Solar Energy 315 (2026) 114743

Contents lists available at ScienceDirect

Solar Energy

journal homepage: www.elsevier.com/locate/solener

Machine learning assisted development of lead-free $Cs_3Bi_2I_9$ perovskite solar cells

Ankit Choudhary [a], Abdul Hamid Rumman [b], Miah Abdullah Sahriar [c], Md Tohidul Islam [d], Saquib Ahmed [e], Haralabos Efstathiadis [a,*]

[a] Department of Nanoscale Science & Engineering, College of Nanotechnology, Science, and Engineering, University at Albany, 1400 Washington Ave, Albany, NY 12222, USA
[b] Department of Materials and Metallurgical Engineering (MME), Bangladesh University of Engineering and Technology (BUET), East Campus, Dhaka 1000, Bangladesh
[c] Department of Chemical Engineering and Materials Science, University of Minnesota, Minneapolis, MN 55455, USA
[d] Department of Materials Design and Innovation, University at Buffalo, Buffalo, NY 14260, USA
[e] Department of Mechanical Engineering Technology, Center for Integrated Studies in Nanoscience and Nanotechnology, SUNY – Buffalo State University, 1300 Elmwood Avenue, Buffalo, NY 14222, USA

ARTICLE INFO

Keywords:
Lead-free
Machine-learning
Bismuth perovskite

ABSTRACT

Perovskite solar cells (PSCs) have emerged as a potential alternative to conventional silicon photovoltaics due to their high-power conversion efficiencies (PCEs); however, their commercialization remains hindered by the toxicity of lead-based active material and short lifetime. This current investigation presents a methodology for machine learning (ML) assisted fabrication of $Cs_3Bi_2I_9$ PSCs. A dataset (about 26,000 data points) from the Perovskite Database Project was utilized to train predictive ML models to optimize device fabrication parameters. Furthermore, the SHapley Additive exPlanations (SHAP) algorithm was used for the feature importance analysis, delineating which fabrication parameters are more important than others in impacting final device performance. ML results were then used in the experimental fabrication of $Cs_3Bi_2I_9$-based perovskite solar cells, achieving a maximum PCE of 3.73%. The ML-predicted PCE of this optimally fabricated device was 4.58%, showcasing an offset of 0.85%. Additionally, local explanations from SHAP analysis revealed individual contributions from numerical, binary, and categorical features, related to device architecture (e.g., perovskite, HTL, ETL material, and bandgap) as well as layer deposition parameters (e.g., deposition procedure, solvents, and annealing temperature). Overall, this study presents an ML-assisted approach for identifying and fabricating lead-free PSCs and advances $Cs_3Bi_2I_9$ as an alternative to its lead-based counterparts.

1. Introduction

Since the first all-solid-state perovskite solar cell surpassed the 10% power conversion efficiency (PCE) mark in 2012 [1], perovskite solar cells have demonstrated great promise, with their PCE increasing beyond 25% [2]. This superior performance of perovskite solar cells is attributed to the unique optical properties of perovskite materials, such as higher optical absorptions [3], low concentrations of non-radiative defects [4], and tunable band gap [5]. Commercialization of these devices is, however, hindered primarily due to the presence of toxic lead in the $APbX_3$ perovskite general structure and the shorter lifetime of these devices compared to their silicon-based counterparts [6]. To address the toxicity concern of lead, Sn-based perovskites are being considered for photovoltaic applications, as tin has a similar electronic configuration to lead with two s-orbital electrons. Due to this, tin-based perovskites are the most explored lead-free perovskites for photovoltaic applications, with a reported PCE of 14.63% [7]. However, these devices still suffer from poor lifetime due to the unstable nature of Sn^{2+}, present in the perovskite structure [8]. Bismuth has been used as an alternative B-site cation to address these challenges, as Bi^{3+} has the electronic configuration of 6 s^2, similar to that of Pb^{2+}. Cs^+ is often used as the A-site cation in Bi^{3+}-based perovskites due to its inorganic nature. $Cs_3Bi_2I_9$-based solar cells have shown significant improvements in lifetime, although the poor morphology of the perovskite film remains a major factor limiting the PCE of these devices [9–13]. Bai et al. used a dissolution–recrystallization method to address the morphology challenge,

* Corresponding author.
E-mail address: hefstathiadis@albany.edu (H. Efstathiadis).

https://doi.org/10.1016/j.solener.2026.114743
Received 22 January 2026; Received in revised form 20 April 2026; Accepted 27 May 2026

achieving a PCE of 3.20% [14]. In this work, we present a detailed machine learning (ML) assisted fabrication of a $Cs_3Bi_2I_9$ perovskite solar cell with an experimental PCE of 3.73%. Once $Cs_3Bi_2I_9$ is selected as the active material, ML methodologies have been employed to study the material system and assist in device fabrication by predicting favorable experimental conditions.

In recent years, ML models have been employed for predicting various fundamental material properties of PSCs, including bandgap [15–18], formability [19–22], stability [18,23], and the performance metrics, i.e., PCE, open-circuit voltage (V_{OC}), short-circuit current (J_{SC}), and fill factor (FF) [24–27]. The application of ML has expanded, enabling the identification of optimal halide concentrations for various non-toxic substitutes [28]. Additionally, advanced deep learning-based algorithms such as Artificial Neural Networks (ANN) and Genetic Algorithms (GA) are being utilized for device structure and fabrication process optimization tasks [29]. However, traditional ML algorithms such as Random Forest (RF), Gradient boosting machines (GBM, XGB, LightGBM), Decision Trees (DT), Support Vector Machines (SVM), K-nearest neighbors (KNN), and so on offer superior interpretability for structured data, which is essential for further experimental validation. These ML algorithms allow researchers to harness the power of the large data pool recorded through extensive experiments. Furthermore, ML models developed using experimentally validated data can provide valuable insights into the impact of material properties on the device performance and uncover relationships among complex physicochemical processes involved. Moreover, compared to the conventional simulation-based software and time-intensive heuristic experiments, the efficiency of ML models is unparalleled.

Nevertheless, most of these studies are based on a limited subset of perovskite materials collected through high-throughput experiments or theoretical simulations such as Density Function Theory (DFT) and SCAPS simulations. One of the notable studies in the context of $Cs_3Bi_2I_9$ PSCs was conducted by S. Sun et al., where a deep neural network-assisted methodology was developed to predict perovskite dimensionality leveraging 75 distinct halide perovskite X-ray diffraction (XRD) peak data [30]. Similarly, a study led by M. S. Islam et al. utilized SCAPS-1D simulation to generate 63,500 unique device data, utilizing Cs-perovskite as the absorber layer, and employed ML algorithms to predict the PCE [31].

In this project, we utilized 26,000 experimental data points filtered from the largest perovskite solar device database, "The Perovskite Database Project" [32], to train machine learning models on a comprehensive set of 29 features to predict the V_{OC}, J_{SC}, FF, and PCE. On the one hand, various quantitative features were utilized, including parameters such as the thicknesses of each layer (e.g., perovskite absorber layer, electron transport layer (ETL), hole transport layer (HTL), and back-contact layer), and thermal annealing temperature and time. On the other hand, various categorical features were also integrated, such as the layer deposition techniques for perovskite, ETL, and HTL. These features include widely used deposition procedures, such as spin-coating, drop-casting, electrospinning, evaporation, and so on. The novelty of the present work, in the context of existing $Cs_3Bi_2I_9$ PSCs lies in the application of feature interpretability. While many researchers use traditional feature importance analysis, we used a game-theoretic algorithm called SHAP to interpret the most important features that can be further optimized during solar cell fabrication. In this work, both global and local SHAP explanations are used to quantify the contribution of individual features, including categorical variables such as HTL/ETL material and deposition procedure, for the specific $Cs_3Bi_2I_9$ configuration. To the best of our knowledge, such device-specific interpretability has not been previously reported. Moreover, the performance measured from the experimentally fabricated $Cs_3Bi_2I_9$ showed similar results with the ML predicted ones, underscoring the potential of data-driven methodologies for solar cell research.

2. Materials and methods

2.1. Machine learning methodology

The methodology for the ML part of this research was previously developed in a previous study by Rumman et. al. [33]. The dataset was initially pre-processed to screen duplicate entries and potential outliers using the interquartile range method applied to the PCE values. Meanwhile, the missing values of numerical features were imputed by employing the KNN imputation algorithm. After the data preprocessing step, 26,440 instances remained in the cleaned dataset, comprising 5 numerical features, 24 categorical features, and four target variables (i.e., V_{OC}, J_{SC}, FF, and PCE). The input $Cs_3Bi_2I_9$ PSC features are shown in Table 1. A detailed description of the features can be found on the database website (https://www.perovskitedatabase.com/Resources). Subsequently, the categorical features were encoded using the target encoding method. While many encoding algorithms are available, such as frequency encoding, label encoding, and one-hot encoding, according to a benchmark study, the target encoding algorithm yields better results with categorical features with a high number of unique labels [34]. Consequently, an encoded dataset was generated for each target variable.

Each encoded dataset was split into a 70:30 ratio as training and test sets. Subsequently, ten widely used traditional regression algorithms (i.e., Random Forest (RF), Gradient Boosting (GB), Xtreme Gradient Boosting (XGB), Decision Tree (DT), etc.) were trained. While these algorithms are robust, their parameters (also known as hyperparameters) need to be tuned to reduce the error of the models. Consequently, "RandomSearchCV" and "GridSearchCV" techniques were used to search the optimal hyperparameters of the models. Both of these techniques iterate through a set of combinations and suggest the best

Table 1
$Cs_3Bi_2I_9$ perovskite features.

Feature List			
Feature Name	Model Input	Feature Name	Model Input
Cell architecture	nip	Perovskite band gap	2.28
Cell flexible	False	Perovskite band gap graded	False
Cell semi-transparent	False	Perovskite deposition procedure	Spin-coating
Substrate stack sequence	SLG | FTO	Perovskite deposition solvents	DMF; DMSO
ETL stack sequence	c-TiO_2	Perovskite deposition quenching induced crystallization	True
ETL thickness	40 nm	Perovskite deposition thermal annealing temperature	150 °C
ETL deposition procedure	Spin-coating	Perovskite deposition thermal annealing time	10 min
Perovskite dimension 2D	False	Perovskite deposition solvent annealing	False
Perovskite dimension 2D3D mixture	False	HTL stack sequence	CuSCN
Perovskite dimension 3D	True	HTL thickness list	220 nm
Perovskite dimension 3D with 2D capping layer	False	HTL deposition procedure	Spin-coating
Perovskite composition perovskite ABC_3 structure	False	Back contact stack sequence	Ag
Perovskite composition long-form	$Cs_3Bi_2I_9$	Back contact thickness list	100 nm
Perovskite thickness	500 nm	Back contact deposition procedure	Evaporation
Perovskite composition inorganic	True		

possible set of model parameters. The model performances were evaluated based on Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and Correlation Score (R^2 Score). Finally, the best-performing model (Random Forest) was used to predict the V_{OC}, J_{SC}, FF, and PCE. The overall ML methodology steps are visually illustrated in Fig. 1.

2.2. Materials

Titanium diisopropoxide bis(acetylacetonate), N,N-Dimethylformamide (DMF), Cesium Iodide (CsI), Bismuth Iodide (BiI_3), and Cuprous thiocyanate (CuSCN) were purchased from Sigma-Aldrich. Ethyl alcohol was purchased from Pharmco. Dimethyl Sulfoxide (DMSO) was purchased from Acros. Silver pellets were purchased from Kurt J. Lesker.

2.3. Substrate Preparation

Fluorine-dopped tin oxide coated glass substrate (Ossila TEC 8, S2001S1, 6–9 Ω/square) was patterned using zinc powder and 2 M HCl solution. After the wet etching step, substrates were rinsed with deionized water and cleaned in the ultrasonic cleaner successively with deionized water, acetone, ethanol, and isopropyl alcohol (IPA) for 15 min each. These substrates were then treated with UV/O_3 for 20 min.

To compound the compact TiO_2 (c-TiO_2) ETL, 2 mL of titanium diisopropoxide bis(acetylacetonate) was dropwise added to 20 mL of anhydrous ethanol and stirred for 30 min at room temperature. The c-TiO_2 layer was then spin-coated on FTO glass substrate at 4000 RPM for 30 s and dried on a hot plate at 150 °C for 10 min, followed by a substrate heat treatment at 500 °C for 1 h. After c-TiO_2 deposition, samples were transferred into the glovebox for the next steps of the fabrication process.

2.4. Perovskite solution preparation and device fabrication

The $Cs_3Bi_2I_9$ was prepared using the stoichiometric molar ratio of CsI and BiI_3. In short, 0.5 M $Cs_3Bi_2I_9$ precursor was prepared by dissolving 2.33829 g CsI and 3.53814 g BiI_3 in 6 mL mixed solvent of DMF/DMSO (7:3). The solution was stirred at 650 rpm for 24 h at 85 °C. The perovskite film was deposited by spin coating 150 µl of the prepared perovskite precursor solution on the FTO/c-TiO_2 substrate at 2000 rpm for 5 s and 6000 rpm for 30 s. The spin-coated sample was then annealed at 80 °C for 15 min. The CuSCN HTL was then deposited by spin-coating cuprous thiocyanate solution in diethyl sulfide (35 mg/mL) at 3000 rpm for 30 s, followed by 10-minute annealing at 150 °C. Afterward, 80 nm silver was deposited by thermal evaporation.

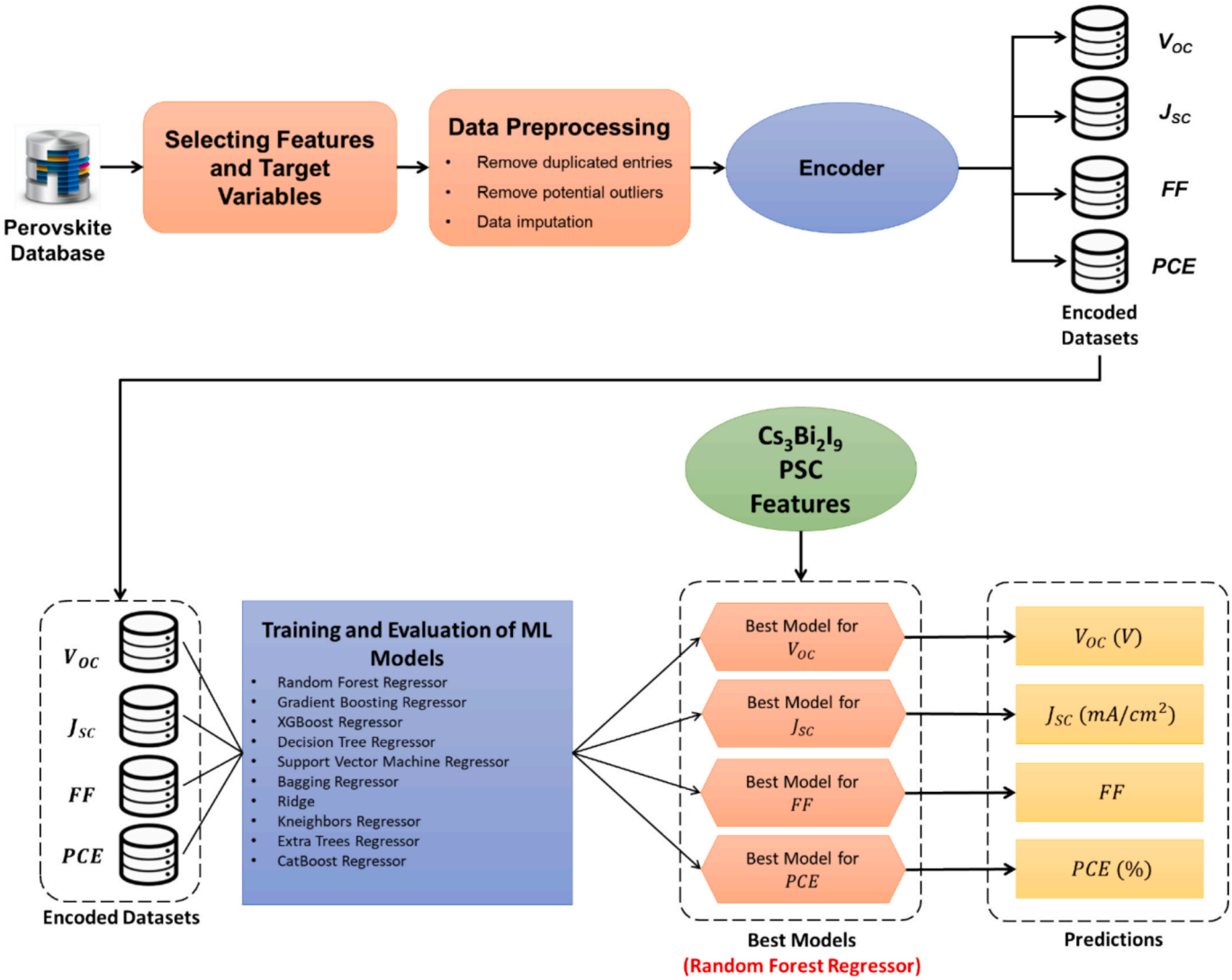

Fig. 1. Dataset preprocessing, encoding, and model development methodology.

2.5. Film and device characterization

Current density–voltage characteristics were studied using PV Measurements Solar Cell I-V measurements system (Model IV16(L)) under the illumination of one sun (100 mW/cm^2) at AM 1.5G. The cells were illuminated for 30 s before measurements. The absorption profile was studied using the PerkinElmer 650 UV/VIS spectrometer. A Bruker D8 Advance diffractometer equipped with a Cu Kα radiation source was used to study the crystallinity of the deposited films by XRD analysis. A scanning electron microscope (SEM, Leo Zeiss 1550 FE-SEM) was used to investigate the morphology of the deposited films.

3. Results and discussion

In this study, extensive data from 1604 unique perovskites were used to train 10 different traditional ML models. Subsequently, the best model was chosen to predict the V_{OC}, J_{SC}, FF, and PCE, which were compared to the values measured from the experimentally fabricated cell. This investigation's uniqueness lies in the incorporation of cell fabrication features with the device's architectural features. As a result, researchers can gain valuable insights into which material to choose for each layer and the optimal experimental fabrication method to use.

3.1. ML model performances and predictions

The most widely used error metrics for regression algorithms include MAE, MSE, RMSE, and R^2 score, which were utilized to assess model performance. The metrics are explained briefly in the following, along with the corresponding mathematical formula:

Mean Absolute Error (MAE): Mean value of absolute differences between actual (y_i) and predicted values ($\hat{y}_i$), expressed in the same units as the actual values by the formula shown in Eq. (1).

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{1}$$

Mean Squared Error (MSE): Mean value of squared differences between actual (y_i) and predicted value ($\hat{y}_i$), which heavily penalizes large errors (outliers) expressed by the formula in Eq. (2).

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{2}$$

Root Mean Squared Error (RMSE): The square root of mean squared error (MSE), measured in the original units, making it more interpretable compared to MSE which is expressed by the formula in Eq. (3).

$$RMSE = \sqrt{MSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{3}$$

R^2-score: The proportion of the variance in the dependent variable explained by the model (0 to 1), calculated from two components: residual sum of squares (SSR—the sum of squared differences between the actual (y_i) and predicted values ($\hat{y}_i$)) and total sum of squares (SST—the sum of squared differences between the actual values (y_i) and mean of the actual values ($\overline{y_i}$). It is expressed by the formula in Eq. (4).

$$R^2 = 1 - \frac{SSR}{SST} = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \overline{y_i})^2} \tag{4}$$

We particularly emphasized the R^2 score (1.0 for perfect prediction) for the PCE since it is the most critical parameter for a solar device. In comparison among various models, the Random Forest model achieved the overall highest scores on both training (R^2 score = 0.881) and test (R^2 = 0.71314) datasets. The detailed performance benchmarking based on MSE, RMSE, MAE, and R^2-score was presented in a separate article in Ref. [33]. The model performance metrics are summarized in Table 2.

3.2. Feature importance analysis using SHAP

SHapley Additive exPlanations (SHAP) is a contemporary algorithm that can be used for tree-based ML models that can quantify the effect of input features on the individual predictions, also known as local explanations [35,36]. The SHAP analysis technique measures these local explanations by assigning credit values to distinct features and consolidating those into an interpretable structure [36]. As previously mentioned, the dataset comprised various high-cardinality categorical features, which required target encoding. These features are not directly interpretable using SHAP importance plots generated from all the model predictions and may often mismatch with experimental findings. Nevertheless, the impact of all of the features can still be analyzed using SHAP, utilizing individual prediction. Therefore, SHAP importance plots were utilized for numerical and binary features (Figs. 2, 3), whilst the individual feature impact on the ML model for our $Cs_3Bi_2I_9$ PSC was shown using SHAP waterfall plots (Fig. 4).

A regression model's prediction ($p(\boldsymbol{x})$) for a feature instance ($\boldsymbol{x}$) can be mathematically expressed as the combination of the base value ($\phi_o(p)$) and the sum of SHAP values ($\phi_i(p,\boldsymbol{x})$) shown in Eq. (5). A positive SHAP value is interpreted as the features push towards predicting a higher target variable, whereas a negative SHAP value corresponds to predicting a lower target variable.

$$p(\boldsymbol{x}) = \phi_o(p) + \sum_{i=1}^{M}\phi_i(p,\boldsymbol{x}) \tag{5}$$

For example, a higher bandgap of the perovskite results in a negative SHAP value, as shown in Fig. 2b, influencing the ML model to predict a lower J_{SC}. In contrast, a higher bandgap of perovskite results in a positive SHAP value (Fig. 2c), leading to a higher V_{OC} prediction. While bandgap controls the movement of charge carriers (i.e., electrons and holes) between the valence band and conduction band, the band alignment of the materials within the layers plays a crucial role [37,38]. The higher the bandgap value, the lower the probability of electrons promoting from the valence band to the conduction band, thereby lowering the J_{SC} and increasing the V_{OC}. These results are also consistent with a recent study conducted on the influence of perovskite bandgap on solar cells using SCAPS-1D and impedance spectroscopy [39]. Meanwhile, the negative SHAP values at higher bandgap suggest that it contributes to the lowering of PCE, as illustrated in Fig. 2a. Notably, the ETL and HTL thickness have a greater influence on V_{OC} and FF compared to the perovskite bandgap, respectively, as shown in Fig. 2c, d.

It is evident from Fig. 3 that the perovskite deposition quenching status is the most important binary (True/False) feature. All performance metrics have a strong dependence on whether the crystallization

Table 2
Random Forest Model metrics for Training and Testing Data.

Metrics	Training Data				Test Data			
	MSE	RMSE	MAE	R^2 Score	MSE	RMSE	MAE	R^2 Score
V_{OC}	0.004	0.065	0.032	0.872	0.010	0.100	0.054	0.690
J_{SC}	5.712	2.390	1.383	0.799	9.298	3.049	1.864	0.669
FF	0.004	0.065	0.043	0.716	0.007	0.085	0.055	0.513
PCE	3.187	1.785	1.184	0.881	7.742	2.782	1.960	0.713

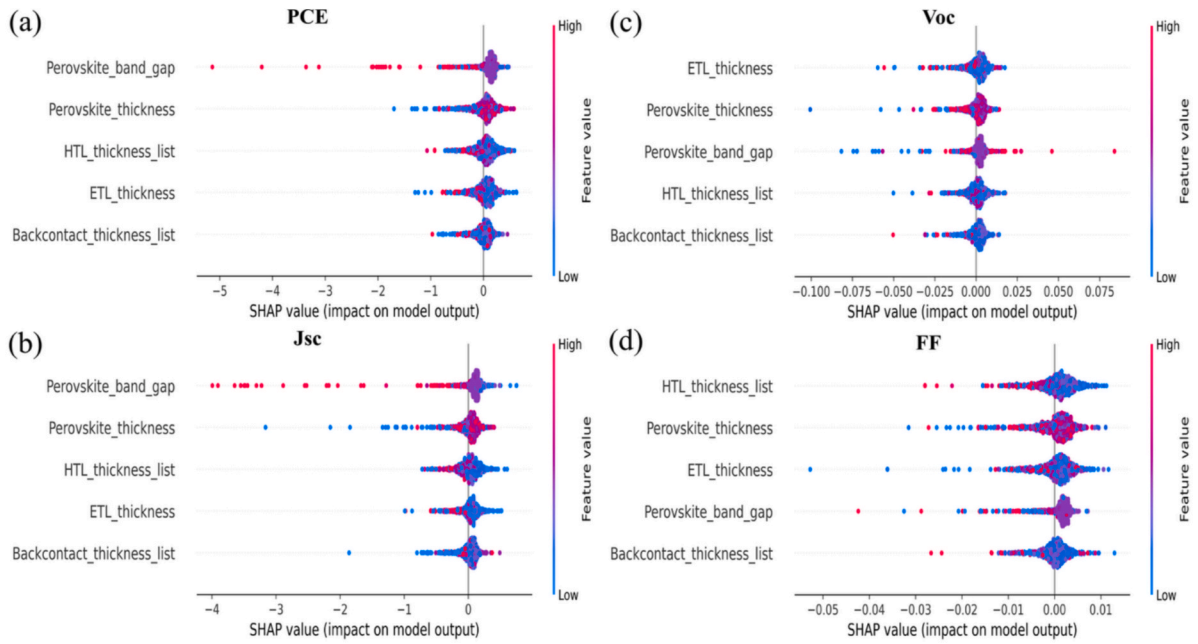

Fig. 2. SHAP feature importance plots for (a) PCE, (b) J_{SC}, (c) V_{OC}, and (d) FF predictions corresponding to numerical features based on 1000 sampled entries. A positive SHAP value suggests that the feature tends to increase the target value and vice versa.

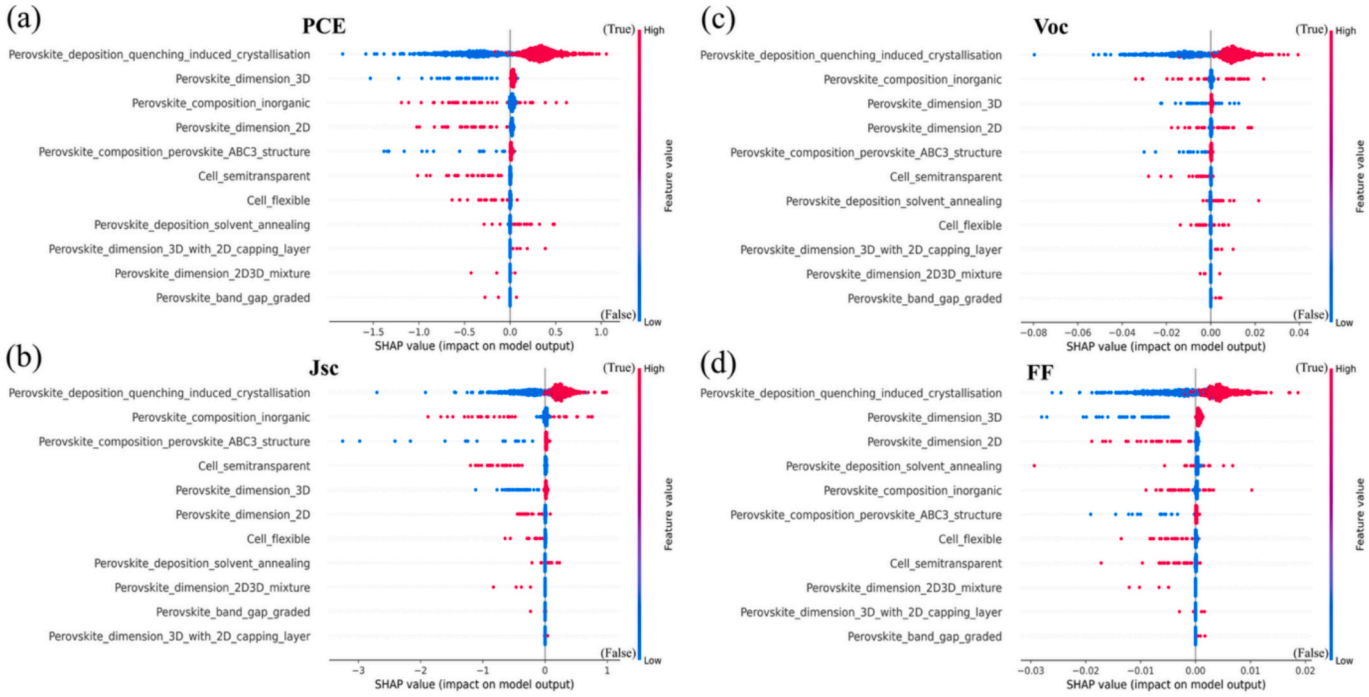

Fig. 3. SHAP feature importance plots for (a) PCE, (b) J_{SC}, (c) V_{OC}, and (d) FF predictions corresponding to binary categorical (True/False) features based on 1000 sampled entries. The red and blue colors represent true and false feature values, respectively. A positive SHAP value suggests that the feature tends to increase the target value and vice versa. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

process was accelerated using quenching techniques (i.e., antisolvent treatment, vacuum treatment, or gas blowing) while maintaining a constant temperature. The three-dimensional structure of the perovskite (represented by the "Perovskite_Dimension_3D" feature) has a minimal influence on the model, as the SHAP value remains close to zero for true entries (highlighted in red). In contrast, a non-3D perovskite structure yields a negative SHAP value (highlighted in blue), suggesting reduced target metrics. Meanwhile, an inverse relationship can be found for certain features, such as "Cell semi-transparent" and "Cell flexible", indicating that the ML model predicts lower PCE, J_{SC}, V_{OC}, and FF if the solar cell structure is semi-transparent and flexible. The impact of individual features can be visualized more intuitively using waterfall plots shown in Fig. 4(a-d). It illustrates a base value $E(p(x))$ and each feature's contribution towards the model's prediction $p(x)$. The impact of categorical features (e.g., perovskite, ETL, HTL material, and deposition parameters) can be intuitively visualized alongside numerical and binary features. For example, choosing $Cs_3Bi_2I_9$ as the perovskite layer has negatively influenced PCE and J_{SC}, whilst positively influenced V_{OC}.

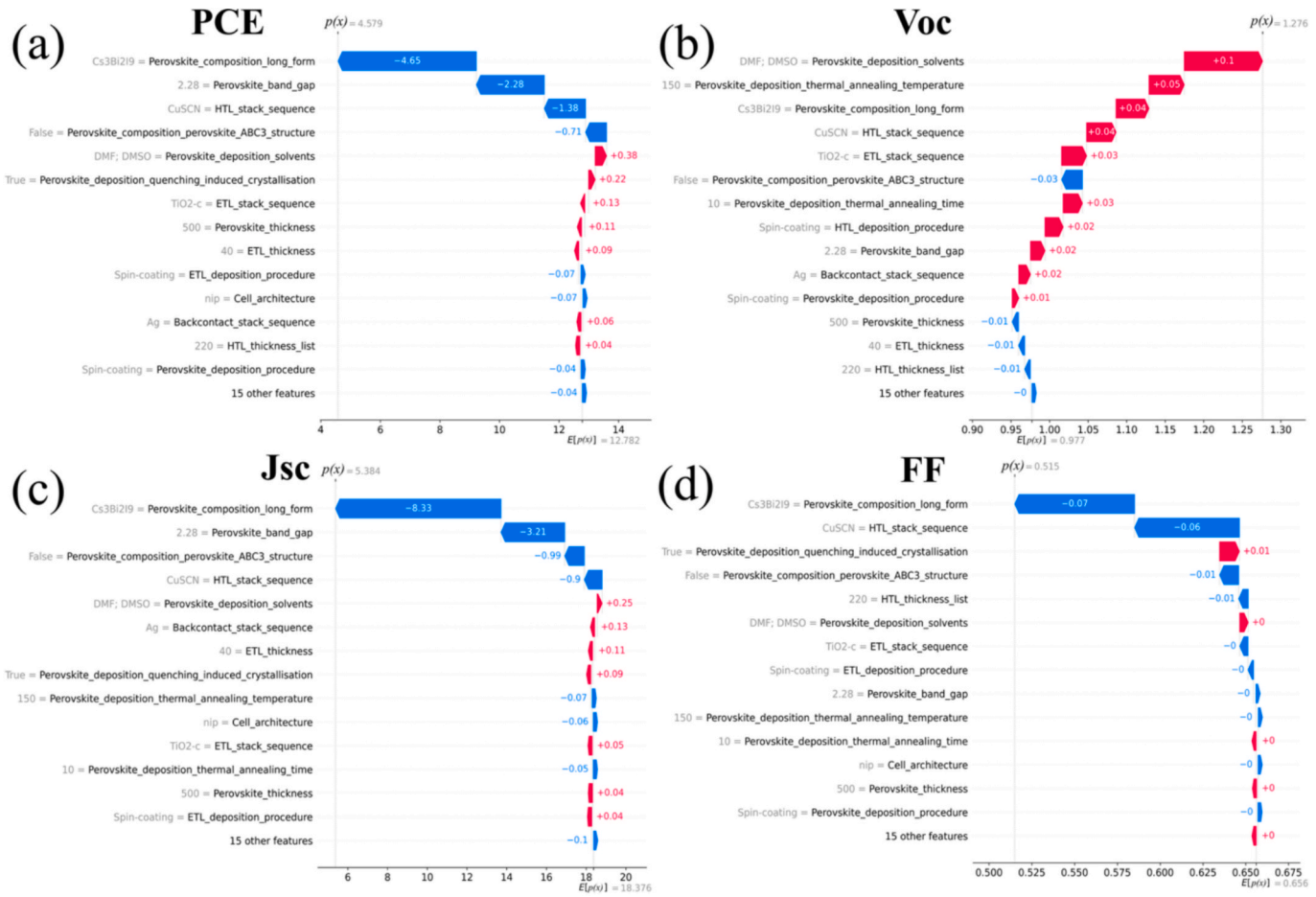

Fig. 4. SHAP waterfall plot for $Cs_3Bi_2I_9$ PSC features. The y-axis shows individual features, and the x-axis represents feature contributions towards the model prediction.

Similar trend can be observed for the HTL layer (CuSCN) as well. Notably, many features had negligible impact on FF predictions, as shown by the zero contributions in Fig. 4(d). Overall, the SHAP analysis provided valuable insights about solar cell architecture and deposition parameters on PSC performance.

3.3. Experimental characterization

Fig. 5(a) shows the optical absorption spectra for the $Cs_3Bi_2I_9$ film collected over 300–800 nm. Films exhibit a good optical absorption profile in the visible region. The bandgap of the film is then calculated using the Tauc plot method; see Fig. 5(b). The direct optical bandgap for the perovskite film is calculated to be 2.28 eV. This value is comparable to previously reported bandgap values of 2.08–2.20 eV for this material [10,11,40–43]. The additional peaks seen in the spectrum have also been reported previously for $Cs_3Bi_2I_9$ films [10,42]. Fig. 5(c) shows the XRD scan of the $Cs_3Bi_2I_9$ perovskite film. The sharp peaks in the spectra indicate the multicrystalline nature of the perovskite film. The highest intensity peak is observed at $2\theta = 25.08°$ (006), indicating the preferred orientation of crystal growth is along the z-axis. Other significant peaks are observed at 2θ = 16.62°, 24.21°, 25.74°, 27.45°, 29.64°, 32.25°, 42.90°, and 51.63°, corresponding to (004), (105), (202), (203), (204), (205), (220), and (0 0 12) planes respectively of the JCPDS #70–0666 card. Film has a hexagonal structure with $P6_3/mmc$ space group. Lattice parameters are then calculated to be a = 8.5 Å and c = 22.6 Å, with an average crystallite size of 15.9 nm, calculated using the Debye Scherrer equation. These numbers are in good agreement with the JCPDS card and confirm the successful fabrication of a multicrystalline $Cs_3Bi_2I_9$ perovskite film. Fig. 5(d) shows the surface morphology of the perovskite film, where hexagon-like thin crystal flakes of $Cs_3Bi_2I_9$ can be observed. Line-shaped crystals seen in the image are the crystal flakes oriented perpendicular to the plane of the image. This shows a uniform deposition of the perovskite film with complete surface coverage.

3.4. Photovoltaic performance of the cell

After the successful fabrication and characterization of $Cs_3Bi_2I_9$ films, n-i-p type solar cells with FTO/c-TiO_2/ $Cs_3Bi_2I_9$/CuSCN/Ag architecture were fabricated. This device architecture is illustrated in Fig. 6(a). TiO_2 and CuSCN are selected as the ETL and HTL, respectively, due to their optimal band alignment with $Cs_3Bi_2I_9$, which assures smooth extraction of the photogenerated charge carriers Fig. 6(b). Fig. 6 (c) shows a cross-section image of a sister device from the same batch as the champion device reported here. The J-V curve of the champion device is demonstrated in Fig. 6(d).

The device showed a maximum PCE of 3.73% with V_{OC} = 0.69 V, J_{SC} = 12.22 mA/cm^2, and FF = 43.94%. The experimental V_{OC}, J_{SC}, and FF values are in good agreement with the values calculated in ML modeling. A comparison of ML predictions and experimental values for cell performance is presented in Table 3. While the ML-predicted V_{OC} and PCE are significantly close to the experimentally observed values, Jsc and FF values exhibit a comparatively larger deviation. This discrepancy can be attributed in part to the limited representation of $Cs_3Bi_2I_9$-based device data within the training dataset, creating a bias toward more conservative predictions. Nevertheless, from a fundamental standpoint, wide-bandgap materials are expected to yield lower J_{SC}. The experimentally

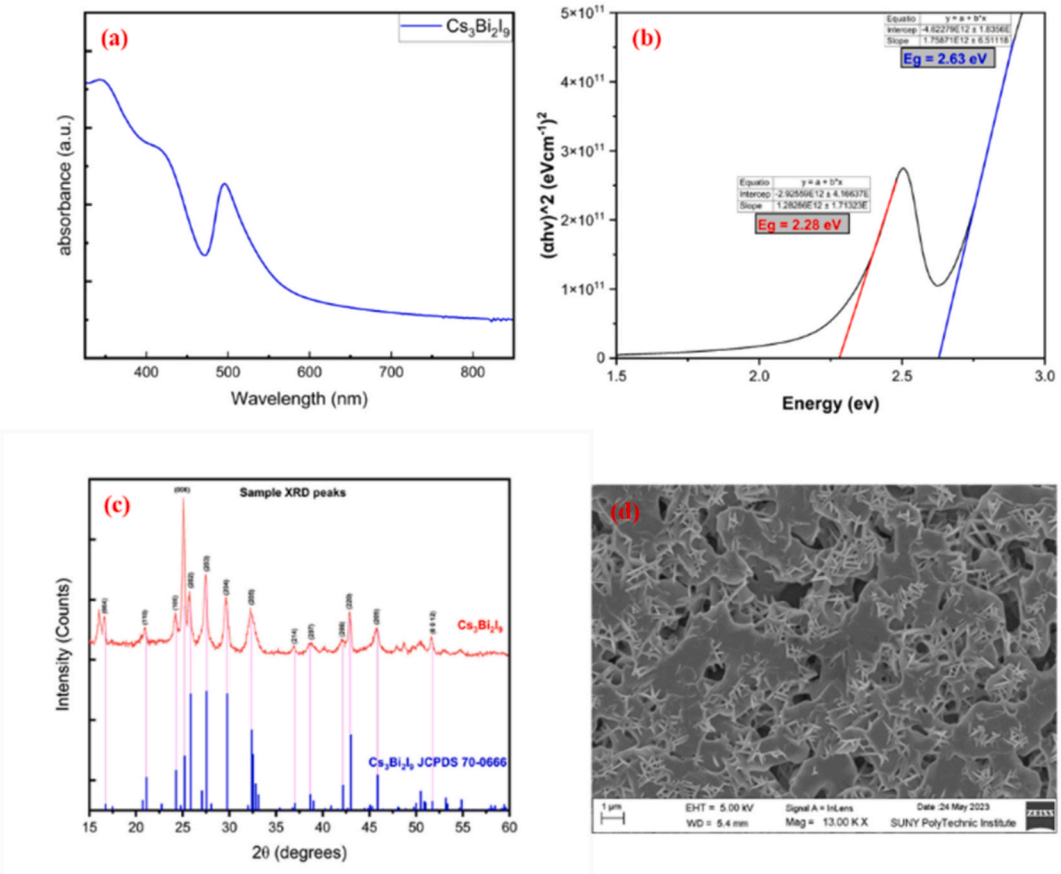

Fig. 5. (a) Optical absorption spectra, (b) Tauc plot calculations, (c) XRD pattern, and (d) Surface morphology of the perovskite film.

observed higher value suggests improved charge collection, potentially due to reduced non-radiative recombination associated with lower defect densities, as well as enhancements in film morphology, crystallinity, and fabrication conditions. Such effects are not fully captured by the input features of the model. As a result, the experimental device can outperform ML predictions.

4. Conclusion and future work

Machine learning models act as effective tools for uncovering complex, nonlinear relationships within high-dimensional datasets, which are often difficult to resolve using traditional analytical methods. In this framework, the models provide not only predictive performance but also data-driven insights that aid in material selection and process optimization. This study presents a combined machine learning and experimental fabrication strategy for developing lead-free perovskite solar cells. $Cs_3Bi_2I_9$ was used as the active material due to its lead-free inorganic nature. We have used a large dataset of 26,000 experimental data points from the Perovskite Database Project to train multiple ML models. The SHAP algorithm has identified device architecture (i.e., the selection of optimal ETL and HTL) and experimental methodologies for perovskite deposition as the primary factors that dictate cell performance. These ML insights were carefully considered in the design of the complete device and its subsequent fabrication. The performance of the experimentally fabricated devices showed a decent agreement with the ML predictions, with a champion device achieving a PCE of 3.73%, V_{OC} of 0.69 V, J_{SC} of 12.22 mA cm^{-2}, and FF of 43.94%. While the PCE and V_{OC} prediction using the ML model is in good agreement with the experimentally measured value with a small offset, J_{SC} and FF values, however, show a large deviation from the MAE of the model. This discrepancy stems from the model operating in an extrapolative mode rather than interpolation within the known data distribution. Incorporating a large amount of experimental data specific to $Cs_3Bi_2I_9$ to form an expanded dataset can be a good approach to enhancing the ML model's generalization capabilities. Initial observations have demonstrated that the color of the perovskite layer remains unchanged for up to 100 days under ambient conditions outside the glovebox; however, further systematic lifetime testing, as well as additional morphological improvements in the active layer, are recommended.

5. Supporting information

The Jupyter Notebook code files supporting findings of this study are openly available in a GitHub repository at https://github.com/AbdulHamidRumman/Cs3Bi2I9_PSC.

Funding sources

This research project was made possible by the generous support of the NYS Center for Advanced Technology in Nanomaterials and Nanoelectronics (CATN2) (Grant number: 1186196-1-92476). We are deeply grateful for their financial contribution and unwavering belief in our work, which has been instrumental in achieving our research goals.

CRediT authorship contribution statement

Ankit Choudhary: Writing – review & editing, Writing – original draft, Visualization, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Abdul Hamid Rumman:** Writing – review & editing, Writing – original draft, Visualization, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Miah Abdullah Sahriar:** Writing – review & editing,

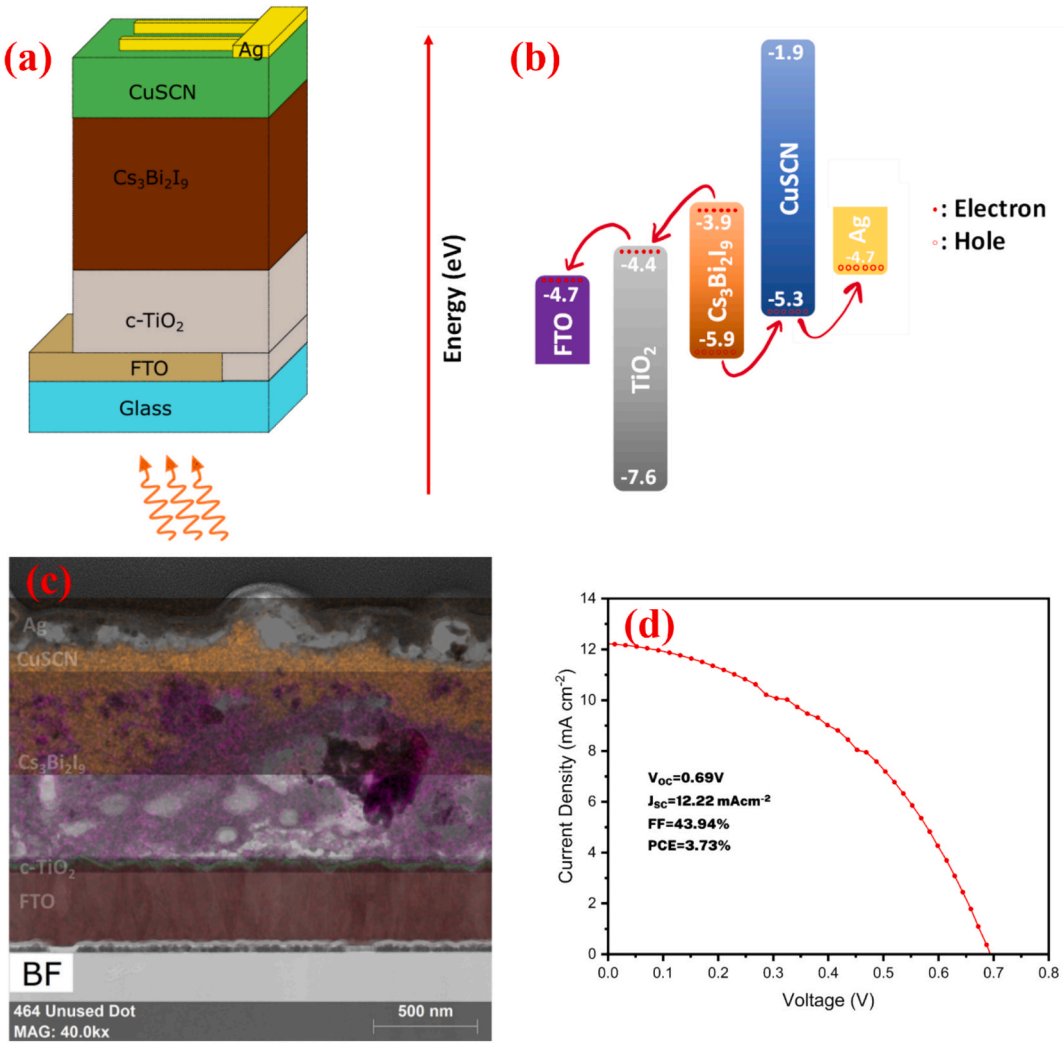

Fig. 6. (a) Schematic of device architecture, (b) band alignment, (c) SEM cross-section of the device with different layers, and (d) J-V curve of the devic.

Table 3
$Cs_3Bi_2I_9$ perovskite solar cell performance.

Target Variables	ML Prediction	Experimental Values
Voc (V)	0.63	0.69
Jsc (mA/cm2)	5.37	12.22
FF (%)	51.49	43.94
PCE (%)	4.58	3.73

Visualization, Software, Methodology, Investigation, Data curation. **Md Tohidul Islam:** Writing – review & editing, Visualization, Software, Methodology, Investigation, Data curation. **Saquib Ahmed:** Writing – review & editing, Validation, Supervision, Resources, Project administration, Methodology, Investigation, Funding acquisition, Conceptualization. **Haralabos Efstathiadis:** Writing – review & editing, Validation, Supervision, Resources, Project administration, Methodology, Investigation, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] M.M. Lee, J. Teuscher, T. Miyasaka, T.N. Murakami, H.J. Snaith, Efficient hybrid solar cells based on meso-superstructured organometal halide perovskites, Science (1979) (338) (2012) 643–647.
[2] 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. Grätzel, J.Y. Kim, Pseudo-halide anion engineering for α-FAPbI3 perovskite solar cells, Nature 592 (2021) 381–385.
[3] W.J. Yin, T. Shi, Y. Yan, Unique properties of halide perovskites as possible origins of the superior solar cell performance, Adv. Mater. 26 (2014) 4653–4658.
[4] W.J. Yin, T. Shi, Y. Yan, Unusual defect physics in CH3NH3PbI3 perovskite solar cell absorber, Appl. Phys. Lett. 104 (2014).
[5] Q. Han, Y. Wei, R. Lin, Z. Fang, K. Xiao, X. Luo, S. Gu, J. Zhu, L. Ding, H. Tan, Low-temperature processed inorganic hole transport layer for efficient and stable mixed Pb-Sn low-bandgap perovskite solar cells, Sci. Bull. (Beijing) 64 (2019) 1399–1401.
[6] J. Han, K. Park, S. Tan, Y. Vaynzof, J. Xue, E.W.G. Diau, M.G. Bawendi, J.W. Lee, I. Jeon, Perovskite solar cells, Nat. Rev. Methods Primers 5 (1) (2025) 3.
[7] X. Jiang, H. Li, Q. Zhou, Q. Wei, M. Wei, L. Jiang, Z. Wang, Z. Peng, F. Wang, Z. Zang, K. Xu, Y. Hou, S. Teale, W. Zhou, R. Si, X. Gao, E.H. Sargent, Z. Ning, One-step synthesis of SnI2·(DMSO)xadducts for high-performance tin perovskite solar, Cells. J. Am. Chem. Soc. 143 (2021) 10970–10976.
[8] Song T. Bin, T. Yokoyama, S. Aramaki, M.G. Kanatzidis, Performance enhancement of lead-free tin- based perovskite solar cells with reducing atmosphere-assisted dispersible additive, ACS Energy Lett. 2 (2017) 897–903.
[9] S.S. Shin, J.P. Correa Baena, R.C. Kurchin, A. Polizzotti, J.J. Yoo, S. Wieghold, M. G. Bawendi, T. Buonassisi, Solvent-engineering method to deposit compact bismuth-based thin films: mechanism and application to photovoltaics, Chem. Mater. 30 (2018) 336–343.
[10] D.B. Khadka, Y. Shirai, M. Yanagida, K. Miyano, Tailoring the film morphology and interface band offset of caesium bismuth iodide-based Pb-free perovskite solar cells, J. Mater. Chem. C Mater. 7 (2019) 8335–8343.

[11] B.W. Park, B. Philippe, X. Zhang, H. Rensmo, G. Boschloo, E.M.J. Johansson, Bismuth based hybrid perovskites A3Bi2I9 (A: methylammonium or cesium) for solar cell application, Adv. Mater. 27 (2015) 6806–6813.
[12] S. Ullah, F. Khan, J.F. Rasheed, Prospects for commercialization of CsPbIBr2-based all-inorganic perovskite solar cells: fabrication, stability, and engineering strategies, Adv. Funct. Mater. 35 (2025) 2503508.
[13] S. Ullah, F. Khan, A. AlZahrani, Comprehensive assessment of all-inorganic CsPbI3–xBrx perovskite-based solar cells: interface engineering, stability, and economic aspects, Coord. Chem. Rev. 516 (2024).
[14] F. Bai, Y. Hu, Y. Hu, T. Qiu, X. Miao, S. Zhang, Lead-free, air-stable ultrathin Cs3Bi2I9 perovskite nanosheets for solar cells, Sol. Energy Mater. Sol. Cells 184 (2018) 15–21.
[15] Y. Liu, W. Yan, S. Han, H. Zhu, Y. Tu, L. Guan, X. Tan, How machine learning predicts and explains the performance of perovskite solar cells, Sol. RRL 6 (2022).
[16] R. Zhao, B. Xing, H. Mu, Y. Fu, L. Zhang, Evaluation of performance of machine learning methods in mining structure-property data of halide perovskite materials, Chin. Phys. B 31 (2022).
[17] S. Rath, G.S. Priyanga, N. Nagappan, T. Thomas, Discovery of direct band gap perovskites for light harvesting by using machine learning, Comput. Mater. Sci. 210 (2022).
[18] Ç. Odabaşı, R. Yıldırım, Machine learning analysis on stability of perovskite solar cells, Sol. Energy Mater. Sol. Cells 205 (2020).
[19] D. Jain, S. Chaube, P. Khullar, S.G. Srinivasan, B. Rai, Bulk and surface DFT investigations of inorganic halide perovskites screened using machine learning and materials property databases, PCCP 21 (2019) 19423–19436.
[20] J. Zhao, X. Wang, Screening perovskites from ABO3 combinations generated by constraint satisfaction techniques using machine learning, ACS Omega 7 (2022) 10483–10491.
[21] X. Yang, L. Li, Q. Tao, W. Lu, M. Li, Rapid discovery of narrow bandgap oxide double perovskites using machine learning, Comput. Mater. Sci. (2021) 196.
[22] Y. Wu, S. Lu, M.G. Ju, Q. Zhou, J. Wang, Accelerated design of promising mixed lead-free double halide organic–inorganic perovskites for photovoltaics using machine learning, Nanoscale 13 (2021) 12250–12259.
[23] B. Yılmaz, Ç. Odabaşı, R. Yıldırım, Efficiency and stability analysis of 2D/3D perovskite solar cells using machine learning, Energ. Technol. 10 (2022) 2100948.
[24] J. Li, B. Pradhan, S. Gaur, J. Thomas, Predictions and strategies learned from machine learning to develop high-performing perovskite solar cells, Adv Energy Mater. 9 (2019) 1901891.
[25] Z. Hui, M. Wang, J. Chen, X. Yin, Y. Yue, J. Lu, Predicting photovoltaic parameters of perovskite solar cells using machine learning, J. Phys. Condens. Matter 36 (2024) 355901.
[26] 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, Machine-learning modeling for ultra-stable high-efficiency perovskite solar cells, Adv. Energy Mater. 12 (2022) 2201463.
[27] W. Liu, Y. Lu, D. Wei, X. Huo, X. Huang, Y. Li, J. Meng, S. Zhao, B. Qiao, Z. Liang, Z. Xu, D. Song, Screening interface passivation materials intelligently through machine learning for highly efficient perovskite solar cells, J. Mater. Chem. A Mater. 10 (2022) 17782–17789.
[28] MdT. Islam, MdR. Jani, S. Rahman, Shorowordi K. Md, S.S. Nishat, D. Hodges, S. Banerjee, H. Efstathiadis, J. Carbonara, S. Ahmed, Investigation of non-Pb all-perovskite 4-T mechanically stacked and 2-T monolithic tandem solar devices utilizing SCAPS simulation, SN Appl. Sci. 3 (2021) 504.
[29] F. Li, X. Peng, Z. Wang, Y. Zhou, Y. Wu, M. Jiang, M. Xu, Machine learning (ML)-assisted design and fabrication for solar cells, Energy Environ. Mater. 2 (2019) 280–291.
[30] 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, Accelerating photovoltaic materials development via high-throughput experiments and machine-learning-assisted diagnosis, Joule 3 (2018) 1437–1451.
[31] MdS. Islam, MdT. Islam, S. Sarker, Jame H. Al, S.S. Nishat, MdR. Jani, A. Rauf, S. Ahsan, Shorowordi K. Md, H. Efstathiadis, J. Carbonara, S. Ahmed, Machine learning approach to delineate the impact of material properties on solar cell device physics, ACS Omega 7 (2022) 22263–22278.
[32] E. Unger, T.J. Jacobsson, The perovskite database project: a perspective on collective data sharing, ACS Energy Lett. 7 (2022) 1240–1245.
[33] A.H. Rumman, M.A. Sahriar, M.T. Islam, K.M. Shorowordi, J. Carbonara, S. Broderick, S. Ahmed, Data-driven design for enhanced efficiency of Sn-based perovskite solar cells using machine learning, APL Mach. Learn. 1 (2023) 46117.
[34] F. Pargent, F. Pfisterer, J. Thomas, B. Bischl, Regularized target encoding outperforms traditional methods in supervised machine learning with high cardinality features, Comput. Stat. 37 (2022) 2671–2692.
[35] S.M. Lundberg, S.I. Lee, A unified approach to interpreting model predictions, Adv. Neural Inf. Process. Syst. (2017) 4766–4775.
[36] S.M. Lundberg, G. Erion, H. Chen, A. DeGrave, J.M. Prutkin, B. Nair, R. Katz, J. Himmelfarb, N. Bansal, S.I. Lee, From local explanations to global understanding with explainable AI for trees, Nat. Mach. Intell. 2 (1) (2020) 56–67.
[37] S. Sarker, M.T. Islam, A. Rauf, H. Al Jame, MdR. Jani, S. Ahsan, MdS. Islam, S. S. Nishat, Shorowordi K. Md, S. Ahmed, A SCAPS simulation investigation of non-toxic MAGeI3-on-Si tandem solar device utilizing monolithically integrated (2-T) and mechanically stacked (4-T) configurations, Sol. Energy 225 (2021) 471–485.
[38] M.T. Islam, M.R. Jani, A.F. Islam, S.K. Md, S. Chowdhury, S.S. Nishat, S. Ahmed, Investigation of $CsSn_{0.5}Ge_{0.5}I_3$ -on-Si tandem solar device utilizing SCAPS simulation, IEEE Trans. Electron. Devices 68 (2021) 618–625.
[39] A. Mortadi, E.M. El Hafidi, H. Nasrellah, M. Monkade, R. El Moznine, Investigation of bandgap grading on performances of perovskite solar cell using SCAPS-1D and impedance spectroscopy, Sol. Energy Adv. 4 (2024).
[40] S.L. Hamukwaya, H. Hao, M.M. Mashingaidze, T. Zhong, S. Tang, J. Dong, J. Xing, H. Liu, Potassium iodide-modified lead-free Cs3Bi2I9 perovskites for enhanced high-efficiency solar cells, Nanomaterials 12 (2022).
[41] Y. Ji, M. She, X. Bai, E. Liu, W. Xue, Z. Zhang, K. Wan, P. Liu, S. Zhang, J. Li, In-depth understanding of the effect of halogen-induced stable 2D bismuth-based perovskites for photocatalytic hydrogen evolution activity, Adv Funct. Mater. 32 (2022).
[42] S.M. Masawa, J. Li, C. Zhao, X. Liu, J. Yao, 0D/2D mixed dimensional lead-free caesium bismuth iodide perovskite for solar cell application, Materials. 15 (2022).
[43] T. Niyitanga, W. Raza, K. Ahmad, A. Alsulmi, H. Kim, Lead-free Cs3Bi2I9 perovskite-like material as efficient photocatalyst for improved hydrogen evolution in presence of platinum co-catalyst, ChemistrySelect 8 (2023).