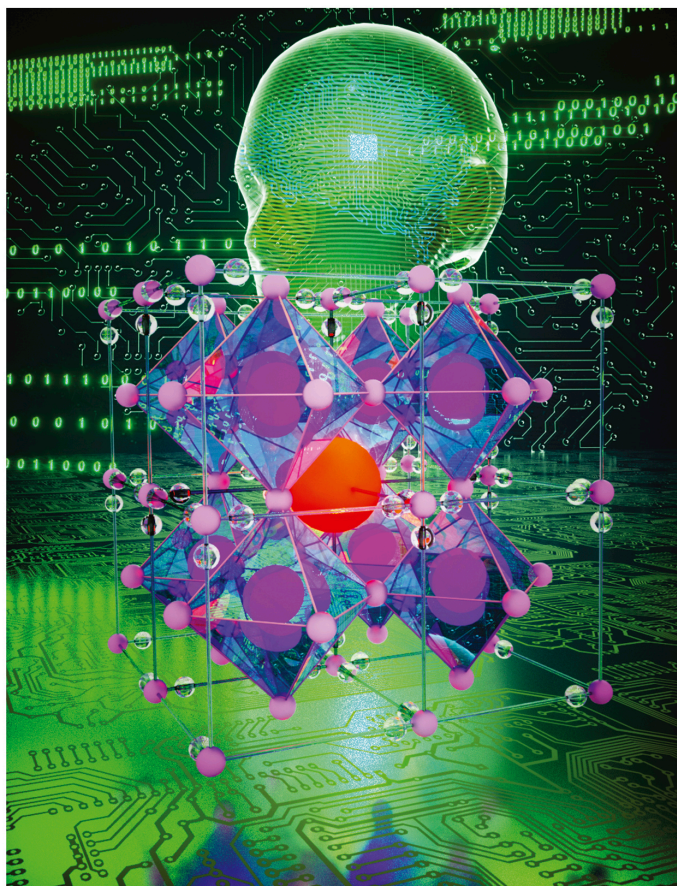


Emerging opportunities for hybrid perovskite solar cells using machine learning

Abigail R. Hering, Mansha Dubey, and Marina S. Leite



Emerging opportunities for hybrid perovskite solar cells using machine learning



Cite as: APL Energy 1, 020901 (2023); doi: 10.1063/5.0146828

Submitted: 16 February 2023 • Accepted: 10 April 2023 •

Published Online: 5 July 2023



Abigail R. Hering,  Mansha Dubey,  and Marina S. Leite^{a)} 

AFFILIATIONS

Department of Materials Science and Engineering, University of California, Davis, 1 Shields Ave., Davis, California 95616, USA

^{a)} Author to whom correspondence should be addressed: mleite@ucdavis.edu

ABSTRACT

While there are several bottlenecks in hybrid organic–inorganic perovskite (HOIP) solar cell production steps, including composition screening, fabrication, material stability, and device performance, machine learning approaches have begun to tackle each of these issues in recent years. Different algorithms have successfully been adopted to solve the unique problems at each step of HOIP development. Specifically, high-throughput experimentation produces vast amount of training data required to effectively implement machine learning methods. Here, we present an overview of machine learning models, including linear regression, neural networks, deep learning, and statistical forecasting. Experimental examples from the literature, where machine learning is applied to HOIP composition screening, thin film fabrication, thin film characterization, and full device testing, are discussed. These paradigms give insights into the future of HOIP solar cell research. As databases expand and computational power improves, increasingly accurate predictions of the HOIP behavior are becoming possible.

© 2023 Author(s). All article content, except where otherwise noted, is licensed under a Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>). <https://doi.org/10.1063/5.0146828>

INTRODUCTION

Over the past decade, hybrid organic–inorganic perovskites (HOIPs) have seen a rapid increase in research interest due to their exceptional optical and electronic properties, which demonstrates their potential for optoelectronic applications, such as photovoltaics (PVs), light-emitting diodes (LEDs), and radiation sensors.¹ Perovskites follow the ABX₃ structure [see Fig. 1(a)], in which A is an organic or inorganic cation [such as methylammonium (MA⁺), formamidinium (FA⁺), or Cs⁺], B is a metal cation (such as Pb²⁺ or Sn²⁺), and X is a halide anion (such as Cl[−], I[−], or Br[−]). There are several possible constituent ions for HOIPs, and each composition has its own unique properties. This adjustability is extremely important as it produces a tunable bandgap, which makes HOIPs an ideal candidate for semiconducting applications. Simultaneously, the innumerable possibilities for chemical compositions and their response to environmental stressors represent a colossal hyperparameter space to be explored.^{2,3} In turn, machine learning (ML) could significantly accelerate the discovery of new HOIP families and their physical properties.^{4,5}

Many applications of HOIPs to PVs have been explored, including dual- and multi-junction devices,^{6–8} which have produced

thousands of scientific publications over the past decade, shown as the blue curve in Fig. 1(b). However, due to the vast parameter space and degrees of freedom encompassing the HOIPs, purely experimental progress is often limited. Machine learning (ML), automated and autonomous high-throughput experimentation, and artificial intelligence can offer solutions to challenges related to HOIP material/device development. The pressing need to identify stable HOIPs has, thus far, led to an exponential growth in the number of peer-reviewed publications applying ML to these materials, as shown by the red curve in Fig. 1(b). This steep “learning curve” demonstrates the scientific community’s interest in the topic, and, if extrapolated to the next five years, confirms a burgeoning interest in the implementation of ML algorithms to effectively identify stable materials, optimize fabrication parameters and device structures, and device operation conditions that maximize device recovery.²

Exceptional physical properties of HOIPs have led to them being considered the “holy grail” for PVs, as they work well in both single- and dual-junction configurations.⁹ Yet, it is imperative that researchers effectively exploit all material- and device-related aspects required for the commercialization of these solar cells. Thus, herein we present our perspective on how ML can be an additional

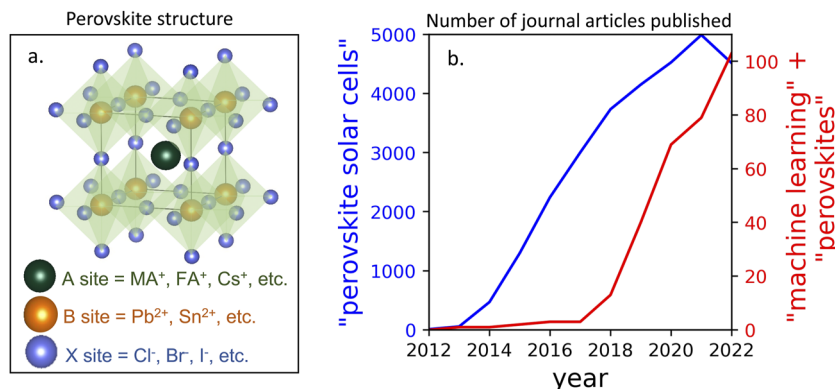


FIG. 1. (a) Schematic of the hybrid organic-inorganic perovskite structure highlighting the broad compositional space that can be achieved by varying the elements and molecules used for each site. (b) Number of scientific papers published on hybrid perovskites and machine learning from 2012 to 2022, per Web of Science. Note that the y axes are orders of magnitude different, indicating the raise of machine learning tools to solve open questions related to this burgeoning material system.

and powerful tool to assist in solving several critical open questions related to this promising class of material. First, we provide a brief overview of five distinct ML models. Second, we discuss examples from the literature applying ML models to different stages of HOIP development. Third, we examine which specific models are most beneficial for solving the distinct HOIP problems. Finally, we share a succinct outlook regarding future experiments that could benefit from ML.

MACHINE LEARNING MODELS

Numerous ML models have been applied to the field of functional materials, and we focus here on the ones that have already been used with HOIPs. We present an overview of significant ML examples from the HOIP literature in this section. The most straightforward one is linear regression, which can elucidate trends in data and suggest where additional analysis may be beneficial.¹⁰ At a more sophisticated level, neural networks (NNs) are multi-layer models that resemble neurons in the brain, and they have successfully been used from text predictions to image analyses.¹¹ Deep learning neural networks entail a series of hidden “black-box” layers, which have the potential to provide extremely accurate predictions in data trends and image analysis.^{12,13} These NN models can remember previous states, which may have lasting historical effects on HOIP degradation, a very relevant topic of research. Finally, statistical forecasting combines advanced statistical models with ML to produce accurate time series predictions of both regular and irregular, non-linear trends.¹⁴

Linear regression

Initial screening for long-term data trends can be made with linear regression before beginning the time intensive task of obtaining adequate training data for advanced ML models. Simple predictions of these data can be made with linear regression, which uses a least-squares curve fitting method [the black solid line in the schematic in Fig. 2(a)] to fit the dependent and independent variables (the blue data points). This model has an extremely low computational cost when compared to more advanced ML algorithms, making it an adequate starting method to detect long-term trends before moving to more advanced procedures, if needed. It is also very useful to use linear regression as a baseline approximation to

which the accuracy of other models can be compared. For example, in Fig. 2(a),¹⁵ the bandgap of Cs_{0.25}FA_{0.75}Pb(Cl_xBr_yI_(1-x-y))₃ HOIPs was estimated with high accuracy, showing a strong correlation between bandgap and halide composition (Pearson’s coefficient, $r > 0.99$) and an overall low root mean squared-error (RMSE, < 0.032 eV).

Nearly linear variation with chemical composition allows predictions of HOIP bandgaps [see Fig. 2(a)] to be calculated quickly, efficiently, and accurately. As mentioned before, the bandgap tunability enables these materials to be used in various optoelectronic devices and, thus, it is critical to screening all potentially useful chemical compositions. Estimating the bandgaps using ML can drastically reduce the time and resource investment to identify stable HOIPs.

Neural networks

Potential applications of neural networks (NNs), which are a subset of ML, include quantitative, categorical, or visual data analysis, making this an extremely versatile approach to investigate HOIP thin films and full devices. The overall NN structure resembles neurons in the brain, as represented by the schematic in Fig. 2(b), which have complex and intricate levels of connections. NNs may have one or several hidden layers between their input and output, which have different weighted computational nodes. In the past decade, with faster graphics processing unit (GPU) more readily available, NNs increased in complexity by adding anywhere from two to thousands of hidden layers. This enhancement in layers is known as deep neural networks or deep learning and is represented by the schematic in Fig. 2(c). There is no definite distinction between NNs and deep neural networks, but most recent studies are taking advantage of the computing power that gives deep neural networks more accurate results. There are numerous types of the latter, including artificial neural networks (ANNs), which are able to predict the potential energy of MAPbI₃ perovskites with a hyperbolic tangent update function [see Fig. 2(b)];¹¹ convolutional neural networks (CNNs), already used to calculate the octahedral tilt angle, lattice constant, and bandgap of 862 perovskite compositions [see Fig. 2(c)];¹⁶ and recurrent neural networks (RNNs),¹⁸ such as long short-term memory (LSTM)⁴ and echo state networks (ESNs),⁴ which can predict the photoluminescence (PL) response of MAPbI₃ and MAPbBr₃

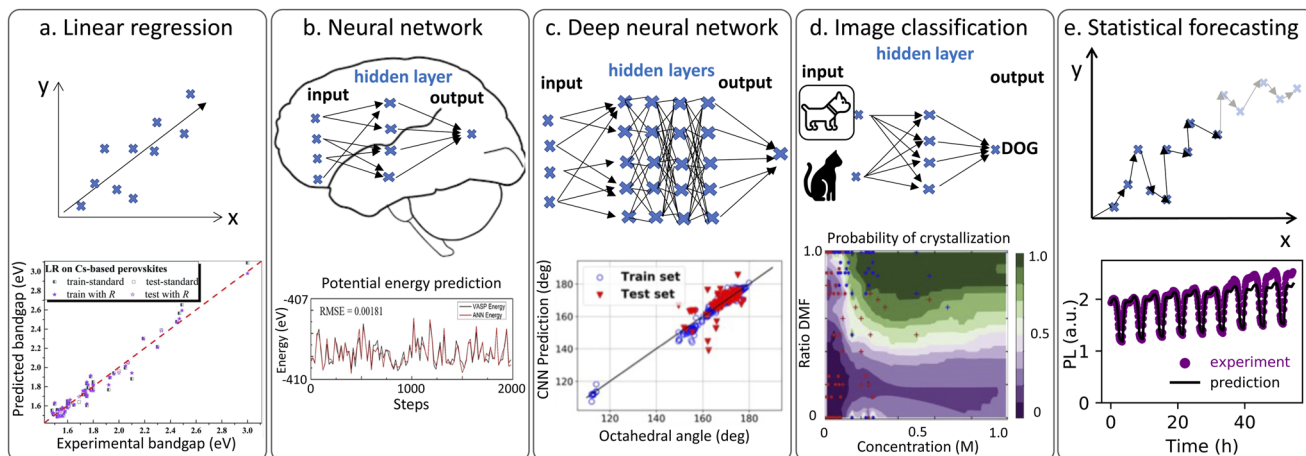


FIG. 2. Selected machine learning algorithms demonstrated to be effective when analyzing data related to hybrid perovskites: (a) linear regression showing bandgap prediction of $\text{Cs}_a\text{FA}_b\text{MA}_{(1-a-b)}\text{Pb}(\text{Cl}_x\text{Br}_y\text{I}_{(1-x-y)})_3$ family of compositions,¹⁵ (b) neural network used to forecast the potential energies and atomic forces in cubic perovskite MAPbI_3 ,¹¹ (c) deep neural network enabled prediction of the octahedral angle tilt in metal halide perovskites,¹⁶ (d) image classification of thin film crystallization in phenethylammonium lead bromide chemical space,¹³ and (e) statistical forecasting of photoluminescence response in a Cs-FA thin film.¹⁷ In all panels, the top and bottom rows refer to illustrations highlighting each concept and one example of published data, respectively. Figures reproduced with permission from (a) Li *et al.*, RSC Adv. 11(26), 15688–15694 (2021), Copyright 2021 RSC Advances; (b) H.-A. Chen and C.-W. Pao, ACS Omega 4(6), 10950–10959 (2019), Copyright 2019 ACS; (c) W. A. Saidi, W. Shadid, and I. E. Castelli, npj Comput. Mater. 6(1), 36 (2020), Copyright 2020 NPJ Computational Materials; (d) Kirman *et al.*, Matter 2(4), 938–947 (2020), Copyright 2020 Matter; and (e) Srivastava *et al.*, ACS Energy Lett. 8, 1716–1722 (2023), Copyright 2023 ACS.

upon their exposure to humidity. After initial predictions with linear regression, Li *et al.* have demonstrated the application of NNs and random forest algorithms (RFAs) to precisely predict bandgaps of the $\text{Cs}_a\text{FA}_b\text{MA}_{(1-a-b)}\text{Pb}(\text{Cl}_x\text{Br}_y\text{I}_{(1-x-y)})_3$ family with an accuracy of at least $\text{RMSE} < 0.145$ and < 0.05 , respectively. The more advanced networks are discussed below.

Image classification

One of the more recent groundbreaking results has been achieved in the field of image classification. As shown in the illustration in Fig. 2(d), with sufficient input information, one can build a trustworthy pathway to distinguish images, for example dogs from cats. This method has rapidly accelerated the classification of visual research data, which previously would either require a scientist with many years of experience (in the simplest cases) or keep material patterns and correlations hidden (in situations where human eyes solely cannot resolve the information). As an effective example of image classification, Kirman *et al.*¹³ adopted a convolutional NN to predict the probability of perovskite thin film crystallization in the phenethylammonium lead bromide family of compositions [see Fig. 2(d)]. Starostin *et al.*¹² utilized deep learning methods to identify Debye–Scherrer rings from grazing incidence x-ray diffraction (GIXRD) images to track the formation of the distinct $(\text{BA})_2(\text{MA})_{n-1}\text{PbI}_{3n+1}$ Ruddlesden–Popper phases. Applications of NNs to identify space groups from XRD data¹⁹ and carrier lifetime from time-resolved PL measurements [see Fig. 3(c-i)]²⁰ are becoming more researched as these techniques offer valuable information related to material stability. In our opinion, deep neural networks and image classification will be adopted more often by research groups across the world and, in the next few years, will become routinely used methods to characterize HOIPs.

Statistical forecasting

While NN-assisted methods are in an intermediate stage of development in the context of HOIPs, statistical forecasting is still in its early stages of application to PV fabrication and device processing. These advanced statistical methods can be implemented to forecast economic and weather trends that may have several unknown confounding variables. For instance, they have the potential to pick up long- and short-term trends and differences in seasons. Srivastava *et al.* showed how a statistical forecasting model reveals photoluminescence trends in Cs-FA thin films for over 50 h while submitting the samples to varying moisture exposure¹⁷ [see Fig. 2(e)] for one example case. Overall, this method outperformed a deep neural network algorithm and achieved greater than 90% accuracy.

MACHINE LEARNING APPLIED TO HOIP MATERIAL SELECTION, CHARACTERIZATION, AND DEVICE TESTING

Effective use of ML in all the steps required to design, test, and manufacture stable HOIP PVs will attract much of the focus of the perovskite industry in next few years. We dedicate this section to demonstrating how ML is useful in each development stage, such as material screening, thin film characterization, and full device testing. Overall, the cornucopia of possible compositions for HOIPs has driven researchers to gather a profusion of data relating to various properties and behaviors of the material combinations^{5,26}. This dataquake has led to the need for the screening of compositions at various levels to achieve the desired performance and stability metrics.²⁷ This methodology is analogous to sifting through a funnel, as depicted in Fig. 3(a-i). Acquiring experimental data to determine materials' properties to then screen each possible

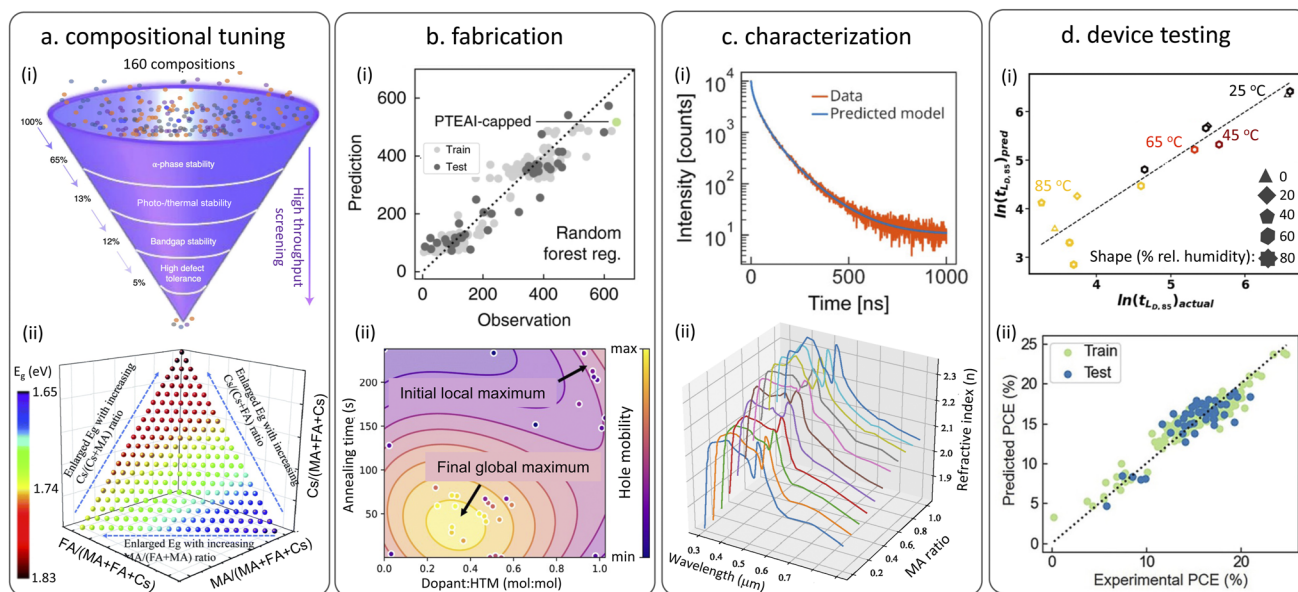


FIG. 3. Machine-learning-assisted examples of the relevant steps required for successfully achieving stable hybrid organic-inorganic perovskite solar cells, including (a) (i) material screening²¹ and (ii) bandgap tuning;¹⁵ (b) fabrication and selection of (i) carrier transport layers²² and (ii) thin film annealing time;²³ (c) characterization of physical properties, such as (i) time-resolved photoluminescence²⁰ and (ii) power conversion efficiency (PCE).²⁵ Figures reproduced with permission from a(i) Zhao *et al.*, *Nat. Energy* **7**(2), 144–152 (2022), Copyright 2021 Springer Nature; a(ii) Li *et al.*, *RSC Adv.* **11**(26), 15688–15694 (2021), Copyright 2021 RSC; b(i) Hartono *et al.*, *Nat. Commun.* **11**(1), 4172 (2020), Copyright 2020 Springer Nature; b(ii) MacLeod *et al.*, *Sci. Adv.* **6**(20) (2020), Copyright 2020 AAAS; c(i) Đorđević *et al.*, *ACS Photonics* **5**(12), 4888–4895 (2018), Copyright 2018 ACS; c(ii) Kim *et al.*, *Nanomaterials* **12**(6), 932 (2022), Copyright 2022 MDPI; d(i) Stoddard *et al.*, *ACS Energy Lett.* **5**(3), 946–954 (2020), Copyright 2020 ACS; and d(ii) Sol. RRL **6**(6), 2101100 (2022), Copyright 2022 Wiley-VCH Verlag.

composition is often extremely laborious and resource intensive. Thus, big data repositories, such as the “Perovskite Database Project,”³⁵ have enabled quick visualization of and information extraction from the existing published data. Beyond PVs, researchers have also developed methods to sort through existing papers using unsupervised ML and natural language processing (NLP)²⁸ methods to extract possible compositions suitable for other optoelectronic applications. In search of new materials, several studies on the application of ML-based methods have provided insights into the structural, electronic, optical, and overall physical properties of HOIPs without the need for fabrication and characterization of each chemical composition,¹⁶ an incredible step forward in the field. Thus, no doubt, we see ML as the impetus of the race to find the ideal composition and optoelectronic properties supported by automated and high-throughput fabrication and characterization.

Material discovery and selection

Regarding material fabrication, changing the ratios of the A site ions, as shown in Fig. 3(a-ii), or the halide ions, can significantly alter the bandgap. For example, increasing the Cs content at the A site is well known for increasing the semiconducting film bandgap, while the addition of FA reduces it. This approach is similar to the situation presented in Fig. 2(c), where two convolutional NNs predicted the perovskite bandgap after calculating the lattice constant and octahedral tilt angle of the structures. Furthermore, Cai

*et al.*²⁹ used LR, support vector regression (SVR), k-nearest neighbor regression (KNNR), RFA, gradient boosting regression (GBR), and NN to predict the bandgaps of mixed Pb–Sn perovskites and they used the predicted bandgaps to estimate device performance in terms of open-circuit voltage (V_{oc}), short-circuit current (J_{sc}), fill factor (FF), and power conversion efficiency (PCE). First-principles calculations³⁰ and density functional theory (DFT) have been employed to judge the accuracy of the output of ML algorithms and indicate robustness through low error-margins and computing costs.

A primary challenge faced by HOIP solar cells is the limited knowledge of environmentally induced degradation. The mechanisms understood so far indicate that improving charge carrier mobility by reducing the density of trap states is required for enhanced device lifetimes. Due to the poor thermal stability of numerous HOIP combinations, defect formation poses a real threat to the lifetime of the devices. Defect-induced localized trap states hinder ionic motion and trap charges at the surface and/or grain boundaries. In PV applications, this behavior is, in general, detrimental to the device open-circuit voltage (V_{oc}) and, hence, the power conversion efficiency (PCE) over a period of time. In addition, exposure to moisture has displayed an exacerbated degradation timeline in MAPbI₃ and is one of the most extensively studied environmental stressors.^{10,11} Solutions to hinder degradation, such as surface passivation using capping layers that modify the material chemistry of the HOIPs, have shown promising results. Due to

the large possibility of chemical species, the selection of appropriate layers has also been subjected to ML regression, assisting the rapid identification of best candidates [see Fig. 3(b-i)]. Here, along with regression models, image classification has been utilized to quickly analyze the color of thin films during degradation experiments within MAPbI₃ for 21 different capping layers. The onset of color change observed was matched against that of a random forest regression with an RMSE of ~100 min.²² Other measures to improve stability entail appropriate annealing times for carrier transport layers, such as spiro-OMeTAD, a commonly used hole transport layer (HTL) in full HOIP devices [see Fig. 3(b-ii)],²³ post-treatment amines, again for which ML has been used to predict options that would be most compatible with perovskite films,³¹ and feature engineering to study the photoelectrochemical properties of MAPbI₃ films in water.^{32,33} We notice that further quantitative studies on the effects of external stressors, such as exposure to oxygen, applied potential, and variations in temperature, are required to fill in the remaining knowledge gaps. Overall, we foresee ML as a key tool for effectively exploring the vast hyperparameter space related to all possible combinations of the relevant stressors.

Lead (Pb) is, by far, the most selected B site metal cation in HOIPs. However, due to environmental and health concerns related to the handling, storage, and use of lead, there has been a strong interest in Pb-free compositions.³⁴ The use of ML for identifying Pb-free families has, thus far, focused on tree-based algorithms that use predictive models for properties such as heat of formation, Debye temperature, and bandgap to determine the suitability of the possible compositions.^{35,36} Sn is a common alternative for Pb-based HOIPs,³⁴ however, the performance and stability of solar cells containing Sn significantly lag behind compared to those with Pb. Again, there is much work to be accomplished with Pb-free HOIPs and the need for ML methods for the discovery of *stable*, Pb-free alternatives remains imperative.

Device fabrication

We also note an emerging interest in higher-throughput fabrication experiments in HOIP research. The fabrication of thin films commonly utilizes techniques such as spin coating, thermal evaporation, and atomic layer deposition. However, these methods are tedious and time-consuming, impeding the pace of characterization and development of new materials.¹⁹ In addition, the occasional poor consistency in the fabrication process has stalled the development of methods for scaling up the production of HOIPs. To overcome these barriers, the automation of the fabrication steps using robotics has already been demonstrated in few publications.^{21,23,27} Uniform coatings of each layer, i.e., the absorber, electron transport layers (ETLs), hole transport layers (HTLs), and electrodes, is required for a successful scale-up. While methods such as spin coating work exceptionally well at the laboratory scale, they cannot be used for fabricating large PV modules. Therefore, there is an impending need for the development of high-throughput fabrication methods that can be scaled successfully to manufacture full PV modules. Fabrication of PV modules is a sequential process. From substrate preparation to deposition of the absorber layer, ETL and HTL, and, finally, electrodes, each step requires optimization in various ways. Liu *et al.*³⁷ applied probabilistic constraints at

each development stage to ascertain the optimal parameters using power conversion efficiency as the target metric. Using the knowledge acquired from one set of experimental data, the next set of experiments is improved and developed. Other methods, such as Bayesian parameter estimation,^{38,39} also utilize the output from previous sets of data to reject unnecessary or unpromising parameters as a means of optimization.

Device characterization

An ascending interest in data-driven approaches to characterize HOIP thin films has led to the development of novel ML-based solutions. The use of automated data diagnosis has significantly reduced the time investment enabling further workflow optimization. High throughput experimentation can often generate very large quantities of data, resulting in time- and computation-intensive methods for analysis, which may leave significant room for error. Material parameters, such as carrier lifetime [see Fig. 3(c-i)] and refractive index [see Fig. 3(c-ii)], have been analyzed and predicted for large datasets using ML models to quickly and accurately report crucial data required for the development of stable HOIPs. For the optimization of complete PV devices, the identification of the main limiting factor in J–V characteristics is essential.⁴⁰ Surface and bulk defects in HOIP solar cells lead to Shockley–Read–Hall (SRH) recombination dominating the loss and reducing Voc. Several publications have indicated the interfaces between the absorber layer and electron transport layer (ETL)/hole transport layer (HTL) to be the main recombination centers in HOIP solar cells. Determining the type of recombination mechanism can help identify the presence of trap states. ML algorithms, such as tree-based methods, have been used to classify recombination processes based on Voc and ideality factor. Predictions for carrier diffusion lengths in MAPbI₃ as shown in Fig. 3(d-i) and the PCE of MA/FA/Cs–Pb/Sn–I₃/Br₃/Cl₃ type compositions in Fig. 3(d-ii) are additional prime examples for the crucial need for the implementation of ML methods in assisting accurate and reliable analysis of large sets of data.

OUTLOOK

Yesterday's technology can no longer sufficiently meet the needs of today's rapid pace of material discovery and development. The vast compositional space and non-standardized fabrication and characterization techniques for HOIPs require ML-facilitated analyses and predictions to keep up with the pace of knowledge being acquired in the field. Various ML models and their applications to solve the numerous stability and performance issues related to HOIP solar cells presented in this *perspective* clearly indicate the rising interest and uptake in utilizing advanced, computational frameworks that are considerably less resource and time intensive. The sharing of ML algorithms and databases is already helping in standardizing the performance metrics for all kinds of HOIPs for energy applications, including LEDs and thermoelectrics. By extending this to robotic applications to high-throughput experiments and fabrication techniques, key environmental factors, optimal processing conditions, and the best suited ML models for various problems can be narrowed down. Specifically, we advocate for high-throughput measurements that do not require data augmentation¹⁷ and can

provide direct information about solar cells' figures of merit, such as ellipsometry, PL, and time resolved PL, to acquire information related to the generation, recombination, and collection of carriers, respectively. In the realm of imaging, the analysis of data produced by sophisticated methods, such as scanning probe and electron microscopies, is now benefiting from ML-assisted methods that entail the conversion of information (images and spectroscopy measurements) into descriptors that enables its quick assessment using statistical methods and graph analysis.⁴¹ Images acquired by more simple methods can also be extremely informative. As an example, dark-field imaging has already been successfully adopted to inform variations in the electrical conductivity of HOIPs, revealing subtle features not distinguishable by human eyes.⁴ Ideally, photographs using cell phones could enable quick, low cost, and vast amount of information about HOIP changes once submitted to distinct environmental stressors, which could then be correlated with the materials' optical properties. Concrete examples include correlating variations in sample color with changes in the material permittivity (i.e., light absorption) or the formation of new compounds on the surface of the HOIP films. In this ideal scenario, the main challenge would be to assure that images acquired from different laboratories around the world are comparable. Thus, a standard set of parameters (camera angle, lighting, etc.) is imperative. Furthermore, our need to access colossal amount of data to train any ML in an accurate manner demonstrates more than ever the need for researchers to follow the findable, accessible, interoperable, and reusable (FAIR) guiding principles.⁴²

CONCLUSIONS

Significant present-day challenges to HOIP solar cell development are long-term stability and reproducibility. This issue is currently being addressed by diverse approaches that integrate ML models. However, ML accuracy is still limited by the availability of adequate training data. To train ML models with high accuracy to forecast HOIP degradation and performance, emphasis on high-throughput, automated, or robotic experimentation is needed. Neural networks can be applied to a diverse set of problems in the HOIP development process, so are an extremely vital approach. Statistical forecasting methods are computationally expensive but can predict long-term behavior with high accuracy. Overall, we presented a survey of the most promising ML models to predict HOIP materials and PV devices. As the number of scientific publications encompassing HOIP solar cells and ML continues to grow (likely exponentially), we anticipate the use of artificial intelligence tools to become more widely used prior to, during, and after experiments.

ACKNOWLEDGMENTS

M.S.L. thanks the financial support from the National Science Foundation (Grant No. 20-23974) and from Sandia National Laboratories. Sandia National Laboratories is a multimission laboratory managed and operated by National Technology and Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International, Inc., for the U.S. Department of Energy's National Nuclear Security Administration under Contract

No. DE-NA0003525. This paper describes objective technical results and analysis. Any subjective views or opinions that might be expressed in this paper do not necessarily represent the views of the U.S. Department of Energy or the United States Government.

AUTHOR DECLARATIONS

Conflict of Interest

The authors have no conflicts to disclose.

Author Contributions

A.R.H. and M.D. contributed equally to this work.

Abigail R. Hering: Data curation (equal); Formal analysis (equal); Visualization (equal); Writing – original draft (equal). **Mansha Dubey:** Data curation (equal); Formal analysis (equal); Visualization (equal); Writing – original draft (equal). **Marina S. Leite:** Conceptualization (lead); Funding acquisition (lead); Project administration (lead); Supervision (lead); Visualization (supporting); Writing – review & editing (supporting).

DATA AVAILABILITY

Data sharing is not applicable to this article as no new data were created or analyzed in this study.

REFERENCES

- ¹C. W. Myung, A. Hajibabaei, J. H. Cha, M. Ha, J. Kim, and K. S. Kim, "Challenges, opportunities, and prospects in metal halide perovskites from theoretical and machine learning perspectives," *Adv. Energy Mater.* **12**(45), 2202279 (2022).
- ²J. M. Howard, E. M. Tennyson, B. R. A. Neves, and M. S. Leite, "Machine learning for perovskites' reap-rest-recovery cycle," *Joule* **3**(2), 325–337 (2019).
- ³M. V. Khenkin, E. A. Katz, A. Abate, G. Bardizza, J. J. Berry, C. Brabec, F. Brunetti, V. Bulović, Q. Burlingame, A. Di Carlo, R. Cheacharoen, Y.-B. Cheng, A. Colmann, S. Cros, K. Domanski, M. Dusza, C. J. Fell, S. R. Forrest, Y. Galagan, D. Di Girolamo, M. Grätzel, A. Hagfeldt, E. von Hauff, H. Hoppe, J. Kettle, H. Köbler, M. S. Leite, S. Liu, Y.-L. Loo, J. M. Luther, C.-Q. Ma, M. Madsen, M. Manceau, M. Matheron, M. McGehee, R. Meitzner, M. K. Nazeeruddin, A. F. Nogueira, Ç. Odabaşı, A. Osherov, N.-G. Park, M. O. Reese, F. De Rossi, M. Saliba, U. S. Schubert, H. J. Snaith, S. D. Stranks, W. Tress, P. A. Troshin, V. Turkovic, S. Veenstra, I. Visoly-Fisher, A. Walsh, T. Watson, H. Xie, R. Yildırım, S. M. Zakeeruddin, K. Zhu, and M. Lira-Cantu, "Consensus statement for stability assessment and reporting for perovskite photovoltaics based on ISOS procedures," *Nat. Energy* **5**(1), 35–49 (2020).
- ⁴M. Srivastava, J. M. Howard, T. Gong, M. Rebello Sousa Dias, and M. S. Leite, "Machine learning roadmap for perovskite photovoltaics," *J. Phys. Chem. Lett.* **12**(32), 7866–7877 (2021).
- ⁵T. J. Jacobsson, A. Hultqvist, A. García-Fernández, A. Anand, A. Al-Ashouri, A. Hagfeldt, A. Crovetto, A. Abate, A. G. Ricciardulli, A. Vijayan, A. Kulkarni, A. Y. Anderson, B. P. Darwich, B. Yang, B. L. Coles, C. A. R. Perini, C. Rehmann, D. Ramirez, D. Fairen-Jimenez, D. Di Girolamo, D. Jia, E. Avila, E. J. Juarez-Perez, F. Baumann, F. Mathies, G. S. A. González, G. Boschloo, G. Nasti, G. Paramasivam, G. Martínez-Denegri, H. Näsström, H. Michaels, H. Köbler, H. Wu, I. Benesperi, M. I. Dar, I. Bayrak Pehlivan, I. E. Gould, J. N. Vagott, J. Dagar, J. Kettle, J. Yang, J. Li, J. A. Smith, J. Pascual, J. J. Jerónimo-Rendón, J. F. Montoya, J.-P. Correa-Baena, J. Qiu, J. Wang, K. Sveinbjörnsson, K. Hirslandt, K. Dey, K. Frohna, L. Mathies, L. A. Castriotta, M. H. Aldamasy, M. Vasquez-Montoya, M. A. Ruiz-Preciado, M. A. Flatken, M. V. Khenkin, M. Grischek, M. Kedia, M. Saliba, M. Anaya, M. Veldhoen, N. Arora, O. Shargaieva, O. Maus, O. S. Game, O. Yudilevich, P. Fassl, Q.

- Zhou, R. Betancur, R. Munir, R. Patidar, S. D. Stranks, S. Alam, S. Kar, T. Unold, T. Abzieher, T. Edvinsson, T. W. David, U. W. Paetzold, W. Zia, W. Fu, W. Zuo, V. R. F. Schröder, W. Tress, X. Zhang, Y.-H. Chiang, Z. Iqbal, Z. Xie, and E. Unger, "An open-access database and analysis tool for perovskite solar cells based on the FAIR data principles," *Nat. Energy* **7**(1), 107–115 (2022).
- ⁶P. V. Kamat, "Hybrid perovskites for multijunction tandem solar cells and solar fuels. A virtual issue," *ACS Energy Lett.* **3**(1), 28–29 (2018).
- ⁷M. T. Hörantner, T. Leijtens, M. E. Ziffer, G. E. Eperon, M. G. Christoforo, M. D. McGehee, and H. J. Snaith, "The potential of multijunction perovskite solar cells," *ACS Energy Lett.* **2**(10), 2506–2513 (2017).
- ⁸Gurudayal, D. Sabba, M. H. Kumar, L. H. Wong, J. Barber, M. Grätzel, and N. Mathews, "Perovskite-hematite tandem cells for efficient overall solar driven water splitting," *Nano Lett.* **15**(6), 3833–3839 (2015).
- ⁹M. A. Ruiz-Preciado, F. Gota, P. Fassl, I. M. Hossain, R. Singh, F. Laufer, F. Schackmar, T. Feeney, A. Farag, I. Allegro, H. Hu, S. Gharibzadeh, B. A. Nejjand, V. S. Gevaerts, M. Simor, P. J. Bolt, and U. W. Paetzold, "Monolithic two-terminal perovskite/CIS tandem solar cells with efficiency approaching 25%," *ACS Energy Lett.* **7**(7), 2273–2281 (2022).
- ¹⁰R. J. Stoddard, W. A. Dunlap-Shohl, H. Qiao, Y. Meng, W. F. Kau, and H. W. Hillhouse, "Forecasting the decay of hybrid perovskite performance using optical transmittance or reflected dark-field imaging," *ACS Energy Lett.* **5**(3), 946–954 (2020).
- ¹¹H.-A. Chen and C.-W. Pao, "Fast and accurate artificial neural network potential model for MAPbI₃ perovskite materials," *ACS Omega* **4**(6), 10950–10959 (2019).
- ¹²V. Starostin, V. Munteanu, A. Greco, E. Kneschaurek, A. Pleli, F. Bertram, A. Gerlach, A. Hinderhofer, and F. Schreiber, "Tracking perovskite crystallization via deep learning-based feature detection on 2D X-ray scattering data," *npj Comput. Mater.* **8**(1), 101 (2022).
- ¹³J. Kirman, A. Johnston, D. A. Kuntz, M. Askerka, Y. Gao, P. Todorović, D. Ma, G. G. Privé, and E. H. Sargent, "Machine-learning-accelerated perovskite crystallization," *Matter* **2**(4), 938–947 (2020).
- ¹⁴S. C. Hillmer and G. C. Tiao, "An ARIMA-model-based approach to seasonal adjustment," *J. Am. Stat. Assoc.* **77**(377), 63–70 (1982).
- ¹⁵Y. Li, Y. Lu, X. Huo, D. Wei, J. Meng, J. Dong, B. Qiao, S. Zhao, Z. Xu, and D. Song, "Bandgap tuning strategy by cations and halide ions of lead halide perovskites learned from machine learning," *RSC Adv.* **11**(26), 15688–15694 (2021).
- ¹⁶W. A. Saidi, W. Shadid, and I. E. Castelli, "Machine-learning structural and electronic properties of metal halide perovskites using a hierarchical convolutional neural network," *npj Comput. Mater.* **6**(1), 36 (2020).
- ¹⁷M. Srivastava, A. R. Hering, Y. An, J.-P. Correa-Baena, and M. S. Leite, "Machine learning enables prediction of halide perovskites' optical behavior with >90% accuracy," *ACS Energy Lett.* **8**(4), 1716–1722 (2023).
- ¹⁸J. M. Howard, Q. Wang, M. Srivastava, T. Gong, E. Lee, A. Abate, and M. S. Leite, "Quantitative predictions of moisture-driven photoemission dynamics in metal halide perovskites via machine learning," *J. Phys. Chem. Lett.* **13**(9), 2254–2263 (2022).
- ¹⁹J.-P. Correa-Baena, K. Hippalgaonkar, J. van Duren, S. Jaffer, V. R. Chandrasekhar, V. Stevanovic, C. Wadia, S. Guha, and T. Buonassisi, "Accelerating materials development via automation, machine learning, and high-performance computing," *Joule* **2**(8), 1410–1420 (2018).
- ²⁰N. Đorđević, J. S. Beckwith, M. Yarema, O. Yarema, A. Rosspeintner, N. Yazdani, J. Leuthold, E. Vauthey, and V. Wood, "Machine learning for analysis of time-resolved luminescence data," *ACS Photonics* **5**(12), 4888–4895 (2018).
- ²¹Y. Zhao, T. Heumueller, J. Zhang, J. Luo, O. Kasian, S. Langner, C. Kupfer, B. Liu, Y. Zhong, J. Elia, A. Osvet, J. Wu, C. Liu, Z. Wan, C. Jia, N. Li, J. Hauch, and C. J. Brabec, "A bilayer conducting polymer structure for planar perovskite solar cells with over 1,400 hours operational stability at elevated temperatures," *Nat. Energy* **7**(2), 144–152 (2022).
- ²²N. T. P. Hartono, J. Thapa, A. Tihihonen, F. Oviedo, C. Batali, J. J. Yoo, Z. Liu, R. Li, D. F. Marrón, M. G. Bawendi, T. Buonassisi, and S. Sun, "How machine learning can help select capping layers to suppress perovskite degradation," *Nat. Commun.* **11**(1), 4172 (2020).
- ²³B. P. MacLeod, F. G. L. Parlange, T. D. Morrissey, F. Häse, L. M. Roch, K. E. Dettelbach, R. Moreira, L. P. E. Yunker, M. B. Rooney, J. R. Deeth, V. Lai, G. J. Ng, H. Situ, R. H. Zhang, M. S. Elliott, T. H. Haley, D. J. Dvorak, A. Aspuru-Guzik, J. E. Hein, and C. P. Berlinguette, "Self-driving laboratory for accelerated discovery of thin-film materials," *Sci. Adv.* **6**(20), eaaz8867 (2020).
- ²⁴S. M. Kim, S. D. Naqvi, M. G. Kang, H. Song, and S. Ahn, "Optical characterization and prediction with neural network modeling of various stoichiometries of perovskite materials using a hyperregression method," *Nanomaterials* **12**(6), 932 (2022).
- ²⁵Y. Liu, W. Yan, S. Han, H. Zhu, Y. Tu, L. Guan, and X. Tan, "How machine learning predicts and explains the performance of perovskite solar cells," *Sol. RRL* **6**(6), 2101100 (2022).
- ²⁶Q. Sun, W.-J. Yin, and S.-H. Wei, "Searching for stable perovskite solar cell materials using materials genome techniques and high-throughput calculations," *J. Mater. Chem. C* **8**(35), 12012–12035 (2020).
- ²⁷Y. Zhao, J. Zhang, Z. Xu, S. Sun, S. Langner, N. T. P. Hartono, T. Heumueller, Y. Hou, J. Elia, N. Li, G. J. Matt, X. Du, W. Meng, A. Osvet, K. Zhang, T. Stubhan, Y. Feng, J. Hauch, E. H. Sargent, T. Buonassisi, and C. J. Brabec, "Discovery of temperature-induced stability reversal in perovskites using high-throughput robotic learning," *Nat. Commun.* **12**(1), 2191 (2021).
- ²⁸L. Zhang and M. He, "Unsupervised machine learning for solar cell materials from the literature," *J. Appl. Phys.* **131**(6), 064902 (2022).
- ²⁹X. 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 MASn_xPb_{1-x}I₃ perovskite materials by machine learning and experimental realization," *Light: Sci. Appl.* **11**(1), 234 (2022).
- ³⁰K. 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 Photonics* **5**(3), 771–775 (2018).
- ³¹Y. 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**(2), 397–404 (2019).
- ³²Z. Pan, Y. Zhou, and L. Zhang, "Photoelectrochemical properties, machine learning, and symbolic regression for molecularly engineered halide perovskite materials in water," *ACS Appl. Mater. Interfaces* **14**(7), 9933–9943 (2022).
- ³³L. Zhang, W. Hu, M. He, K. Xu, and Z. Pan, "Interpretable machine learning for investigating photoelectrochemical properties of cosensitizer-based CH₃NH₃PbI₃/TiO₂ films in water," *J. Phys. Chem. C* **126**(14), 6482–6490 (2022).
- ³⁴P. Jiang, D. Acharya, G. Volonakis, M. Zacharias, M. Kepenekian, L. Pedesseau, C. Katan, and J. Even, "Pb-free halide perovskites for solar cells, light-emitting diodes, and photocatalysts," *APL Mater.* **10**(6), 060902 (2022).
- ³⁵J. Im, S. Lee, T.-W. Ko, H. W. Kim, Y. Hyon, and H. Chang, "Identifying Pb-free perovskites for solar cells by machine learning," *npj Comput. Mater.* **5**(1), 37 (2019).
- ³⁶X. Cai, Y. Zhang, Z. Shi, Y. Chen, Y. Xia, A. Yu, Y. Xu, F. Xie, H. Shao, H. Zhu, D. Fu, Y. Zhan, and H. Zhang, "Discovery of lead-free perovskites for high-performance solar cells via machine learning: Ultrabroadband absorption, low radiative combination, and enhanced thermal conductivities," *Adv. Sci.* **9**(4), 2103648 (2022).
- ³⁷Z. Liu, N. Rolston, A. C. Flick, T. W. Colburn, Z. Ren, R. H. Dauskardt, and T. Buonassisi, "Machine learning with knowledge constraints for process optimization of open-air perovskite solar cell manufacturing," *Joule* **6**(4), 834–849 (2022).
- ³⁸R. E. Brandt, R. C. Kurchin, V. Steinmann, D. Kitchaev, C. Roat, S. Levchenko, G. Ceder, T. Unold, and T. Buonassisi, "Rapid photovoltaic device characterization through Bayesian parameter estimation," *Joule* **1**(4), 843–856 (2017).
- ³⁹Q. Song, Y. Bai, and Q. Chen, "The spring of processing chemistry in perovskite solar cells—Bayesian optimization," *J. Phys. Chem. Lett.* **13**(46), 10741–10750 (2022).
- ⁴⁰V. M. Le Corre, T. S. Sherkar, M. Koopmans, and L. J. A. Koster, "Identification of the dominant recombination process for perovskite solar cells based on machine learning," *Cell Rep. Phys. Sci.* **2**(2), 100346 (2021).
- ⁴¹M. Ziatdinov, A. Ghosh, C. Y. Wong, and S. V. Kalinin, "AtomAI framework for deep learning analysis of image and spectroscopy data in electron and scanning probe microscopy," *Nat. Mach. Intell.* **4**(12), 1101–1112 (2022).
- ⁴²M. D. Wilkinson, M. Dumontier, I. J. Aalbersberg, G. Appleton, M. Axton, A. Baak, N. Blomberg, J.-W. Boiten, L. B. da Silva Santos, P. E. Bourne, J. Bouwman, A. J. Brookes, T. Clark, M. Crosas, I. Dillo, O. Dumon, S. Edmunds, C. T.

Evelo, R. Finkers, A. Gonzalez-Beltran, A. J. G. Gray, P. Groth, C. Goble, J. S. Grethe, J. Heringa, P. A. C. 't Hoen, R. Hooft, T. Kuhn, R. Kok, J. Kok, S. J. Lusher, M. E. Martone, A. Mons, A. L. Packer, B. Persson, P. Rocca-Serra, M. Roos, R. van Schaik, S.-A. Sansone, E. Schultes, T. Sengstag, T. Slater, G. Strawn, M. A. Swertz,

M. Thompson, J. van der Lei, E. van Mulligen, J. Velterop, A. Waagmeester, P. Wittenburg, K. Wolstencroft, J. Zhao, and B. Mons, "The FAIR guiding principles for scientific data management and stewardship," [Sci. Data](#) **3**(1), 160018 (2016).