion when screening perovskite materials for PSCs. Although traditional theoretical calculations can provide bandgaps of perovskite materials, obtaining precise analyses for large quantities of complex materials remains challenging. ML can be integrated with traditional computations, such as DFT calculations, to predict the bandgap of perovskites, thereby accelerating the screening of perovskite materials and advancing research in the field of PSCs. The bandgaps of perovskites is determined by factors such as perovskite composition (atomic radii, electron affinities) and structure (perovskite dimensionality, Goldschmidt tolerance factor, and the octahedral factor).^[85] Theoretical calculations are employed to obtain bandgaps and corresponding influencing parameters. Then, ML models are trained using the data produced by theoretical calculations to predict bandgaps

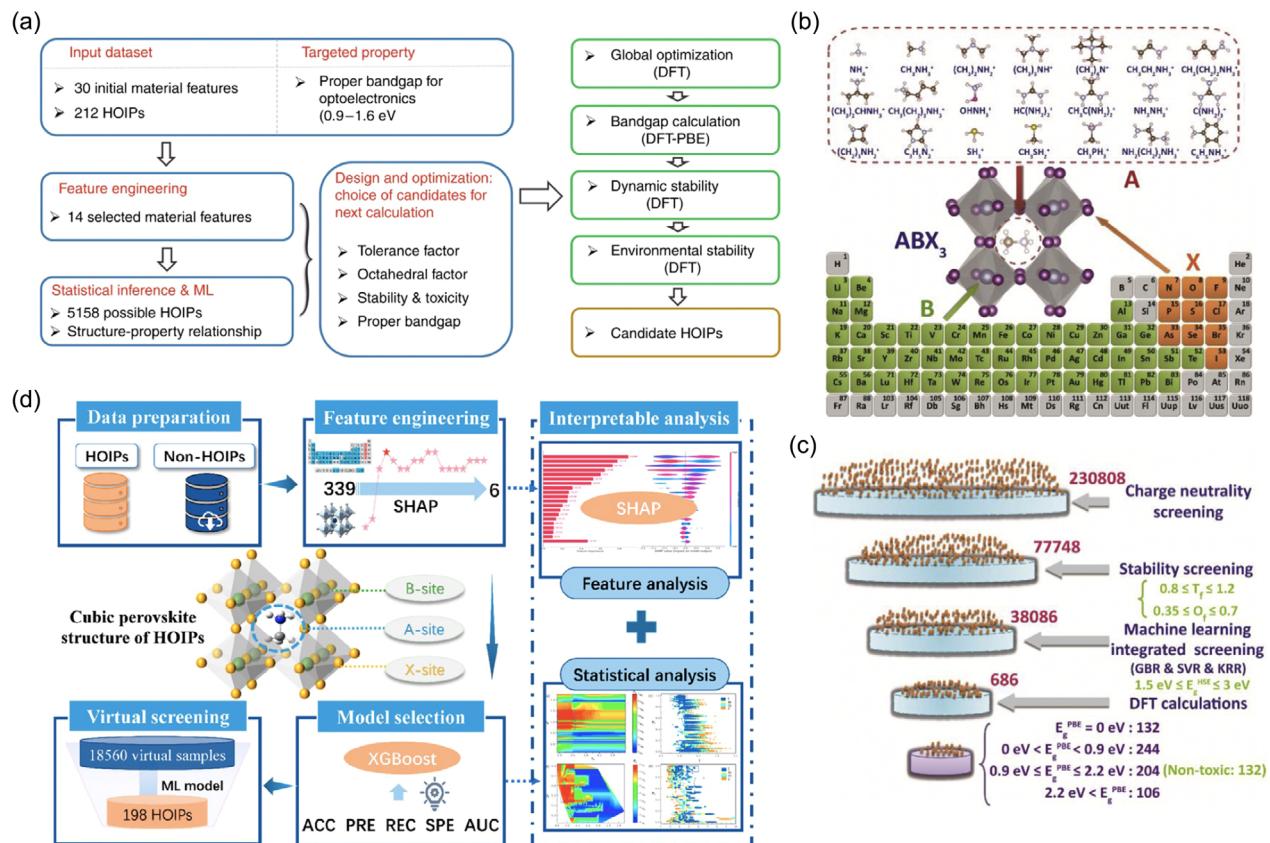


Figure 2. a) The material design framework combined with ML and density-functional theory (DFT) to discovery novel Pb-free hybrid organic–inorganic perovskites (HOIPs). Adapted with permission.^[77] Copyright 2018, Springer Nature. b,c) Schematic design framework combined with ML and DFT for predicting novel HOIPs. Adapted with permission.^[78] Copyright 2019, Elsevier. d) Flowchart used to predict the formability of HOIPs. Adapted with permission.^[81] Copyright 2021, American Chemical Society.

and identify promising candidates for PSCs. Most of studies on bandgap prediction using ML are based on computational data. For example, Rath et al.^[86] used XGBoost to classify bandgap types in inorganic perovskites. Their SHAP analysis revealed that the absence of transition metals and elements from groups IIIA to VIIIA with atomic number greater than 20 enhances of the direct bandgap perovskite. Park et al.^[87] employed ML to establish relationships between octahedral deformation and bandgap, as well as created a similarity map among all the computed compounds. Moreover, the data points are also collected from published articles. Cai et al.^[88] established the asymmetrically bowing relationship between bandgap and Sn composition using a novel ML approach based on forward-reverse framework. An optimized Sn:Pb composition ratio of approximately 0.6 is finally obtained for high-performance PSCs.

3.2.3. Stability

The ML studies focus on predicting the thermodynamic stability and decomposition energies of perovskite materials.^[89–91] Talapatra et al.^[92] explored the large chemical space of single and double oxide perovskites to identify compositions that could form stable perovskites. They employed a multipronged

approach to examine the formability and thermodynamic stability of oxide perovskites. The intersection of this extensive chemical space led to the identification of 414 compositions, which are proposed as the most promising candidates for future experimental synthesis of novel oxide perovskites. In the investigation of PSC stability, MLs are employed to establish correlations between influential factors (e.g. capping layer, grain size) and the outcomes of stability testing, thereby offering valuable guidance for the development of stable PSCs.^[93]

3.2.4. PCE

The efficiency of PSCs is commonly assessed by calculating the PCE based on photovoltaic performance metrics, including open-circuit voltage (V_{OC}), short-circuit current density (J_{SC}), and fill factor (FF). These metrics are highly dependent on the quality of materials, interfaces, and device structure. Liu et al.^[94] developed ML models using experimental data from the literature to predict the performance of PSCs. These models effectively predict PCE and demonstrate a strong correlation with experimental results. This exceptional prediction performance of ML can enhance the efficiency of PSC fabrication and reduce the probability of experimental errors. Additionally, these models coupled with the

SHAP offer insights into the factors influencing PSC efficiency. Among the 13 examined features, FA^+ content emerges as the most crucial contributor to enhancing PCE. Nakanishi et al.^[95] employed a set of 133 distinct $\text{A}_2\text{Sn}(\text{IV})\text{I}_6$ 0D pseudoperovskites to predict the PCE of $\text{ASn}(\text{II})\text{I}_3$, with A representing a combination of ternary and quaternary organic cations. The strong correlation between the model and experimental validation enables the discovery of a novel Sn-based perovskites, $(\text{FA}_{0.92}\text{IM}_{0.08})_{0.9}\text{PEA}_{0.1}\text{SnI}_3$, boasting an impressive PCE of 7.22%.

3.2.5. Other Properties

ML have also been applied to predict properties thermodynamic properties such as atomization energies^[96] and formation energies.^[97,98] Electronic properties, such as carrier effective mass,^[99] conductivity,^[100] and density of states,^[101] have also been the predicted results of ML. Moreover, ML also plays a crucial role in predicting the ion adsorption energy,^[102] hysteresis and reproducibility,^[103] area-specific resistance,^[104] etc.

4. Implementation of ML for the Advancement of Efficient and Stable PSCs

It should be emphasized that the performance and stability of PSCs depend not only on optoelectronic properties of perovskites, but also on the entire device structure. The optoelectronic properties (such as light absorption, charge-extraction capabilities) and stability of perovskite materials are strongly related to chemical composition, microstructure, and processing. As the device investigations often involve electron-transport layers (ETL), hole-transport layers (HTL), and interfacial layers,^[105–107] the applications of ML in developing efficient and stable PSCs can be grouped into four categories: screening materials, optimizing fabrication processes, optimizing device structures, and understanding the mechanism.

4.1. Application of ML in Rational Design of Perovskite Materials

Despite the rapid development of PSCs, their instability remains a significant obstacle to their industrial applications. Stable perovskite materials play a pivotal role in achieving exceptional stability for PSCs. Numerous works have been put into exploring stable, efficient, and environment friendly perovskite materials.^[108–110] The huge compositional space of perovskites enable a series of optoelectronic properties including optical bandgap, absorption coefficients, exciton binding energy, and charge-carrier transport properties. For example, the bandgaps of perovskites are mainly determined by the metal B site, and halide X site. In the MAPbI_3 structure, the 6s orbital of Pb element and the 5p orbital of I element form the antibonding state that determines the valence band maximum, while the conduction band minimum (CBM) is mainly contributed by the 6p states of Pb. In this case, the bandgaps can be tuned over a wide range by compositional substitution of the B site metal cation and X site halide anion.^[111,112] A-site cations can also indirectly

influence the bandgaps of perovskites by affecting the lattice structure (tilting the BX_6 octahedra, contracting or expanding the lattice isotropically).^[113] Furthermore, composition of perovskite will affect the lattice constant and Goldschmidt tolerance factor of the material, thus changing the optoelectrical properties of perovskites and the performance of corresponding PSCs.^[114,115]

Therefore, it is of great importance to further understanding the relationship between the composition and perovskite properties (e.g. bandgap, formability, thermal stability, etc.) to achieve high-performance and stable PSCs. A wide range of structural, electronic, and optical properties is derived through DFT calculations. To rationally design perovskites for PSCs, these material properties serve as training data and ML is trained on the dataset to perform target prediction. This section provides a review of the application of ML in the rational design of various perovskite materials, including 3D HOIPs, 3D all-inorganic perovskites, 3D double perovskites, and 2D-layered perovskites.

4.1.1. 3D HOIPs

Halide perovskites can be classified into two categories, namely HOIPs and all-inorganic perovskites, based on the composition of A-site cation in the ABX_3 crystal structure. It is crucial to explore new nontoxic HOIPs with high efficiency and high stability. Theoretically, there are a huge number of possible HOIPs with numerous organic cations and metal cations to be selected.^[116] ML plays a key role in discovering high-performance 3D HOIPs for photovoltaics since the traditional trial-and-error methods are usually inefficient especially when thousands of candidates are treated. Lu et al.^[77] proposed a target-driven approach that combines ML with DFT to screen for undiscovered HOIPs based on their bandgap. The innovative method aims to address the challenges of toxicity and poor environmental stability in HOIPs. Six kinds of ML algorithms (i.e., multilayer perceptron (MLP), GBR, KRR, SVR, GPR, and DT) were applied to predict bandgaps, among which GBR showed the best predictions on the true bandgap values. A structure–property relationship mapping the HOIPs bandgap was established and six orthorhombic lead-free HOIPs with proper bandgap and high stability were screened out from 5158 unexplored 3D HOIPs. Bandgap prediction is of vital importance in screening potential undiscovered HOIPs. Saidi et al.^[117] used convolutional neural network (CNN) model to predict the bandgaps of perovskite materials and found that the lattice constant and octahedral tilt angle play an important role in the prediction of bandgaps of MHPs (Figure 3a–c). The bandgap and structural parameters of 862 candidates were calculated using DFT. A hierarchical CNN model was then constructed to predict the bandgaps. The carefully designed hierarchical ML model proves to be highly efficient in handling uneven dataset and demonstrates superior accuracy in predicting properties of perovskites and designing new perovskite materials compared to straightforward methods. Furthermore, the bandgaps of 3D mixed cation perovskites predicted by ML fluctuates within ± 0.5 eV. To achieve higher accuracy in the bandgap prediction of hybrid materials, the variation originating from the local deformation and a crystal field effect should be studied. Park et al. investigated the influence of

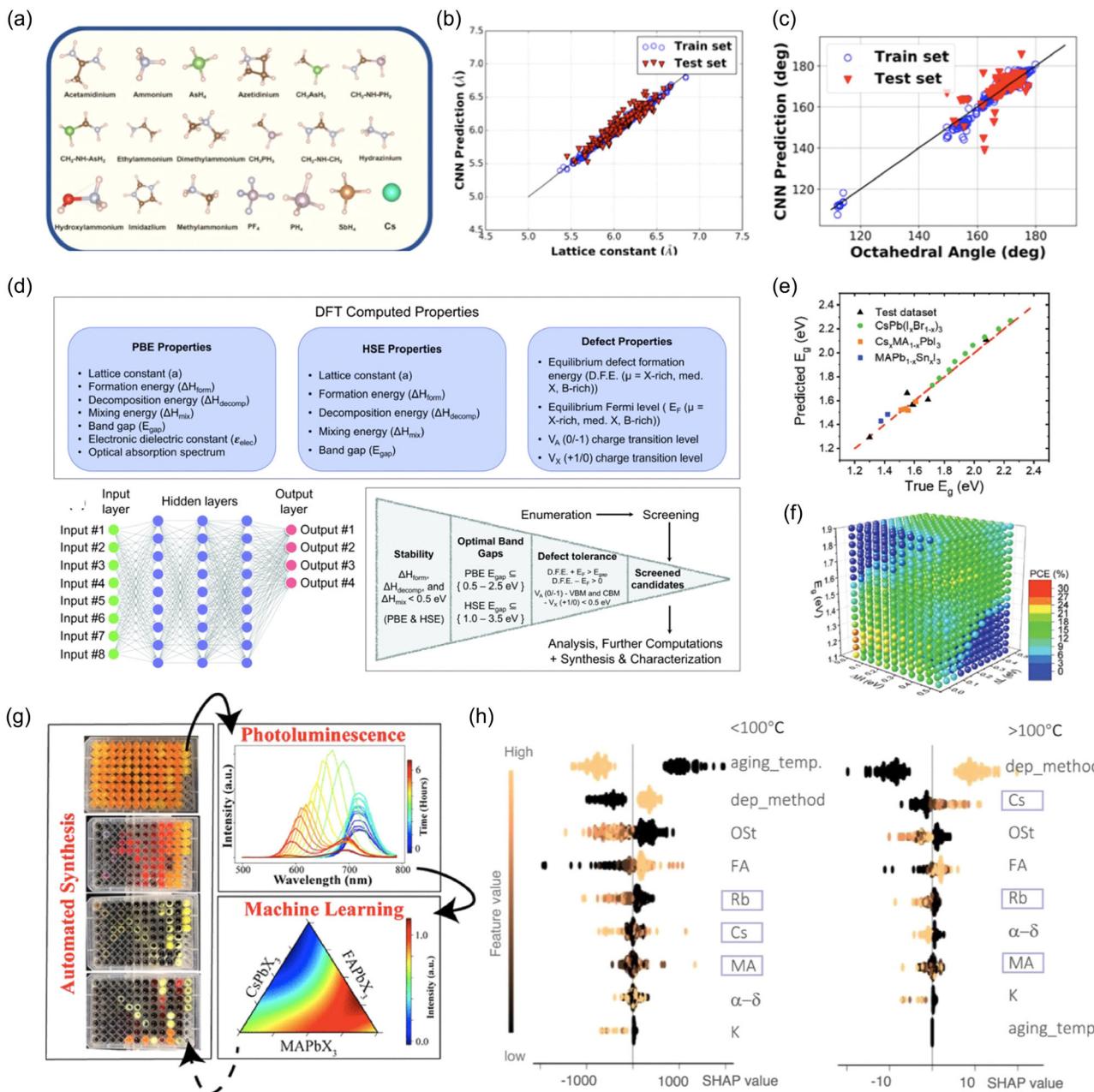


Figure 3. a) The material design space of A site, including Cs and 18 different organic molecules. Representative plots for the prediction of convolutional neural networks (CNNs) for b) lattice constant and c) octahedral angle. Adapted with permission.^[117] Copyright 2020, Springer Nature. d) The screening process of 17 955 perovskites using neural network framework. The 16 properties of 229 perovskites calculated by DFT are used as labels for the ML models. Stability, optimal bandgap, and defect tolerance are used as filtering conditions. Adapted with permission.^[89] Copyright 2022, Royal Society of Chemistry. e) Performance of the artificial neural network (ANN) model showing the correlation between the true data and predicted results. f) The 4D plot of power conversion efficiency (PCE) distribution with respect to bandgap, ΔH , and ΔL . Adapted with permission.^[44] Copyright 2019, Wiley-VCH. g) Schematic of the automated workflow for the combinatorial discovery of perovskite materials. Adapted with permission.^[119] Copyright 2020, American Chemical Society. h) The ranking of feature importance using gradient boosting decision tree regression and shapley additive explanation (SHAP) assessment, showing the impact of each cation and processing condition in descending order of importance. Adapted with permission.^[124] Copyright 2021, Springer Nature.

structural deformation on the bandgap of perovskite materials and developed a highly accurate predictor to predict the bandgap.^[118]

Most 3D HOIPs use mixed compositions with several monovalent cations or anions simultaneously. Therefore, ML can be

employed to tune the composition of mixed perovskites and predict stable and high-performance 3D HOIPs for photovoltaic absorbers. Chan and coworkers^[89] developed a framework powered by high-throughput DFT calculations and ML to predict

mixed-cation perovskites (Figure 3d). DFT calculations were applied to compute the structural properties (lattice constants), optoelectronic properties (bandgaps, optical absorption), decomposition energies, formation energies, and defect properties of 229 HOIPs. The NN models were used to predict the structural, energetic, electronic, and optical properties of an extensive dataset of 17 955 compounds, and screened 392 potential photovoltaic absorbers in terms of stability, bandgap, and defect tolerance. Data-driven design framework is promising for designing novel mixed compositions, which can be extended to a wider perovskite chemical space. Li et al.^[44] used ML models to optimize material composition, and predict the performance of PSCs. The ML model was built to predict the bandgaps of perovskite materials and guide synthesizing new perovskite material compositions (Figure 3e,f). Then, three kinds of new perovskite compositions, i.e., $\text{Cs}_x\text{MA}_{1-x}\text{PbI}_3$, $\text{MAPb}_{1-x}\text{Sn}_x\text{I}_3$, and $\text{CsPb}(\text{I}_x\text{Br}_{1-x})_3$, were experimentally synthesized to test the practicability of the model. In addition, different ML models were applied to discover the relationship between perovskite properties and PCE using experimentally obtained data from the literature. The small value of the RMSE and the good Pearson's coefficient suggest that the ML model provides guidelines for designing new perovskite compositions and new PSC devices. These findings show that ML is an efficient way to guide the prediction and fabrication of high-performing PSCs.

Moreover, ML provides deeper understanding of perovskite properties associated with the PSC which is almost impossible for a human to find from random data points. Higgins et al.^[119] used interpolative regression to explore long-term stability of four model perovskite systems: $\text{MA}_x\text{FAyCs}_{1-x-y}\text{PbBr}_3$, $\text{MA}_x\text{FA}_y\text{Cs}_{1-x-y}\text{PbI}_3$, $\text{Cs}_x\text{FA}_y\text{MA}_{1-x-y}\text{Pb}(\text{Br}_{x+y}\text{I}_{1-x-y})_3$, and $\text{Cs}_x\text{MAyFA}_{1-x-y}\text{Pb}(\text{I}_{x+y}\text{Br}_{1-x-y})_3$ in ambient conditions (Figure 3g). This study establishes automated experimental workflow based on combinatorial synthesis and rapid throughput characterization for accelerated discovery of large compositional space. Yan et al.^[120] used ML to predict five compositions of $(\text{FAPbI}_3)_x(\text{MAPbBr}_{2.8}\text{Cl}_{0.2})_{1-x}$ perovskites on the performance of solar cells for experimental guidance. The bandgap, the J_{SC} , and V_{OC} were predicted by the ML. The measured and predicated bandgaps of these perovskites show excellent consistency with the relative errors less than 2%. In addition, the experimentally measured J_{SC} and V_{OC} are consistent with the ML predictions, demonstrating the effectiveness of ML in experimental guidance. Jame et al.^[121] utilized an ML model to explore the optimum bromine doping concentration in the single-junction $\text{MASnI}_3_{-x}\text{Br}_x$ devices, yielding the highest predicted PCEs of 20.72% with a bromine concentration of 22.43%. Data-driven optimizations were performed on a total of 42 000 distinct devices, which were constructed using a solar cell capacitance simulator. The devices were investigated through variating bromine doping, bandgap, electron affinity, series resistance, back-contact metal, and acceptor concentration. This work shows that the ML-assisted analysis has great potentials in optimizing the doping concentration of mixed perovskites for PSCs.

The effect of composition on the stability of perovskites was systematically analyzed with ML model. Park et al.^[122] developed a series of ML models to investigate the impact of the A-site cation on the phase stability of the perovskite structure and identified the possibility of known cations incorporated in the

perovskite formulation. The effective atomic radius and the number of lone pairs residing on the A-site cation are sufficient features to describe the perovskite phase stability. This work demonstrates that conjunction of advanced electronic structure theory and ML analysis is an effective strategy for exploring a broad composition space. Ali et al.^[123] used a perovskite structure stability predictor (PSSP) ML model to predict the cubic phase stability of perovskites and synthesized a series of mixed-cation perovskites $\text{Cs}_x\text{MA}_{0.85-x}\text{DMA}_{0.15}\text{PbI}_3$ to confirm the findings. As the guidance of ML model, the cubic crystal structure of the double A-site cation perovskite ($\text{MA}_{0.85}\text{DMA}_{0.15}\text{PbI}_3$), which suffers from severe instability at room temperature, can potentially achieve good stability by adjusting the optimal mixture of A-site cations. This prediction-to-lab-scale synthesis of cubic phase lead halide perovskites guided by ML greatly speeds up the experimental process and reduces cost. Lu et al.^[64] established ML model to predict the relationship between various parameters (fabrication process, compositions) and device performance based on a dataset including 1072 devices from previously published articles. The ML model accurately predicts the PCE with an RMSE of 1.28% and a Pearson coefficient r of 0.768. Moreover, the feature importance was analyzed by SHAPs, and A-site cation has the most significant influence on device performance. According to DFT calculations and experiments, A-site cation alloying can effectively enhance the crystallinity of perovskite films, leading to reduced defect density and significantly improved device performance, consistent with the predictions from the ML model. Zhao et al.^[124] coupled high-throughput experimentation with ML to study the effect of A cations on stability and discovered a stability-reversal behavior at high/low-temperature (Figure 3h). Through switching the ageing temperature regime from high temperatures to low temperatures, the impact of the ratio of organic to inorganic cations is reversed. Specifically, organic cation (e.g., MA) is in fact a stability enhancer while inorganic cation (e.g. Cs/Rb) is a stability killer below 100 °C. This underlying mechanism is governed by both change in activation energy and rate constant of decomposition by incorporation of multiple cations. They identified that incorporating at least 10 mol% MA and up to 5 mol% Cs/Rb is beneficial for the device stability at device-operating temperature (<100 °C). Teunissen et al.^[125] investigated the specific role of non-covalent bonds in contributing to the stability through a combination of first-principles calculations and the sure independent screening and sparsifying operator (SISSO) algorithm. The design of perovskites with suppressed ion migration is guided by non-covalent bonding interactions. They demonstrated that the factors affecting ion migration, in order of importance, are the type of halogen ions, the presence of π -X bonds, and the number of hydrogen bonds donors including C-H. Furthermore, the inclusion of F^- ions can significantly improve stability. This work emphasized the important role of composition engineering in improving the stability of perovskites from the perspective of non-covalent bonds.

4.1.2. 3D All-Inorganic Perovskites

All-inorganic perovskites have been a research focus because of their excellent thermal stability. In the case of 3D all-inorganic

perovskites, the A-site cation is an inorganic alkali metal ion (e.g., Cs⁺) instead of organic cations. Recently, the efficiency of all-inorganic PSCs have achieved as high as $\approx 22\%$,^[126] indicating high potential for future commercialization. However, all-inorganic CsPbX₃ perovskite suffers from phase instability under humidity, light soaking, and electrical field conditions, which are unable to meet the demand of photovoltaic applications.^[127–129] Therefore, MLs have been used extensively to investigate stable and high-performance inorganic perovskites.

MLs have been successfully used to predict the bandgap and stability of all-inorganic perovskites as well as determine the applicability for photovoltaic applications. As shown in Figure 4a, an ML approach employing a generalized element-agnostic fingerprint was used to rapidly screen inorganic perovskites based on bandgap, formation energy, and convex hull distance.^[97] A new database of 344 all-inorganic perovskites was generated by DFT calculations. The possible ions in these

inorganic perovskites were given by A = Cs⁺, Rb⁺, K⁺, or Na⁺, B = Sn²⁺ or Ge²⁺, and X = I⁻, Br⁻, or Cl⁻. The ML model was utilized to predict trends in entirely different chemical spaces, and perform rapid composition and configuration space sampling. Lu et al.^[130] developed a multistep screening method combining high-throughput DFT calculations and ML to explore all-inorganic perovskites with excellent stability and appropriate bandgaps. This method greatly reduces computational cost and ensures good calculation accuracy compared to the traditional trial-and-error methods in vast chemical space. After multiple-step predictions, 151 stable ferroelectric perovskites with suitable bandgap were screened out from 19 841 candidate compositions. In addition, the accuracy of ML-predicted results on both energy difference and bandgap regressions is over 90%, which is also verified by DFT calculations. This research demonstrates that ML models can effectively predict many properties of material at once and screen novel materials with multiple steps.

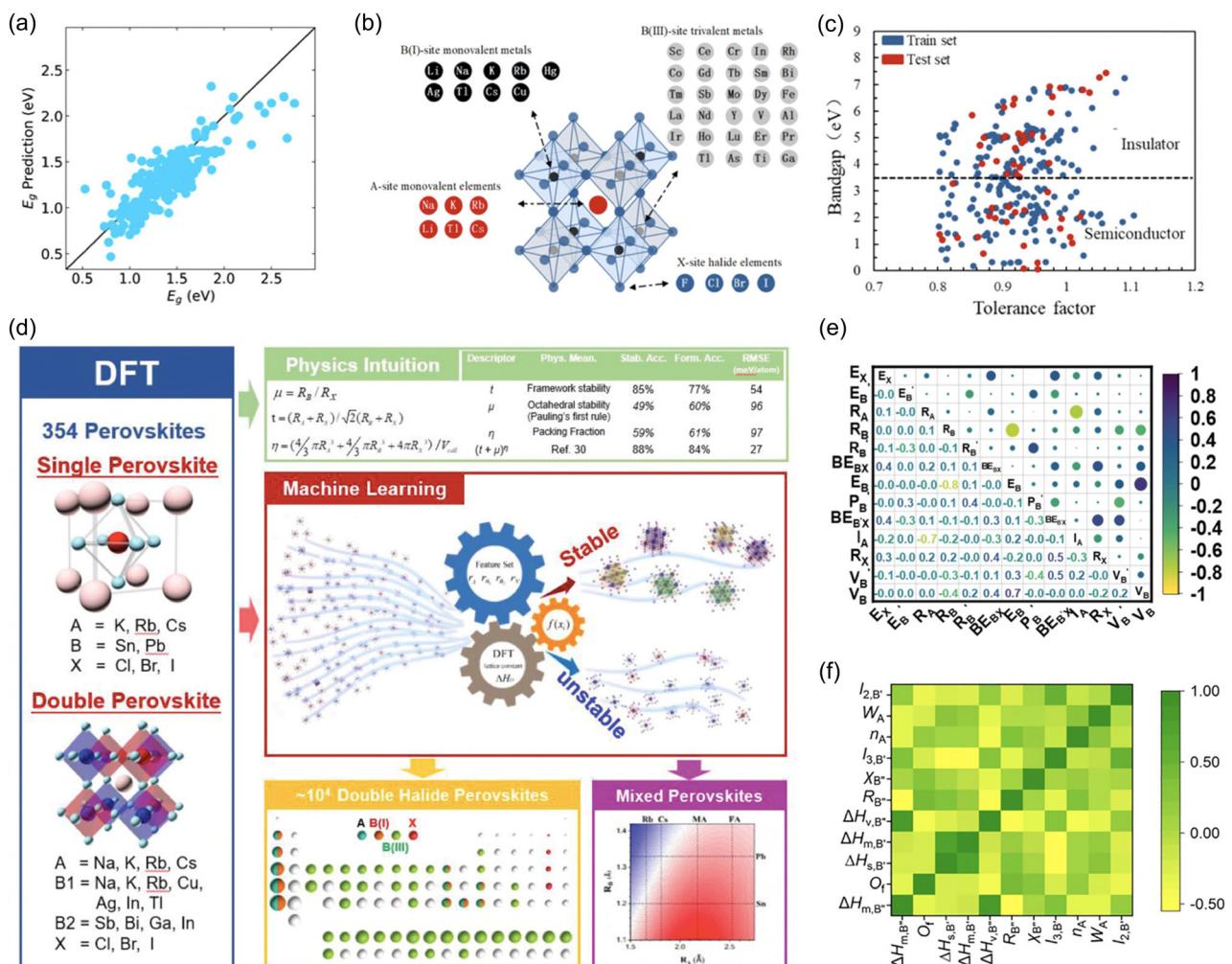


Figure 4. a) Bandgap prediction for caesium perovskites. Adapted with permission.^[97] Copyright 2020, Wiley-VCH. b) A total of 272 double-perovskite structures, including 6A-site elements, 9 B(I)-site elements, 29 B(III)-site elements, and 4X-site halogen elements. c) Data visualization of the tolerance factor and the bandgap. Adapted with permission.^[134] Copyright 2021, American Chemical Society. d) Schematic strategy of ML based on DFT for high-throughput stability engineering of perovskites. Adapted with permission.^[91] Copyright 2019, Wiley-VCH. e) Pearson correlation heat map of 13 features. Adapted with permission.^[135] Copyright 2021, Elsevier. f) Correlation of 11 features selected in SVR. Adapted with permission.^[136] Copyright 2021, Elsevier.

ML models are also developed to investigate the structural instability of inorganic perovskites induced by ionic defects. Yang et al.^[131] developed a deep learning potential to study the effect of ionic defects on the stability of CsPbI₃ perovskite. The deep learning potential was trained on DFT-calculated data. Furthermore, MD simulations with a 5000-atom system were performed to study the effects of defects on the stability of CsPbI₃ perovskite. In the case of surface defects, Cs defects is dominant factor to cause the instability of CsPbI₃ when the defect concentration is less than 15 %, while Pb defects take over when the defect concentration is over 20%. This work provides a new method to explore the correlation between ionic defect and structural stability of perovskites, establishing a theoretical foundation for designing high stable perovskites.

4.1.3. 3D Double Perovskites

Recently, halide double perovskites have gained popularity as potential light-absorbing materials for PSCs due to their high stability and low toxicity. Double perovskites (A₂BB'X₆) can be obtained by replacing B-site divalent metal cation with monovalent (e.g. Ag⁺) and trivalent (e.g. Bi³⁺) metal cations.^[128,132] However, the PCE of 3D double PSCs is much lower than conventional lead-based PSCs due to the wide bandgaps, low optical absorption coefficient, and high density of defects.^[133] Therefore, exploring new 3D double perovskites with superb optoelectronic properties is an effective strategy to improve the performance of PSCs.

To efficiently discover suitable double perovskite for photovoltaic application, researchers combine high efficiency ML and high accuracy DFT calculations to compensate for their respective shortcomings. For example, Yang and coworkers^[134] combined ML and DFT calculations to develop a goal-driven method to search for double-perovskite materials with suitable photoelectric properties and thermal stability (Figure 4b,c). Through a series of processes such as training excellent ML models, multi-condition combination screening, and DFT calculations, 10 promising double-perovskite materials (including Cs₂AgVBr₆, K₂AgVBr₆, Na₂AgCrBr₆, Rb₂AgBiBr₆, etc.) were selected from 16 400 double-perovskite candidates to guide subsequent experimental synthesis. Yao et al.^[135] employed a search strategy combining ML and DFT calculations to predict promising inorganic double-perovskite materials from 5,796 inorganic double-perovskites candidates (Figure 4e). They found that XGBoost regression exhibited higher prediction accuracy than ANN and SVR algorithm. Two novel double perovskites (i.e., Na₂MgMnI₆ and K₂NaInI₆) with suitable direct bandgaps were successfully obtained by XGBoost regression. Furthermore, DFT calculations were used to compute the electronic band structures of Na₂MgMnI₆ and K₂NaInI₆. DFT calculations also confirmed that the two new double perovskites have good optical properties and high thermal stability.

Moreover, ML models are employed to evaluate the formability and thermodynamic stability of double perovskites. Through high-throughput DFT calculations, Yin and coworkers^[91] established a database for ML training containing the decomposition energies of 354 double-halide perovskite candidates (Figure 4d). Then, a well-functioned ML model was trained over this

theory-based database to explore the underlying relationship between the structure, chemistry features, and the decomposition energies. To further validate ML-predicted results, they randomly calculated the decomposition energies of 74 unexplored double perovskites out of the 14 190 perovskites by applying DFT calculations, which were in good agreement with ML-predicted results. These studies suggest that combining ML and DFT calculations can rapidly and efficiently screen stable double perovskites from thousands of candidates.

In addition to halide double perovskites, researcher also employed ML approaches to screen new chalcogenide double perovskites for photovoltaic applications. Yang et al.^[136] successfully screened 60 promising oxide double perovskites for photovoltaic applications with narrow bandgaps (1.0–1.6 eV) from 6529 perovskite candidates using ML algorithms (support vector classification [SVC] and SVR). Among these materials, 19 double perovskite oxides have the bandgaps of 1.25–1.45 eV. The ML results show that Fe-, Ni-, Sc-, and Co-occupying B'-site and Bi-, Ta-, Nb-, Sb-, V-, and Mn-occupying B'-site are most likely to form narrow-bandgap oxide double perovskites. This study presents an example for the design and discovery of new double-perovskite oxides for PSCs via ML (Figure 4f). Recently, Min and coworkers^[137] employed ML and high-throughput DFT calculations to predict the synthesizability of double-perovskite oxides. They constructed the dataset by extracting material properties from the MP database, and then employed a regression model and a classification model to predict the formation energy and convex hull energy of general inorganic materials. Both regression and classification models are employed to estimate the possible synthesizability of 11 763 double-perovskite oxides, and their accuracy is further validated by high-throughput DFT calculation. This work demonstrates the potential of ML methods in facilitating the exploration and identification of synthesizable double-perovskite oxide structures.

4.1.4. 2D-Layered Perovskites

The 2D-layered halide perovskites are considered more stable than 3D halide perovskites. According to the spacer cation, 2D-layered halide perovskite can be divided into Ruddlesden-Popper-phase perovskite with monovalent ion and chemical formula of A'₂A_{n-1}B_nX_{3n+1} as well as divalent-ion-based A'A_{n-1}B_nX_{3n+1}-type Dion-Jacobson-phase perovskite.^[19,112] Even though the environmental stability of 2D-layered perovskites is much improved, the bandgaps are generally larger than ≈ 1.7 eV.^[138] Screening stable 2D perovskite materials with appropriate bandgaps is urgently needed for developing highly efficient 2D PSCs.

The dimensionality of halide perovskites is significantly influenced by the molecular structures of organic cations. Recently, ML models are employed to identify the dimensionality of HOIPs by predicting the structure-directing effect of organic cations. Lyu et al.^[139] reported an ML-assisted approach to study the effects of ammonium cations on the dimensionality of HOIPs (Figure 5a–c). They selected 86 reported ammonium cations from literature and classified them into two categories ("2D"-forming and "non-2D"-forming) based on the dimensionality

of their perovskites. After ML models were trained and tested, four structural features (steric effect index, largest ring size, eccentricity, and hydrogen-bond donor) were identified as the key factors. Subsequently, a prediction equation was created to evaluate the possibility of forming 2D perovskite on the basis of the four features. Finally, several untested ammonium cations were selected to verify ML predictions by the synthesis of perovskites single crystals from these ammonium cations.

ML method is also an efficient tool to rapidly predict the bandgaps of 2D-layered perovskites. Marchenko et al.^[140] developed an open-access database with 515 experimentally investigated 2D HOIPs, which contains 180 different organic cations, 10 metal cations (Pb, Sn, Bi, Cd, Cu, Fe, Ge, Mn, Pd,

and Sb), and three halogens (I, Br, and Cl). The database contains a geometrical and crystal chemical analysis of the structures (such as space group symmetry, type of layered perovskite structure, bond lengths and angles, bandgaps, etc.), which are useful for studying quantitative structure–property relationships (Figure 5d,e). The ML model is trained on the database for predicting bandgaps with accuracy within 0.1 eV. In addition, the predicted bandgaps decrease with an increase of the number of inorganic layers and M–X–M angles.

In addition to 2D HOIPs, 2D all-inorganic perovskites have attracted great research attention recently. ML models are also employed to screen 2D-layered inorganic perovskites as promising photovoltaic absorbers. Hu et al.^[141] combined

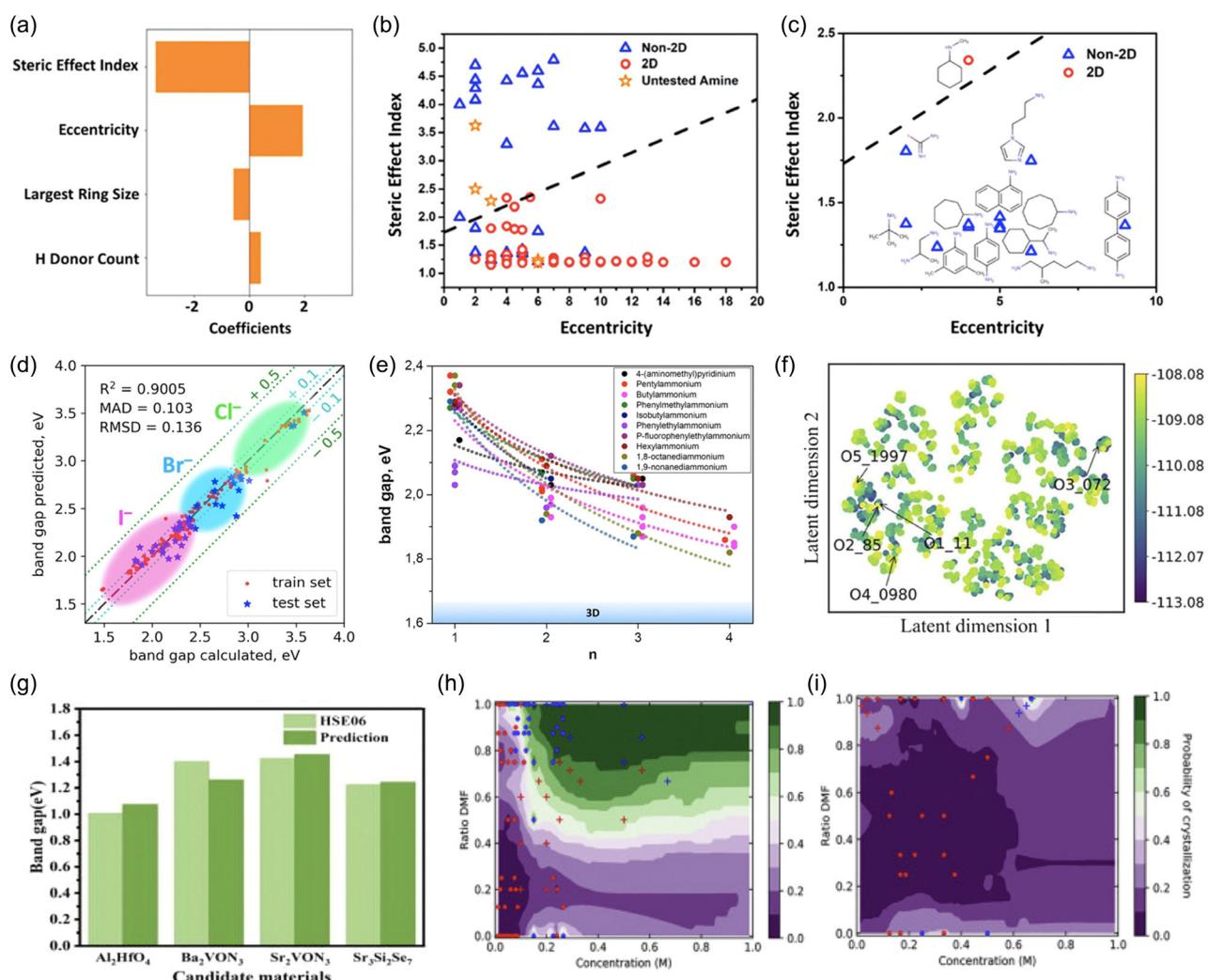


Figure 5. a) Feature coefficients from the linear regression (LR) model with ℓ_1 penalty. b) Data distribution based on steric effect index and Ec with a borderline curve of $P = 0.5$. c) Structures of unseparated amines. Adapted with permission.^[139] Copyright 2021, American Chemical Society. d) Bandgaps of 2D hybrid halide perovskites calculated by DFT and predicted by the ML model. e) Bandgaps calculated by the first ML model for a series of 2D hybrid halide perovskites (with different spacer organic cations). Adapted with permission.^[140] Copyright 2020, American Chemical Society. f) The t-distributed stochastic neighbor embedding representation of more than 3000 disorder structures. Adapted with permission.^[141] Copyright 2020, American Institute of Physics. g) The comparison of the bandgaps for the four predicted materials between ML prediction and Heyd–Scuseria–Ernzerh DFT calculations. Adapted with permission.^[142] Copyright 2021, American Institute of Physics. Results of the parameter exploration for the h) $(\text{PEA})_2\text{PbBr}_4$ chemical space and i) $(3\text{-PLA})_2\text{PbBr}_4$ chemical space. The blue and red dots indicate crystallization and no crystallization, respectively, and were used in the training of the algorithm, whereas the crosses are from a new experiment. Adapted with permission.^[41] Copyright 2020, Elsevier.

high-throughput DFT calculations and an ML model with ANN algorithm to design a 2D all-inorganic perovskite $\text{Ca}_6\text{Sn}_4\text{S}_{14-x}\text{O}_x$ ($x=1-5$) for solar cell applications (Figure 5f). Based on the arrangement of mixed S/O atoms, more than 3000 derivative structures of $\text{Ca}_6\text{Sn}_4\text{S}_{14-x}\text{O}_x$ ($x=1-5$) were selected to screen promising 2D perovskites with appropriate direct bandgaps. They found that $\text{Ca}_6\text{Sn}_4\text{S}_{14-x}\text{O}_x$ ($x=4$ and 5) have high absorption coefficients and appropriate bandgap values (1.19–1.64 and 1.02–1.47 eV), suggesting that the 2D perovskites are potential candidates as photovoltaic absorbers. Feng et al.^[142] applied ML and DFT calculations to accelerate the discovery of 2D perovskites as photovoltaic materials by predicting the PCEs of 2D PSCs. Based on the ML prediction, stability test, optical absorption, and theoretical PCE evaluation, two promising photovoltaics, i.e., Sr_2VON_3 and Ba_2VON_3 , are discovered to be suitable photovoltaic materials with PCE as high as 30.35% and 26.03%, respectively (Figure 5g).

ML was used to combine with high-throughput experimental framework to discover new perovskite single crystals. Kirman et al.^[41] synthesized a new perovskite single crystals of $(3\text{-PLA})_2\text{PbCl}_4$ with the aid of high throughput experimentation and ML method (Figure 5h,i). The authors perform high-throughput syntheses of perovskite single crystals with a protein crystallization robot and characterize the outcomes using CNN-based image recognition. The ML model was then used to predict the optimal conditions for the synthesis of the new perovskite single crystal of $(3\text{-PLA})_2\text{PbCl}_4$. The perovskite crystal exhibits strong blue emission, demonstrating the applicability of combining high-throughput experimentation and ML in discovering new optoelectronic materials.

The emergence of 3D/2D hybrid perovskite, which combine the high efficiency of a 3D perovskite with the superior stability of 2D perovskite, present a promising avenue for PSCs. Yilmaz et al.^[143] constructed a dataset containing 599 data points from 146 publications on 2D/3D PSCs. This dataset was then analyzed for PCE and stability using ML algorithms of XG Boost regression, RF regression, ANN, and association rule mining (ARM) to identify the descriptors that lead to highly efficient and stable 2D/3D device. According to the ARM analysis for the stability, the devices with layered perovskite structures are more stable, while butylammonium and FA-Cs-mixed cations are shown to be the most stable 2D and 3D cations, respectively.

4.2. ML-Assisted Optimization of Perovskite Fabrication Process

Fabrication process of perovskite thin film is a major determinant of the morphology and crystallization quality of the film, which are the most important factors affecting device performance and stability. Recent studies have shown that perovskite thin films with full coverage, larger grains (fewer grain boundaries), and high crystallization are beneficial for realizing high-performance and high-stability PSCs.^[144,145] Full coverage with large grain size of perovskite indicates lower intrinsic defect density, slower moisture-induced degradation, and lesser diffusion of ionic defects. In addition, structure and crystallinity control the optoelectronic quality (trap state density, charge-carrier lifetime, and mobilities) of perovskite thin films.^[146–148] Because of the nature of the solution engineering, it is crucial to regulate the solvent, deposition method, annealing conditions, and additives

to achieve desirable crystallization (nucleation and growth) of perovskite thin films.

The ML methods were also used by Chen et al. to calculate the crystal size of the perovskite film deposited on TiO_2 ETL.^[149] After locating the edges of the crystal grains, a flood-fill algorithm can be applied to determine the distribution of crystal grain areas (Figure 6e). The ML is superior to the conventional non-ML Canny edge detection technique as the ML exhibits an 85.58% accuracy and produces an edge mask that clearly outlines the crystals visually.

Solvent has a directly influence on the nucleation and growth of crystals because of the interaction between perovskites and solvents. Antisolvent crystallization methods are frequently used to fabricate high-quality MHP films.^[150,151] Higgins et al.^[152] developed a high-throughput experimental workflow combining chemical robotic synthesis, automated characterization, and ML to systematic explore the effect of specific antisolvent on the intrinsic stability of binary perovskite systems. Fifteen combinatorial libraries were synthesized using various combinations of perovskites, resulting in approximately 1100 unique compositions. Two different antisolvents (toluene and chloroform) were used twice for fabricating each library (Figure 6d). Once synthesized, photoluminescence spectroscopy is automatically carried out every 5 min for about 6 h. Then, the time- and compositional-dependent optoelectronic properties are mapped using nonnegative matrix factorization. It is found that the choice of antisolvent is essential to the inherent stability of perovskites in ambient conditions. The effect of various solvents on the stability of PSCs was studied using ARM and DTs.^[153] As shown in Figure 6f, the x-axis is the days for stable operation and the y-axis represents the lift, which is the measure of the factor appearance in high-stability PSCs relative to entire data set (greater is better). The lift for the DMF/DMSO mixture increases over time, indicating the positive effect of mixed solvent.

4.3. ML-Assisted Optimization of Device Structures

ML is used to optimize device structure-related aspects including ETL and HTL, interfacial layer, etc., and consider the synergistic effects of these aspects.

4.3.1. Charge-Transport Layers

The charge-transport layers, including ETL and HTL, are critical for achieving high PCE and stability of PSCs. The suitable ETL and HTL should exhibit favorable energy-level alignment with the perovskite for efficient charge extraction, exhibit good charge mobility for efficient charge transport, low optical absorption, and minimal interfacial defects to reduce recombination losses.^[138,154] Optimization of the charge-transport layer usually involves energy-level matching, electron/hole-extraction efficiency, and electron/hole-transport properties. To date, the commonly used strategies include designing novel ETL, doping, and inserting an interlayer between ETL and perovskite layer.^[155–157]

Yildirim et al.^[158] prepared a library of metal-oxide-conjugated polymer composites to be used as an HTL for PSCs. An ML approach was used to predict and compare the performance

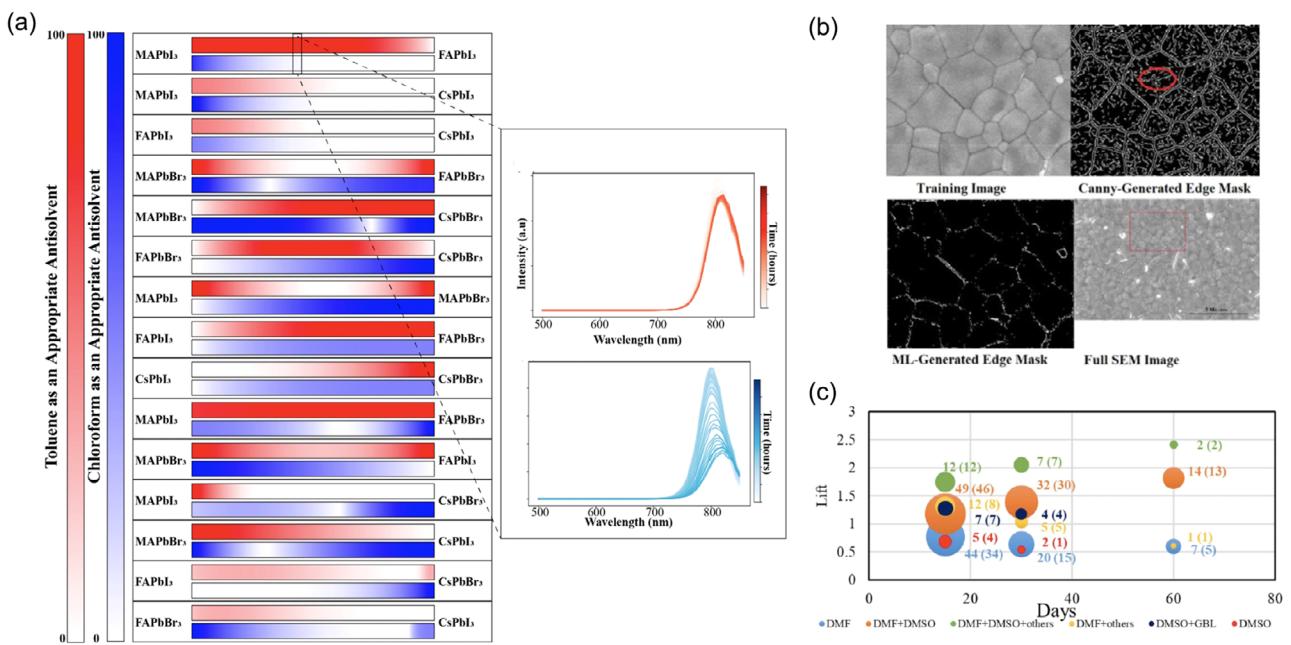


Figure 6. a) Guide for which antisolvent is more applicable for each binary metal halide perovskites system. White or light red/blue indicates that the antisolvent is not ideal for that section of the phase diagram. As the color darkens, this indicates the antisolvent is an appropriate choice for the portion of the phase diagram. Adapted with permission.^[152] Copyright 2021, American Chemical Society. b) Edge mask of perovskite scanning electron microscope images produced by ML program. Adapted with permission.^[149] Copyright 2019, Springer Nature. c) Change of lift with storage time for precursor solution in association rule mining analysis for regular (n-i-p) type cells. Adapted with permission.^[153] Copyright 2020, Elsevier.

of PSCs with the developed WO₃ and its composites. For evaluation of PSCs performance, a decision tree model was ideal for the WO₃-PEDOT composite, while a random forest model was also found to be suitable. The enhanced PCE was attributed to high FF values resulting from faster charge transport. ML can be a pioneering prediction model for the PSCs performance and its validation. She et al.^[159] used a two-step ML approach to develop some heuristics for high-efficiency PSC and PCE improvement induced by doping of the ETL (Figure 7a,b). The dataset contains 2006 PSC experimental data points extracted from 880 articles published between 2013 and 2020. It is shown that the PCE can be further improved by ETL doping to adjust the CBM, Fermi level, and conductivity of the ETL. Moreover, they predicted that an FA-MA-based PSC with a Cs-doped TiO₂ ETL and a Cs-FA-MA-based PSC with S-doped SnO₂ ETL exhibit PCEs of as high as 30.47% and 28.54%, respectively. Odabasi et al.^[76] conducted an analysis of a database consisting of 1921 device performance data points from the PSCs published between 2013 and 2018. They employed ML to develop heuristics and predictive models to assessing the PSC performance (Figure 7c). After ARM analysis of factors related to the fabrication methods of PSCs with ML, such as the composition, solvent engineering, anti-solvent, spinning process, and device architecture, more detailed heuristics and models for high efficiency are provided. For example, lithium bis(trifluoromethanesulfonyl) imide (LiTFSI)+4-tert-butylpyridine (TBP)+tris(2-(1H-pyrazol-1-yl)-4-tert-butylpyridine) cobalt(III) (FK209) as HTL additive and SnO₂ as ETL layer were detected as significant factors for high efficiency in regular (n-i-p) devices. The use of

poly-triarylamine (PTAA) as HTL, mixed cation perovskite, and bathocuproine (BCP) as ETL interlayer also appeared to be important for inverted (p-i-n) cells. The best efficiencies reported by NREL were also analyzed, and it was found that the evolution of maximum efficiency obey S-shape logistic curve.

To detect correlations between hole-transporting materials (HTM) properties and efficiencies of the PSCs, Cueto et al.^[160] proposed a predictive ML model that was trained on a dataset of 269 PSCs to screen possible candidate HTMs. They found that some chemical fragments, including arylamine and aryloxy groups, present a relatively large positive correlation with the PCE of the cell, whereas other groups, like thiophene groups, show a negative correlation with PCE.

4.3.2. Interface Layers

There are a large number of defects at the surface and grain boundaries, which traps photogenerated charges and results in additional non-radiative energy loss, limiting the V_{OC} and overall performance of the solar cell.^[5,161–163] Suitable interface layers can not only reduce defects at the surface and grain boundaries, but also optimize the energy level at the perovskite/transport layer interface. As a result, the carrier extraction can be improved and the FF of the device can be increased. Therefore, developing suitable interfacial materials is beneficial to improve both V_{OC} and FF, and thus effectively improve PCE.^[1,164,165]

Liu et al.^[166] used ML to screen interfacial materials by studying the correlation between PCE and interfacial passivation materials at the atomic level (Figure 7d). Based on SHAP values,

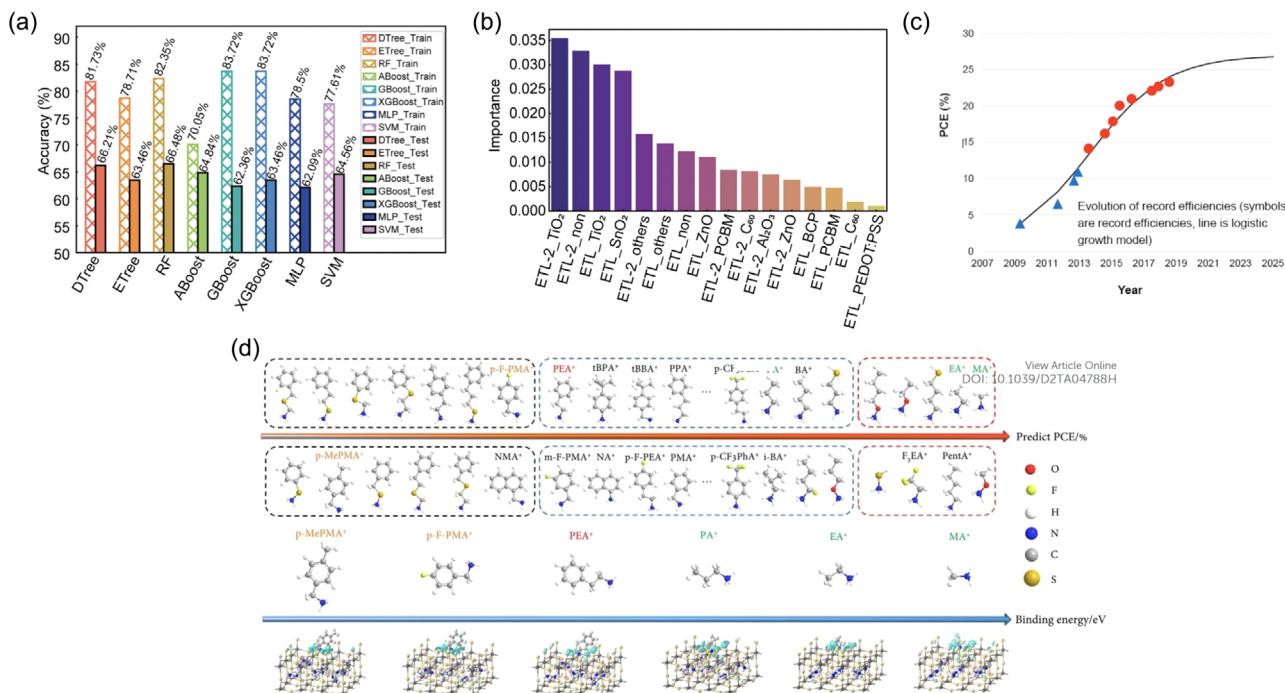


Figure 7. a) Comparison of the classification accuracies of various models on the training and test sets. d) Relative importance of the electron-transport layer-related features given by the random forest (RF) classification model. Adapted with permission.^[159] Copyright 2021, Royal Society of Chemistry. c) Evolution of efficiency for regular (n-i-p) structure. Adapted with permission.^[76] Copyright 2019, Elsevier. d) Predicted PCEs from different interface materials and the DFT calculation results based on several representative materials. Adapted with permission.^[166] Copyright 2022, Royal Society of Chemistry.

feature importance, and correlation analysis, the most critical 15 features were selected from over 300 features. High-throughput predictions were further made and rationalized using DFT calculations. They provide general screening rules for interface materials: 1) the material should have at least one NH^{3+} terminus; 2) preferably with a short C chain and without a benzene ring; 3) other functional groups are not as effective as NH^{3+} . The PCE can be increased by about 1.5% with the introduction of interfacial material, revealing the reliability of the predictions. Wu and coworkers^[62] applied ML to study the trend of reactivity of different types of amines, which were used as interfacial materials. By analyzing the ML models, it was found that amines with fewer hydrogen bond donors and acceptors, more steric bulk, secondary, tertiary amines, and pyridine derivatives tend to have high compatibility with perovskite films. Liu et al.^[167] used a data-driven ML approach to investigate the relationship between the molecular features of ammonium salts and the PCE improvement after passivation. The most important features of ammonium salt for passivation are identified by SHAP analysis. The ML is further used to screen suitable candidates from a pool of 112 salts in the PubChem database.

4.4. The Utilization of ML in Understanding the Mechanism

MLs are used to understand the role of the recombination and hysteresis in device performance, as well as explore the factors

affecting the long-term stability, which will guide the experimental fabrication of efficient and stable PSCs.

4.4.1. Charge-Carriers Recombination in PSCs

The PCE of PSCs has rapidly improved through improving perovskite absorber layer quality and optimizing device configurations. Nevertheless, the best efficiency is still below the theoretical efficiency because of non-radiative recombination losses. These non-radiative recombination center not only reduce the V_{OC} and PCE, but also accelerates the degradation process of perovskite absorbers.^[168-171] To further boost device performances and simultaneously enhance the long-term stability of PSCs, it is important to minimize non-radiative recombination losses and hysteresis.^[106,172]

The introduction of additives can improve the crystallization of perovskite and passivate defects at surfaces, leading to reduced nonradiative recombination. For example, alkali metal doping has been reported as an effective method to decrease the defect density in the perovskite film, as a proper amount of alkali metal in the perovskite crystal lattice could increase the crystallinity and reduce the charge recombination.^[173–175] Additionally, excess iodide ions could compensate for halide vacancies at the grain boundaries and passivate defects. Jiang et al.^[176] utilized ML approach to aid in the optimization of KI-doped MAPbI₃-based PSCs. It discovered that 3% KI doping can contribute to

enhanced device efficiency of 20.91%. Additive engineering is a promising strategy for further enhancing perovskite device performance and a key challenge is to choose suitable additive for perovskite. Zhang et al.^[177] demonstrated a deep-learning method that can predict the effectiveness of additives in perovskite LEDs by using a small dataset of 132 molecules (**Figure 8a,b**). This model can considerably reduce the duplicated problem that frequently occurred with earlier ML models for molecular screening while maximizing the information contained in the molecules. Based on the predicated additive, very high-efficiency PeLEDs with a peak external quantum efficiency up to 22.7 % can be achieved.

Le Corre et al.^[178] developed ML model for PSCs to get a better understanding of the light intensity dependence of the V_{OC} and identify the dominant recombination process. They trained the drift-diffusion simulations in an ML model based on DTs to classify the light intensity measurement output and identify the dominant recombination. After training, the model demonstrated the ability to classify the dominant recombination

process (band-to-band, interface, or grain boundaries trap recombination) with over 80% accuracy in just a few seconds (Figure 8c). In addition, the role of ions and doping density on the recombination process were also defined by the trained ML model (Figure 8d,e). Zhou et al.^[179] applied ML to nonadiabatic molecular dynamics simulation to explore the non-radiative charge recombination in MAPbI₃. They found that the I–I–I angle is the crucial structural factor controlling the bandgap and nonadiabatic electron–phonon coupling of MAPbI₃. Even though MA does not contribute to electron or hole wave functions, it has been shown that its rotational and center-of-mass motions will affect charge recombination. These results highlight inorganic lattice deformation and MA rotation during polaron production and provide an explanation for the unusual temperature dependency of carrier lifetimes in halide perovskites.

ML are also used to analyze the relationships between hysteresis, reproducibility, PCE, and long-term stability of PSCs.^[103] The hysteresis was analyzed using ARM and a dataset containing

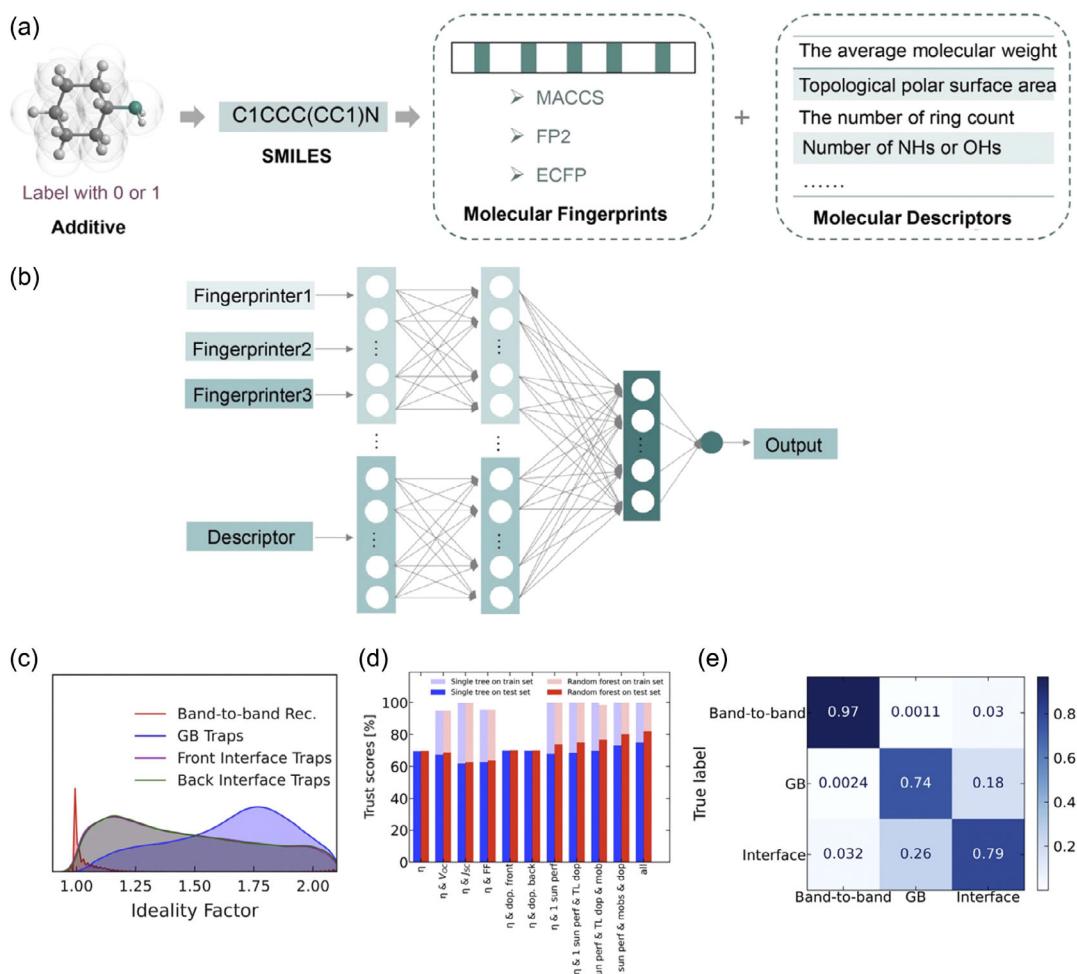


Figure 8. The training process of enhanced molecular information model (EMIM). a) Additive transfer to canonical simplified molecular-input line-entry system (SMILES), then convert SMILES to fingerprints and descriptors. b) The schematic plot of the EMIM. Both the fingerprints/descriptors and labels are input to EMIM. Adapted with permission.^[177] Copyright 2022, Wiley-VCH. c) Density distribution of the ideality factors under different recombination processes. d) Evolution of the trust score on training and test dataset depending on the features used to train single tree or RF. e) Confusion matrix of the optimal classification model based on RF algorithm. Adapted with permission.^[178] Copyright 2021, Elsevier.

387 cells from 194 articles in the literature, while the reproducibility was analyzed using pooled variance of 24 142 cells. It is found that for n-i-p cell, mixed cation perovskites, two-step spin coating or multiple spin coating in one step, DMF + DMSO as solvent, PTAA as HTL, LiTFSI + TBP + FK209 as HTL dopant, and carbon as back contact are beneficial for both low hysteresis and high reproducibility. In contrast, the toluene as antisolvent, BCP as ETL interlayer, and Ag as back contact are found to have positive impacts in p-i-n cells. It is to be noted that those factors are also highly relevant for PCE and stability.

4.4.2. The Degradation of PSCs

Currently, the instability of PSCs becomes a bottleneck limiting the industrialization and the studies on stability of PSCs are still lagging behind. Therefore, it is highly important to use ML to deeply study the causes of instability and strategies to improve the stability.

Hu et al.^[63] employed ML to predict the influences of five factors (grain size, defect density, bandgap, fluorescence lifetime, and surface roughness) on the efficiency and stability of PSCs. The results indicate that the efficiency is significantly affected by the bandgap, while the stability is most influenced by the surface roughness and grain size. Based on the predictions of ML models, different annealing temperatures were employed to obtain different grain sizes in the perovskite films, while different organic molecules were applied to modify the surface roughness. The target with largest grain size and smallest surface roughness exhibit the excellent long-term stability with 97.6% of the best PCE maintained after 3288 h of aging under relative humidity (RH) of 25% and 25 °C, which is the best stability reported under the same conditions. The experimental results match well with the predictions of the ML model. First-principles calculations were also carried out to provide a deeper explanation for the superb ambient stability in the atom scale. Hartono et al.^[180] utilized ML to optimize the capping layer for MAPbI₃ perovskite to increase the environmental stability. A total of 21 different organic halide salts as capping layers were applied onto MAPbI₃ films to test the stability under accelerated conditions. Supervised ML and Shapley values were used to determine features governing stability. The results showed that low number of hydrogen bond donors and small topological polar surface area of organic molecule are beneficial for increasing the film stability. Phenyltriethylammonium iodide (PTEAI), the best-performing organic halide, successfully increases the stability lifetime of bare MAPbI₃ by 4 ± 2 times. The PTEAI-based stability lifetime is even 1.3 ± 0.3 times over the cutting-edge octylammonium bromide. They then found that the capping layer stabilizes perovskite layer by changing the surface chemistry and suppressing methylammonium loss through characterization. Higgins et al.^[181] studied the electrochemical reactions at the interface of MAPbBr₃ perovskite and Au electrode using ML in situ time-of-flight secondary ion mass spectroscopy. The ML workflow combined with the Hough transform, non-negative matrix factorization, and non-negative tensor decomposition was developed, which could interpret the multi-dimensional data, extract salient features of associated chemical changes, and separate the light- and voltage-dependent

dynamics. The comprehensive information on the chemical nature of moving species, ion accumulation, and interfacial electrochemical reactions in PSCs could be provided.

Stoddard et al.^[182] used the ML greedy feature selection model to explore the evolution of material-level optoelectronic properties as MAPbI₃ degrades and discovered that the carrier diffusion length (LD) decays before quasi-Fermi-level splitting (ΔE_F) in every case (oxygen, humidity, thermal stress, or a combination). ML model was used to predict the time taken to reduce LD to fall 85% of its initial value. This model reveals a strong relation between the initial rate of transmittance change and the time until loss of transport. This study reveals that the optoelectronic properties changes occur during MAPbI₃ degradation and provide a simple framework for forecasting perovskite materials and device lifetimes.

5. Conclusion and Outlook

We briefly summarize the fundamental process of the ML in material design and comprehensively review the application of ML in rational design of perovskite materials and fabrication of PSCs. Compared to time-consuming theoretical calculations and trial-and-error experiments, ML has great advantages in finding new materials and revealing the relationship between structural, compositional, and performance, as well as optimizing fabrication processes and device structures. However, there are still some challenges that need to be handled. 1) The data extracted from experimental or computational databases are much easier suffering from heterogeneity and large noises. One of the solutions is developing experimental procedures like combinatorial approaches as reported by Saliba et al.^[183] The majority of feature selection procedures are decided by the intuition and experience of researchers, leading to the omission of some useful features or redundancy of features. Automated feature engineering could help experimenters to train models with higher prediction accuracy and lower artificial errors. In addition, the physical and chemical knowledge is lacking in ML-predicted results and need to be interpreted for further research. Developing more accurate and interpretable features is an effect for improving the interpretability of ML models. The relationship between the features and perovskite property should also be studied to understand the hidden physical laws behind the characteristics. Apart from these, developing more accurate and advanced ML algorithms is necessary to deal with the challenge of small samples, such as meta-learning. For example, the Bayesian learning framework can reach the level of human experience through one-shot learning with insufficient data.^[184] 2) While small-area PSCs have demonstrated remarkable efficiency, the inadequate performance of large-area PSCs becomes a significant obstacle for the scale-up and commercialization. In solution-based perovskite modules fabrication, the quality of modules is sensitive to a multitude of control parameters, including complex compositions, solution characteristics, RH, process conditions. Moreover, some commonly uncontrolled variables such as the residual atmospheric solvent vapor also have an impact on product quality. Therefore, ML should be employed to identify optimal processing conditions for perovskite modules, which can lead to reduced labor and expenses. ML should also integrate

with robotics to autonomously fabricate and optimize perovskite modules, establishing a correlation between input synthesis and device performance. When uncontrolled variables change, these variations will be detected and seamlessly integrated into the ML-based process to fine-tune the fabrication procedure accordingly. Furthermore, understanding the kinetic processes of crystallization and physical mechanism of PSCs is crucial for ensuring production quality and achieving the successful scaling up of perovskite modules. The atomistic, optical, electrical, and physical properties, including chemical bonds, absorption coefficients, defect tolerance, charge mobility, and carrier diffusion length, are pivotal factors influencing the performance of PSCs. Therefore, these properties should be applied into ML models to better guide the synthetic procedure. 3) Stability is pivotal for the industrialization of perovskites, and currently, the stability of PSCs significantly lags behind that of Si solar cells. Currently, ML studies on predicting the stability of PSCs are still at an early stage. Many ML studies focus on predicting perovskite materials with high thermodynamic stability that are suitable for PSCs. However, the realistic working condition (e.g. water penetration, thermal stress, and light-induced migration) is not considered. To solve this problem, a correlation between designed materials and long-term stability testing should be established using ML. In addition, the degradation mechanism of PSCs at atomic level (e.g. octahedra tilting, lattice strain) under various conditions should be understood.

In summary, with the continuous development of computer technology, ML will be an economical and effective method in the fields of materials science, chemistry, and physics because of its strong analytical and computational capabilities. Although challenges exist, there is still room for improving the reliability and interpretability of ML approaches. The application of ML in the design and fabrication of PSCs has gained tremendous achievement recently. Further advancement will surely contribute toward cost-effective, highly efficient, and stable PSCs.

Acknowledgements

This work was supported by Young Scientists Fund of the Natural Science Foundation Program of Shaanxi Province (grant no. 2023-JC-QN-0143) and the Key Research and Development Program of Shaanxi (grant no. 2020GY-247). W.C. acknowledges the General Research Fund (grant nos. 17201819, 17211220, and 17200021) and Collaborative Research Fund (grant no. C7035-20G) from the Research Grants Council (RGC) of Hong Kong Special Administrative Region, China.

Conflict of Interest

The authors declare no conflict of interest.

Keywords

design, fabrications, machine learnings, perovskite solar cells, power conversion efficiencies, stabilities

Received: August 17, 2023

Revised: September 9, 2023

Published online: October 6, 2023

- [1] F. H. Isikgor, S. Zhumagali, L. V. T. Merino, M. De Bastiani, I. McCulloch, S. De Wolf, *Nat. Rev. Mater.* **2023**, *8*, 89.
- [2] X. Li, W. Zhang, X. Guo, C. Lu, J. Wei, J. Fang, *Science* **2022**, *375*, 434.
- [3] G. Li, Z. Su, L. Canil, D. Hughes, M. H. Aldamasy, J. Dagar, S. Trofimov, L. Wang, W. Zuo, J. J. Jerónimo-Rendon, M. M. Byranvand, C. Wang, R. Zhu, Z. Zhang, F. Yang, G. Nasti, B. Naydenov, W. C. Tsoi, Z. Li, X. Gao, Z. Wang, Y. Jia, E. Unger, M. Saliba, M. Li, A. Abate, *Science* **2023**, *379*, 399.
- [4] T. Bu, L. K. Ono, J. Li, J. Su, G. Tong, W. Zhang, Y. Liu, J. Zhang, J. Chang, S. Kazaoui, F. Huang, Y.-B. Cheng, Y. Qi, *Nat. Energy* **2022**, *7*, 528.
- [5] Z. Guo, A. K. Jena, G. M. Kim, T. Miyasaka, *Energy Environ. Sci.* **2022**, *15*, 3171.
- [6] Z.-W. Gao, Y. Wang, W. C. H. Choy, *Adv. Energ. Mater.* **2022**, *12*, 2104030.
- [7] F. Zhang, S. Y. Park, C. Yao, H. Lu, S. P. Dunfield, C. Xiao, S. Uličná, X. Zhao, L. Du Hill, X. Chen, X. Wang, L. E. Mundt, K. H. Stone, L. T. Schelhas, G. Teeter, S. Parkin, E. L. Ratcliff, Y.-L. Loo, J. J. Berry, M. C. Beard, Y. Yan, B. W. Larson, K. Zhu, *Science* **2022**, *375*, 71.
- [8] <https://www.nrel.gov/pv/cell-efficiency.html> (accessed: September 2023).
- [9] A. Kojima, K. Teshima, Y. Shirai, T. Miyasaka, *J. Am. Chem. Soc.* **2009**, *131*, 6050.
- [10] T. Liu, J. Zhang, M. Qin, X. Wu, F. Li, X. Lu, Z. Zhu, A. K. Y. Jen, *Adv. Funct. Mater.* **2021**, *31*, 2009515.
- [11] M. A. Mahmud, T. Duong, J. Peng, Y. Wu, H. Shen, D. Walter, H. T. Nguyen, N. Mozaffari, G. D. Tabi, K. R. Catchpole, K. J. Weber, T. P. White, *Adv. Funct. Mater.* **2022**, *32*, 2009164.
- [12] X. Lin, H. Su, S. He, Y. Song, Y. Wang, Z. Qin, Y. Wu, X. Yang, Q. Han, J. Fang, Y. Zhang, H. Segawa, M. Grätzel, L. Han, *Nat. Energy* **2022**, *7*, 520.
- [13] T. Wu, L. K. Ono, R. Yoshioka, C. Ding, C. Zhang, S. Mariotti, J. Zhang, K. Mitrofanov, X. Liu, H. Segawa, R. Kabe, L. Han, Y. Qi, *Energy Environ. Sci.* **2022**, *15*, 4612.
- [14] P. Wu, S. Wang, X. Li, F. Zhang, *Matter* **2022**, *5*, 1137.
- [15] J. C. Yu, B. Li, C. J. Dunn, J. Yan, B. T. Diroll, A. S. R. Chesman, J. J. Jasieniak, *Adv. Sci.* **2022**, *9*, 2201487.
- [16] T. Matsui, T. Yamamoto, T. Nishihara, R. Morisawa, T. Yokoyama, T. Sekiguchi, T. Negami, *Adv. Mater.* **2019**, *31*, 1806823.
- [17] D.-H. Kang, N.-G. Park, *Adv. Mater.* **2019**, *31*, 1805214.
- [18] Y. Che, Z. Liu, Y. Duan, J. Wang, S. Yang, D. Xu, W. Xiang, T. Wang, N. Yuan, J. Ding, S. Liu, *Angew. Chem. Int. Ed.* **2022**, *61*, e202205012.
- [19] G. Wu, R. Liang, M. Ge, G. Sun, Y. Zhang, G. Xing, *Adv. Mater.* **2022**, *34*, 2105635.
- [20] W. Zhao, H. Lin, Y. Li, D. Wang, J. Wang, Z. Liu, N. Yuan, J. Ding, Q. Wang, S. Liu, *Adv. Funct. Mater.* **2022**, *32*, 2112032.
- [21] Y. Dong, W. Shen, W. Dong, C. Bai, J. Zhao, Y. Zhou, F. Huang, Y.-B. Cheng, J. Zhong, *Adv. Energy Mater.* **2022**, *12*, 2200417.
- [22] X. Zhou, X. Luan, L. Zhang, H. Hu, Z. Jiang, Y. Li, J. Wu, Y. Liu, J. Chen, D. Wang, C. Liu, S. Chen, Y. Zhang, M. Zhang, Y. Peng, P. A. Troshin, X. Wang, Y. Mai, B. Xu, *ACS Nano* **2023**, *17*, 3776.
- [23] H. Xu, Y. Miao, N. Wei, H. Chen, Z. Qin, X. Liu, X. Wang, Y. Qi, T. Zhang, Y. Zhao, *Adv. Energy Mater.* **2022**, *12*, 2103151.
- [24] L. Yang, J. Feng, Z. Liu, Y. Duan, S. Zhan, S. Yang, K. He, Y. Li, Y. Zhou, N. Yuan, J. Ding, S. Liu, *Adv. Mater.* **2022**, *34*, 2201681.
- [25] M. Wang, Y. Zhao, X. Jiang, Y. Yin, I. Yavuz, P. Zhu, A. Zhang, G. S. Han, H. S. Jung, Y. Zhou, W. Yang, J. Bian, S. Jin, J.-W. Lee, Y. Yang, *Joule* **2022**, *6*, 1032.
- [26] X. Zhang, Z. Yu, D. Zhang, Q. Tai, X.-Z. Zhao, *Adv. Energy Mater.* **2022**, *13*, 2201320.

- [27] S. Ma, G. Yuan, Y. Zhang, N. Yang, Y. Li, Q. Chen, *Energy Environ. Sci.* **2022**, *15*, 13.
- [28] S. H. Reddy, F. Di Giacomo, A. Di Carlo, *Adv. Energy Mater.* **2022**, *12*, 2103534.
- [29] C. Liu, K. Huang, B. Hu, Y. Li, L. Zhang, X. Zhou, Y. Liu, Z. Liu, Y. Sheng, S. Chen, X. Wang, B. Xu, *Adv. Funct. Mater.* **2023**, *33*, 2212698.
- [30] B. Yilmaz, R. Yıldırım, *Nano Energy* **2021**, *80*, 105546.
- [31] M. Srivastava, J. M. Howard, T. Gong, M. Rebello Sousa Dias, M. S. Leite, *J. Phys. Chem. Lett.* **2021**, *12*, 7866.
- [32] C. W. Myung, A. Hajibabaei, J.-H. Cha, M. Ha, J. Kim, K. S. Kim, *Adv. Energy Mater.* **2022**, *12*, 2202279.
- [33] N. Li, X. Niu, Q. Chen, H. Zhou, *Chem. Soc. Rev.* **2020**, *49*, 8235.
- [34] Y. Cheng, L. Ding, *Energy Environ. Sci.* **2021**, *14*, 3233.
- [35] Q. Dong, M. Chen, Y. Liu, F. T. Eickemeyer, W. Zhao, Z. Dai, Y. Yin, C. Jiang, J. Feng, S. Jin, S. Liu, S. M. Zakeeruddin, M. Grätzel, N. P. Padture, Y. Shi, *Joule* **2021**, *5*, 1587.
- [36] F. Dinic, K. Singh, T. Dong, M. Rezazadeh, Z. Wang, A. Khosrozadeh, T. Yuan, O. Voznyy, *Adv. Funct. Mater.* **2021**, *31*, 2104195.
- [37] C. Gao, X. Min, M. Fang, T. Tao, X. Zheng, Y. Liu, X. Wu, Z. Huang, *Adv. Funct. Mater.* **2022**, *32*, 2108044.
- [38] Y. Liu, X. Tan, J. Liang, H. Han, P. Xiang, W. Yan, *Adv. Funct. Mater.* **2023**, *33*, 2214271.
- [39] C. Liu, L. Lüer, V. M. L. Corre, K. Forberich, P. Weitz, T. Heumüller, X. Du, J. Wortmann, J. Zhang, J. Wagner, L. Ying, J. Hauch, N. Li, C. J. Brabec, *Adv. Mater.* **2023**, 2300259.
- [40] X. Du, L. Lüer, T. Heumueller, J. Wagner, C. Berger, T. Osterrieder, J. Wortmann, S. Langner, U. Vongsaysy, M. Bertrand, N. Li, T. Stubhan, J. Hauch, C. J. Brabec, *Joule* **2021**, *5*, 495.
- [41] J. Kirman, A. Johnston, D. A. Kuntz, M. Askerka, Y. Gao, P. Todorović, D. Ma, G. G. Privé, E. H. Sargent, *Matter* **2020**, *2*, 938.
- [42] Q. Tao, P. Xu, M. Li, W. Lu, *NPJ Comput. Mater.* **2021**, *7*, 23.
- [43] Z. Liu, N. Rolston, A. C. Flick, T. W. Colburn, Z. Ren, R. H. Dauskardt, T. Buonassisi, *Joule* **2022**, *6*, 834.
- [44] J. Li, B. Pradhan, S. Gaur, J. Thomas, *Adv. Energy Mater.* **2019**, *9*, 1901891.
- [45] M. I. Jordan, T. M. Mitchell, *Science* **2015**, *349*, 255.
- [46] A. L. Blum, P. Langley, *Artif. Intell.* **1997**, *97*, 245.
- [47] G. L. W. Hart, T. Mueller, C. Toher, S. Curtarolo, *Nat. Rev. Mater.* **2021**, *6*, 730.
- [48] Y. Liu, R. Long, W.-H. Fang, *J. Phys. Chem. Lett.* **2022**, *13*, 8537.
- [49] Y. Wu, D. Liu, W. Chu, B. Wang, A. S. Vasenko, O. V. Prezhdo, *ACS Appl. Mater. Interfaces* **2022**, *14*, 55753.
- [50] E. C. Gok, M. O. Yildirim, M. P. U. Haris, E. Eren, M. Pegu, N. H. Hemasiri, P. Huang, S. Kazim, A. Uygur Oksuz, S. Ahmad, *Sol. RRL* **2022**, *6*, 2100927.
- [51] M. Ahmadi, M. Ziatdinov, Y. Zhou, E. A. Lass, S. V. Kalinin, *Joule* **2021**, *5*, 2797.
- [52] A. Mahmood, J.-L. Wang, *Energy Environ. Sci.* **2021**, *14*, 90.
- [53] C. Lampe, I. Kouroudis, M. Harth, S. Martin, A. Gagliardi, A. S. Urban, *Adv. Mater.* **2023**, *35*, 2208772.
- [54] X. Wan, W. Feng, Y. Wang, H. Wang, X. Zhang, C. Deng, N. Yang, *Nano Lett.* **2019**, *19*, 3387.
- [55] C. R. Groom, I. J. Bruno, M. P. Lightfoot, S. C. Ward, *Acta Crystallogr. Sect. B: Struct. Sci. Cryst. Eng. Mater.* **2016**, *72*, 171.
- [56] S. Gražulis, A. Daškevič, A. Merkys, D. Chateigner, L. Lutterotti, M. Quiros, N. R. Serebryanaya, P. Moeck, R. T. Downs, A. Le Bail, *Nucl. Acids Res.* **2012**, *40*, D420.
- [57] M. Hellenbrandt, *Crystallogr. Rev.* **2004**, *10*, 17.
- [58] S. Kim, J. Chen, T. Cheng, A. Gindulyte, J. He, S. He, Q. Li, B. A. Shoemaker, P. A. Thiessen, B. Yu, L. Zaslavsky, J. Zhang, E. E. Bolton, *Nucl. Acids Res.* **2021**, *49*, D1388.
- [59] A. Jain, S. P. Ong, G. Hautier, W. Chen, W. D. Richards, S. Dacek, S. Cholia, D. Gunter, D. Skinner, G. Ceder, K. A. Persson, *APL Mater.* **2013**, *1*, 011002.
- [60] S. Kirklin, J. E. Saal, B. Meredig, A. Thompson, J. W. Doak, M. Aykol, S. Rühl, C. Wolverton, *NPJ Comput. Mater.* **2015**, *1*, 15010.
- [61] C. Draxl, M. Scheffler, *MRS Bull.* **2018**, *43*, 676.
- [62] Y. Yu, X. Tan, S. Ning, Y. Wu, *ACS Energy Lett.* **2019**, *4*, 397.
- [63] Y. Hu, X. Hu, L. Zhang, T. Zheng, J. You, B. Jia, Y. Ma, X. Du, L. Zhang, J. Wang, B. Che, T. Chen, S. Liu, *Adv. Energy Mater.* **2022**, *12*, 2201463.
- [64] Y. Lu, D. Wei, W. Liu, J. Meng, X. Huo, Y. Zhang, Z. Liang, B. Qiao, S. Zhao, D. Song, Z. Xu, *J. Energy Chem.* **2023**, *77*, 200.
- [65] S.-A. N. Alexandropoulos, S. B. Kotsiantis, M. N. Vrahatis, *Knowl. Eng. Rev.* **2019**, *34*, e1.
- [66] T. Zhou, Z. Song, K. Sundmacher, *Engineering* **2019**, *5*, 1017.
- [67] A. Jain, G. Hautier, S. P. Ong, K. Persson, *J. Mater. Res.* **2016**, *31*, 977.
- [68] J. Schmidhuber, *Neural Networks* **2015**, *61*, 85.
- [69] C. Chen, Y. Zuo, W. Ye, X. Li, Z. Deng, S. P. Ong, *Adv. Energy Mater.* **2020**, *10*, 1903242.
- [70] J. Schmidt, M. R. G. Marques, S. Botti, M. A. L. Marques, *NPJ Comput. Mater.* **2019**, *5*, 83.
- [71] K. T. Butler, D. W. Davies, H. Cartwright, O. Isayev, A. Walsh, *Nature* **2018**, *559*, 547.
- [72] Z. Xiong, Y. Cui, Z. Liu, Y. Zhao, M. Hu, J. Hu, *Comput. Mater. Sci.* **2020**, *171*, 109203.
- [73] V. Yaghoubi, M. K. Vakilzadeh, T. J. S. Abrahamsson, *Mech. Syst. Signal Process.* **2018**, *100*, 289.
- [74] Y. Cui, P. Zhu, X. Shi, X. Liao, Y. Chen, *J. Phys. Chem. C* **2021**, *125*, 10250.
- [75] Y. Cui, P. Zhu, X. Liao, Y. Chen, *J. Mater. Chem. C* **2020**, *8*, 15920.
- [76] Ç. Odabaşı, R. Yıldırım, *Nano Energy* **2019**, *56*, 770.
- [77] S. Lu, Q. Zhou, Y. Ouyang, Y. Guo, Q. Li, J. Wang, *Nat. Commun.* **2018**, *9*, 3405.
- [78] T. Wu, J. Wang, *Nano Energy* **2019**, *66*, 104070.
- [79] P. De Luna, J. Wei, Y. Bengio, A. Aspuru-Guzik, E. Sargent, *Nature* **2017**, *552*, 23.
- [80] F. Laufer, S. Ziegler, F. Schackmar, E. A. M. Viteri, M. Götz, C. Debus, F. Isensee, U. W. Paetzold, *Sol. RRL* **2023**, *7*, 2201114.
- [81] S. Zhang, T. Lu, P. Xu, Q. Tao, M. Li, W. Lu, *J. Phys. Chem. Lett.* **2021**, *12*, 7423.
- [82] Y. Rong, Y. Hu, A. Mei, H. Tan, M. I. Saidaminov, S. I. Seok, M. D. McGehee, E. H. Sargent, H. Han, *Science* **2018**, *361*, eaat8235.
- [83] A. Rajagopal, K. Yao, A. K. Y. Jen, *Adv. Mater.* **2018**, *30*, 1800455.
- [84] S. Sun, N. T. P. Hartono, Z. D. Ren, F. Oviedo, A. M. Buscemi, M. Layurova, D. X. Chen, T. Ogunfunmi, J. Thapa, S. Ramasamy, C. Settens, B. L. DeCost, A. G. Kusne, Z. Liu, S. I. P. Tian, I. M. Peters, J.-P. Correa-Baena, T. Buonassisi, *Joule* **2019**, *3*, 1437.
- [85] Z. Hu, Z. Lin, J. Su, J. Zhang, J. Chang, Y. Hao, *Sol. RRL* **2019**, *3*, 1900304.
- [86] S. Rath, G. Sudha Priyanga, N. Nagappan, T. Thomas, *Comput. Mater. Sci.* **2022**, *210*, 111476.
- [87] H. Park, R. Mall, F. H. Alharbi, S. Sanvito, N. Tabet, H. Bensmail, F. El-Mellouhi, *Phys. Chem. Chem. Phys.* **2019**, *21*, 1078.
- [88] X. Cai, F. Liu, A. Yu, J. Qin, M. Hatamvand, I. Ahmed, J. Luo, Y. Zhang, H. Zhang, Y. Zhan, *Light: Sci. Appl.* **2022**, *11*, 234.
- [89] A. Mannodi-Kanakkithodi, M. K. Y. Chan, *Energy Environ. Sci.* **2022**, *15*, 1930.
- [90] J. Schmidt, J. Shi, P. Borlido, L. Chen, S. Botti, M. A. L. Marques, *Chem. Mater.* **2017**, *29*, 5090.
- [91] Z. Li, Q. Xu, Q. Sun, Z. Hou, W.-J. Yin, *Adv. Funct. Mater.* **2019**, *29*, 1807280.
- [92] A. Talapatra, B. P. Uberuaga, C. R. Stanek, G. Pilania, *Chem. Mater.* **2021**, *33*, 845.

- [93] D. Kim, E. S. Muckley, N. Creange, T. H. Wan, M. H. Ann, E. Quattrochi, R. K. Vasudevan, J. H. Kim, F. Ciucci, I. N. Ivanov, S. V. Kalinin, M. Ahmadi, *Adv. Sci.* **2021**, *8*, 2002510.
- [94] Y. Liu, W. Yan, S. Han, H. Zhu, Y. Tu, L. Guan, X. Tan, *Sol. RRL* **2022**, *6*, 2101100.
- [95] E. Nakanishi, R. Nishikubo, F. Ishiwari, T. Nakamura, A. Wakamiya, A. Saeki, *ACS Mater. Lett.* **2022**, *4*, 1124.
- [96] V. A. Aristizabal-Ferreira, J. M. Guevara-Vela, A. Sauza-de la Vega, Á. M. Pendás, G. Fuentes-Pineda, T. Rocha-Rinza, *Theor. Chem. Acc.* **2022**, *141*, 19.
- [97] J. C. Stanley, F. Mayr, A. Gagliardi, *Adv. Theory Simul.* **2020**, *3*, 1900178.
- [98] R. Zhao, B. Xing, H. Mu, Y. Fu, L. Zhang, *Chin. Phys. B* **2022**, *31*, 056302.
- [99] S. Kanno, Y. Immura, M. Hada, *Phys. Rev. Mater.* **2019**, *3*, 075403.
- [100] P. Priya, N. R. Aluru, *NPJ Comput. Mater.* **2021**, *7*, 90.
- [101] A. Mannodi-Kanakkithodi, J.-S. Park, N. Jeon, D. H. Cao, D. J. Gosztola, A. B. F. Martinson, M. K. Y. Chan, *Chem. Mater.* **2019**, *31*, 3599.
- [102] W. Hu, L. Zhang, Z. Pan, *ACS Appl. Mater. Interfaces* **2022**, *14*, 21596.
- [103] Ç. Odabaşı, R. Yıldırım, *Energy Technol.* **2020**, *8*, 1901449.
- [104] S. Zhai, H. Xie, P. Cui, D. Guan, J. Wang, S. Zhao, B. Chen, Y. Song, Z. Shao, M. Ni, *Nat. Energy* **2022**, *7*, 866.
- [105] J. Xia, M. Sohail, M. K. Nazeeruddin, *Adv. Mater.* **2023**, *35*, 2211324.
- [106] Z. Zhang, L. Qiao, K. Meng, R. Long, G. Chen, P. Gao, *Chem. Soc. Rev.* **2023**, *52*, 163.
- [107] P. Holzhey, M. Prettl, S. Collavini, N. L. Chang, M. Saliba, *Joule* **2023**, *7*, 257.
- [108] J. Y. Kim, J.-W. Lee, H. S. Jung, H. Shin, N.-G. Park, *Chem. Rev.* **2020**, *120*, 7867.
- [109] J.-P. Correa-Baena, M. Saliba, T. Buonassisi, M. Grätzel, A. Abate, W. Tress, A. Hagfeldt, *Science* **2017**, *358*, 739.
- [110] M. L. Petrus, J. Schlipf, C. Li, T. P. Gujar, N. Giesbrecht, P. Müller-Buschbaum, M. Thelakkat, T. Bein, S. Hüttner, P. Docampo, *Adv. Energy Mater.* **2017**, *7*, 1700264.
- [111] H. Lu, A. Krishna, S. M. Zakeeruddin, M. Grätzel, A. Hagfeldt, *Iscience* **2020**, *23*, 101359.
- [112] C. C. Stoumpos, C. M. M. Soe, H. Tsai, W. Nie, J.-C. Blancon, D. H. Cao, F. Liu, B. Traoré, C. Katan, J. Even, A. D. Mohite, M. G. Kanatzidis, *Chem* **2017**, *2*, 427.
- [113] G. J. Man, C. Kamal, A. Kalinko, D. Phuyal, J. Acharya, S. Mukherjee, P. K. Nayak, H. Rensmo, M. Odelius, S. M. Butorin, *Nat. Commun.* **2022**, *13*, 3839.
- [114] C. Shi, Q. Song, H. Wang, S. Ma, C. Wang, X. Zhang, J. Dou, T. Song, P. Chen, H. Zhou, Y. Chen, C. Zhu, Y. Bai, Q. Chen, *Adv. Funct. Mater.* **2022**, *32*, 2201193.
- [115] Y. Wang, G. Chen, D. Ouyang, X. He, C. Li, R. Ma, W.-J. Yin, W. C. H. Choy, *Adv. Mater.* **2020**, *32*, 2000186.
- [116] N. Rolston, A. D. Printz, J. M. Tracy, H. C. Weerasinghe, D. Vak, L. J. Haur, A. Priyadarshi, N. Mathews, D. J. Slotcavage, M. D. McGehee, R. E. Kalan, K. Zielinski, R. L. Grimm, H. Tsai, W. Nie, A. D. Mohite, S. Gholipour, M. Saliba, M. Grätzel, R. H. Dauskardt, *Adv. Energy Mater.* **2018**, *8*, 1702116.
- [117] W. A. Saidi, W. Shadid, I. E. Castelli, *NPJ Comput. Mater.* **2020**, *6*, 36.
- [118] H. Park, R. Mall, A. Ali, S. Sanvito, H. Bensmail, F. El-Mellouhi, *Comput. Mater. Sci.* **2020**, *184*, 109858.
- [119] K. Higgins, S. M. Valletti, M. Ziatdinov, S. V. Kalinin, M. Ahmadi, *ACS Energy Lett.* **2020**, *5*, 3426.
- [120] W. Yan, Y. Liu, Y. Zang, J. Cheng, Y. Wang, L. Chu, X. Tan, L. Liu, P. Zhou, W. Li, Z. Zhong, *Nano Energy* **2022**, *99*, 107394.
- [121] H. A. Jame, S. Sarker, M. S. Islam, M. T. Islam, A. Rauf, S. Ahsan, S. S. Nishat, M. R. Jani, K. M. Shorowordi, J. Carbonara, S. Ahmed, *ACS Appl. Mater. Interfaces* **2022**, *14*, 502.
- [122] H. Park, R. Mall, F. H. Alharbi, S. Sanvito, N. Tabet, H. Bensmail, F. El-Mellouhi, *J. Phys. Chem. A* **2019**, *123*, 7323.
- [123] A. Ali, H. Park, R. Mall, B. Aïssa, S. Sanvito, H. Bensmail, A. Belaidi, F. El-Mellouhi, *Chem. Mater.* **2020**, *32*, 2998.
- [124] Y. Zhao, J. Zhang, Z. Xu, S. Sun, S. Langner, N. T. P. Hartono, T. Heumueller, Y. Hou, J. Elia, N. Li, G. J. Matt, X. Du, W. Meng, A. Osvet, K. Zhang, T. Stubhan, Y. Feng, J. Hauch, E. H. Sargent, T. Buonassisi, C. J. Brabec, *Nat. Commun.* **2021**, *12*, 2191.
- [125] J. L. Teunissen, F. Da Pieve, *J. Phys. Chem. C* **2021**, *125*, 25316.
- [126] S. Tan, B. Yu, Y. Cui, F. Meng, C. Huang, Y. Li, Z. Chen, H. Wu, J. Shi, Y. Luo, D. Li, Q. Meng, *Angew. Chem. Int. Ed.* **2022**, *61*, e202201300.
- [127] J. Zhang, G. Hodes, Z. Jin, S. Liu, *Angew. Chem. Int. Ed.* **2019**, *58*, 15596.
- [128] J. Song, H. Xie, E. L. Lim, A. Hagfeldt, D. Bi, *Adv. Energy Mater.* **2022**, *12*, 2201854.
- [129] X. Chu, Q. Ye, Z. Wang, C. Zhang, F. Ma, Z. Qu, Y. Zhao, Z. Yin, H.-X. Deng, X. Zhang, J. You, *Nat. Energy* **2023**, *8*, 372.
- [130] S. Lu, Q. Zhou, L. Ma, Y. Guo, J. Wang, *Small Methods* **2019**, *3*, 1900360.
- [131] W. Yang, J. Li, X. Chen, Y. Feng, C. Wu, I. D. Gates, Z. Gao, X. Ding, J. Yao, H. Li, *ChemPhysChem* **2022**, *23*, e202100841.
- [132] G. Longo, S. Mahesh, L. R. V. Buizza, A. D. Wright, A. J. Ramadan, M. Abdi-Jalebi, P. K. Nayak, L. M. Herz, H. J. Snaith, *ACS Energy Lett.* **2020**, *5*, 2200.
- [133] B. Wang, N. Li, L. Yang, C. Dall'Agnese, A. K. Jena, S.-I. Sasaki, T. Miyasaka, H. Tamiaki, X.-F. Wang, *J. Am. Chem. Soc.* **2021**, *143*, 2207.
- [134] Z. Yang, Y. Liu, Y. Zhang, L. Wang, C. Lin, Y. Lv, Y. Ma, C. Shao, *J. Phys. Chem. C* **2021**, *125*, 22483.
- [135] Z. Gao, H. Zhang, G. Mao, J. Ren, Z. Chen, C. Wu, I. D. Gates, W. Yang, X. Ding, J. Yao, *Appl. Surf. Sci.* **2021**, *568*, 150916.
- [136] X. Yang, L. Li, Q. Tao, W. Lu, M. Li, *Comput. Mater. Sci.* **2021**, *196*, 110528.
- [137] J. Kim, E. Kim, K. Min, *Adv. Theory Simul.* **2021**, *4*, 2100263.
- [138] J. Yang, S. Xiong, J. Song, H. Wu, Y. Zeng, L. Lu, K. Shen, T. Hao, Z. Ma, F. Liu, C. Duan, M. Fahlman, Q. Bao, *Adv. Energy Mater.* **2020**, *10*, 2000687.
- [139] R. Lyu, C. E. Moore, T. Liu, Y. Yu, Y. Wu, *J. Am. Chem. Soc.* **2021**, *143*, 12766.
- [140] E. I. Marchenko, S. A. Fateev, A. A. Petrov, V. V. Korolev, A. Mitrofanov, A. V. Petrov, E. A. Goodilin, A. B. Tarasov, *Chem. Mater.* **2020**, *32*, 7383.
- [141] J. Hu, C. Wang, Q. Li, R. Sa, P. Gao, *APL Mater.* **2020**, *8*, 111109.
- [142] H.-J. Feng, P. Ma, *Appl. Phys. Lett.* **2021**, *119*, 231902.
- [143] B. Yilmaz, Ç. Odabaşı, R. Yıldırım, *Energy Technol.* **2022**, *10*, 2100948.
- [144] H. Zhang, K. Darabi, N. Y. Nia, A. Krishna, P. Ahlawat, B. Guo, M. H. S. Almaliki, T.-S. Su, D. Ren, V. Bolnykh, L. A. Castriotta, M. Zendehdel, L. Pan, S. S. Alonso, R. Li, S. M. Zakeeruddin, A. Hagfeldt, U. Rothlisberger, A. Di Carlo, A. Amassian, M. Grätzel, *Nat. Commun.* **2022**, *13*, 89.
- [145] D. P. McMeekin, P. Holzhey, S. O. Fürer, S. P. Harvey, L. T. Schelhas, J. M. Ball, S. Mahesh, S. Seo, N. Hawkins, J. Lu, M. B. Johnston, J. J. Berry, U. Bach, H. J. Snaith, *Nat. Mater.* **2023**, *22*, 73.
- [146] X. Liu, D. Luo, Z.-H. Lu, J. S. Yun, M. Saliba, S. I. Seok, W. Zhang, *Nat. Rev. Chem.* **2023**, *7*, 462.
- [147] T. Wang, J. Yang, Q. Cao, X. Pu, Y. Li, H. Chen, J. Zhao, Y. Zhang, X. Chen, X. Li, *Nat. Commun.* **2023**, *14*, 1342.

- [148] L. Luo, H. Zeng, Z. Wang, M. Li, S. You, B. Chen, A. Maxwell, Q. An, L. Cui, D. Luo, J. Hu, S. Li, X. Cai, W. Li, L. Li, R. Guo, R. Huang, W. Liang, Z.-H. Lu, L. Mai, Y. Rong, E. H. Sargent, X. Li, *Nat. Energy* **2023**, *8*, 294.
- [149] T. Chen, Y. Zhou, M. Rafailovich, *MRS Adv.* **2019**, *4*, 793.
- [150] S. Yang, Y. Duan, Z. Liu, S. Liu, *Adv. Energy Mater.* **2022**, *13*, 2201733.
- [151] L. Chao, T. Niu, W. Gao, C. Ran, L. Song, Y. Chen, W. Huang, *Adv. Mater.* **2021**, *33*, 2005410.
- [152] K. Higgins, M. Ziatdinov, S. V. Kalinin, M. Ahmadi, *J. Am. Chem. Soc.* **2021**, *143*, 19945.
- [153] Ç. Odabaşı, R. Yıldırım, *Sol. Energy Mater. Sol. Cells* **2020**, *205*, 110284.
- [154] L. Lin, T. W. Jones, T. C.-J. Yang, N. W. Duffy, J. Li, L. Zhao, B. Chi, X. Wang, G. J. Wilson, *Adv. Funct. Mater.* **2021**, *31*, 2008300.
- [155] X. Jiang, F. Wang, Q. Wei, H. Li, Y. Shang, W. Zhou, C. Wang, P. Cheng, Q. Chen, L. Chen, Z. Ning, *Nat. Commun.* **2020**, *11*, 1245.
- [156] H. Min, D. Y. Lee, J. Kim, G. Kim, K. S. Lee, J. Kim, M. J. Paik, Y. K. Kim, K. S. Kim, M. G. Kim, T. J. Shin, S. Il Seok, *Nature* **2021**, *598*, 444.
- [157] Y. Yao, C. Cheng, C. Zhang, H. Hu, K. Wang, S. De Wolf, *Adv. Mater.* **2022**, *34*, 2203794.
- [158] M. O. Yildirim, E. C. Gok, N. H. Hemasiri, E. Eren, S. Kazim, A. U. Oksuz, S. Ahmad, *ChemPlusChem* **2021**, *86*, 785.
- [159] C. She, Q. Huang, C. Chen, Y. Jiang, Z. Fan, J. Gao, *J. Mater. Chem. A* **2021**, *9*, 25168.
- [160] M. del Cueto, C. Rawski-Furman, J. Aragó, E. Ortí, A. Troisi, *J. Phys. Chem. C* **2022**, *126*, 13053.
- [161] X. Liu, J. Min, Q. Chen, T. Liu, G. Qu, P. Xie, H. Xiao, J.-J. Liou, T. Park, Z.-X. Xu, *Angew. Chem. Int. Ed.* **2022**, *61*, e202117303.
- [162] Z. Guo, S. Zhao, N. Shibayama, A. Kumar Jena, I. Takei, T. Miyasaka, *Adv. Funct. Mater.* **2022**, *32*, 2207554.
- [163] X. Zhang, S. Wang, W. Zhu, Z. Cao, A. Wang, F. Hao, *Adv. Funct. Mater.* **2022**, *32*, 2108832.
- [164] R. He, W. Wang, Z. Yi, F. Lang, C. Chen, J. Luo, J. Zhu, J. Thiesbrummel, S. Shah, K. Wei, Y. Luo, C. Wang, H. Lai, H. Huang, J. Zhou, B. Zou, X. Yin, S. Ren, X. Hao, L. Wu, J. Zhang, J. Zhang, M. Stolterfoht, F. Fu, W. Tang, D. Zhao, *Nature* **2023**, *618*, 80.
- [165] S. You, F. T. Eickemeyer, J. Gao, J.-H. Yum, X. Zheng, D. Ren, M. Xia, R. Guo, Y. Rong, S. M. Zakeeruddin, K. Sivula, J. Tang, Z. Shen, X. Li, M. Grätzel, *Nat. Energy* **2023**, *8*, 515.
- [166] W. Liu, Y. Lu, D. Wei, X. Huo, X. Huang, Y. Li, J. Meng, S. Zhao, B. Qiao, Z. Liang, Z. Xu, D. Song, *J. Mater. Chem. A* **2022**, *10*, 17782.
- [167] C. Zhi, S. Wang, S. Sun, C. Li, Z. Li, Z. Wan, H. Wang, Z. Li, Z. Liu, *ACS Energy Lett.* **2023**, *8*, 1424.
- [168] D. Luo, R. Su, W. Zhang, Q. Gong, R. Zhu, *Nat. Rev. Mater.* **2020**, *5*, 44.
- [169] P. Caprioglio, M. Stolterfoht, C. M. Wolff, T. Unold, B. Rech, S. Albrecht, D. Neher, *Adv. Energy Mater.* **2019**, *9*, 1901631.
- [170] Y. Yang, M. Yang, David T. Moore, Y. Yan, Elisa M. Miller, K. Zhu, M. C. Beard, *Nat. Energy* **2017**, *2*, 16207.
- [171] Y. Cui, P. Zhu, H. Hu, X. Xia, X. Lu, S. Yu, H. Tempeld, R.-A. Eichel, X. Liao, Y. Chen, *Angew. Chem. Int. Ed.* **2023**, e202304931.
- [172] M. Stolterfoht, C. M. Wolff, J. A. Márquez, S. Zhang, C. J. Hages, D. Rothhardt, S. Albrecht, P. L. Burn, P. Meredith, T. Unold, D. Neher, *Nat. Energy* **2018**, *3*, 847.
- [173] M. Abdi-Jalebi, Z. Andaji-Garmaroudi, S. Cacovich, C. Stavrakas, B. Philippe, J. M. Richter, M. Alsari, E. P. Booker, E. M. Hutter, A. J. Pearson, S. Lilliu, T. J. Savenije, H. Rensmo, G. Divitini, C. Ducati, R. H. Friend, S. D. Stranks, *Nature* **2018**, *555*, 497.
- [174] Y. Chen, N. Li, L. Wang, L. Li, Z. Xu, H. Jiao, P. Liu, C. Zhu, H. Zai, M. Sun, W. Zou, S. Zhang, G. Xing, X. Liu, J. Wang, D. Li, B. Huang, Q. Chen, H. Zhou, *Nat. Commun.* **2019**, *10*, 1112.
- [175] A. Kausar, A. Sattar, C. Xu, S. Zhang, Z. Kang, Y. Zhang, *Chem. Soc. Rev.* **2021**, *50*, 2696.
- [176] S. Jiang, C.-C. Wu, F. Li, Y.-Q. Zhang, Z.-H. Zhang, Q.-H. Zhang, Z.-J. Chen, B. Qu, L.-X. Xiao, M.-L. Jiang, *Rare Metals* **2021**, *40*, 1698.
- [177] L. Zhang, N. Li, D. Liu, G. Tao, W. Xu, M. Li, Y. Chu, C. Cao, F. Lu, C. Hao, J. Zhang, Y. Cao, F. Gao, N. Wang, L. Zhu, W. Huang, J. Wang, *Angew. Chem. Int. Ed.* **2022**, *61*, e202209337.
- [178] V. M. Le Corre, T. S. Sherkar, M. Koopmans, L. J. A. Koster, *Cell Rep. Phys. Sci.* **2021**, *2*, 100346.
- [179] G. Zhou, W. Chu, O. V. Prezhdo, *ACS Energy Lett.* **2020**, *5*, 1930.
- [180] N. T. P. Hartono, J. Thapa, A. Tiihonen, F. Oviedo, C. Batali, J. J. Yoo, Z. Liu, R. Li, D. F. Marrón, M. G. Bawendi, T. Buonassisi, S. Sun, *Nat. Commun.* **2020**, *11*, 4172.
- [181] K. Higgins, M. Lorenz, M. Ziatdinov, R. K. Vasudevan, A. V. Levlev, E. D. Lukosi, O. S. Ovchinnikova, S. V. Kalinin, M. Ahmadi, *Adv. Funct. Mater.* **2020**, *30*, 2001995.
- [182] R. J. Stoddard, W. A. Dunlap-Shohl, H. Qiao, Y. Meng, W. F. Kau, H. W. Hillhouse, *ACS Energ. Lett.* **2020**, *5*, 946.
- [183] M. Saliba, *Adv. Energy Mater.* **2019**, *9*, 1803754.
- [184] B. M. Lake, R. Salakhutdinov, J. B. Tenenbaum, *Science* **2015**, *350*, 1332.
- [185] K. Takahashi, L. Takahashi, I. Miyazato, Y. Tanaka, *ACS Photonics* **2018**, *5*, 771.
- [186] V. Gladkikh, D. Y. Kim, A. Hajibabaei, A. Jana, C. W. Myung, K. S. Kim, *J. Phys. Chem. C* **2020**, *124*, 8905.
- [187] M.-H. Jao, S.-H. Chan, M.-C. Wu, C.-S. Lai, *J. Phys. Chem. Lett.* **2020**, *11*, 8914.
- [188] Y. Li, Y. Lu, X. Huo, D. Wei, J. Meng, J. Dong, B. Qiao, S. Zhao, Z. Xu, D. Song, *RSC Adv.* **2021**, *11*, 15688.
- [189] M. L. Agiorgousis, Y.-Y. Sun, D.-H. Choe, D. West, S. Zhang, *Adv. Theory Simul.* **2019**, *2*, 1800173.
- [190] X.-Y. Ma, J. P. Lewis, Q.-B. Yan, G. Su, *J. Phys. Chem. Lett.* **2019**, *10*, 6734.
- [191] J. Im, S. Lee, T.-W. Ko, H. W. Kim, Y. Hyon, H. Chang, *NPJ Comput. Mater.* **2019**, *5*, 37.
- [192] Y. Wu, S. Lu, M.-G. Ju, Q. Zhou, J. Wang, *Nanoscale* **2021**, *13*, 12250.
- [193] Q. Xu, Z. Li, M. Liu, W.-J. Yin, *J. Phys. Chem. Lett.* **2018**, *9*, 6948.
- [194] D. Jain, S. Chaube, P. Khullar, S. G. Srinivasan, B. Rai, *Phys. Chem. Chem. Phys.* **2019**, *21*, 19423.
- [195] G. Pilania, P. V. Balachandran, C. Kim, T. Lookman, *Front. Mater.* **2016**, *3*, 19.
- [196] F. Oviedo, Z. Ren, S. Sun, C. Settens, Z. Liu, N. T. P. Hartono, S. Ramasamy, B. L. DeCost, S. I. P. Tian, G. Romano, A. Gilad Kusne, T. Buonassisi, *NPJ Comput. Mater.* **2019**, *5*, 60.
- [197] S. R. Xie, P. Kotlarz, R. G. Hennig, J. C. Nino, *Comput. Mater. Sci.* **2020**, *180*, 109690.
- [198] F. Massuyeau, T. Broux, F. Coulet, A. Demessence, A. Mesbah, R. Gautier, *Adv. Mater.* **2022**, *34*, 2203879.



Tiantian Liu is an associate professor at Xi'an University of Architecture and Technology. She received her Ph.D. degree from City University of Hong Kong in Prof. Alex Jen's group. Her research interests focus on the field of materials science, artificial intelligence manufacturing, and structure optimization of perovskite solar cells.



Jun Zhou is a professor at Xi'an University of Architecture and Technology. He is vice dean of the School of Chemistry and Chemical Engineering. He has published more than four books (including three books as chief editor), taken charge of a great variety of scientific research projects, won five provincial- and ministerial-level science and technology awards, and received the nomination award of the second "National University Metallurgical Dean Award". His research interests mainly focus on efficient conversion and utilization technology of carbon-based energy, exploitation and application of new energy sources, and artificial intelligence algorithms, etc.



Wallace C. H. Choy is a full professor in the Department of Electrical and Electronic Engineering, the University of Hong Kong. His research interests cover organic/inorganic photonics and optoelectronics, plasmonic structures, metal/metal oxides nanomaterials, and nanomaterial devices. He has published more than 260 peer-reviewed papers, several book chapters, patents, and edited one book. He was recognized as the top 1% of most-cited scientists in Thomson Reuter's Essential Science Indicators from 2014 to 2022. He is an elected fellow of OSA/Optica.