

# Using machine learning for prediction of spray coated perovskite solar cells efficiency: From experimental to theoretical models



Reisya Ichwani<sup>a,b,1</sup>, Stephen Price<sup>c,1</sup>, Oluwaseun K. Oyewole<sup>a</sup>, Rodica Neamtu<sup>c</sup>, Winston O. Soboyejo<sup>a,\*</sup>

<sup>a</sup> Department of Mechanical Engineering, Worcester Polytechnic Institute, Worcester, MA 01609, USA

<sup>b</sup> Department of Chemistry, Faculty of Mathematics and Natural Sciences, Universitas Sumatera Utara, Medan 20155, Indonesia

<sup>c</sup> Department of Computer Science, Worcester Polytechnic Institute, Worcester, MA 01609, USA

## ARTICLE INFO

### Keywords:

Perovskite solar cells  
Spray process optimization  
Machine learning  
Experimental design  
Regression analysis

## ABSTRACT

Low-cost perovskite solar cells (PSCs) have experienced unprecedented gains in power conversion efficiency (PCE) of up to 25% of lab-scale devices. To be realized in the market, however, PSCs are not only required to be efficient but also scalable in production. While spray coating has viability as an industrial manufacturing process for perovskite photovoltaics scaling, optimizing the spray conditions is often seen as a challenging and time-consuming process due to its complex and multidimensional parameters. Herein, we use a machine learning (ML) approach to capture the relationship between spray parameter settings to the resultant photoconversion efficiency (PCE) of PSCs from experimental collected data points. This data-driven approach has the potential to accurately predict PCE values given the manufacturing parameters, enabling optimization and resulting in an increased experimentally recorded PCE. Furthermore, we also used a Convolutional Neural Network (CNN) to predict defect size distributions in the PSC structures to improve the understanding of defect formation mechanism at given spray parameters. The implications of the results are discussed for optimizing spray manufacturing process of efficient perovskite photovoltaics.

## 1. Introduction

Perovskites have been favored to potentially replace the widespread market-adapted silicon-based solar photovoltaics (PV) [1] and shown potential for efficient, low-cost, lightweight solar cells due to their solution-processing and tunable optoelectrical properties [2–4]. Perovskite-based solar cells (PSCs) first came to the scene in 2009 with 3.8% [5] power conversion efficiency (PCE) and rose to over 25% [6] in 2020, a game-changer in the PV community. While PSCs have undergone an unprecedented rise in efficiency, most high-efficiency PSCs are limited to laboratory scale devices with an active area of less than 0.1 cm<sup>2</sup> by spin-coating techniques, which is poorly suited to industrial manufacturing [7]. This suggests that scalable coating techniques are needed to produce perovskite films with consistent properties at the manufacturing scale.

Spray-coating is a scalable coating technique that has been applicable in many coating industries e.g. automotive parts [8], catalysts [9], and batteries [10], and could, in principle, be used for perovskite

manufacturing. Although spray-coating is very suitable for scale-up, the physics behind this technique are multiscale and complex [11]. Without careful control of spray conditions, deposited films would contain structural defects, resulting to poor performance of PSCs [10]. Furthermore, optimization of spray conditions is often carried out through a *trial-and-error* method that is time-consuming and laborious [12,13].

As the *trial-and-error* experimental approaches are challenging, a data-driven approach can be beneficial by reducing the time and cost required. Using previously collected data from trial-and-error optimization, Machine Learning (ML) can scan the data, recognize patterns/relationships, predict outcomes with minimal human intervention, and reduce the need for such a large dataset [14,15]. In the case of high-dimensional spray coating techniques, supervised learning, such as a regression model, have been used to map the relationships and trends through available data to optimize the coating process and its experimental design of direct ink printing [16], rapid spray plasma processing [17], and flame spray pyrolysis [10]. This approach has shown the

\* Corresponding author.

E-mail address: [wsoboyejo@wpi.edu](mailto:wsoboyejo@wpi.edu) (W.O. Soboyejo).

<sup>1</sup> These authors contributed equally.

utility of ML in accelerating film scaling and in reducing the experimental burden for exploring the high-dimensional parameter space of spray coating.

Other than the advancement of coating techniques, various factors affect the performances of PSCs in the real world. For example, the nature of perovskites instability in the air [18], stresses during spray fabrication [6,19], and poor controlled spray coating procedure [8], may contribute to creating defects in perovskite itself or solar cell structures that are detrimental to their performances. These can be defects in grain or grain boundaries, interfacial and defects, and interlayer defects. Quantification is important to identify the optimally performing processing technique.

Recent work within the field of Material Science has shown the applications of Convolutional Neural Networks (CNNs) for image analysis. Using a trained Mark R-CNN model [20], images collected from a Scanning Electron Microscope (SEM), satellite instances (deformations in metallic powder particles) can be automatically detected, segmented, and analyzed [21–23]. In addition to highlighting the applicability of regression models for analyzing, predicting, and optimizing machining parameters used in PSC manufacturing, this work seeks to prove that similar techniques can be applicable for PSC manufacturing to detect and quantify the previously mentioned defects. We aim to show that a Mark R-CNN [20] can learn to identify deformations in the cross-sectional view PSCs.

Herein, we demonstrate the efficacy of supervised learning in capturing the relation between perovskite film spray parameter settings and the resulting power conversion efficiencies (PCEs) of assembled perovskite-based solar cells. The initial step consists of optimizing perovskite film fabrication using spray coating. Four input variables are explored: spray nozzle speed, substrate temperature, nozzle/substrate height, and pressure added after spraying. Second, the efficiency of sprayed-perovskites solar cells are measured under the solar simulator, and the resulting current–voltage (*I*-*V*) curves are analyzed to extract the incident solar energy that can convert it into electrical energy. Third, a regression model is trained to learn the process variables–solar cell efficiency correlation and used to predict the optimal PCE relative to the obtained PCE. In addition to the training and usage of regression models, this study explores the efficacy of computer vision for detecting deformations in PSCs, and further work will explore the joining of these two techniques for a more in-depth optimization problem. Manually identifying the ideal spray parameters for manufacturing PSCs is expensive and can take months to years. Thus, combining supervised learning in spray process optimization and understanding defect size distributions in solar cells is crucial to suggest new experiment sets to achieve high-efficiency solar cells. This study can reduce the time and cost to achieve the process control and reproducibility of solar cells using spray technology. The results of this study could guide the design of spray manufacturing technologies and provide the rational basis for processing parameters for sprayed PSCs optimization.

## 2. Background

### 2.1. Theoretical Shockley-Queisser limit in single junction perovskite solar cells

The theoretical maximum light to electric power conversion efficiency,  $\eta$ , of a single p-n junction solar cell illuminated by the sun is known as the detailed balance limit or *Shockley-Queisser* (SQ) limit [24]. It was first estimated by *William Shockley*, and *Hans-Joachim Queisser* in 1961 in which the black body emission approximated sun with a surface temperature of 6000 K. Only radiative combination was considered in the ideal diode, and nonradiative processes (e.g., Auger effects) were neglected. In terms of electron-hole pair production, a photon with energy below the energy band gap is assumed to not interact with the solar cell, while a photon with higher energy than the band gap is converted to electron-hole pair with a quantum efficiency of 100%. This limit gave

a maximum efficiency of ~33% at 1.1 eV [25].

## 2.2. Regression models

At a glance, regression models capture the relationship between the independent variable(s) and a dependent variable and are used across many domains [26–28], including material science [29]. They establish a mathematical estimation of each variable's impact on the dependent variable based on the Least Squares Method [30]. Given that our preliminary experiments for measuring solar cells efficiency use four independent variables (temperature, speed, distance, and pressure), the most suitable models are multiple linear regression and multiple polynomial regression. While linear models are simpler to use and easier to understand, they depend on a linear correlation between the data points. If such a relationship does not exist, then polynomial regression can better capture the potential non-linear relationships. In recent years, regression models have been at the core of data-driven interdisciplinary material science research, improving the analysis and insight gained from collected data [12]. This transition to data-driven research has led to many breakthroughs in material science, including but not limited to analyzing and predicting atomic properties that otherwise would have been near-impossible by hand [31], predicting energy and atomic forces orders of magnitude faster [32], predicting materials properties [33].

## 3. Experimental evaluation

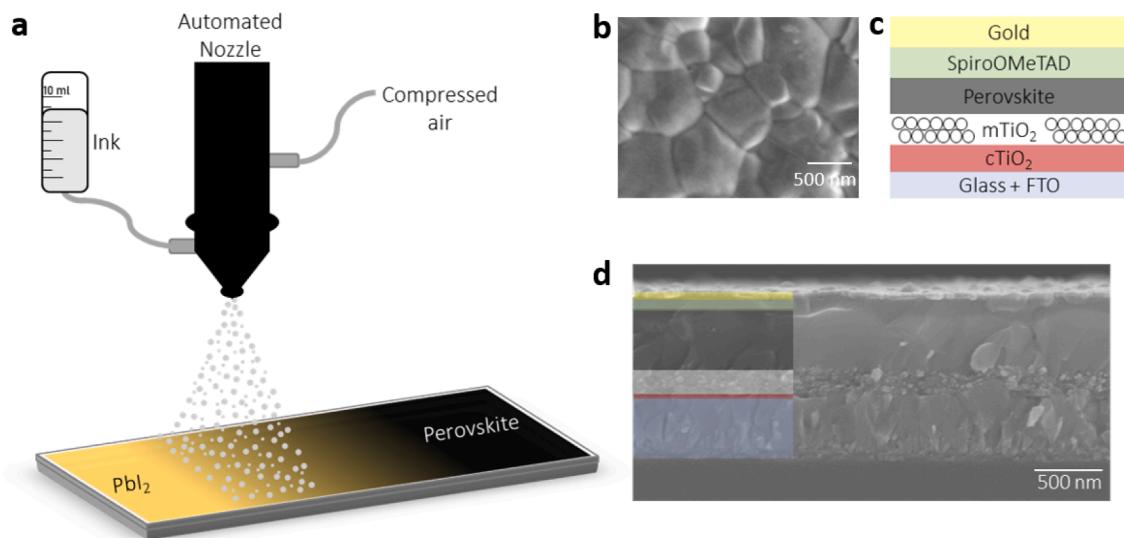
### 3.1. Spray processing and device fabrication

The FTO substrates were sequentially washed in Decon-90, distilled water, acetone, ethanol, and IPA; and were treated under UV Ozone. The electron transporting layer (ETL) was coated on the substrates with 0.15 and 0.3 M TiO<sub>2</sub> solution of titanium diisopropoxide bis(acetylacetone) (Sigma) in n-butanol (Sigma). The solution was spin-coated at 2000 rpm for 30 s and then annealed at 125 °C for 5 min. It is followed by spin coating of 0.3 M solution and annealing at 500 °C for 30 min. For mesoporous TiO<sub>2</sub>, a 1:5 ratio of titania paste:ethanol was spun at 4000 rpm for 30 s and sintered at 500 °C [6].

A two-step sequential deposition technique was used to fabricate the perovskite films. For every 1 ml of precursor solution, a mixture of 599.3 mg PbI<sub>2</sub> (> 98.9 % purity) (Sigma) was dissolved in a mixture of 9.5:0.5 ration of DMF:DMSO and stirred for 2 h. The PbI<sub>2</sub> solution was spin-coated onto the ETL layer at 1500 rpm for 30s and annealed at 70 °C for 1 min [19].

Formamidinium (FA)-rich precursor with the following quantities: 60 mg FAI, 6 mg of MABr and 6 mg of MACl (Sigma) was used for every 1 ml of IPA and was sprayed on PbI<sub>2</sub>-coated substrate using automated spray coating system (MTI Corporation, Richmond, CA) (Fig. 1a). The atomizer nozzle was 40 kHz. The precursor solution was initially fed to the ink chamber, and the spray configurations were programmed to move across the substrate in a single pass at a speed of 100–300 mm/s, a head-substrate height of 5–8 cm, and a flow rate of 1 μm/min via compressed air. The compressed air controlled the solution's spray pattern and velocity at the nozzle, with pressure held constant at 10 psi. The FTO/TiO<sub>2</sub>/PbI<sub>2</sub> substrate was put on a hot plate and heated at the desired temperature for 1 min before the organic precursor was sprayed. The spray coating was done in a single pass for eight samples. The substrate temperature ( $T_{sub}$ ) was also varied from 50–90 °C. As-sprayed films were left for 30s for even distribution of solution and annealed at 130 °C for 30 min to promote complete conversion from PbI<sub>2</sub> to FA-rich perovskite crystal. Here, the yellow film of PbI<sub>2</sub> completely turned to the dark color of perovskite film.

To complete the solar cells assembly, the hole transport layer (HTL), SpiroOMeTAD (2,2',7,7'-Tetrakis[N,N-di(4-methoxyphenyl) amino]-9,9'-spirobifluorene), was prepared by dissolving 72 mg of SpiroOMeTAD, 30 μL of 4-tert-butylpyridine (tBP) and 35 μL of lithium bis(trifluoromethylsulphonyl) imide (Li-TFSI) (260 mg of Li-TFSI in 1 ml of



**Fig. 1.** Schematic of spray coating method on perovskite film (a), Morphology of sprayed perovskite film (b), Device architectures of PSCs (c), Cross-sectional SEM image representative of PSCs (d) Functional layers of a solar cell separated by color.

acetonitrile) dopants in 1 ml of chlorobenzene. We spin-coated SpiroOMeTAD at 4000 rpm for 30 s. The final 90-nm gold back contact was thermally evaporated to complete PSCs using a thermal evaporator (Edwards E306 A, Easton, PA) under a vacuum pressure of  $10^{-6}$  Torr at a deposition rate of  $0.1 \text{ nm s}^{-1}$ .

### 3.2. Pressure application

Complete stack PSCs were subjected to external pressure ranging from 0–10 MPa using a 5848 Instron MicroTester (Instron, Norwood, MA, USA) with cured Polydimethylsiloxane (PDMS) anvil layer placed between the device and the fixture. PDMS was made using a mixture ratio (10:1) by weight of Sylgard 184 silicon elastomer base and curing agent (Dow Corning Corporation, Midland, MI). The mixture was cured at 65 °C for two hours in a mold. The fabricated solar cells cut the PDMS anvil into the desired dimension. The Instron was set to compress the PSCs devices at a displacement rate of  $-1.0 \text{ mm min}^{-1}$  and hold for 10 min.

### 3.3. Material and solar cell characterization

To view the morphology of sprayed perovskites, the top-view of the scanning electron microscope (SEM) was collected for each perovskite film, while the cross-sectional SEM was collected for the complete stack of solar cells. The photovoltaic performance of PSCs was measured using a solar simulator (Oriel, Newport Corporation, Irvine, CA) with a source meter (Keithley, Tektronix, Newark, NJ). Light intensity was calibrated at  $90 \text{ mW/cm}^2$  using a silicon reference cell (MKS/Newport Instruments, Newport Beach, CA, USA). The current density–voltage (*I*–*V*) curves were obtained by scanning in the range of –0.4 to 1.2 V, with a device exposed area of  $0.1 \text{ cm}^2$ .

### 3.4. Datasets and model training

The experimental data collected, relating machine parameters to resultant efficiency, contained 106 unique combinations that were multi-variate in nature. Specifically, as shown in Table 1, the data contained four independent variables: Substrate temperature (measured in °C), the linear speed of the spray nozzle (measured in mm/s), height between head and heated substrate (measured in cm), and Pressure (Measured in MPa), and one dependent (or target) variable: Efficiency (measured in %). To train a model and validate results, the dataset was broken into an 80:20 split for training and validation sets such that a

**Table 1**  
Ranges of possible values for each machining parameter.

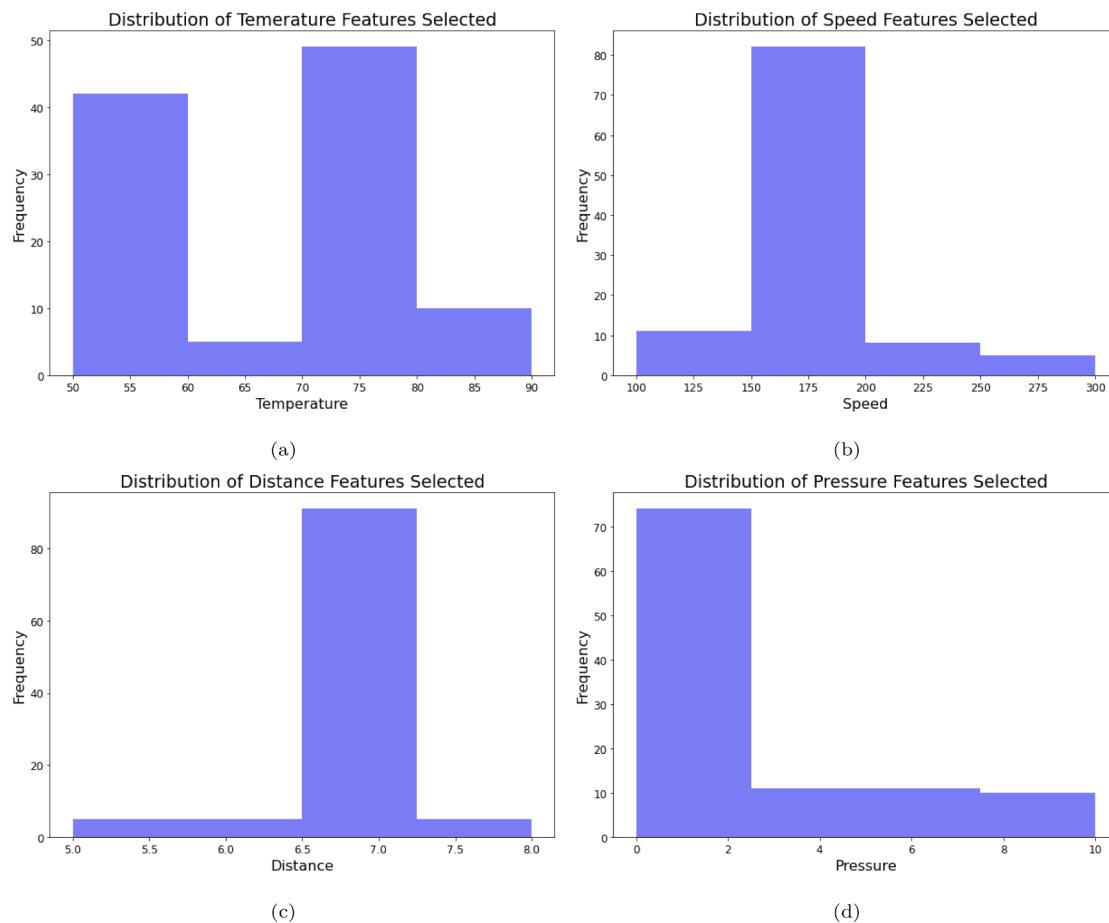
Process Variables	Total Range (units)
Substrate Temperature	50–90 (°C)
Linear Nozzle Speed	100–300 (mm/s)
Head/substrate Height	5–8 (cm)
Applied Pressure	0–10 (MPa)

model would be trained on 80% of the data. Once trained, a model would be used to predict efficiencies for the remaining 20%, comparing against experimentally collected results, to evaluate model performance.

The dataset used here was collected experimentally, attempting to optimize performance manually by adjusting one parameter at a time. Due to this, the dataset suffers from imbalance, and a limited training size, a problem commonly encountered in material science [34]. Figs. 2, showing the distribution of collected values for each machining parameter, provides a visualization of the imbalance of experimentally collected parameters. As can be seen, the temperature distribution was bimodal (with most values within two distinct ranges), and the distributions of speed, distance, and pressure were unimodal (with most values within a single subset of machining values). To ensure that no inherent bias was introduced into the model through a selection of training and validation, Five-Fold Cross Validation [35] was implemented. Generically, K-Fold Cross Validation is a technique in which a dataset is separated into a variable number of equal-length sections. Models are then iteratively trained on all but one section, with the remaining section used for evaluation. Using 5-Fold Cross Validation specifically, the dataset was separated into five equal sections such that a model was trained on four sections and evaluated on the fifth. Lastly, to ensure a fair comparison of regression models at each degree, the same Cross Validation splits were used on each regression model, such that the only factor that varied was the degree of the model.

### 3.5. Convolutional neural networks (CNNs) for defect identification

In order to detect and segment deformations in PSCs, a Mask R-CNN was implemented on images collected via a Scanning Electron Microscope. These images were taken of cross-sectional views of the PSCs, using a diverse set of cells composed of different machining parameters to prevent overfitting a model [36]. Here, there were 25 images



**Fig. 2.** Distribution of experimentally collected data for each machine parameter.

collected and annotated, which were split 80:20 such that the training set had 20 images and the validation set had 5 images and each image contained approximately 20–30 annotations per image. These images were evaluated for performance using precision and recall, as well as a visual inspection.

#### 4. Results and discussion

##### 4.1. Sprayed perovskite structures

In this study, we used the spray coating system to deposit organic precursor on the underlying  $\text{PbI}_2$  to form perovskite active film for solar cells shown in Fig. 1a. The spray process includes the FA-rich precursor droplet generation, droplet transportation toward the heated  $\text{PbI}_2$ , droplet coalescence into a wet film, and drying process of perovskite wet film. The yellow film of  $\text{PbI}_2$  completely turned to a dark color of perovskite film once the FA-rich precursor was sprayed. Process control in spray deposition techniques is crucial in PSCs fabrication. Many processing parameters, including nozzle speed, substrate temperature, nozzle/substrate height, and pressure added after spraying, must be optimized to achieve the full potential of sprayed PSCs. An informed understanding of each of these parameters' impact upon sprayed PSCs will lead to the production of compact and defect-free perovskite layers. Defects formed in perovskite film lead to the degradation of the perovskite film, contributing to PSCs' decreased performance efficiency [37]. By tuning those variables, perovskites deposited by spray coating have compact structures that consist of large grains shown in Fig. 1b.

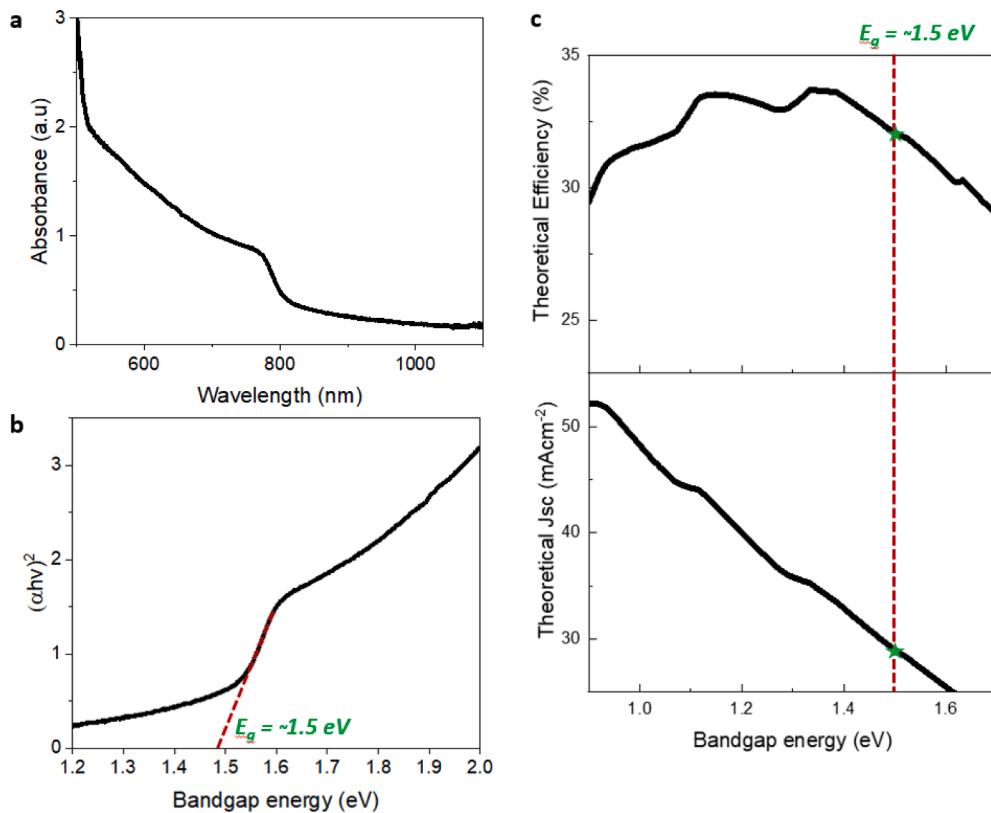
In assembly, the mesoscopic solar cell architecture was used (Fig. 1c), consisting of a structure with  $\text{TiO}_2$  ETL, a perovskite layer formed via the spray-assisted method, spiroOMeTAD HTL, and gold

thermally evaporated top contacts. Incorporating sprayed-perovskite as the active absorber layer, Fig. 1d shows a cross-sectional image of solar cells stack representative with different functional layers, in which the perovskite was infiltrated into the porous oxide mesostructure. Over these length scales, the optimized perovskite films are relatively uniform, with an approximate thickness of 500 nm.

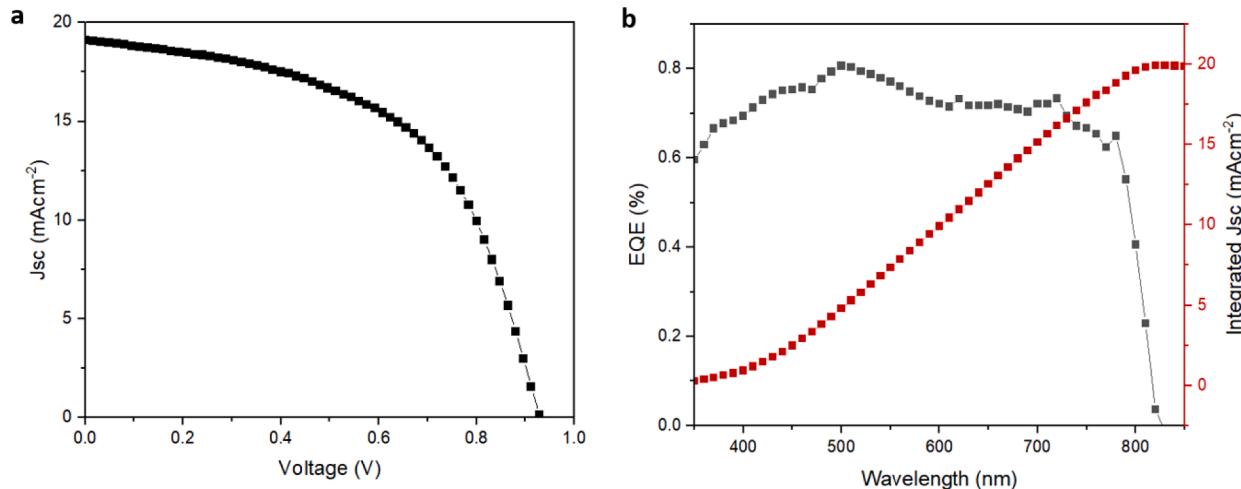
##### 4.2. Performances of sprayed perovskite solar cells

To estimate the bandgap energy of the absorber material, we measured UV-vis absorption spectrum (Fig. 3a) of the optimized sprayed-perovskite film. The UV absorption demonstrates the absorption coefficient ( $\alpha$ ) at a different visible wavelength ( $\lambda$ ), and the optical bandgap of perovskite is calculated by plotting Tauc plot from UV absorption spectra. The resulting Tauc plot has a distinct linear regime which denotes the onset of absorption. According to the Tauc plot (Fig. 3b), the sprayed perovskites obtained the bandgap energy,  $E_g$ , of ~1.5 eV for 500 nm-thick of film. Lying in the peak range of the SQ limit (Fig. 3c), a single-junction solar cell with sprayed perovskite absorber layer (1.5 eV bandgap energy) is capable of achieving a theoretical maximum efficiency of 32.21% corresponding to a theoretical short-circuit current density,  $J_{sc}$ , of 29.38 mA/cm<sup>2</sup>.

After tuning spray parameters, mesoscopic perovskite solar cells were fabricated using the optimized sprayed perovskites. External pressure was physically added to the complete solar cells. Fig. 4a gives the current density–voltage ( $I$ - $V$ ) characteristics of the best-performed solar cells. The best-performed PSCs were obtained from the substrate temperature of 50 °C, linear nozzle speed of 150 mm/s, nozzle/substrate height of 7 cm, and applied pressure of 7 MPa after spraying. The power conversion efficiency (PCE) of 9.86 (9.03 ± 0.74)% was achieved with



**Fig. 3.** UV–vis absorption spectra of sprayed perovskite films (a), Tauc plot method, which extracts the bandgap from absorption coefficient data (b), and the Shockley–Queisser theoretical limit for power conversion efficiency (PCE) and short-circuit current density  $J_{sc}$  as a function of bandgap energy. The dashed line shows the bandgap energy of the optimized sprayed perovskites obtained in this study (c).



**Fig. 4.** Current density–voltage ( $I$ - $V$ ) curve for the best performing sprayed PSC with AM 1.5 radiation under ambient condition (a), and external quantum efficiency (EQE) spectrum and integrated  $J_{sc}$  for solar cells based on the 500 nm-thick sprayed perovskite film (b).

$J_{sc}$  of 19.11  $\text{mA/cm}^2$ ,  $V_{oc}$  of 0.91 V, and fill factor,  $FF$ , of 0.56, which is still far behind the theoretical SQ limit for solar cells with 1.5 eV bandgap of single-junction perovskites. To confirm the charge collection, the external quantum efficiency (EQE) of sprayed perovskite solar cells was measured in Fig. 4b. Integrating the EQE spectrum yields a current density of about 20  $\text{mA/cm}^2$ , which was in agreement with the current density–voltage measurements (Fig. 4a). This low current density–voltage characteristic is typically due to current losses through the devices, such as the remaining defects in the bulk of perovskite active layers and at interfaces with the charge transport layers [18,38], and

leakage around the edges of devices [39]. Thus, an improvement is needed to fabricate perovskite solar cells via spray coating to exceed the obtained performances.

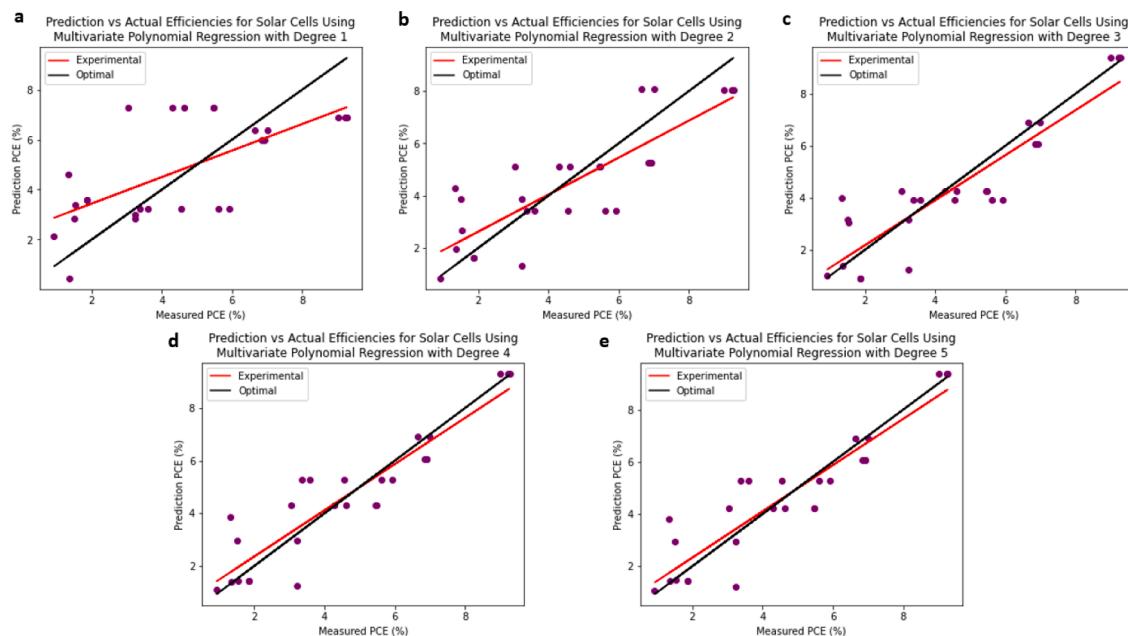
#### 4.3. Regression model performance

To further optimize the sprayed perovskite solar cells fabrication, a regression model was employed to capture the relationship and predict the impact of various processing variables of sprayed perovskite fabrication on the efficiency of PSCs based on the *trial-and-error* optimization

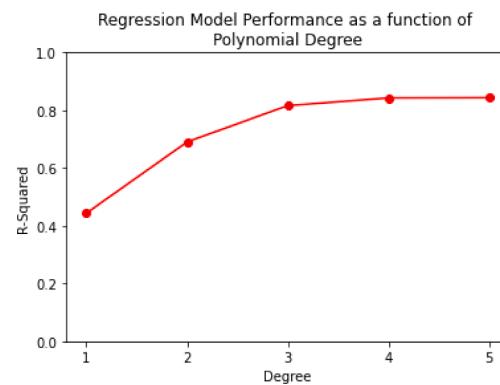
dataset. Models of increasing polynomial degrees were trained to best capture a relationship between the parameters and resultant efficiency. Fig. 5, with graphs comparing measured PCE to predicted PCE, shows that as the polynomial degree increased, so did the model's performance. We used  $R^2$ , a statistical measure of how close the data is to the fitted regression line, for model evaluation [40]. Models were trained up to degree 5 because, as shown in Fig. 6 based on the associated  $R^2$  value for models trained at each degree, model performance appears to be nearing an asymptote. Thus, training models with a higher degree polynomial would not add additional value.

Two key considerations were made to compare these models and select an optimal one: Performance and Generalizability. As shown in Figs. 5 and 6, models trained between degrees 3–5 all performed quite similarly and each performed better than models with degrees 1 and 2. To evaluate the model's generalizability, efficiency predictions were made for each possible parameter combination. These efficiencies were then compared with their expected value based on domain knowledge. For example, PCE shouldn't be less than 0% and has a theoretical SQ maximum of 33% efficiency for single-junction perovskite solar cells, so a model should not be predicting values outside of this range. Lastly, based on previously collected experimental, sprayed perovskite solar cells reached peaks of 19% [11], so a well-trained model would be expected to predict the PCE of most values under that range.

Distributions of predicted values for all combinations are presented in Fig. 7 for trained models with degrees from 1 to 5. The model trained with a degree of 1 (Fig. 7a) was the only model for which all values fell within the thresholds (greater than 0 and less than 33% PCE) identified by domain experts. However, the model failed to predict PCE values over 8%, even when training and validation sets included values close to 10%. Based on this, it is believed that the regression model trained with a degree of 1 was *underfitting*. The model trained with a degree of 2 (Fig. 7b) and has a few of values that fall outside of the scope (less than 0% predicted PCE), but nearly all values were within the range, and with a maximum value of 15%, closer to experimentally collected values. The model trained with a degree of 3 (Fig. 7c) performed closest to what would be expected. Most values were within the 0–10%, some in the 10–20% range, with a very small subset approaching the theoretical SQ limit maximum of 33% and none exceeding it. Similarly, some values were below 0, which is believed to be a result of the training set's imbalanced nature and limited size. Models trained with a degree of 4



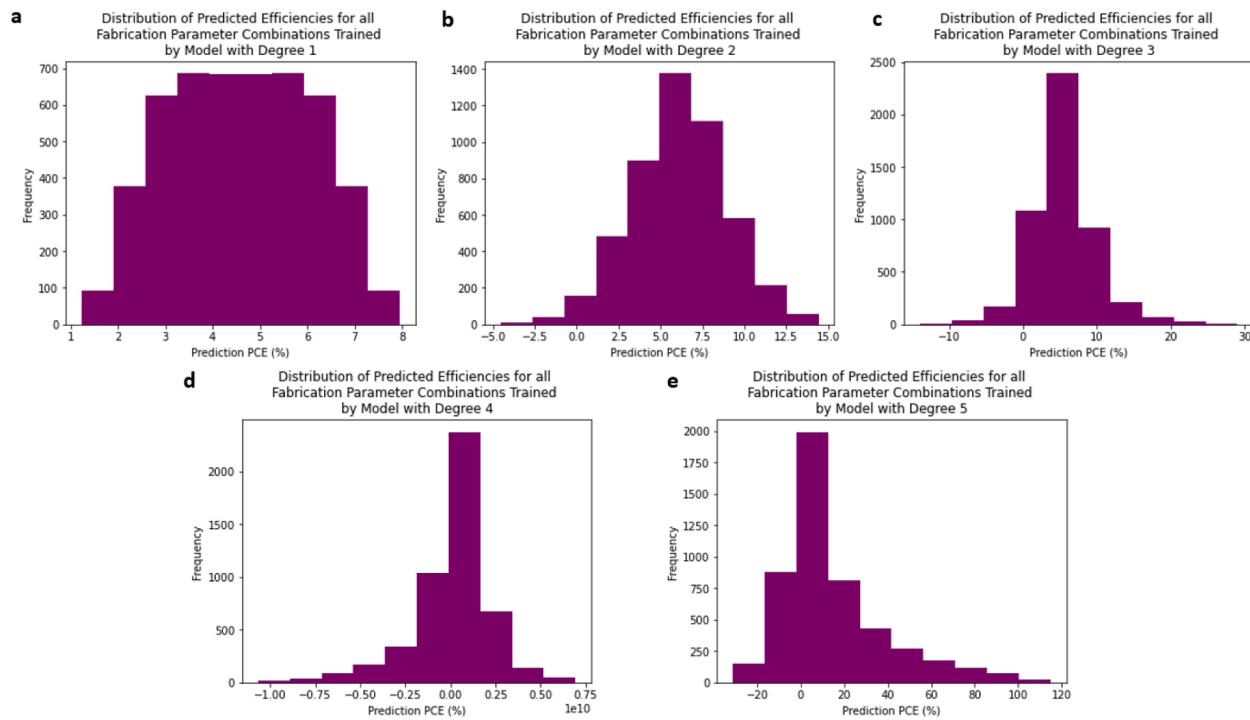
**Fig. 5.** Comparison of measured vs predicted PCE of sprayed PSCs by polynomial regression model from degree 1 through 5.



**Fig. 6.** Comparison of  $R^2$  values for polynomial regression models trained using degrees ranging from 1 to 5 for predicting predicting power conversion efficiency (PCE) for perovskite solar cells (PSCs).

and 5 (Fig. 7d-e), did not adhere to expected values as indicated by domain experts. With respect to the predicted values less than 0 by models with a degree of two or greater, this is not an indicator that those parameters will result in a solar cell with negative efficiency. It is instead a result of the limited dataset and the way in which the equations were trained. On data similar to the training set, all values are greater than zero, as can be seen in Fig. 5. However, the less similar machining parameters are to those found in the training set, the more possible it is for the resultant prediction to be negative, as it is outside of what was considered for training the model. It is possible their resultant prediction are negative. This is one of the leading motivations for collecting a larger dataset moving forward, enabling us to predict on a broader range of values with a higher degree of confidence.

Considering performance, models trained with degrees 3, 4, and 5 equally had the highest performance, as shown in Fig. 6. However, when also considering generalizability, models trained with degrees 4 and 5 suffered significantly from *overfitting*, as presented in Fig. 7. As a result, the polynomial regression model trained with a degree of three, scoring an  $R^2$  of 0.816, is chosen for our problem.



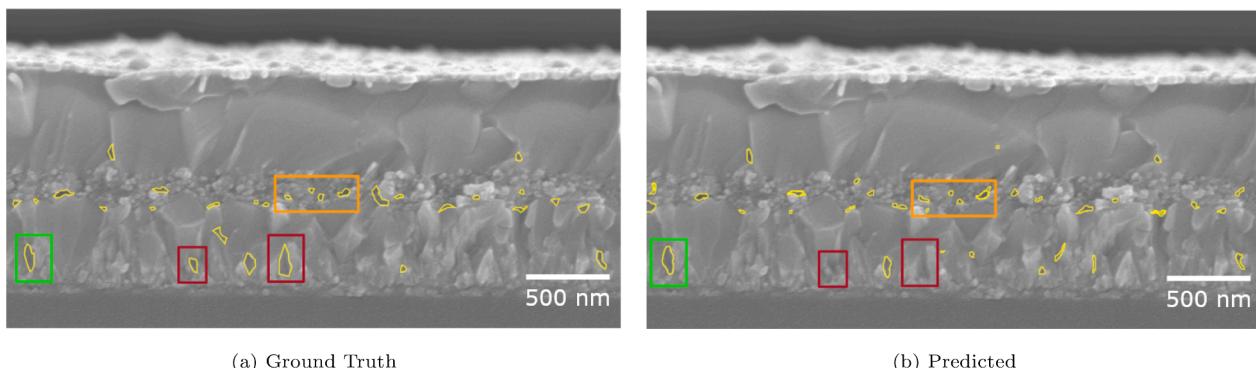
**Fig. 7.** Distribution of predicted PCE by polynomial regression model of degree 1 to 5.

#### 4.4. Defect identification in PSCs structures

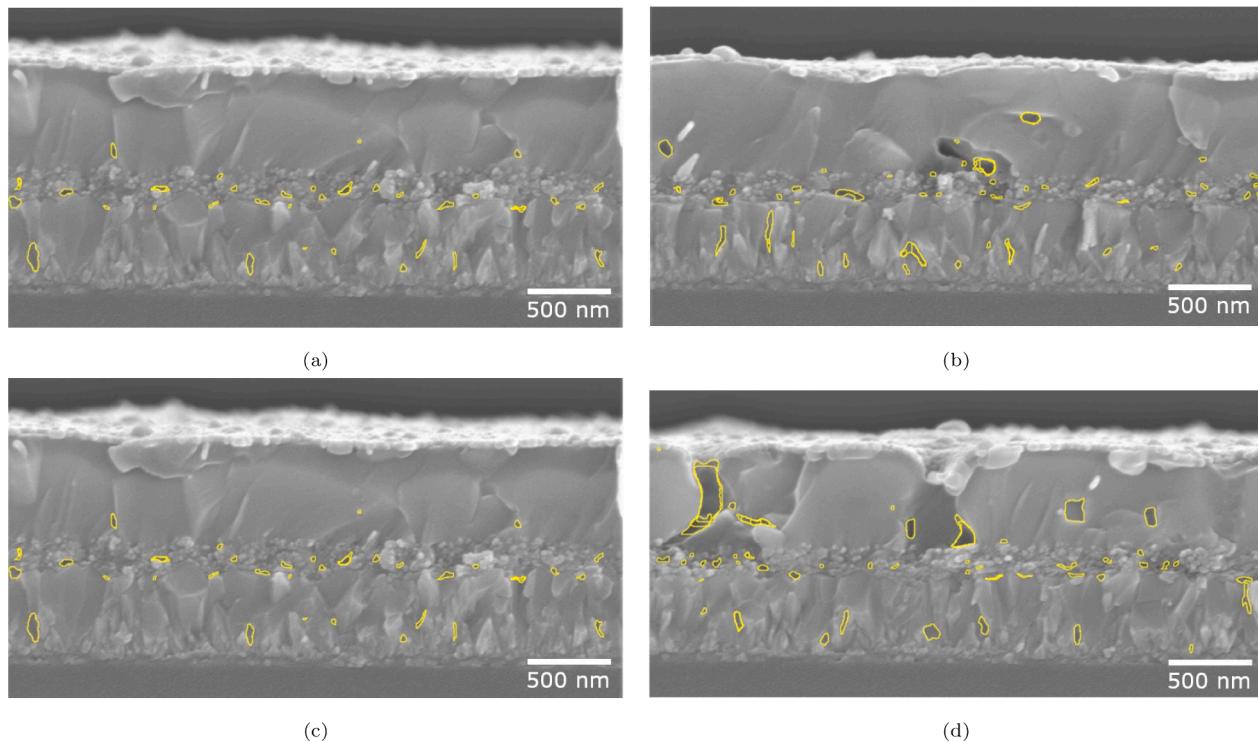
As time has progressed and computer vision algorithms improved, Material Science has proven a ripe field for its applications [21–23], and the prediction of PSC deformations is no exception. Trained on a small dataset of only 20 images, the Mask R-CNN model learned to predict deformations with a high degree of accuracy. As shown in Fig. 8, showing a comparison between ground truth and predicted results on an SEM image, the Mask R-CNN model performed quite well, scoring a precision of 83.3% and recall of 86.2% for segmented deformations. Highlighting specific annotations for discussion on performance, Fig. 8 have been labeled with green, orange, and red boxes. In the green boxes, we see predicted annotations, found in Fig. 8b, are more detailed than human-annotated ones, found in Fig. 8a, supporting the idea a Mask R-CNN provides utility moving forward. However, in the orange boxes, we see a region in which predicted regions are approximately equivalent to human-annotated ones, and in the section highlighted red, defects identified by a human were missed by the trained model, indicating that additional data-annotation and training are necessary. Moving forward, with an expanded training set, this model can be used as additional features to be considered beyond processing parameters when

predicting efficiency. While efficiency is the easiest metric to evaluate and one of the most relevant, device lifetime is also an important factor. A standardized method of quantifying defects can help improve the lifetime of solar cells and thus, sustainability.

Defect size distribution in PSCs predicted by R-CNN model at given spray parameters improves the understanding of the presence of deformations in solar cell structures. Poor spray processing controls lead to the poor formation of perovskite absorber film, which relates to the PSCs' performances and lifetimes. For example, at fixed substrate heating temperature, nozzle/substrate height, and pressure added after spraying, velocity of the nozzle head across the substrate surfaces contributes to the different defect formation in solar cell structures. In Fig. 9a, dense and compact grain perovskite structures are seen at a nozzle moving speed of 150 mm/s, while perovskite film consists of some voids at a higher speed of 250 mm/s (Fig. 9b). Defective sites in surfaces or interfaces generally accelerate the degradation of perovskite film, limiting the PSCs' performances and reducing PSCs' lifetimes [41]. Pressure effects after spraying, as shown in Fig. 9c and Fig. 9d, also show different defect distributions in PSC structures. Pressure added also distributes to different defects formation in solar cell structures. Excess pressure of 10 MPa leads to the formation of interlayer defects in solar



**Fig. 8.** Comparison of Ground Truth to Predicted values of Perovskite Solar Cells Identified by a Mask R-CNN model.



**Fig. 9.** The comparison of predicted deformations identified by a Mask R-CNN model for solar cells fabricated with linear nozzle moving speed of (a) 150 mm/s and (b) 250 mm/s. Other example for defect distribution of solar cells at different applied pressure of (c) 0 MPa and (d) 10 MPa.

cell structures which is detrimental to the efficiency of PSCs.

These results indicate the first step toward using a computer vision model to consider both efficiency and structural integrity when optimizing PSCs. Thus, we are preparing to collect a more extensive dataset that contains more entries and a balanced composition. In addition to evaluating the PSCs in this new dataset, the Mask R-CNN presented here will be used to evaluate the structural integrity. Once quantified, future optimization models can consider efficiency and structural integrity.

#### 4.5. Tuning regression equations

This work shows that a regression equation has the potential for learning relationships between processing parameters and PCE for solar cells. In this case, a third-degree polynomial was identified as the optimal model considering performance and overfitting. However, this does not mean that third-degree regression models are optimal for all circumstances similar to this. A different dataset, even if in a similar domain, may possess a different relationship, one that might require a different degree regression model [42]. For example, while a third-degree model performed better than the quadratic model, as shown in Fig. 6 and Fig. 7, a quadratic model might perform better on a different experiment. In addition to tuning to a specific degree, a set of degrees may perform better than others based on the equation itself, especially when generalizing to unseen data. For example, even functions, generally defined as symmetrical across the y-axis, behave more similarly to each other than an even function compared to an odd function and vice versa. Recognizing this can allow for better tuning and selecting of a regression equation. Specific to solar cells, Fig. 7 shows how both degree three and degree five (both odd functions) predicted relatively similarly on all possible machining parameters, whereas degree four predicted some values as high as  $0.75 \times 10^{10}$ . Ultimately, an understanding that different polynomial degrees may capture the relationship of a dataset better or worse, and a comprehensive tuning procedure will result in improved performance with a lower likelihood of overfitting.

#### 5. Implications

In this work, data-driven approaches are explored to improve the performances of perovskite solar cells. The results suggest that polynomial regression models possess the capability to capture relationships between machine parameters and resultant efficiencies. They also suggest that such models can be used to engineer the effects of key variables on the performances of spray-coated perovskite solar cells. Furthermore, a Mask R-CNN is capable of accurately predicting defects size distributions, allowing for future analysis between manufacturing parameters and defects in PSCs. This prediction may be used to guide the design of manufacturing parameters, or identification of defect distributions associated with quality control and device degradation. Since the measured efficiencies of PSCs are far below the theoretical Shockley-Queisser limit, machine learning approach could provide us with the rational basis of the design of processing parameters for optimization of spray PSCs. Finally, the current work suggest that machine learning can be used to design and optimize the performance characteristics of PSCs. It can also be used to minimize the time and resources required for data collection during the development of optimal conditions for solar cell manufacturing.

#### 6. Conclusion

In summary, the experimental PCE and predicted PCE of sprayed perovskite solar cells obtained from a third-degree polynomial regression model has a relatively high correlation with an  $R^2$  of 0.816. By recognizing trends in the experimentally collected data, this ML algorithm shows excellent prediction of solar cell performances that can assist the fabrication of PSCs, especially for optimizing high-dimensional parameters of spray processing. Our initial exploration yielded promising results that hint to the possibility of PCEs as high as 29%, approaching the theoretical SQ limit for single-junction solar cells. Therefore, devising a novel ML-guided experimental strategy, finding optimal spray values for best-performed PCE, and using CNNs to detect

and classify defects in PSCs structures can accelerate the exploration of high-dimensional parameters space of sprayed perovskite solar cells, leading to scalable, stable, and high-efficiency solar devices.

## 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.

## Data availability

The raw data and processed data required to reproduce these findings are available to download from <https://github.com/Data-Driven-Materials-Science/solarCells>.

## References

- [1] M. Zhang, D. Xin, X. Zheng, Q. Chen, W.H. Zhang, Toward greener solution processing of perovskite solar cells, *ACS Sustain. Chem. Eng.* 8 (2020) 13126–13138.
- [2] Rong, Y., Hu, Y., Mei, A., Tan, H., Saidaminov, M.I., Seok, S.I., McGehee, M.D., Sargent, E.H., Han, H. Challenges for commercializing perovskite solar cells, 2018. *Science* 361, eaat8235.
- [3] G. Xing, N. Mathews, S. Sun, S.S. Lim, Y.M. Lam, M. Grätzel, S. Mhaisalkar, T. C. Sum, Long-range balanced electron-and hole-transport lengths in organic-inorganic  $\text{ch}_3\text{nh}_3\text{pbI}_3$ , *Science* 342 (2013) 344–347.
- [4] Zuo, C., Bolink, H.J., Han, H., Huang, J., Cahen, D., Ding, L. Advances in perovskite solar cells, 2016. *Adv. Sci.* 3, 1500324.
- [5] M. Zhang, Z. Li, Z. Gong, Z. Li, C. Zhang, Perspectives on the mechanical robustness of flexible perovskite solar cells, *Energy Adv.* 2 (2023) 355–364.
- [6] R. Ichwani, V. Uzonwanne, A. Huda, R. Koech, O. Oyewole, W. Soboyejo, Adhesion in perovskite solar cell multilayer structures, *ACS Appl. Energy Mater.* 5 (2022) 6011–6018.
- [7] Z. Yang, S. Zhang, L. Li, W. Chen, Research progress on large-area perovskite thin films and solar modules, *J. Materomics* 3 (2017) 231–244.
- [8] H. Huang, J. Shi, L. Zhu, D. Li, Y. Luo, Q. Meng, Two-step ultrasonic spray deposition of  $\text{ch}_3\text{nh}_3\text{pbI}_3$  for efficient and large-area perovskite solar cell, *Nano Energy* 27 (2016) 352–358.
- [9] C. Guild, S. Biswas, Y. Meng, T. Jafari, A.M. Gaffney, S.L. Suib, Perspectives of spray pyrolysis for facile synthesis of catalysts and thin films: An introduction and summary of recent directions, *Catal. Today* 238 (2014) 87–94.
- [10] Paulson, N.H., Libera, J.A., Stan, M. Flame spray pyrolysis optimization via statistics and machine learning, 2020. *Mater. Des.* 196, 108972.
- [11] J.E. Bishop, C.D. Read, J.A. Smith, T.J. Routledge, D.G. Lidzey, Fully spray-coated triple-cation perovskite solar cells, *Sci. Rep.* 10 (2020) 1–8.
- [12] Wei, L., Xu, X., Gurudayal, Bullock, J., Ager, J.W. Machine learning optimization of p-type transparent conducting films, 2019. *Chem. Mater.* 31, 7340–7350.
- [13] J. Ling, M. Hutchinson, E. Antono, S. Paradiso, B. Meredig, High-dimensional materials and process optimization using data-driven experimental design with well-calibrated uncertainty estimates, *Integr. Mater. Manuf. Innov.* 6 (2017) 207–217.
- [14] Sun, W., Zheng, Y., Yang, K., Zhang, Q., Shah, A.A., Wu, Z., Sun, Y., Feng, L., Chen, D., Xiao, Z. Machine learning-assisted molecular design and efficiency prediction for high-performance organic photovoltaic materials, 2019. *Sci. Adv.* 5, eaay4275.
- [15] C.M. Bishop, N.M. Nasrabadi, *Pattern recognition and machine learning*, vol. 4, Springer, 2006.
- [16] H. Zhang, S.K. Moon, Reviews on machine learning approaches for process optimization in noncontact direct ink writing, *ACS Appl. Mater. Interfaces* 13 (2021) 53323–53345.
- [17] Z. Liu, N. Rolston, A.C. Flick, T.W. Colburn, Z. Ren, R.H. Dauskardt, T. Buonassisi, Machine learning with knowledge constraints for process optimization of open-air perovskite solar cell manufacturing, *Joule* 6 (2022) 834–849.
- [18] Olanrewaju, Y.A., Koech, R.K., Oyelade, O.V., Ahmed, R.A., Ichwani, R., Ebunu, A. I., Cromwell, J., Bello, A., Anye, V.C., Oyewole, O.K., et al. Thermally induced failure mechanisms in double and triple cations perovskite solar cells, 2022. *AIP Adv.* 12, 085014.
- [19] Ichwani, R., Koech, R., Oyewole, O.K., Huda, A., Oyewole, D.O., Cromwell, J., Martin, J.L., Grimm, R.L., Soboyejo, W.O. Interfacial fracture of hybrid organic-inorganic perovskite solar cells, 2022. *Extreme Mech. Lett.* 50, 101515.
- [20] He, K., Gkioxari, G., Dollár, P., Girshick, R., 2017. Mask r-cnn, in: *Proceedings of the IEEE international conference on computer vision*, pp. 2961–2969.
- [21] S.E. Price, M.A. Gleason, B.C. Sousa, D.L. Cote, R. Neamtu, Automated and refined application of convolutional neural network modeling to metallic powder particle satellite detection, *Integr. Mater. Manuf. Innov.* 10 (2021) 661–676.
- [22] R. Cohn, I. Anderson, T. Prost, J. Tiarks, E. White, E. Holm, Instance segmentation for direct measurements of satellites in metal powders and automated microstructural characterization from image data, *JOM J. Miner. Metals Mater. Soc.* 73 (2021) 2159–2172.
- [23] S. Price, R. Neamtu, Identifying, evaluating, and addressing nondeterminism in mask r-cnns, *International Conference on Pattern Recognition and Artificial Intelligence*, Springer (2022) 3–14.
- [24] W. Shockley, H.J. Queisser, Detailed balance limit of efficiency of p-n junction solar cells, *J. Appl. Phys.* 32 (1961) 510–519.
- [25] S. Rühle, Tabulated values of the shockley–queisser limit for single junction solar cells, *Solar Energy* 130 (2016) 139–147.
- [26] S. Weichenthal, K. Van Ryswyk, A. Goldstein, S. Bagg, M. Shekharifard, M. Hatzopoulou, A land use regression model for ambient ultrafine particles in montreal, canada: A comparison of linear regression and a machine learning approach, *Environ. Res.* 146 (2016) 65–72.
- [27] Rong, S., Bao-Wen, Z., 2018. The research of regression model in machine learning field, in: *MATEC Web of Conferences*, EDP Sci. p. 01033.
- [28] Forkor, G., Hounkpatin, O.K., Welp, G., Thiel, M. High resolution mapping of soil properties using remote sensing variables in south-western burkina faso: a comparison of machine learning and multiple linear regression models, 2017. *PLoS one* 12, e0170478.
- [29] G. Han, Y. Sun, Y. Feng, G. Lin, N. Lu, Machine learning regression guided thermoelectric materials discovery—a review, *ES Mater. Manuf.* 14 (2021) 20–35.
- [30] Tobias, R.D., et al., 1995. An introduction to partial least squares regression, in: *Proceedings of the twentieth annual SAS users group international conference*, Citeseer. pp. 1250–1257.
- [31] V.L. Deringer, A.P. Bartók, N. Bernstein, D.M. Wilkins, M. Ceriotti, G. Csányi, Gaussian process regression for materials and molecules, *Chem. Rev.* 121 (2021) 10073–10141.
- [32] Mishin, Y. Machine-learning interatomic potentials for materials science, 2021. *Acta Materialia* 214, 116980.
- [33] B. Meredig, E. Antono, C. Church, M. Hutchinson, J. Ling, S. Paradiso, B. Blaiszik, I. Foster, B. Gibbons, J. Hattrick-Simpers, et al., Can machine learning identify the next high-temperature superconductor? examining extrapolation performance for materials discovery, *Mol. Syst. Des. Eng.* 3 (2018) 819–825.
- [34] Y. Zhang, C. Ling, A strategy to apply machine learning to small datasets in materials science, *Npj Comput. Mater.* 4 (2018) 1–8.
- [35] T. Fushiki, Estimation of prediction error by using k-fold cross-validation, *Stat. Comput.* 21 (2011) 137–146.
- [36] Ying, X., 2019. An overview of overfitting and its solutions, in: *Journal of physics: Conference series*, IOP Publishing. p. 022022.
- [37] S. Bera, A. Saha, S. Mondal, A. Biswas, S. Mallick, R. Chatterjee, S. Roy, Review of defect engineering in perovskites for photovoltaic application, *Mater. Adv.* 3 (2022) 5234–5247.
- [38] Koech, R.K., Ichwani, R., Martin, J.L., Oyewole, D.O., Oyelade, O.V., Olanrewaju, Y.A., Sanni, D.M., Adeniji, S.A., Grimm, R.L., Bello, A. A study of the effects of a thermally evaporated nanoscale csbr layer on the optoelectronic properties and stability of formamidinium-rich perovskite solar cells, 2021. *AIP Adv.* 11, 095112.
- [39] C. Ma, N.G. Park, A realistic methodology for 30% efficient perovskite solar cells, *Chem* 6 (2020) 1254–1264.
- [40] Acharya, M.S., Armaan, A., Antony, A.S., 2019. A comparison of regression models for prediction of graduate admissions, in: *2019 international conference on computational intelligence in data science (ICCIDIS)*, IEEE. pp. 1–5.
- [41] Wang, M., Fei, C., Uddin, M.A., Huang, J. Influence of voids on the thermal and light stability of perovskite solar cells, 2022. *Sci. Adv.* 8, eab05977.
- [42] F. Yao, H.G. Müller, Functional quadratic regression, *Biometrika* 97 (2010) 49–64.