



Machine Learning-Assisted Defect Analysis and Optimization for P-I-N-Structured Perovskite Solar Cells

Seongtak Kim¹ · Younghun Jeong^{1,2} · Dong-Woon Han¹ · Chan Bin Mo¹

Received: 30 November 2022 / Accepted: 26 May 2023 / Published online: 22 June 2023
© The Minerals, Metals & Materials Society 2023

Abstract

A solar cell capacitance simulator (SCAPS-1D) was used to prepare 3611 cell data with different defect densities in the bulk and interface of p-i-n-structured perovskite solar cells. The training was conducted using four machine learning algorithm models. The random forest algorithm had an accuracy and root mean square error of 0.999 and 0.00306, respectively. As per the explanatory Shapley additive explanations (SHAP) analysis, the bulk defects of the perovskite and the hole transfer layer/perovskite and perovskite/electron transfer layer interface defects greatly affected the power conversion efficiency of the solar cells. It was also confirmed that when the perovskite defect density was low, the cell performance was more sensitive to the interface defect densities. Based on the predictive analysis of machine learning, a strategy to improve the solar cell efficiency of the p-i-n structure was presented, and the efficiency was improved from 17.97% to 24.66% in the poly(triarylamine)/perovskite/phenyl-C61-butric acid methyl ester structure by optimizing the defect density and resistance. It is expected that this methodology will not only help in identifying the factors affecting the efficiency of perovskite solar cells but also in optimizing the structure of solar cells during the manufacturing process.

Keywords Bulk defect · interfacial defect · machine learning · perovskite solar cells · SCAPS · SHAP analysis

Introduction

Commercialization of perovskite solar cells (PSCs) is increasing owing to their recently recorded efficiency of 25.7%, which is close to the efficiency of silicon solar cells.¹ Unlike silicon solar cells, PSCs can be manufactured as a thin film of ~ 1 μm thickness; hence, they can overcome the limitations of applicability of rigid silicon solar cells and can be easily manufactured as tandem solar cells by combining two or more solar cells.^{2,3} PSCs typically have a device structure in which various perovskite (PVSK) light absorbers, comprising organic and inorganic electron/hole

transport layers (ETLs/HTLs), are used depending on the n-i-p and p-i-n structures.^{4–10} In particular, most of the reported p-i-n-structured solar cells can be used for flexible substrate devices because they can be processed at a low temperature of less than 150°C, thereby enabling cost-effective manufacturing processes such as large-scale roll-to-roll processes. Consequently, various applications are expected as next-generation solar cells.^{11,12} However, for the commercialization of PSCs, it is necessary to study the effect of various materials on the device performance and further research is needed to improve the efficiency.

Artificial intelligence (AI), which has received significant attention, especially since the emergence of AlphaGo in 2016, is rapidly developing and being applied to various fields, thereby surpassing human capabilities such as in advanced material finding, manufacturing process optimization, predictive maintenance, and autonomous driving for next-generation mobility.^{13–18} Most AI-based prediction processes perform optimal judgment and prediction via inductive reasoning based on extensive data obtained from theoretical and experimental results, rather than performing deductive reasoning. However, because the method of inductive reasoning cannot explain why such predictions

✉ Seongtak Kim
seongtak@kitech.re.kr

Chan Bin Mo
cbmo@kitech.re.kr

¹ Functional Materials and Components R&D Group, Korea Institute of Industrial Technology (KITECH), 47 Hojeo-ro, Wonju-si, Gangwon-do 26336, Republic of Korea

² Department of Materials Science and Engineering, Yonsei University, 50 Yonsei-ro, Seodaemun-gu, Seoul 03722, Republic of Korea

are made, explanatory AI models such as Shapley additive explanations (SHAP) and local interpretable model-agnostic explanation (LIME) have been recently developed to interpret the prediction results.^{19,20}

Research and development using AI are also being conducted in the field of PSCs. Various studies on PVSK material formability, capping layer optimization for minimizing PVSK degradation, prediction of amine types, and bandgap and formation energies of undiscovered PVSks using various machine learning (ML) algorithms have been reported.^{21–25} In addition, recently, studies have been conducted on the prediction and optimization of PSCs through a solar cell capacitance simulator (SCAPS) using ML for optimization of lead-free, Cs-based, $\text{TiO}_2/\text{SnO}_2$ ETL-structured PSCs.^{26–28}

Notably, however, only a few studies have investigated the effect of defects on the performance of PSCs through the inductive method using ML. Thus, in this study, we analyzed the effect of defects on efficiency through ML and explainable AI models based on 3611 cell data from various bulk and interfacial defect conditions obtained from SCAPS

simulations. Based on our results, we propose a method to improve the efficiency through defect optimization of p-i-n-structured PSCs.

Materials and Methods

SCAPS Simulation Conditions

The simulation of the PSCs with p-i-n structure was carried out using SCAPS software (version 3.3.09).²⁹ Figure 1a shows the structure of the PSC modeled in this study. Standard parameters were input based on poly(triarylamine) (PTAA), mixed PVSK (including formamide, methylammonium, lead, iodine, and bromine), and phenyl-C61-butric acid methyl ester (PCBM), which are presented in Table I and Fig. 1b. The manufacturing thickness of the solar cell was used, and the other input parameters were based on those from previous papers.^{30–34} The absorption coefficient of each layer in the simulation is shown in Fig. 1c. All simulations were performed under one-sun (100 mW

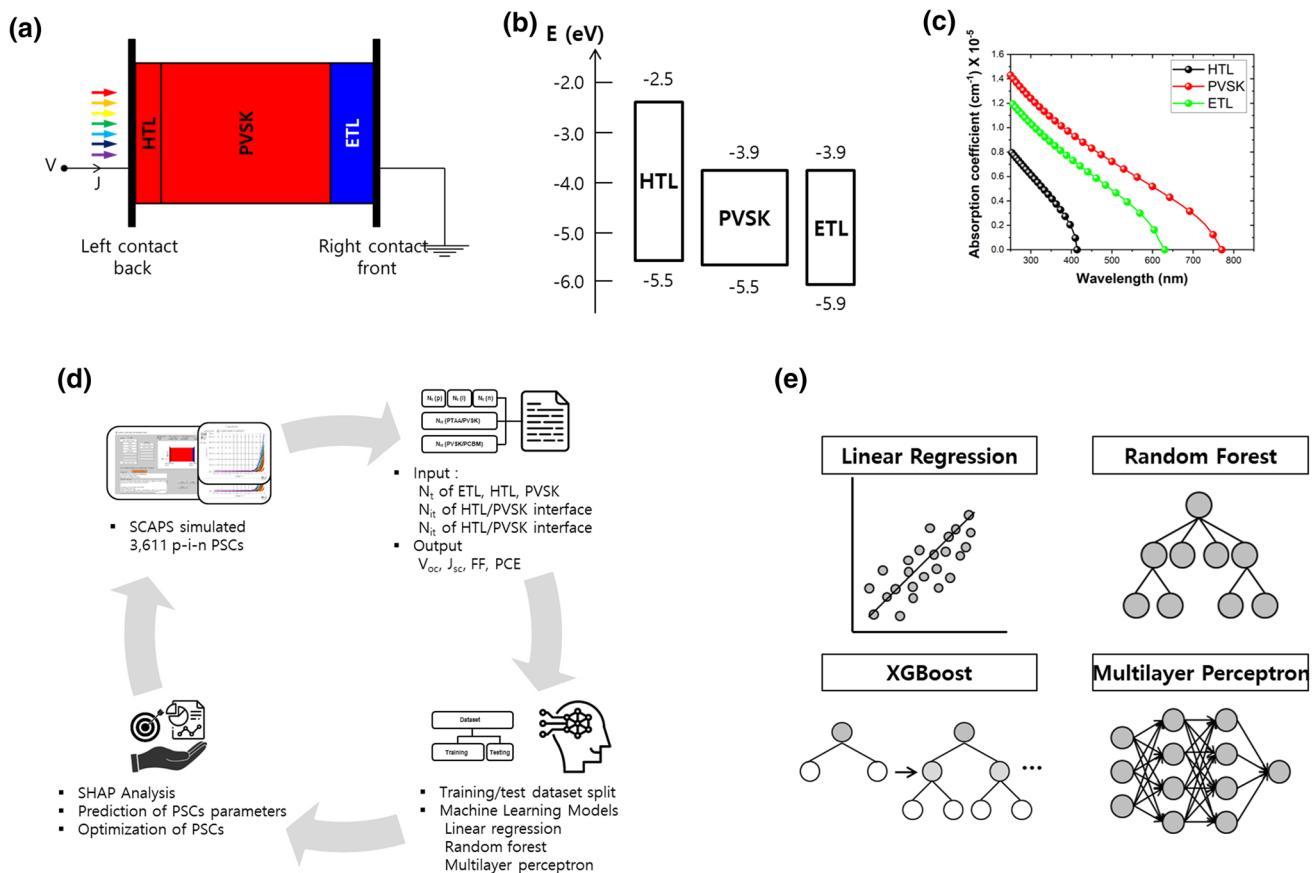


Fig. 1 (a) PSC architecture for SCAPS; (b) energy band diagram of p-i-n-structured PSCs; (c) absorption coefficients of HTL, PVSK, and ETL as input parameters for SCAPS; (d) overall process schematic

for defect analysis and optimization of PSCs using ML training and prediction; and (e) the four ML algorithms used for training, test, and prediction.

Table I Input parameters used for SCAPS simulation

		HTL	PVSK	ETL	HTL/PVSK	PVSK/ETL
Thickness	nm	20	500	30	—	—
Bandgap	eV	3.0	1.6	2.0	—	—
Electron affinity	eV	2.5	3.9	3.9	—	—
Dielectric permittivity		3.5	8.0	3.9	—	—
Conduction band effective density of states	cm ⁻³	1×10 ²⁰	2.2×10 ¹⁸	2.5×10 ²¹	—	—
Valence band effective density of states	cm ⁻³	1×10 ²⁰	2.2×10 ¹⁸	2.5×10 ²¹	—	—
Electron thermal velocity	cm s ⁻¹	1×10 ⁷	1×10 ⁷	—	—	—
Hole thermal velocity	cm s ⁻¹	1×10 ⁷	1×10 ⁷	—	—	—
Electron mobility	cm ² V ⁻¹ s ⁻¹	1.5×10 ⁻⁴	2	1×10 ⁻²	—	—
Hole mobility	cm ² V ⁻¹ s ⁻¹	1.5×10 ⁻⁴	2	1×10 ⁻²	—	—
Shallow uniform donor density	cm ⁻³	0	0	2.9×10 ¹⁷	—	—
Shallow uniform acceptor density	cm ⁻³	1×10 ¹⁸	1×10 ¹⁵	0	—	—
Capture cross section electron	cm s ⁻¹	1×10 ¹⁵	1×10 ¹⁶	1×10 ¹⁵	1×10 ¹⁹	1×10 ¹⁹
Capture cross section hole	cm s ⁻¹	1×10 ¹⁵	1×10 ¹⁶	1×10 ¹⁵	1×10 ¹⁹	1×10 ¹⁹
Energetic distribution	—	Gaussian	Gaussian	Gaussian	Gaussian	Gaussian
Energy with respect to reference	eV	Above Ev 0.6				
Defect density	cm ⁻³	1×10 ¹¹ –1×10 ¹⁹	1×10 ¹² –1×10 ²⁰	1×10 ¹¹ –1×10 ¹⁹	1×10 ¹¹ –1×10 ¹⁹	1×10 ¹¹ –1×10 ¹⁹

cm⁻²) illumination with air mass (AM) 1.5G spectrum to extract current–voltage (I–V) characteristics. To obtain I–V characteristics with the changes in bulk and interface defects of the p-i-n-structured PSCs, five types of defects were classified as follows: bulk defects (density, N_t) of HTL, PVSK, and ETL and HTL/PVSK and PVSK/ETL interfacial defects (density, N_{it}). As shown in Table I, 3611 data were formed using SCAPS in a certain range of defect densities.

ML, Evaluation, and Analysis

Figure 1d shows a schematic for the ML process using cell data, prediction, and optimization of the efficiency of p-i-n-structured PSCs. For ML, five types of defects with various defect densities were used as the input parameters, while the cell performance—power conversion efficiency (PCE), open-circuit voltage (V_{oc}), short-circuit current density (J_{sc}), and fill factor (FF)—was the output target value. The 3611 cell data were divided into 80% training set and 20% test set. Four ML models were used for training, i.e., linear regression (LIR), random forest regression (RFR), extreme gradient boosting (XGB), and multilayer perceptron (MLP) (Fig. 1e).^{35–38} The LIR model is a widely used representative regression algorithm that models the linear correlation between the dependent variable and one or more independent variables. This model can estimate unknown parameters by modeling a regression equation using a linear prediction function. The RFR model is a bagging ensemble method, which

divides the entire training data set into several small training data sets and trains each. An optimal model of the entire data set can be obtained by combining each trained model. The XGB model is a boosting ensemble model. During the process of dividing several training sets and training, weights can be provided according to the prediction accuracy from the learning results of the preceding weak learner. Each time a new model is added, the prediction error is reduced because it learns with the minimum error. The MLP consists of an artificial network structure in which several intermediate layers (hidden layers) exist between the input and output layers. The output value is used as the input value for the next step. An optimal AI learning model can be formed by adjusting the number of hidden layers and nodes. Among them, the best-learned model was selected and analyzed. Learning accuracy and loss metrics were evaluated by R squared (R^2), root mean square error (RMSE), and K-fold cross validation (CV). R^2 is an indicator of how the independent variables describe the dependent variable. Meanwhile, RMSE is the root value obtained by dividing the sum of the squares of the residual of the actual and predicted values. The smaller the RMSE value, the higher the precision of the model. Moreover, the CV method is used to ensure that all data is used as a test set at least once to prevent overfitting of the model.³⁹ This is because the trained model is suitable for the train set; however, the accuracy may decrease on the test set. In this model, data for the training and validation was shuffled for 10 iterations. The average CV values

were then displayed, where the closer the CV value to 1, the higher the accuracy. The scikit-learn library was used for the ML and evaluation algorithms used in this study.

The SHAP analysis was used to analyze the defects that affect the performance of p-i-n-structured PSCs. This analysis technique shows the importance and influence of each feature on the target value. It is one of the explanatory AI algorithms mainly used to interpret the “black box” model, which is difficult to understand using the exact result derivation process. Notably, it is possible to check the impact of each feature on the output from the SHAP value. Moreover, this value can be used to intuitively predict the positive and negative effects of each feature on the final output values.

Material Information, Device Fabrication, and Characterization

Lead (II) iodide (99.99%) and lead (II) bromide were purchased from TCI. Formamidinium iodide (FAI), methylammonium bromide (MABr), and methylammonium chloride (MACl) were purchased from Greatcell Solar Materials, whereas PCBM was obtained from Nano-C. All other materials were purchased from Sigma-Aldrich. Patterned indium tin oxide (ITO) glass substrates ($10 \Omega \text{ sq}^{-1}$) were cleaned with acetone, ethanol, and isopropanol for 10 min. The PTAA was deposited onto the ITO substrates by spin-coating of a PTAA diluted solution (1.5 mg mL^{-1} in chlorobenzene), followed by drying at 100°C for 10 min on a hot plate. Furthermore, $1.3 \text{ M} (\text{FAPbI}_3)_{0.92}(\text{MAPbBr}_3)_{0.08}$ and 0.8 M MACl in dimethyl sulfoxide (DMSO)/dimethyl-formamide (DMF) solutions were deposited onto the PTAA substrates by two-step spin-coating (1000 rpm for 10 s followed by 6000 rpm for 30 s). During the spin-coating process, $500 \mu\text{m}$ diethyl ether was dropped onto the substrate. After spin-coating, the substrates were annealed at 120°C for 30 min. The PCBM solution (20 mg mL^{-1} in chlorobenzene) was deposited onto the PVK by spin-coating, which was followed by spin-coating of 0.5 mL mL^{-1} bathocuproine (BCP) solution. Finally, the contacts were deposited by thermal evaporation of 100 nm Au in a vacuum chamber. The light I–V performance of the PSCs with an active area of 0.075 cm^2 was measured using a Keithley 2400 source meter with an AM1.5G one-sun solar simulator (WACOM WXS-155S-10 class AAA). The I–V curves were obtained using a delay time of 200 ms at each point in the reverse scan from the V_{oc} to the J_{sc} .

Optimization of the PTAA/PVK/PCBM PSCs

Based on the I–V characteristics of the fabricated PTAA/PVK/PCBM solar cells, the cell parameters were modeled using SCAPS software. To optimize and improve the cell

performance, the bulk and interfacial defect densities of PVK, PTAA/PVK, and PVK/PCBM together with the resistance of the devices were considered.

Results and Discussion

Model Training and Test Using Four ML Algorithms

Five types of defect densities and the PCEs as the representative cell characteristic were set as the input and output (target value), respectively. Learning and prediction were carried out using the four ML models. Fig. 2 shows a graph comparing the predicted value obtained using ML and the simulated value obtained using SCAPS. Regardless of bulk or interfacial defects, the cell performance degrades with increasing defect density. In particular, when the defect density is higher than $1 \times 10^{18} \text{ cm}^{-3}$, the PCE decreases rapidly. The SCAPS condition used in this study shows that when the PVK defect density is $1 \times 10^{18} \text{ cm}^{-3}$, the carrier lifetime and diffusion length are approximately 1 ns and 72 nm, respectively. In general, the PCE of sensitized-type solar cells displayed a carrier lifetime of a few nanoseconds, which was less than 10% of PCE; hence, this simulation result can be considered reasonable.^{40,41} By comparing the simulated and predicted data in Fig. 2a, b, c and d, it was observed that most of the PCE values predicted according to each defect density by the four ML models fit well. A comparison between the training and test set using the four ML models and the actual simulated value is displayed in the last column of Fig. 2. Among the four ML models, LIR and MLP did not match well while RFR and XGB matched well in terms of the simulated and predicted values, regardless of the training and test data. Table II shows the accuracy and loss metrics using the R^2 and RMSE methods for comparing the accuracy of this ML model quantitatively. In terms of R^2 , the RFR and XGB values were close to 1, indicating the high accuracy of these models. In the loss metrics result (RMSE), RFR showed loss values of 2.173×10^{-15} in the training set and 0.00306 in the test set, both of which are very low compared to those obtained with the XGB model. Because high accuracy can cause overfitting, comparative verification was conducted using CV. Additionally, the RFR model showed high accuracy of 0.999 in the CV, indicating that reliable learning was performed.

Cell Performance Prediction Model Using RFR

The changes in the cell performance (PCE, V_{oc} , J_{sc} , and FF) due to the defect densities were studied using RFR, which presented the best performance among the four tested ML models, as discussed in the “[Model Training](#)

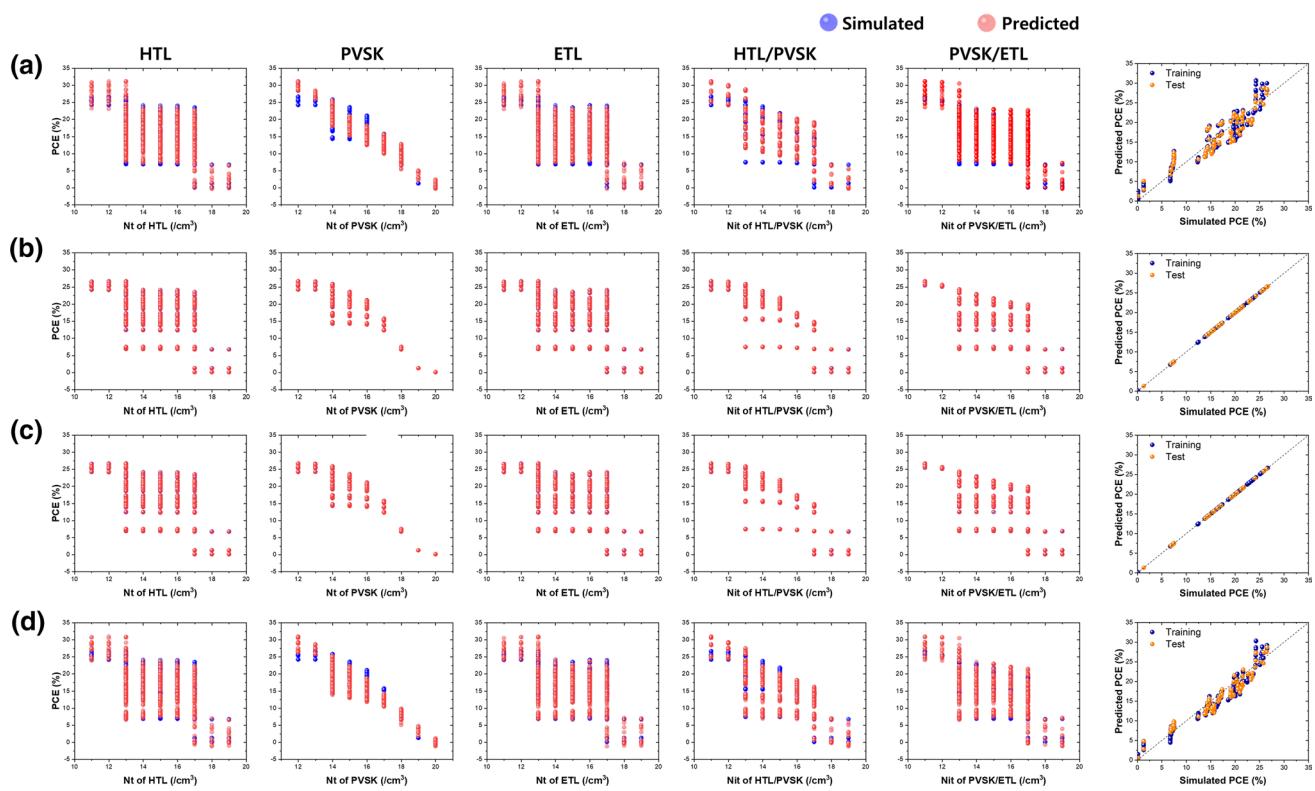


Fig. 2 Simulation and prediction of the PCE with bulk and interface defect density using four ML algorithms: (a) LIR, (b) RFR, (c) XGB, and (d) MLP; the last column represents the comparison results between the PCE simulation and prediction.

Table II Summary of model accuracy (R^2), loss metrics (RMSE), and CV of the four trained ML models

Model	R^2		RMSE		CV
	Training	Test	Training	Test	
LIR	0.843	0.848	2.480	2.493	0.843
RFR	0.999	0.999	2.173×10^{15}	0.00306	0.999
XGB	0.999	0.999	0.00338	0.00528	0.999
MLP	0.913	0.917	1.838	1.846	0.926

and Test Using Four MI Algorithms” section. Figure 3 shows a graph comparing the simulated and predicted values by the RFR model with the five types of defects as input parameters. The V_{oc} , J_{sc} , FF, and PCE showed a high accuracy of 0.999 and a very low RMSE loss value, indicating good learning of the model. Because all the data for this learning were obtained from the SCAPS results, the accuracy and loss metrics are quite high compared to those attained when the general experimental values are used. Similar results were reported in other studies in which the accuracy was high and the loss metrics were very low when performing ML using SCAPS data.^{26–28}

SHAP Analysis for Investigating the Relative Importance of Each Defect

Based on the well-learned ML results, SHAP analysis, as explainable AI, was performed to study the influence of defects on the cell performance. The SHAP analysis was also used to study the importance of the defect type. Figure 4 shows the SHAP summary plot that can be used for determining the SHAP value and feature impact. The feature value indicates the defect density distribution, whereas the SHAP value indicates the effect of each feature on the output. Figure 4a shows that as the five types of defect feature values increase, all the SHAP values become negative. This means that an increase in the defect concentration has a negative effect on the output value (cell performance). The V_{oc} , J_{sc} , and FF were the most affected by the PVK defect density. Specifically, as the PVK bulk defect increases, the lifetime and diffusion length decrease; therefore, the saturation current increases, causing a decrease in the cell parameters. PVK defects are known to form deep defect sites due to the interstitial spaces of lead and iodine and shallow defect sites due to the vacancies of these two elements.⁴² In addition, the defect concentration may increase depending on the manufacturing method and external environment, and thus, defect management for manufacturing high-efficiency

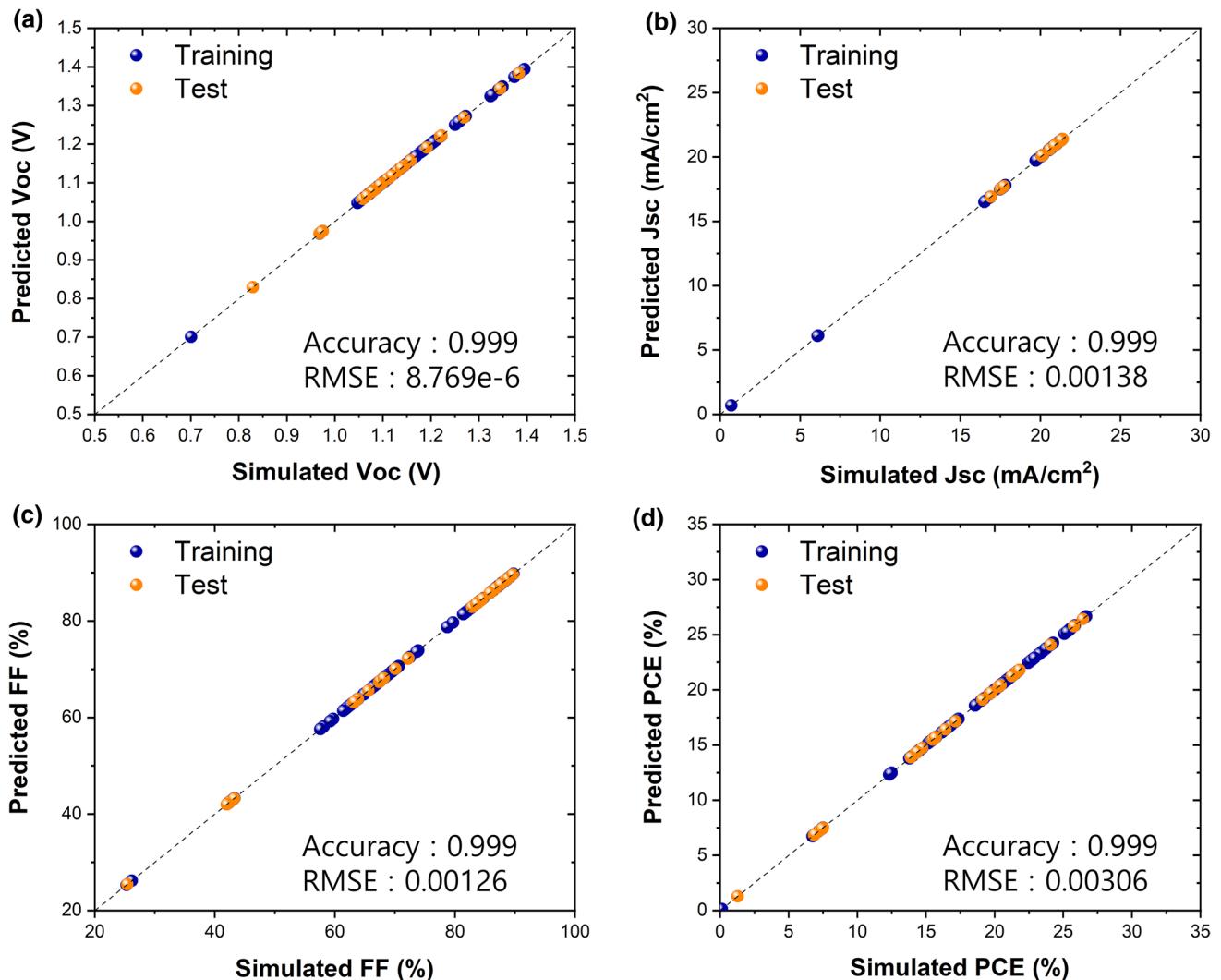


Fig. 3 Comparison between the simulated and predicted I–V characteristics of the training and test sets: (a) V_{oc} , (b) J_{sc} , (c) FF, and (d) PCE.

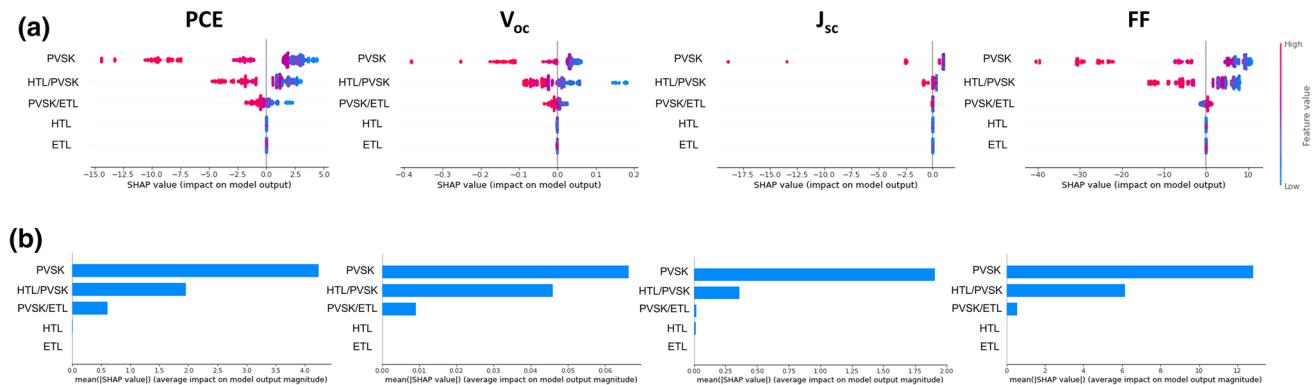


Fig. 4 SHAP summary plot of the I–V characteristics obtained by RFR algorithm and SHAP analysis: (a) SHAP value of each feature and (b) mean SHAP value which has an average impact on the model output magnitude.

solar cells is important. From the SHAP analysis results, it appears that the increase in the defects of PVSK and the interface causes a decrease in the voltage and FF of the cell. The average SHAP value is shown in Fig. 4b. Defects in the HTL and ETL appear to have little effect on cell performance changes. The second most influential factor that determines the PCE is the interfacial defect between the HTL and PVSK, because the collection probability of carriers formed in the PVSK adjacent to the HTL is affected by the incident light in the HTL direction. In the case of the n-i-p structure, the defect density of the PVSK/ETL interface is thought to have a greater effect than that of the PVSK/HTL interface. However, it can be seen that a higher defect density at the PVSK/ETL interface lowers the V_{oc} more drastically than does the defect at the HTL/PVSK interface.

Comparison Analysis of Low and High Defect Density of PVSK

To verify the influence of interfacial defects on the cell performance, the PVSK defect density section was divided into two ranges—low defect density (1×10^{12} to $1 \times 10^{16} \text{ cm}^{-3}$) and high defect density (1×10^{16} to $1 \times 10^{20} \text{ cm}^{-3}$)—and SHAP analysis was then conducted (Fig. 5a and b). It was found that the influence of interface defects on the PCE is higher in the low than in the high defect density range of PVSK. These results are confirmed by the SHAP force plot shown in Fig. 5c. When the defect concentration was less than $1 \times 10^{14} \text{ cm}^{-3}$, the influence of the interfacial defects was large. Additionally, the change in PCE due to the defects of PVSK was remarkable when the defect concentration was more than $1 \times 10^{17} \text{ cm}^{-3}$. When the PVSK defect density was $1 \times 10^{14} \text{ cm}^{-3}$, $1 \times 10^{16} \text{ cm}^{-3}$, and $1 \times 10^{18} \text{ cm}^{-3}$, the V_{oc} and PCE changed according to the change in the interfacial defect density, as shown in Fig. 5d and e, respectively. When the interfacial defect density of the HTL/PVSK and PVSK/ETL interfaces increased from $1 \times 10^{11} \text{ cm}^{-3}$ to $1 \times 10^{20} \text{ cm}^{-3}$, the V_{oc} and PCE at $1 \times 10^{14} \text{ cm}^{-3}$ PVSK defect density decreased by 23.5% and 46.7%, respectively; however, the decrease at $1 \times 10^{18} \text{ cm}^{-3}$ PVSK defect density was only 0.71% and 10.17%, respectively. This indicates that the lower the defect density of PVSK, the greater is the effect of the interfacial defect density on the PCE, which corresponds to the SHAP analysis result. In addition, the HTL/PVSK interfacial defect density only affects the PCE above $1 \times 10^{15} \text{ cm}^{-3}$, whereas the PVSK/ETL interfacial defect density affects the PCE even under $1 \times 10^{15} \text{ cm}^{-3}$. In a previous study, the interfacial defect densities of the HTL/PVSK and PVSK/ETL interfaces were reported to be $1.2 \times 10^{17} \text{ cm}^{-3}$ and $1.1 \times 10^{17} \text{ cm}^{-3}$ in PTAA/PVSK and PVSK/C60, respectively.⁴³ Therefore, we supposed that to manufacture

high-efficiency PSCs, it is important to minimize the interfacial defects. Interfacial engineering, such as the insertion of a passivation layer that can control interfacial defects, is important and is consistent with the recently reported high-efficiency solar cell research process.^{44–46} Because the defect level within PVSK can affect cell performance, the cell performances were displayed with altered defect energy levels ranging from -0.5 eV to 2.2 eV at varied defect densities ranging from $1 \times 10^{12} \text{ cm}^{-3}$ to $1 \times 10^{20} \text{ cm}^{-3}$, as shown in Fig. 5f. The PCE decreases by non-radiative recombination, owing to defects at the deep level rather than at the shallow level. Considering the change in cell performance, it is supposed that the shallow and deep defect level ranges were $0\text{--}0.2 \text{ eV}$, $1.4\text{--}1.6 \text{ eV}$, and $0.6\text{--}1.0 \text{ eV}$. In the PVSK with deep defect levels, the PCE dropped significantly when the PVSK defect density was above $1 \times 10^{16} \text{ cm}^{-3}$. In general, it is known that the defect density of single-crystal PVSK is $1 \times 10^{10} \text{ cm}^{-3}$ and that of the thin film PVSK is $1 \times 10^{16} \text{ cm}^{-3}$.^{47,48} Hence, to minimize the PCE loss, it is important to control the defect density of PVSK to $1 \times 10^{16} \text{ cm}^{-3}$ or less.

Optimization of p-i-n PSCs Through Defect and Resistance Modification

Figure 6 and Table III show the optimization process of the PCE according to the defect density and resistance modification of the PTAA/PVSK/PCBM-structured PSCs. The device modeled by the SCAPS of an experimentally fabricated solar cell has the following defect characteristics: N_t of PVSK = $1 \times 10^{14} \text{ cm}^{-3}$, N_{it} of PTAA/PVSK = $1 \times 10^{15} \text{ cm}^{-3}$ and of PVSK/PCBM = $1 \times 10^{17} \text{ cm}^{-3}$, series resistance (R_s) = $3.0 \Omega \text{ cm}^2$, and shunt resistance (R_{sh}) = $7 \times 10^2 \Omega \text{ cm}^2$. Because it was confirmed that the PVSK/PCBM interfacial defect density through the modeling of the experimentally fabricated solar cell was high, we first attempted to improve the solar cell efficiency by optimizing N_{it} . As a result of this optimization, there was a significant improvement in the PCE. Similar to the optimization results, several studies on high-efficiency PSCs have reported an improved PCE with PVSK/ETL interface defect control.^{49,50} As shown in Fig. 5, for high-efficiency solar cells, the minimization of interfacial defects is considered more important than the effect of the N_t of PVSK. Finally, through resistance optimization, the PCE (24.66%) of the optimized device was enhanced by 137% compared to that of the initial device (17.97%). The analysis using the explainable AI of SHAP proved that the bulk defects of PVSK and HTL/PVSK and PVSK/ETL interfacial defects significantly affect the cell performance. Conversely, the bulk defects of HTL and ETL have little effect on the solar cell characteristics. We confirmed that

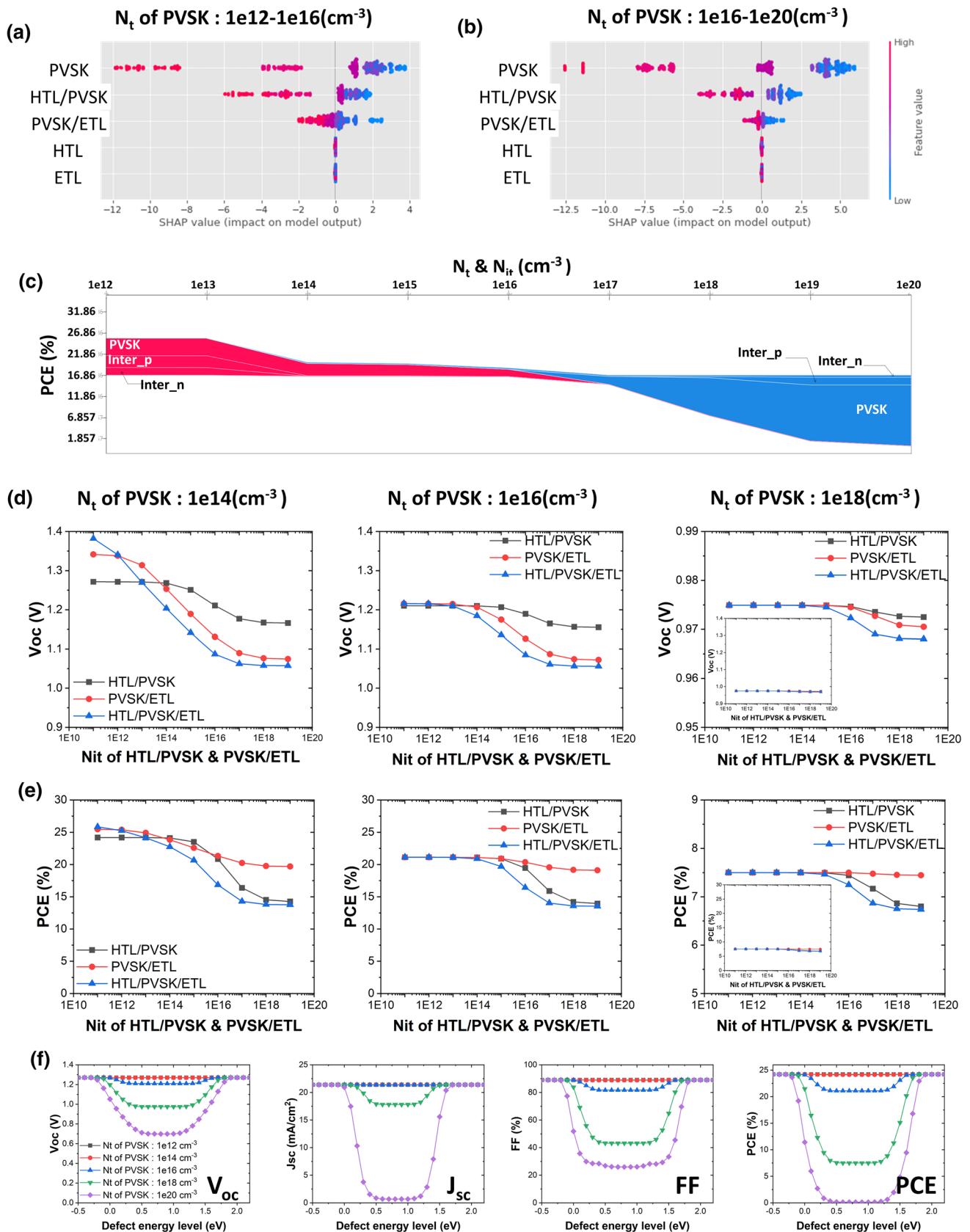


Fig. 5 SHAP summary plots for (a) N_t of PVK between $1 \times 10^{12} \text{ cm}^{-3}$ and $1 \times 10^{16} \text{ cm}^{-3}$ and (b) N_t of PVSK between $1 \times 10^{16} \text{ cm}^{-3}$ and $1 \times 10^{20} \text{ cm}^{-3}$. (c) SHAP force plot according to the bulk and interfacial defects and PSC performance variation with N_t of PVSK. (d) V_{oc} and (e) PCE changes with interfacial defect density when N_t of PVSK is $1 \times 10^{14} \text{ cm}^{-3}$, $1 \times 10^{16} \text{ cm}^{-3}$, and $1 \times 10^{18} \text{ cm}^{-3}$. (f) Changes in V_{oc} , J_{sc} , FF, and PCE with different defect positions and densities in the PVSK layer. The obtained solar cell parameters (V_{oc} , J_{sc} , FF, and PCE) are all extracted from simulated I-V characteristics under one-sun irradiation.

the influence of the interfacial defect concentration becomes more significant as the solar cell efficiency increases and that the effect of the deep level becomes more pronounced when the defect density is greater than $1 \times 10^{16} \text{ cm}^{-3}$.

In the methodology using ML, we attempted to confirm the importance of a feature (defects) that affects the results (cell efficiency) using a model that learns extensive data. Each defect affects the solar cell characteristics by a complex relationship, and thus, the degree to which defects affect the result was analyzed for optimizing the cell performance. As a result, the efficiency improved from 17.97% to 24.66% in the PTAA/PVSK /PCBM structure by optimizing the defect density and resistance. It is thought that this methodology can not only be applied to actual solar cells but also to materials and processes in other devices.

Conclusions

In this study, the effect of bulk and interfacial defects of p-i-n-structured PSCs on the cell performance was investigated using ML. Cell data with different defect concentrations within a certain range were used. Among the four ML algorithms employed, RFR showed the best results, with a high accuracy of 0.999 and a low loss metric of 0.00306. In the trials to learn V_{oc} , J_{sc} , and FF values according to defect conditions using RFR, all the models showed a prediction accuracy close to 1. Through the SHAP analysis, which is an AI algorithm that can be explained based on the learned model, it was confirmed that the bulk defect concentration of PVSK had the greatest effect on the PCE. In addition, it was confirmed that the cell performance was more sensitive to interfacial defects as the PVSK defect concentration decreased. Future improvement of the device performance could be achieved through the optimization of bulk and interfacial defects together with the resistance of the device. As a result, the PCE remarkably improved from 17.97% to 24.66%. Using this methodology, which is based on AI, the analysis and optimization of the important factors affecting solar cell efficiency can be used to provide insights into the development of solar cells.

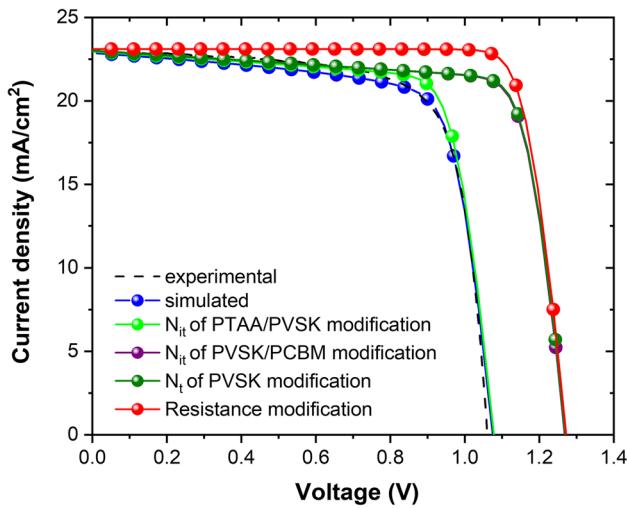


Fig. 6 Optimization progress of PTAA/PVSK/PCBM-structured PSCs.

Table III Experimental, simulated, and optimized I-V results of PTAA/PVSK/PCBM solar cells

Modification process	V_{oc} (V)	J_{sc} (mA cm^{-2})	FF (%)	PCE (%)
Experimental	1.06	23.03	73.58	17.97
Simulated	1.08	22.88	73.49	18.09
N_{it} of PTAA/PVSK modification	1.08	22.96	76.33	18.90
N_{it} of PVSK/PCBM modification	1.27	23.01	78.30	22.88
N_t of PVSK modification	1.27	23.01	78.45	22.94
Resistance modification	1.27	23.11	83.76	24.66

Acknowledgments This work was supported by the Korea Institute of Industrial Technology (UR-22-0042)

Author Contributions SK developed the ML algorithms, wrote the original manuscript, and supervised the experiments. YJ fabricated and evaluated the PSCs. DWH performed the SCAPS simulations using the prepared parameters. CBM supervised all the experiments and simulations and performed an overall review of the manuscript. All authors read and approved the final manuscript.

Conflict of interest The authors declare no known competing financial interests or personal relationships that could have influenced the work reported in this paper.

References

- NREL: Best Research-Cell Efficiencies. <https://www.nrel.gov/pv/assets/pdfs/best-research-cell-efficiencies.pdf>. Accessed 4 May 2023
- X. Meng, Z. Cai, Y. Zhang, X. Hu, Z. Xing, Z. Huang, Z. Huang, Y. Cui, T. Hu, M. Su, X. Liao, L. Zhang, F. Wang, Y. Song, and Y. Chen, Bio-inspired vertebral design for scalable and flexible perovskite solar cells. *Nat. Commun.* 11, 1 (2020).
- D. Kim, H.J. Jung, I.J. Park, B.W. Larson, S.P. Dunfield, C. Xiao, J. Kim, J. Tong, P. Boonmongkolras, S.G. Ji, F. Zhang, S.R. Pae, M. Kim, S.B. Kang, V. Dravid, J.J. Berry, J.Y. Kim, K. Zhu, D.H. Kim, and B. Shin, Efficient, stable silicon tandem cells enabled by anion-engineered wide-bandgap perovskites. *Science* 368, 155 (2020).
- Q. Jiang, L. Zhang, H. Wang, X. Yang, J. Meng, H. Liu, Z. Yin, J. Wu, X. Zhang, and J. You, Enhanced electron extraction using SnO_2 for high-efficiency planar-structure $\text{HC}(\text{NH}_2)_2\text{PbI}_3$ -based perovskite solar cells. *Nat. Energy* 2, 1 (2016).
- J. Jeong, M. Kim, J. Seo, H. Lu, P. Ahlawat, A. Mishra, Y. Yang, M.A. Hope, F.T. Eickemeyer, M. Kim, Y.J. Yoon, I.W. Choi, B.P. Darwich, S.J. Choi, Y. Jo, J.H. Lee, B. Walker, S.M. Zakeeruddin, L. Emsley, U. Rothlisberger, A. Hagfeldt, D.S. Kim, M. Gratzel, and J.Y. Kim, Pseudo-halide anion engineering for $\alpha\text{-FAPbI}_3$ perovskite solar cells. *Nature* 592, 381 (2021).
- H.S. Kim, C.R. Lee, J.H. Im, K.B. Lee, T. Moehl, A. Marchioro, S.J. Moon, R. Humphry-Baker, J.H. Yum, J.E. Moser, M. Gratzel, and N.G. Park, Lead iodide perovskite sensitized all-solid-state submicron thin film mesoscopic solar cell with efficiency exceeding 9%. *Sci. Rep.* 2, 1 (2012).
- F. Li, X. Deng, F. Qi, Z. Li, D. Liu, D. Shen, M. Qin, S. Wu, F. Lin, S.H. Jang, J. Zhang, X. Lu, D. Lei, C.S. Lee, Z. Zhu, and A.K.-Y. Jen, Regulating surface termination for efficient inverted perovskite solar cells with greater than 23% efficiency. *J. Am. Chem. Soc.* 142, 20134 (2020).
- J.Y. Jeng, Y.F. Chiang, M.H. Lee, S.R. Peng, T.F. Guo, P. Chen, and T.C. Wen, $\text{CH}_3\text{NH}_3\text{PbI}_3$ perovskite/fullerene planar-heterojunction hybrid solar cells. *Adv. Mater.* 25, 3727 (2013).
- S. Cacovich, G. Vidon, M. Degani, M. Legrand, L. Gouda, J.B. Puel, Y. Vaynzof, J.F. Guillemoles, D. Ory, and G. Grancini, Imaging and quantifying non-radiative losses at 23% efficient inverted perovskite solar cells interfaces. *Nat. Commun.* 13, 1 (2022).
- X. Lin, D. Cui, X. Luo, C. Zhang, Q. Han, Y. Wang, and L. Han, Efficiency progress of inverted perovskite solar cells. *Energy Environ. Sci.* 13, 3823 (2020).
- D. Angmo, G. Deluca, A.D. Scully, A.S.R. Chesman, A. Seeber, C. Zuo, D. Vak, U. Bach, and M. Gao, A lab-to-fab study toward roll-to-roll fabrication of reproducible perovskite solar cells under ambient room conditions. *Cell Rep. Phys. Sci.* 2, 100293 (2021).
- A. Guchhait, G.K. Dalapati, P. Sonar, S. Gopalan, F.B. Suhaimi, T. Das, V.G.V. Dutt, N. Mishra, C. Mahata, A. Kumar, and S. Ramakrishna, p-i-n Structured semitransparent perovskite solar cells with solution-processed electron transport layer. *J. Electron. Mater.* 50, 5732 (2021).
- D. Silver, A. Huang, C.J. Maddison, A. Guez, L. Sifre, G.V.D. Driessche, J. Schrittwieser, I. Antonoglou, V. Panneershelvam, M. Lanctot, S. Dieleman, D. Grewe, J. Nham, N. Kalchbrenner, I. Sutskever, T. Lillicrap, M. Leach, K. Kavukcuoglu, T. Graepel, and D. Hassabis, Mastering the game of Go with deep neural networks and tree search. *Nature* 529, 484 (2016).
- A. Talapatra, B.P. Uveruaga, C.R. Stanek, and G. Pilania, A machine learning approach for the prediction of formability and thermodynamic stability of single and double perovskite oxides. *Chem. Mater.* 33, 845 (2021).
- D. Weichert, P. Link, A. Stoll, S. Ruping, S. Ihlenfeldt, and S. Wrobel, A review of machine learning for the optimization of production processes. *J. Adv. Manuf. Technol.* 104, 1889 (2019).
- T.P. Carvalho, F.A.A.M.N. Soares, R. Vita, R.D.P. Fancisco, J.P. Basto, and S.G.S. Alcala, A systematic literature review of machine learning methods applied to predictive maintenance. *Comput. Ind. Eng.* 137, 106024 (2019).
- H. Fujiyoshi, T. Hirakawa, and T. Yamashita, Deep learning-based image recognition for autonomous driving. *IATSS Res.* 43, 244 (2019).
- X. Cai, F. Liu, A. Yu, J. Qin, M. Hatamvand, I. Ahmed, J. Luo, Y. Zhang, H. Zhang, and Y. Zhan, Data-driven design of high-performance $\text{MASn}_x\text{Pb}_{1-x}\text{I}_3$ perovskite materials by machine learning and experimental realization. *Light Sci. Appl.* 11, 234 (2022).
- S.M. Lundberg and S.I. Lee, A unified approach to interpreting model predictions. *Adv. Neural. Inf. Process Syst.* 31, 4768 (2017).
- M. T. Ribeiro, S. Singh, C. Guestrin, "Why should I trust you?" Explaining the predictions of any classifier. In Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining 1135 (2016).
- Q. Xu, Z. Li, M. Liu, and W.J. Yin, Rationalizing perovskite data for machine learning and materials design. *J. Phys. Chem. Lett.* 9, 6948 (2018).
- Y. Yu, X. Tan, S. Ning, and Y. Wu, Machine learning for understanding compatibility of organic–inorganic hybrid perovskites with post-treatment amines. *ACS Energy Lett.* 4, 397 (2019).
- N.T.P. Hartono, J. Thapa, A. Tiilinen, F. Oviedo, C. Batali, J.J. Yoo, Z. Liu, R. Li, D.F. Marron, M.G. Bawendi, T. Buonassisi, and S. Sun, How machine learning can help select capping layers to suppress perovskite degradation. *Nat. Commun.* 11, 1 (2020).
- K. Takahashi, L. Takahashi, I. Miyazato, and Y. Tanaka, Searching for hidden perovskite materials for photovoltaic systems by combining data science and first principle calculations. *ACS Photon.* 5, 771 (2018).
- X. Li, Y. Dan, R. Dong, Z. Cao, C. Niu, Y. Song, S. Li, and J. Hu, Computational screening of new perovskite materials using transfer learning and deep learning. *Appl. Sci.* 9, 5510 (2019).
- M.S. Islam, M.T. Islam, S. Sarker, H.A. Jame, S.S. Nishat, M.R. Jani, A. Rauf, S. Ahsan, K.M. Shorowordi, H. Efstathiadis, J. Carbonara, and S. Ahmed, Machine learning approach to delineate the impact of material properties on solar cell device Physics. *ACS Omega* 7, 22263 (2022).
- 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, and S. Ahmed, Supervised machine learning-aided SCAPS-based quantitative analysis for the discovery of optimum bromine doping in methylammonium tin-based perovskite ($\text{MASnI}_{3-x}\text{Br}_x$). *ACS Appl. Mater. Interfaces* 14, 502 (2021).

28. I.O. Oboh, Y.H. Offor, and N.D. Okon, Artificial neural network modeling for potential performance enhancement of a planar perovskite solar cell with a novel $\text{TiO}_2/\text{SnO}_2$ electron transport bilayer using nonlinear programming. *Energy Rep.* 8, 973 (2022).
29. M. Burgelman, P. Nollet, and S. Degrave, Modelling polycrystalline semiconductor solar cells. *Thin Solid Films* 361, 527 (2000).
30. S.Z. Haider, H. Anwar, and M. Wang, A comprehensive device modeling of perovskite solar cell with inorganic copper iodide as hole transport material. *Semicond. Sci. Technol.* 33, 035001 (2018).
31. N. Lakhdar and A. Hima, Electron transport material effect on performance of perovskite solar cells based on $\text{CH}_3\text{NH}_3\text{GeI}_3$. *Opt. Mater.* 99, 109517 (2020).
32. M.S. Chowdhury, S.A. Shahahmadi, P. Chelvanathan, S.K. Tiong, N. Amin, K. Techato, N. Nuthammachot, T. Chowdhury, and M. Suklueng, Effect of deep-level defect density of the absorber layer and n/i interface in perovskite solar cells by SCAPS-1D. *Results Phys.* 16, 102839 (2020).
33. Y.H. Khattak, F. Baig, A. Shuja, L. Atourki, K. Riaz, and B.M. Soucase, Device optimization of PIN structured perovskite solar cells: impact of design variants. *ACS Appl. Electron. Mater.* 3, 3509 (2021).
34. L.K. Ono, S. Liu, and Y. Qi, Reducing detrimental defects for high-performance metal halide perovskite solar cells. *Angew. Chem. Int. Ed.* 59, 6676 (2020).
35. P. Refaeilzadeh, L. Tang, and H. Liu, Cross-validation, *Encyclopedia of database systems*. ed. L. Liu, and M.T. Özsü (Boston: Springer, 2009), p. 532.
36. S. Ray, A quick review of machine learning algorithms. In 2019 International conference on machine learning, big data, cloud and parallel computing (COMITCon) IEEE 35 (2019).
37. L. Breiman, Bagging predictors. *Mach. Learn.* 24, 123 (1996).
38. T. Chen, C. Guestrin, Xgboost: A scalable tree boosting system. In Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining. Association for Computing Machinery, 785 (2016).
39. S.K. Pal and S. Mitra, Multilayer perceptron, fuzzy sets, classification. *IEEE trans. neural netw.* 3, 683 (1992).
40. J. Bisquert, F. Fabregat-Santiago, I. Mora-Sero, G. Garcia-Belmonte, and S. Gimenez, Electron lifetime in dye-sensitized solar cells: theory and interpretation of measurements. *J. Phys. Chem. C.* 113, 17278 (2009).
41. A. Kojima, K. Techima, Y. Shirai, and T. Miyasaka, Organometal halide perovskites as visible-light sensitizers for photovoltaic cells. *J. Am. Chem. Soc.* 131, 6050 (2009).
42. M.L. Petrus, J. Schlipf, C. Li, T.P. Gujar, N. Giesbrecht, P. Müller-Buschbaum, M. Thelakkat, T. Bein, S. Huttner, and P. Docampo, Capturing the sun: a review of the challenges and perspectives of perovskite solar cells. *Adv. Energy Mater.* 7, 1700264 (2017).
43. Z. Ni, C. Bao, Y. Liu, Q. Jiang, W.Q. Wu, S. Chen, X. Dai, B. Chen, B. Hartweg, and J. Huang, Resolving spatial and energetic distributions of trap states in metal halide perovskite solar cells. *Science* 367, 1352 (2020).
44. F. Zhang, S. Ye, H. Zhang, F. Zhou, Y. Hao, H. Cai, J. Song, and J. Qu, Comprehensive passivation strategy for achieving inverted perovskite solar cells with efficiency exceeding 23% by trap passivation and ion constraint. *Nano Energy* 89, 106370 (2021).
45. J. Peng, Y. Wu, W. Ye, D.A. Jacobs, H. Shen, X. Fu, Y. Wan, T. Duong, N. Wu, C. Barugkin, H.T. Nguyen, D. Zhong, J. Li, T. Lu, Y. Liu, M.N. Lockrey, K.J. Weber, K.R. Catchpole, and T.P. White, Interface passivation using ultrathin polymer–fullerene films for high-efficiency perovskite solar cells with negligible hysteresis. *Energy Environ. Sci.* 10, 1792 (2017).
46. P. Boonmongkolras, S.D.H. Naqvi, D. Kim, S.R. Pae, M.K. Kim, S. Ahn, and B. Shin, Universal passivation strategy for the hole transport layer/perovskite interface via an alkali treatment for high-efficiency perovskite solar cells. *Sol. RRL* 5, 2000793 (2021).
47. V. Adinolfi, M. Yuan, R. Comin, E.S. Thibau, D. Shi, M.I. Saidaminov, P. Kanjanaboons, D. Kopilovic, S. Hoogland, Z.H. Lu, O.M. Bakr, and E.H. Sargent, The in-gap electronic state spectrum of methylammonium lead iodide single-crystal perovskites. *Adv. Mater.* 28, 3406 (2016).
48. T.M. Brenner, D.A. Egger, L. Kronik, G. Hodes, and D. Cahen, Hybrid organic–inorganic perovskites: low-cost semiconductors with intriguing charge-transport properties. *Nat. Rev. Mater.* 1, 1 (2016).
49. 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, and S.I. Seok, Perovskite solar cells with atomically coherent interlayers on SnO_2 electrodes. *Nature* 598, 444 (2021).
50. P.W. Liang, C.C. Chueh, S.T. Williams, and A.K.Y. Jen, Roles of fullerene-based interlayers in enhancing the performance of organometal perovskite thin-film solar cells. *Adv. Energy Mater.* 5, 1402321 (2015).

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.