THE JOURNAL OF PHYSICAL CHEMISTRY LETTERS
A JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

pubs.acs.org/JPCL

Perspective

Machine Learning Roadmap for Perovskite Photovoltaics

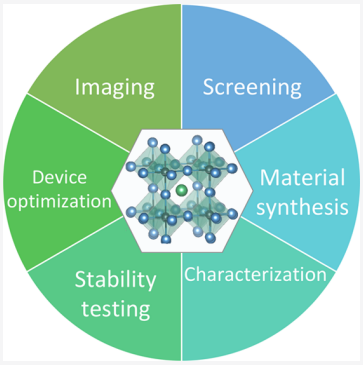
Meghna Srivastava, John M. Howard, Tao Gong, Mariama Rebello Sousa Dias, and Marina S. Leite*

Cite This: *J. Phys. Chem. Lett.* 2021, 12, 7866–7877

Read Online

ACCESS | Metrics & More | Article Recommendations

ABSTRACT: Perovskite solar cells (PSC) are a favorable candidate for next-generation solar systems with efficiencies comparable to Si photovoltaics, but their long-term stability must be proven prior to commercialization. However, traditional trial-and-error approaches to PSC screening, development, and stability testing are slow and labor-intensive. In this Perspective, we present a survey of how machine learning (ML) and autonomous experimentation provide additional toolkits to gain physical understanding while accelerating practical device advancement. We propose a roadmap for applying ML to PSC research at all stages of design (compositional selection, perovskite material synthesis and testing, and full device evaluation). We also provide an overview of relevant concepts and baseline models that apply ML to diverse materials problems, demonstrating its broad relevance while highlighting promising research directions and associated challenges. Finally, we discuss our outlook for an integrated pipeline that encompasses all design stages and presents a path to commercialization.



Recent advances in perovskite solar cells (PSC) have yielded power conversion efficiencies (PCEs) exceeding 25% for single-junction devices and 29% for tandem perovskite-Si cells.[1,2] The types of perovskite absorber have also diversified, from the baseline methylammonium lead triiodide (MAPI) system to more exotic alternatives such as lead-free, mixed-cation/anion, and all-inorganic perovskites.[3] As PSCs approach commercialization, several experimental challenges slow their advancement, including the vast compositional parameter space available and the limited stability of perovskite materials, which restricts photovoltaic (PV) device lifetimes to 3000 h.[4] For meaningful commercial deployment, these lifetimes must increase to ≥25 years for both all-perovskite and perovskite-Si tandem devices.[5,6] The current PSC design process involves careful compositional tuning,[7] systematic fabrication and/or characterization for each active layer,[8,9] and time-consuming stability and aging tests.[10,11] The high-dimensional perovskite parameter space and lengthy testing times motivate the use of machine learning (ML) to resolve the integral effects of environmental factors (light, temperature, bias, oxygen, and humidity) on the stability and performance of PSC without traversing every single stressor combination for all possible compositions. ML comprises a growing set of versatile techniques with a wide range of applications, including traffic speed prediction,[12] medical image screening,[13] and wind speed forecasting.[14] Thoughtful implementations of ML algorithms for materials science have already shown potential to alter research paradigms and drastically quicken the material design process. In energy generation, ML has been leveraged for Si PVs to predict future system performance metrics or environmental conditions in real time.[15,16] Furthermore, ML has applications in related fields such as energy storage, where various research groups have implemented models to forecast the remaining useful lifetime in batteries and fuel cells.[17−20] Likewise, the PSC community will benefit from novel research approaches involving ML, which greatly mitigate current experimental impediments.

The high-dimensional perovskite parameter space and lengthy testing times motivate the use of machine learning (ML) to resolve the integral effects of environmental factors (light, temperature, bias, oxygen, and humidity) on the stability and performance of PSC without traversing every single stressor combination for all possible compositions.

In this Perspective, we describe how ML is an ideal toolkit to accelerate PSC research and development to an unprecedented

Received: June 18, 2021
Accepted: July 22, 2021

https://doi.org/10.1021/acs.jpclett.1c01961
J. Phys. Chem. Lett. 2021, 12, 7866–7877

pace. We begin with a brief overview of relevant ML concepts, then detail our framework for applying learning algorithms to perovskites. We dedicate a large portion of the discussion to three important stages in the PSC design pipeline: compositional screening, perovskite layer synthesis and analysis, and full device fabrication and testing. In each of these sections, we delineate the key questions to answer and provide literature examples that apply ML to these questions. Next, we discuss ML for spatially resolved characterization and imaging projects. Throughout the Perspective, we also present three original proof-of-concept examples, which apply baseline models to data from the scientific community. These demonstrative examples include time-series forecasting, predicting trends in device degradation, and extracting information from images, with all associated data and code publicly available at https://github.com/mgsrivastava/ML-perovskites for easy reference. Finally, we conclude with our vision for an integrated ML pipeline that spans multiple stages of PSC design and highlight the most promising avenues for future work in ML for perovskites.

We begin with a concise description of key ML-related concepts mentioned in this Perspective. ML algorithms for materials science research are often grouped into two broad categories: supervised and unsupervised learning. Supervised learning processes involve input data that is labeled with a known "correct answer" and include regression and classification.[21] In regression, the model predicts a continuous valued output based on one or multiple inputs. Classification also involves labeled input data, but the model output is noncontinuous and instead sorts the input data points into known categories.[22] Note that many algorithms can perform both regression and classification. Unsupervised learning uses data that is not previously labeled and includes clustering techniques such as k-means.[23] Like classification, clustering sorts input data into groupings.

Deep learning and neural networks are powerful techniques to make complex, accurate predictions and are excellent candidates for perovskite PV applications. However, these models lack the interpretability of simpler algorithms such as linear regression or decision trees. Artificial neural networks (ANNs), which are often referred to simply as "neural networks (NNs)", functionally mimic neurons in the brain.[24] Layers of neurons (or "nodes") perform calculations on input data and then transfer the computed values to other layers. One common type of neural network is the convolutional neural network (CNN), which is often used for image analysis tasks such as object recognition or image classification.[25] Another class of NN is recurrent neural networks (RNNs), which are unique in that they can retain a "memory" of previous events in the data. Thus, RNNs are particularly well-suited for time-series prediction tasks.

Implementation of an ML model requires thoughtful data acquisition, preprocessing, visualization, and feature engineering. Normalization is common, especially if there are many input features with differing ranges of values. Obtaining enough data to train a model is vital, as we illustrate in later sections. Baseline models should be trained on the information available and then evaluated to determine if the data is sufficient or if more must be acquired. In cases where training data is lacking, augmentation can create a larger training set.[26] This may involve interpolation using linear combinations, adding noise, or image rotation and mirroring.[27,28] Recent work has also developed physics-informed data augmentation strategies. For example, Oviedo et al. took advantage of physics domain knowledge to augment an XRD data set through peak scaling, peak elimination, and pattern shifting.[27] Such physics-driven augmentation is a rational, robust method to expand the sparse data sets that are common in materials research problems.

Finally, to optimize an ML model, hyperparameters (parameters that control the learning process and must be set prior to training) are tuned to yield the best predictive performance on a validation data set. This process is often automated using iterative or grid-search approaches, which train many models with slightly different hyperparameters to pinpoint the optimal values. After the model is finalized and hyperparameters are set, it should be evaluated using additional data. Note that the data for this step (test set) is distinct from both the data used to tune the model hyperparameters (validation set) and the data used to train the model (train set). Outside of the succinct overview presented here, we direct readers to several recent papers for more detailed information on ML best practices focusing on materials science.[29−31]

The overarching goal of applying ML to PSC research is to elucidate novel physical insights and drastically reduce the time to develop, characterize, and optimize devices. With this in mind, we show how ML aids PSC design at three distinct stages: compositional screening (Figure 1a), material fabrication and stability analysis (Figure 1b), and full device development and testing (Figure 1c), as described in the following sections. At the compositional level, elemental site occupancies are selected based on chemical and lattice information. This elemental tuning is a time-consuming process, which ML expedites. For example, Saidi and colleagues used a hierarchical CNN to predict the lattice constant and octahedral tilt angle for ABX_3 perovskites and then used these features as the input for a second CNN to predict the material band gap.[32] Model input features were elemental and structural descriptors such as ionization energies, electron affinities, and tolerance factor. Using this network architecture, the authors predict band gap values varying from 0.2 to 6.0 eV, relevant for the design of optoelectronic devices such as tandem photovoltaics and LEDs. The CNN performed remarkably well, providing band gaps with a root-mean-square error (RMSE) of only 0.02 eV (Figure 1d) compared to density functional theory (DFT) results.

The overarching goal of applying ML to PSC research is to elucidate novel physical insights and drastically reduce the time to develop, characterize, and optimize devices.

With suitable elemental compositions established, PSC development reaches the materials level (Figure 1b). At this stage, perovskite fabrication parameters are set with critical absorber layer properties (such as light absorption and carrier mobility) in mind. Given perovskites' facile degradation, this focus also includes crucial stability testing to investigate the underlying physical mechanisms. ML models can use past data to predict future perovskite performance under various

https://doi.org/10.1021/acs.jpclett.1c01961
J. Phys. Chem. Lett. 2021, 12, 7866−7877

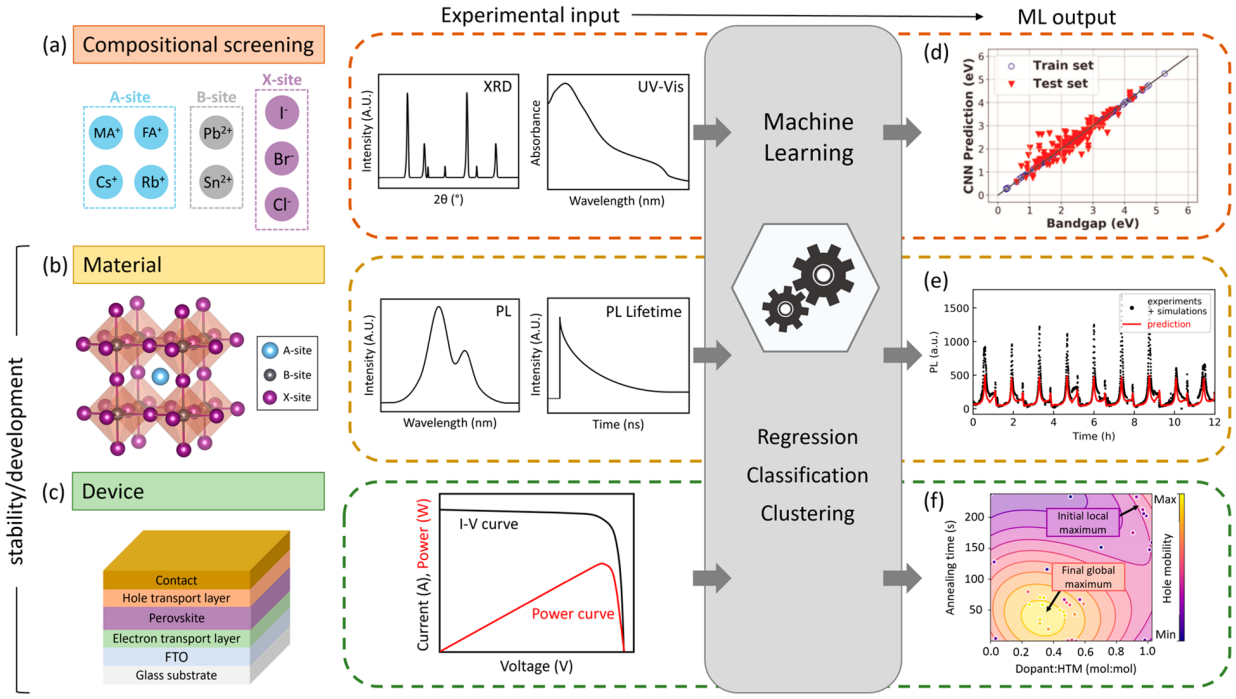


Figure 1. Machine learning for perovskite solar cell development. (a) Compositional screening, (b) material, and (c) device development/stability testing. At each stage, experimental data from techniques such as X-ray diffraction (XRD), ultraviolet−visible spectroscopy (UV−vis), photoluminescence (PL), PL lifetime, and current−voltage (I−V) measurements are fed into ML models to obtain key physical quantities and device figures of merit or to generate predictions. Specific examples of model outputs from literature are shown in (d−f). (d) Band gaps for various perovskite compositions predicted using a convolutional neural network (CNN) with elemental, structural, and precursor-based data as the experimental inputs. Adapted with permission from reference 32. Copyright 2020 The Authors, distributed under Creative Commons CC BY 4.0 license (https://creativecommons.org/licenses/by/4.0/). (e) Time-series prediction of humidity-dependent PL intensity from an echo state network (ESN) using past PL and rH data as inputs. Reprinted with permission from reference 28. Copyright 2020 The Authors, distributed under arXiv.org nonexclusive license. (f) Map of fabrication parameters (annealing time and dopant ratio) to maximize hole mobility in spiro-OmeTAD, an organic hole-transport material (HTM), from automated iterative experiments using the Phoenics global Bayesian optimization algorithm. Adapted with permission from reference 34. Copyright 2020 The Authors, distributed under Creative Commons CC BY-NC 4.0 license (https://creativecommons.org/licenses/by-nc/4.0/).

environmental stressors, including temperature, bias, humidity, light, and oxygen.[33] For instance, Howard et al. implemented an echo state network (ESN) to generate time-series predictions of humidity-dependent photoluminescence (PL) intensity of MAPI thin films (Figure 1e) and demonstrate a 12-h prediction window with <11% normalized root-mean-square error (NRMSE).[28] Extensions of this work to other combinations of environmental stressors can leverage ML to identify the leading factors associated with degradation under a given set of conditions.

At the full device level, investigations expand to include the remaining device layers, such as the electron- (ETL) and hole-transport layers (HTL). MacLeod et al. recently demonstrated a self-driving laboratory to select fabrication parameters for spiro-OMeTAD, a common HTL used in PSC devices (Figure 1f).[34] At this stage, characterizing stability under environmental stressors and standard operating conditions is again vital and often yields different results than for the perovskite layer alone due to interfacial effects such as charge carrier recombination and contact resistance.[35]

Table 1 summarizes the key descriptors and common figures of merit for each development level. ML models at a given level may use characteristic descriptors to predict figures of merit, although they can also generate one descriptor using another or supply completely different information—for instance to classify failure modes in full devices. At each level, we highlight fruitful ML approaches and suggestions for future work. Some strategies, such as high-throughput experimentation, improved data curation, and physics-based ML, apply equally to all levels and should be prioritized accordingly.

Given the extensive variability of questions that ML can address, we present a generalized roadmap that applies to all three levels of PSC development. The steps of this paradigm are

(1) identify the material question of interest;
(2) obtain sufficient data for model training;
(3) preprocess the data;
(4) apply feature engineering as needed;
(5) optimize and test the model.

ML model selection is a vital step in the framework and may occur at various points. The specifics of the selection process depend completely on the goals of the project. In the PSC field, there are opportunities for both numerical prediction models, as in time-series forecasting for PSC stability prediction, or classification algorithms, such as to classify types of defects in a failure analysis task. As stated earlier, one common approach is to train several models on the same data (step 5) and compare their predictive performance.[36−38] Another option is to choose an algorithm, or class of algorithms, immediately after determining the specific materials question of interest (step 1). At this stage, model

https://doi.org/10.1021/acs.jpclett.1c01961
J. Phys. Chem. Lett. 2021, 12, 7866−7877

selection requires a thorough understanding of the input data and of the desired output.

Step (2) underscores the importance of high-throughput systems, which involve drastically shortened experiment times. Conducting such research often necessitates automated data acquisition. Therefore, Figure 1 includes examples of suitable characterization methods for efficient collection of large amounts of data. At the compositional level, X-ray diffraction (XRD) and ultraviolet−visible spectroscopy (UV−vis) are well-known techniques yielding key crystallographic and optical properties that then inform compositional selection.[39,40] At the perovskite material level, PL spectroscopy is a standard method to quickly characterize charge carrier dynamics in thin films without requiring a full device or electrical contacts.[41] Analysis of subtle patterns in spectroscopic PL or PL lifetime races is nearly impossible for the human eye but is attainable using computer vision, enabling deeper physical insight. Device-level techniques include optical transmittance and reflectance, optical imaging, and current−voltage ($I-V$) measurements. These techniques are leveraged to calculate device parameters such as fill factor (FF), maximum power point, and power conversion efficiency (η).[42] The creation of shared data repositories can ease the burden of extensive data collection, supporting rapid ML model training without the need for time-intensive experimentation. In fact, large sets of crowd-sourced data have already been utilized for ML training in materials science and chemistry, for example to predict reaction outcomes.[43] Similarly, the baseline examples that we show in later sections rely on data shared by members of the perovskite scientific community, illustrating the importance of collaborative efforts. Searchable, centralized databases will streamline model training further and likewise depend heavily on community contributions.

The creation of shared data repositories can ease the burden of extensive data collection, supporting rapid ML model training without the need for time-intensive experimentation.

Steps (3) and (4) involve data preprocessing and feature engineering and are crucial in many ML pipelines. These steps often include data normalization, data augmentation, and feature selection and transformation.[27,28] The final step of the process is (5) model optimization and testing, which includes hyperparameter tuning and evaluation on an unseen test set. We provide detailed overviews for how this paradigm applies to each stage of PSC development in this Perspective, illustrated using specific examples from the literature and our own work. Additionally, we explore the question of establishing correlations between easily acquired characterization data and fundamental properties, which yields useful baseline indicators for promising performance. ML is particularly well-suited for this task because of its proven ability to quickly extract features from data that humans cannot, a capability that has a variety of practical applications (for example, to analyze medical scans).[44] In the following sections, we discuss the question of correlations in detail and provide our outlook for how ML approaches at each stage can be holistically integrated.

ML for Compositional Screening. The central goal at the composition level is to efficiently screen for material compositions suitable for PV that are thermodynamically and chemically stable. Traditional material discovery approaches of trial-and-error and incremental improvement are not sufficient given the vast parameter space of hybrid perovskite chemical compositions. Simulation-based approaches such as DFT are an option to narrow the parameter space but are slow and computationally expensive. One common approach is combining simulation data with ML, which extracts figures of merit such as intrinsic dielectric breakdown strength,[36] band gap,[45] and thermodynamic stability[37] from the data. In these studies, the ML component generates additional information from a time-consuming DFT process. Alternatively, ML algorithms can predict parameters such as lattice constants, octahedral angle, and band gaps from atomic data alone.[32,46−49] Carefully selected models also operate significantly faster and with less computational cost than DFT.

Another important facet of the compositional screening stage is high-throughput or automated experiments, which enable rapid perovskite synthesis and data acquisition.[50] Fewer studies have coupled ML models with automated experimentation, developing fully autonomous, self-driving laboratories.[40,51,52] Further work centered on high-throughput structural characterization and analysis is needed to establish a coherent pipeline from broad materials down-selection to precise compositional tuning and experimental validation.

Finally, we emphasize the importance of large data repositories, which greatly expedite the compositional screening process. ML models require sufficient data to make high-quality predictions, and the amount of data needed generally increases with model complexity. In practice, the data size N needs to satisfy the condition for a good generalization: $N \approx O(W/\varepsilon)$, where W denotes the number of free parameters in a model or neural network and ε signifies the fraction of permitted test error.[24] Many of the ML studies cited here used data from large online repositories, including databases from the National Institute of Standards and Technology (NIST) and the National Renewable Energy Laboratory (NREL).[53] The existence of these repositories is one reason why current ML work in the PSC scientific community is largely limited to compositional screening, as the material and device level lack a similar data-sharing infrastructure and reporting protocols to support the effort.[1] In this Perspective, we therefore focus on the underdeveloped material and device levels, highlighting several fundamental research questions that may benefit from an ML-based approach.

ML for Material Development. The perovskite absorber material is the focus of a large proportion of investigations in the PSC field. Key research questions at this level involve finding optimal fabrication parameters while understanding and preventing material degradation. Numerous studies have probed perovskite degradation patterns and mechanisms under various environmental conditions,[54−59] and there are many Review articles[42,60−62] exploring this topic at both the material and full device stages. However, far fewer investigations apply high-throughput experimentation or ML models to perovskite material design, which can reduce laboratory hours and inform avenues of future study. Additionally, analysis of model features and weights can present unique physical insight.[38]

Like fabrication parameter selection, stability testing represents another major bottleneck in the PSC development

https://doi.org/10.1021/acs.jpclett.1c01961
J. Phys. Chem. Lett. 2021, 12, 7866−7877

process, especially as devices near commercialization and the perovskite absorber stability is thoroughly vetted. Once trained, predictive ML models overcome this holdup by drastically reducing the time required to assess degradation in perovskite materials. In recent work, Stoddard et al. used a linear regression model to forecast perovskite decay time under variable environmental conditions and achieved an average error of 12.8% while decreasing the required testing time by 1 order of magnitude.[63] Such single-point prediction, regression-based models have great utility and can achieve low error even with a limited data set. Larger data sets enable the use of complex models with more powerful predictive ability, ultimately supporting long-term PSC performance forecasting that tracks variations in figures of merit over time periods of interest extending hours, days, or weeks ahead.

Larger data sets enable the use of complex models with more powerful predictive ability, ultimately supporting long-term PSC performance forecasting that tracks variations in figures of merit over time periods of interest extending hours, days, or weeks ahead.

To demonstrate the power of a high-complexity model, we implement an echo state network (ESN), one class of RNN, to predict the PL intensity from a triple-cation $Cs_{0.05}FA_{0.79}MA_{0.16}Pb(I_{0.83}Br_{0.17})_3$ thin film (Figure 2) while applying the five-step roadmap introduced earlier. The film was subject to cycling from 0 to 70% relative humidity (rH), while the absolute light emission was measured every 15 s. Environmental humidity is known to dynamically influence the PL response of perovskite films as well as their underlying phase and structure.[57,60,64] The rH-dependent curve (black dots in Figure 2) shows clear cycling of the PL intensity in response to the 0−70% humidity cycles. Because this PL response represents a dynamic system containing physical information from the triple-cation film, there is strong indication that the response could be predictable, potentially using ML. Analogously, the PL response depends on both the ambient humidity at the precise moment of the measurement and on the changes that have already happened within the material—specifically, the partially reversible hydration of the perovskite film over time.[57] In other words, the material has "memory" of what has been experienced due to the history of the environment. We choose ESN for this task, as this model maps an input property (rH) to an output (PL), where the output is a function of both the input property and the intrinsic historical state of the material itself.[65] This "historical state" manifests in the data through unique patterns over time, rather than a completely linear on−off PL response, and thus, an algorithm with a "memory" is desirable.

When implementing complex ML algorithms such as neural networks, it is vital to use sufficient training data. Our raw data set consists of five 4-h runs on samples with identical compositions, where each run contains three humidity cycles and 240 data points. The PL traits observed are unique to the perovskites, following the on−off cycle of rH. Measurements under identical conditions were performed on a GaAs control solar cell, and no significant changes in the PL signal were observed (not shown). We augment this data using a linear

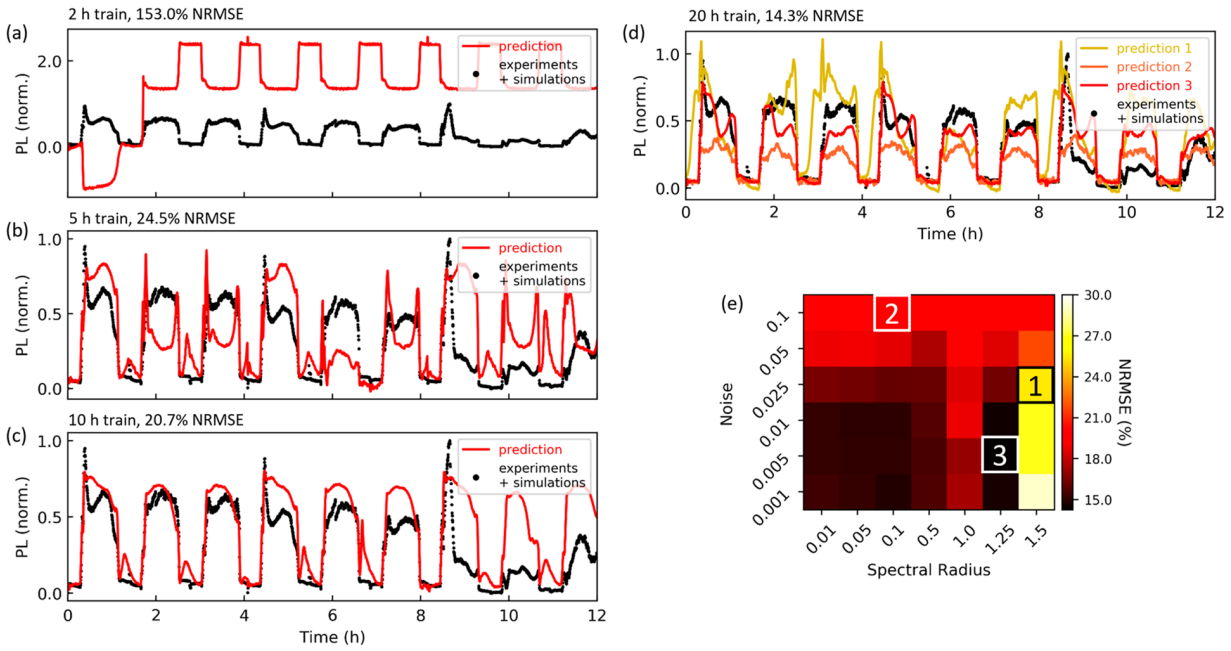


Figure 2. Predicting the optical response of perovskite thin films. Echo state network (ESN) predictions of photoluminescence (PL) for a metal halide perovskite ($Cs_{0.05}FA_{0.79}MA_{0.16}Pb(I_{0.83}Br_{0.17})_3$) thin film after (a) 2 h, (b) 5 h, (c) 10 h, and (d) 20 h of training data. Augmented triple-cation data (black dots) are compared with the ESN prediction (red line). As the model receives longer lengths of training data, its predictions stabilize, and the normalized root-mean-square error (NRMSE) decreases. Predictions 1, 2, and 3 in (d) are obtained after 20 h of training data with various ESN hyperparameters (noise and spectral radius). (e) NRMSE for the three predictions shown in (d) as hyperparameters are optimized using a grid-search approach. The minimum value found is 14.3%.

https://doi.org/10.1021/acs.jpclett.1c01961
J. Phys. Chem. Lett. 2021, 12, 7866−7877

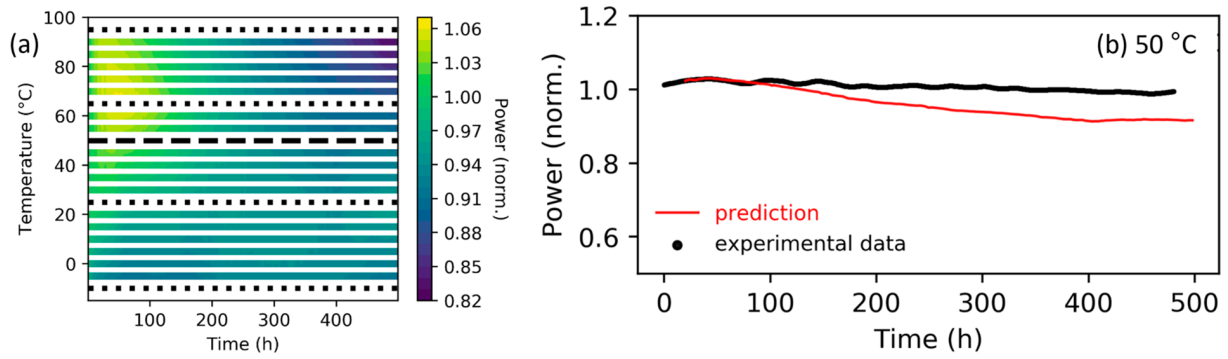


Figure 3. Predicting power output for MAPI devices at unseen temperatures. (a) Augmented data set of 500-h runs at temperatures ranging from −10 to 95 °C, tracking power output over time for the $MAPbI_3$ PV device. Black lines indicate ground truth experimental data (from reference 77), and colored lines indicate augmented data. Four experimental data sets (black, dotted) were used to generate the augmented training data, while the final set (black, dashed) was used for testing. (b) LSTM prediction (red) for the experimental data run at 50 °C (black). The data is normalized by its first point.

interpolation method[66,67] to expand the data set to 60 runs (14 400 points) total, which are then stitched together randomly to generate one coherent time-series. Note that such an augmentation routine assumes that the five collected runs are a fair representation of the underlying material behavior, assumes that they encompass sample-to-sample variance, and neglects effects from long-term degradation. The key assumption here is that the physical properties (e.g., PL and the state of the material) are a continuous function of parameters in the feature space (e.g., rH); thus, similar parameters should yield alike output. This is a critical conjecture that applies to the use of most NN models to approximate solutions to complex problems. Researchers in materials science and beyond have adopted data augmentation methods using a similar strategy, as in the XRD peak-shifting example discussed toward the beginning of this Perspective.[27,28,68] For more information on data augmentation, we direct readers to several references that discuss this topic at length.[26,66,67,69] We then supply the ESN with time-series training data tracking rH and PL from 2 h (Figure 2a), 5 h (Figure 2b), 10 h (Figure 2c), and 20 h (Figure 2d) of the expanded data set, holding out the remaining hours for testing. At 2 h, the ESN diverges from the data, and the NRMSE is over 100%. As the model receives additional training data, the NRMSE decreases steadily, and the prediction (red line) visually shows the improved fit. By 20 h, the ESN has learned details about the data sequence, such as the horizontal, low-PL intensity regions between each humidity cycle, the slope of the PL increases and decreases, and the spiked responses that occur at the beginning of some rH cycles. We then optimize the 20-h model, setting the network's noise and spectral radius through a grid-search approach (Figure 2e). While the exact hyperparameters vary between different algorithms, this tuning step is crucial. The purposely poorly selected hyperparameters (Figure 2d, Predictions 1 and 2) lead to high prediction errors of 25.5 and 20.1%, respectively. By providing sufficient training data and tuning our model hyperparameters, we ultimately achieve an NRMSE of 14.3% over a 12-h prediction window (Figure 2d, Prediction 3). Demonstrative code for this example, and all other example projects, is available on our GitHub repository at https://github.com/mgsrivastava/ML-perovskites. The repository includes both the Python code necessary to replicate our analysis and a detailed outline of all computational steps written in plain English.

We envision further extensions of ML to several facets of the perovskite material stage, where it serves as a powerful tool for accelerated stability testing and rational fabrication of the perovskite layer. Future studies may involve integration with autonomous laboratories to streamline the workflow from compositional selection to thin film fabrication to long-term material stability testing. For the latter, the end goal is a model that generalizes to unseen combinations of environmental conditions. ML can also enhance imaging-based investigations into degradation and defects at the material level, an exciting application which we discuss in-depth in a later section.

ML for Device Development. Much of the work and key questions at the device-level mirror those at the material level. However, the degree of complexity multiplies with each added layer. Understanding degradation under standard operating conditions and optimizing layer fabrication are essential goals. Therefore, work at the device stage typically involves engineering interfaces and architectures[70−73] and investigating PSC stability (under stressors such as light,[74,75] temperature,[76,77] bias,[62] oxygen,[78] and humidity[79]). As with the material level, ML can quickly determine effective processing conditions for various layers.[34] Furthermore, by employing large shared data repositories that include fabrication methodology and *in situ* chemical information, ML techniques at the material and device levels can incorporate additional training features and potentially yield more accurate metrics to indicate degradation far in advance. Clarifying which fundamental properties and environmental conditions exert the most influence over long-term performance will, in turn, lead to more targeted compositional engineering.

As an example, we use data from reference 77 to predict temperature-dependent power output for MAPI cells at unseen temperatures, effectively using ML forecasting to interpolate to untested environmental conditions. The raw experimental data consists of five 500-h runs at −10, 20, 50, 65, and 95 °C, which track variations in the device power. Each run contains one data point per hour, for a total of 2500 points. We set aside one run (at 50 °C) to serve as our test set, leaving only 2000 points for training. To enlarge this sparse data set, we apply a linear interpolation technique to generate runs for temperatures between −10 and 95 °C, with a 5 °C step size. The resulting data set consists of 22 temperatures and 11 000 individual data points. Figure 3a shows a contour plot of the augmented data set, with the ground truth experimental data marked with black

https://doi.org/10.1021/acs.jpclett.1c01961
J. Phys. Chem. Lett. 2021, 12, 7866−7877

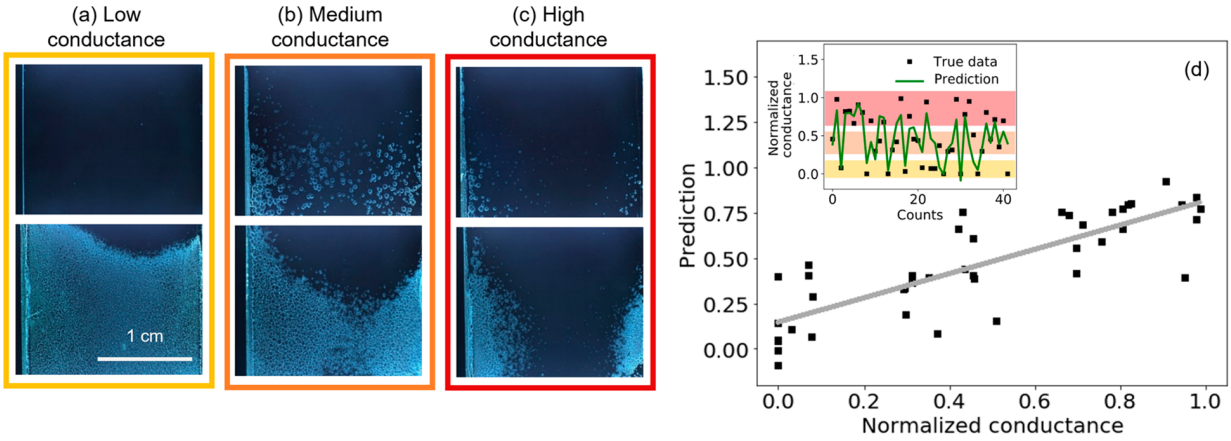


Figure 4. Correlating electrical conductance with dark-field images. A CNN model is trained to predict conductance of spiro-OMeTAD thin films based on their dark-field images. Six examples of DF images are shown for samples with (a) low, (b) medium, and (c) high conductance values, which are normalized between 0 and 1 for simplicity without the loss of generality. No visually distinguishable patterns can be found in these images with different conductance value ranges. (d) The predicted conductance versus the true values using the trained CNN model. Inset shows how these predicted values compare with the true value for each dark-field image in the test data set.

lines. We again use augmentation as a proof-of-concept in the ML workflow and assume that the experimental runs encompass property (power) data over the parameter (temperature) space, such that we can linearly interpolate to obtain sufficient training data and improve prediction accuracy.[69] In practice, we need experimental data with more finely resolved temperature steps to ensure that the data captures the full range of device behavior before augmentation. We then train a long short-term memory (LSTM) model—another class of RNN—on the expanded data set and attempt to predict cell behavior at the unseen temperature. LSTM processes data sequentially to learn trends and, like the ESN, retains a history of the data over time. Practically, the models differ in that the LSTM is more effective in learning long-term trends but is also more computationally expensive. The model architecture consists of an LSTM with five output units, a leaky rectified linear unit (ReLU) as the activation function, a dropout layer to reduce overfitting, and a dense layer to output the prediction.[80]

Using LSTM, we predict the full 500-h power trace for the MAPI device at 50 °C with an NRMSE of only 5.5%. We show the visual fit of our model in Figure 3b, which compares the LSTM prediction (red line) to ground truth experimental data (black dots). Because the model uses the 20 previous hours to generate a prediction, the red line begins at time = 20 h. The LSTM learns the general trend of decreasing power and forecasts the temperature-dependent behavior at 50 °C almost perfectly for the first 70+ predicted hours. Even as the prediction diverges over the course of hundreds of hours, the overall NRMSE remains under 6%. Our example here illustrates the potential that lies ahead in using ML for time-series forecasting of device performance. We envision extensions to additional test conditions (and combinations of conditions) to attain a systematic understanding of how environmental stressors impact PSC over even longer time periods (years to decades).

Visualizing Correlations through Imaging. Finally, we discuss ML for spatially resolved characterization and imaging of perovskites, which is highly relevant to both the material and device stages of development. At the nanoscale, spatially resolved information is commonly used for defect analysis and investigations of structure−property relationships. Here, ML can illustrate powerful connections between global device performance and morphological or structural information such as surface roughness, grain inhomogeneities, and crystallinity. At the micro- and macroscopic levels, imaging methods have been extensively applied in inorganic PV to quickly diagnose *where* device drawbacks occur (e.g., cracks on encapsulation, loss of electrical contact). Characterization methods exist for several key device figures of merit, including carrier lifetime, band gap, external voltage, reflectance, and resistance.[81] These methods involve both scanning-based mapping and full-area capture techniques. Such approaches yield large, information-rich data sets, which are difficult for humans to parse but are efficiently analyzed using ML and multivariate statistics. There are then tremendous opportunities for combining ML methodology with 2D imaging/characterization techniques, which have garnered increasing attention lately in perovskite-related research.[82] Pairing ML with optical microscopy studies is another potentially valuable approach, as machine vision analyzes images faster than humans and quickly resolves trends. Recent work has used image recognition to automate the identification of perovskite single-crystal formation[83] and to classify defects in spiro-OMeTAD thin films.[84] Both of these studies use a 2D CNN as a classifier. As stated earlier, CNNs perform sophisticated image analysis tasks and can establish compelling connections between images and key figures of merit while extracting hidden trends. Compared to a fully connected deep neural network, CNN has a much smaller number of free parameters available for adjustment and is more efficient in capturing key features in an image.[24]

While CNN is commonly used for classification, it can work equally well for regression to produce a numerical prediction. Nevertheless, CNN-based regression has rarely been attempted on hybrid perovskite materials. To illustrate the power of CNN in capturing underlying trends from images that are all but intractable for humans, we construct a CNN model to predict electrical conductance using dark-field (DF) images of spin-coated and thermally annealed spiro-OMeTAD thin films (a common HTM for PSCs). There is an essential underlying correlation between a material's electrical conductivity and optical properties *via* the dielectric function,[85] and consequently, this is a potentially learnable problem. We use data from a previous reference, where the DF images were acquired

https://doi.org/10.1021/acs.jpclett.1c01961
J. Phys. Chem. Lett. 2021, 12, 7866−7877

in an automated sample synthesis and characterization experiment pipeline, and the conductance was measured using the traditional four-probe approach.[34] In total, we separate 210 original DF images into training and testing sets with an 80:20% split. We further divide the training data into train and validation sets with another 80:20% split. Each image has seven corresponding conductance measurements acquired at different locations using $I-V$ data. We average all measurements into one value per image and then normalize the data between 0 and 1. During the training process, image data is augmented through random rotations, width and height shifting, flipping, and zooming to ensure the model is fed with representative images. Our CNN architecture consists of three 2D convolution layers ($3 \times 3 \times 32$, $3 \times 3 \times 64$, and $3 \times 3 \times 64$, respectively), each of which is followed by a max pooling layer (2×2). To mitigate overfitting, the convolutional layers are L2 regularized. The network ends with two fully connected layers (100 and 64 in size, respectively) prior to the output layer, which predicts a single conductance value for each image. The weights are updated using an ADAM optimizer at a learning rate of 0.001.

Figure 4a–c shows examples of DF images on spiro-OMeTAD thin films. We emphasize that our close examination of all 210 DF images indicates that they do not possess visually distinguishable patterns for samples with high, medium, or low conductance. This renders impossible even qualitative visual correlation with conductance value using human visual analysis alone. Yet, as Figure 4d shows, the CNN model can analyze the underlying features of the images and successfully predicts the conductance values with a reasonable uncertainty (20% NRMSE). Considering the limited amount of image data tested here, the model has notably acquired the "knowledge" or trend of the correlation between the DF images and electrical conductance to an extent that would otherwise not be possible by human visual analysis. Our example here demonstrates the remarkable potentiality of using ML-based analysis to identify correlations hidden within microscope images and the opportunity that lies ahead in PSC imaging/spatially resolved characterization.

While ML serves as a powerful toolkit applicable to many problems in the PSC design process, we emphasize that it is not applicable to *every* problem. Researchers must carefully weigh the benefits of an ML approach and acknowledge the limitations to their results. The first item to consider is data sparsity. As mentioned in previous sections, ML models require adequate data for appropriate training, which is often a major precluding factor in materials research. In the perovskite field, we face a lack of large, well-represented, and cleaned experimental databases, which cover the full parameter space of PSC materials and devices, including more than simple compositional information (i.e., morphological information, fabrication methods, and aging conditions). Automated experimentation and physics-driven ML are two promising pathways to address this issue. A standard supplemental data file that includes the previously mentioned parameters could also initiate the data-sharing infrastructure that is currently lacking. Researchers should also emphasize characterization techniques that yield large, quickly acquired data sets. For example, in imaging, optical microscopy methods are better suited for use in ML rather than advanced, time-consuming techniques such as electron microscopy. Additionally, models will only provide high-quality predictions if they are trained on high-quality data. Biased data collection and improper preprocessing are common pitfalls in practical ML implementation and will negatively affect the validity of the resulting model. It is equally important to train NN on negative examples—meaning material and device data showing poor performance—as well as positive, but negative examples are seldom reported in the literature. We again direct readers to recent papers discussing specific techniques and best practices for applying ML to research in materials science.[29−31] Ideally, NN and other models should be built to be robust to common data issues, such as excessive noise or biased reporting. Some algorithms and techniques have been widely adopted for these purposes, such as regularization penalty terms, drop-out layers and other structural simplification, and Hessian-based pruning approaches.[24]

The information gained from ML also often presents limitations. Models that predict promising perovskite compositions (through band gap or other key figures of merit) do not guarantee that these compositions will produce optimal devices nor do they account for ease of material synthesis. Instead, ML at the compositional level narrows the parameter space for more focused research at the material and device levels. By a similar token, optimizing parameters for a single device layer does not ensure optimized behavior in a complete device, and experimental verification is crucial. We also note that ML models developed using full devices will not inherently disentangle effects from each layer but may be trained to do so if features from separate layers are incorporated. Finally, advanced deep learning techniques produce models with convoluted architectures and parameters that tend to lack simple physical interpretation. This reflects the general trade-off between interpretability and prediction power seen in many ML algorithms, though analysis of global model predictions and use of a physics-driven model design may yield greater interpretability. Before applying ML, researchers must understand these and other limitations and develop a careful plan for data acquisition, management, and processing.

In this Perspective, we outlined an ML roadmap for PSC development and demonstrated the principles using baseline models trained on currently available data. We included the steps from start-to-finish for a given perovskite ML project, including (1) identify the material question of interest, (2) obtain sufficient data for model training, (3) preprocess the data, (4) apply feature engineering as needed, and (5) optimize and test the model. We presented three examples showcasing this five-step framework in action using MAPI and triple-cation perovskites. First, we applied ML to time-series forecasting and predict PL from thin films with an NRMSE of 14.3%. Second, we used ML to extrapolate device performance to unseen temperature conditions (NRMSE = 5.5%). Finally, we extract film conductivity from dark-field images alone (NRMSE = 19%). These example projects demonstrate the versatility of ML and its broad applicability to the perovskite field, and we refer readers to our GitHub repository to view all associated data and code, along with plain English descriptions of all computational steps. Future ML models will identify leading factors for perovskite degradation, integrate compositional discovery with long-term device stability, and leverage imaging data sets to understand fundamental properties. Our framework is also applicable to fields involving other renewable energy materials, such as thermoelectrics, batteries, and catalysis.

ML, coupled with high-throughput or automated experimentation, offers a new paradigm for conducting research in

https://doi.org/10.1021/acs.jpclett.1c01961
J. Phys. Chem. Lett. 2021, 12, 7866−7877

Future ML models will identify leading factors for perovskite degradation, integrate compositional discovery with long-term device stability, and leverage imaging data sets to understand fundamental properties.

materials science and chemistry, with greater efficiency and more in-depth physical insight than traditional trial-and-error methods. Materials researchers in diverse fields ranging from phase mapping[86,87] to quantum dots[88] to metallic glasses[89] to photonics[90,91] have begun transitioning to this novel approach. In the perovskite scientific community, we propose a completely integrated ML pipeline that encompasses all three levels of PSC development (composition, material, and device) and includes screening, fabrication, characterization, and stability testing. Such a paradigm would include multiple models to establish feedback loops linking compositional discovery with long-term material and device stability. Use of extensive data repositories, standard supplemental data files, and physics-based ML strategies will facilitate tuning and testing steps for these models. There is a particular need for repositories of standardized perovskite stability testing data, which could inform sophisticated time-series forecasting models that extrapolate to unseen combinations of environmental stressors. This PSC pipeline takes advantage of our current technological capabilities in ML, data science, and autonomous experimentation and presents a rational pathway to accelerate the PSC commercialization process.

AUTHOR INFORMATION

Corresponding Author

Marina S. Leite − *Department of Materials Science and Engineering, University of California, Davis, California 95616, United States;* orcid.org/0000-0003-4888-8195; Email: mleite@ucdavis.edu

Authors

Meghna Srivastava − *Department of Materials Science and Engineering, University of California, Davis, California 95616, United States;* orcid.org/0000-0002-7469-8769

John M. Howard − *Institute for Research in Electronics and Applied Physics, University of Maryland, College Park, Maryland 20742, United States;* orcid.org/0000-0002-3990-5478

Tao Gong − *Department of Materials Science and Engineering, University of California, Davis, California 95616, United States; Department of Electrical and Computer Engineering, University of California, Davis, California 95616, United States*

Mariama Rebello Sousa Dias − *Department of Physics, University of Richmond, Richmond, Virginia 23173, United States;* orcid.org/0000-0001-9090-0018

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jpclett.1c01961

Notes

The authors declare no competing financial interest.

Data Availability Statement: All data and code used to develop the proof-of-concept example projects in this paper are available at https://github.com/mgsrivastava/ML-perovskites. Code is written in Python and presented in Jupyter notebook format.

Biographies

Meghna Srivastava received her B.S. degree in Materials Science & Engineering from Cornell University and is currently a Ph.D. Student at the University of California, Davis. Her research investigates degradation in perovskite photovoltaics and predicts their performance over time using machine learning models.

John M. Howard received his Ph.D. in Materials Science and Engineering from the University of Maryland, College Park, in 2020. His research focuses on understanding the causes of environmentally induced optical and electrical dynamics in metal halide perovskites using *in situ* microscopy methods and machine learning methods.

Tao Gong is a Postdoc researcher at the University of California, Davis. He obtained his Ph.D. degree in electrical engineering at the University of Maryland, College Park, in 2016 and worked as an engineer at MathWorks thereafter. His current research interest covers plasmonic and nanophotonic materials, perovskite photovoltaics, and the Casimir effect.

Mariama Rebello Sousa Dias is an Assistant Professor in the Physics Department at the University of Richmond. She received her Ph.D. from The Federal University of Sao Carlos, Brazil with visiting researcher appointments at the Ohio University, The Free University of Berlin, and The University of Wuerzburg. Before arriving at UR, she was a postdoctoral researcher at the University of Maryland holding the Schlumberger Faculty for the Future Fellowship. The Dias lab currently researches topics that focus on metallic, oxide, and semiconductor nanostructures with potential applications in optics, plasmonics, nanoelectronics, and energy harvesting devices.

Marina S. Leite is an Associate Professor in the Department of Materials Science & Engineering at the University of California, Davis. She received her Ph.D. degree in Physics at Campinas State University in Brazil, followed by a postdoc at Caltech. Her group is engaged in fundamental and applied research on functional materials for renewable energy, including hybrid perovskites for photovoltaics and for photonics.

ACKNOWLEDGMENTS

M.S.L. thanks the financial support from the National Science Foundation (ECCS awards #20-23974 and #16-10833) and UC Davis. M.S. acknowledges funding support from the National Science Foundation Graduate Research Fellowships Program, the UC Davis' Towards Outstanding Postgraduate Students (TOPS), the UC Davis' Graduate Assistance in Areas of National Need (GAANN), and the UC Davis' GEMS Fellowships. M.R.S.D. thanks the financial support of the School of Arts and Sciences, University of Richmond.

REFERENCES

(1) Khenkin, M. V.; Katz, E. A.; Abate, A.; Bardizza, G.; Berry, J. J.; Brabec, C.; Brunetti, F.; Bulović, V.; Burlingame, Q.; Di Carlo, A.; et al. Consensus Statement for Stability Assessment and Reporting for Perovskite Photovoltaics Based on ISOS Procedures. *Nat. Energy* **2020**, *5*, 35−49.

(2) Green, M.; Dunlop, E.; Hohl-Ebinger, J.; Yoshita, M.; Kopidakis, N.; Hao, X. Solar Cell Efficiency Tables (Version 57). *Prog. Photovoltaics* **2021**, *29*, 3−15.

(3) Jena, A. K.; Kulkarni, A.; Miyasaka, T. Halide Perovskite Photovoltaics: Background, Status, and Future Prospects. *Chem. Rev.* **2019**, *119*, 3036−3103.
(4) Wang, Q.; Phung, N.; Di Girolamo, D.; Vivo, P.; Abate, A. Enhancement in Lifespan of Halide Perovskite Solar Cells. *Energy Environ. Sci.* **2019**, *12*, 865−886.
(5) Song, Z.; McElvany, C. L.; Phillips, A. B.; Celik, I.; Krantz, P. W.; Watthage, S. C.; Liyanage, G. K.; Apul, D.; Heben, M. J. A Technoeconomic Analysis of Perovskite Solar Module Manufacturing with Low-Cost Materials and Techniques. *Energy Environ. Sci.* **2017**, *10*, 1297−1305.
(6) Sofia, S. E.; Wang, H.; Bruno, A.; Cruz-Campa, J. L.; Buonassisi, T.; Peters, I. M. Roadmap for Cost-Effective, Commercially-Viable Perovskite Silicon Tandems for the Current and Future PV Market. *Sustainable Energy Fuels* **2020**, *4*, 852−862.
(7) Jeon, N. J.; Noh, J. H.; Yang, W. S.; Kim, Y. C.; Ryu, S.; Seo, J.; Seok, S. I. Compositional Engineering of Perovskite Materials for High-Performance Solar Cells. *Nature* **2015**, *517*, 476−480.
(8) Sharenko, A.; Toney, M. F. Relationships Between Lead Halide Perovskite Thin-Film Fabrication, Morphology, and Performance in Solar Cells. *J. Am. Chem. Soc.* **2016**, *138*, 463−470.
(9) Petrus, M. L.; Schutt, K.; Sirtl, M. T.; Hutter, E. M.; Closs, A. C.; Ball, J. M.; Bijleveld, J. C.; Petrozza, A.; Bein, T.; Dingemans, T. J.; et al. New Generation Hole Transporting Materials for Perovskite Solar Cells: Amide-Based Small-Molecules with Nonconjugated Backbones. *Adv. Energy Mater.* **2018**, *8*, 1801605.
(10) Bella, F.; Griffini, G.; Correa-Baena, J.; Saracco, G.; Graetzel, M.; Hagfeldt, A.; Turri, S.; Gerbaldi, C. Improving Efficiency and Stability of Perovskite Solar Cells with Photocurable Fluoropolymers. *Science* **2016**, *354*, 203−206.
(11) Arora, N.; Dar, M. I.; Hinderhofer, A.; Pellet, N.; Schreiber, F.; Zakeeruddin, S. M.; Grätzel, M. Perovskite Solar Cells with CuSCN Hole Extraction Layers Yield Stabilized Efficiencies Greater Than 20%. *Science* **2017**, *358*, 768−771.
(12) Ma, X.; Tao, Z.; Wang, Y.; Yu, H.; Wang, Y. Long Short-Term Memory Neural Network for Traffic Speed Prediction Using Remote Microwave Sensor Data. *Transp. Res. Part C Emerg. Technol.* **2015**, *54*, 187−197.
(13) Titano, J. J.; Badgeley, M.; Schefflein, J.; Pain, M.; Su, A.; Cai, M.; Swinburne, N.; Zech, J.; Kim, J.; Bederson, J.; et al. Automated Deep-Neural-Network Surveillance of Cranial Images for Acute Neurologic Events. *Nat. Med.* **2018**, *24*, 1337−1341.
(14) Liu, D.; Wang, J.; Wang, H. Short-Term Wind Speed Forecasting Based on Spectral Clustering and Optimised Echo State Networks. *Renewable Energy* **2015**, *78*, 599−608.
(15) Sun, Y. C.; Szucs, G.; Brandt, A. R. Solar PV Output Prediction from Video Streams Using Convolutional Neural Networks. *Energy Environ. Sci.* **2018**, *11*, 1811−1818.
(16) Ruffing, S. M.; Venayagamoorthy, G. K. Short to Medium Range Time Series Prediction of Solar Irradiance Using an Echo State Network. *2009 15th International Conference on Intelligent System Applications to Power Systems* **2009**, 1−6.
(17) Morando, S.; Jemei, S.; Gouriveau, R.; Zerhouni, N.; Hissel, D. Fuel Cells Remaining Useful Lifetime Forecasting Using Echo State Network. *2014 IEEE Vehicle Power and Propulsion Conference (VPPC)* **2014**, 1−6.
(18) Liu, D.; Xie, W.; Liao, H.; Peng, Y. An Integrated Probabilistic Approach to Lithium-Ion Battery Remaining Useful Life Estimation. *IEEE Trans. Instrum. Meas.* **2015**, *64*, 660−670.
(19) Severson, K. A.; Attia, P. M.; Jin, N.; Perkins, N.; Jiang, B.; Yang, Z.; Chen, M. H.; Aykol, M.; Herring, P. K.; Fraggedakis, D.; et al. Data-Driven Prediction of Battery Cycle Life Before Capacity Degradation. *Nat. Energy* **2019**, *4*, 383−391.
(20) Zhang, Y.; Tang, Q.; Zhang, Y.; Wang, J.; Stimming, U.; Lee, A. A. Identifying Degradation Patterns of Lithium Ion Batteries from Impedance Spectroscopy Using Machine Learning. *Nat. Commun.* **2020**, *11*, 1706.
(21) Alpaydin, E. *Introduction to Machine Learning*, 4th ed.; The MIT Press: Cambridge, MA, 2020.
(22) Kotsiantis, S. B. Supervised Machine Learning: A Review of Classification Techniques. *Informatica* **2007**, *31*, 249−268.
(23) Hinton, G.; Sejnowski, T. J. *Unsupervised Learning: Foundations of Neural Computation*; The MIT Press: Cambridge, MA, 1999.
(24) Haykin, S. *Neural Networks and Learning Machines*, 3rd ed.; Pearson Prentice Hall: Upper Saddle River, NJ, 2008.
(25) Yamashita, R.; Nishio, M.; Do, R. K. G.; Togashi, K. Convolutional Neural Networks: An Overview and Application in Radiology. *Insights Imaging* **2018**, *9*, 611−629.
(26) DeVries, T.; Taylor, G. W. Dataset Augmentation in Feature Space; arXiv, 2017.
(27) Oviedo, F.; Ren, Z.; Sun, S.; Settens, C.; Liu, Z.; Hartono, N. T. P.; Ramasamy, S.; DeCost, B. L.; Tian, S. I. P.; Romano, G.; et al. Fast and Interpretable Classification of Small X-ray Diffraction Datasets Using Data Augmentation and Deep Neural Networks. *npj Comput. Mater.* **2019**, *5*, 60.
(28) Howard, J. M.; Wang, Q.; Lee, E.; Lahoti, R.; Gong, T.; Srivastava, M.; Abate, A.; Leite, M. S. *Quantitative Predictions of Photo-Emission Dynamics in Metal Halide Perovskites via Machine Learning*; arXiv, 2020.
(29) Butler, K. T.; Davies, D. W.; Cartwright, H.; Isayev, O.; Walsh, A. Machine Learning for Molecular and Materials Science. *Nature* **2018**, *559*, 547−555.
(30) Wang, A. Y.-T.; Murdock, R. J.; Kauwe, S. K.; Oliynyk, A. O.; Gurlo, A.; Brgoch, J.; Persson, K. A.; Sparks, T. D. Machine Learning for Materials Scientists: An Introductory Guide Toward Best Practices. *Chem. Mater.* **2020**, *32*, 4954−4965.
(31) Chen, C.; Zuo, Y.; Ye, W.; Li, X.; Deng, Z.; Ong, S. P. A Critical Review of Machine Learning of Energy Materials. *Adv. Energy Mater.* **2020**, *10*, 1903242.
(32) Saidi, W. A.; Shadid, W.; Castelli, I. E. Machine-Learning Structural and Electronic Properties of Metal Halide Perovskites Using a Hierarchical Convolutional Neural Network. *npj Comput. Mater.* **2020**, *6*, 36.
(33) Howard, J. M.; Tennyson, E. M.; Neves, B. R. A.; Leite, M. S. Machine Learning for Perovskites' Reap-Rest-Recovery Cycle. *Joule* **2019**, *3*, 325−337.
(34) MacLeod, B. P.; Parlane, F. G. L.; Morrissey, T. D.; Hase, F.; Roch, L. M.; Dettelbach, K. E.; Moreira, R.; Yunker, L. P. E.; Rooney, M. B.; Deeth, J. R.; et al. Self-Driving Laboratory for Accelerated Discovery of Thin-Film Materials. *Sci. Adv.* **2020**, *6*, No. eaaz8867.
(35) Lira-Cantú, M. Perovskite Solar Cells: Stability Lies at Interfaces. *Nat. Energy* **2017**, *2*, 17115.
(36) Kim, C.; Pilania, G.; Ramprasad, R. Machine Learning Assisted Predictions of Intrinsic Dielectric Breakdown Strength of ABX_3 Perovskites. *J. Phys. Chem. C* **2016**, *120*, 14575−14580.
(37) Schmidt, J.; Shi, J.; Borlido, P.; Chen, L.; Botti, S.; Marques, M. A. L. Predicting the Thermodynamic Stability of Solids Combining Density Functional Theory and Machine Learning. *Chem. Mater.* **2017**, *29*, 5090−5103.
(38) Yu, Y.; Tan, X.; Ning, S.; Wu, Y. Machine Learning for Understanding Compatibility of Organic−Inorganic Hybrid Perovskites with Post-Treatment Amines. *ACS Energy Lett.* **2019**, *4*, 397−404.
(39) Christians, J. A.; Miranda Herrera, P. A.; Kamat, P. V. Transformation of the Excited State and Photovoltaic Efficiency of $CH_3NH_3PbI_3$ Perovskite upon Controlled Exposure to Humidified Air. *J. Am. Chem. Soc.* **2015**, *137*, 1530−1538.
(40) Sun, S.; Hartono, N. T. P.; Ren, Z. D.; Oviedo, F.; Buscemi, A. M.; Layurova, M.; Chen, D. X.; Ogunfunmi, T.; Thapa, J.; Ramasamy, S.; et al. Accelerating Photovoltaic Materials Development via High-Throughput Experiments and Machine-Learning-Assisted Diagnosis. *Joule* **2019**, *3*, 1437−1451.
(41) Kirchartz, T.; Márquez, J. A.; Stolterfoht, M.; Unold, T. Photoluminescence-Based Characterization of Halide Perovskites for Photovoltaics. *Adv. Energy Mater.* **2020**, *10*, 1904134.
(42) Asghar, M. I.; Zhang, J.; Wang, H.; Lund, P. D. Device Stability of Perovskite Solar Cells − A Review. *Renewable Sustainable Energy Rev.* **2017**, *77*, 131−146.

https://doi.org/10.1021/acs.jpclett.1c01961
J. Phys. Chem. Lett. 2021, 12, 7866−7877

(43) Raccuglia, P.; Elbert, K. C.; Adler, P. D. F.; Falk, C.; Wenny, M. B.; Mollo, A.; Zeller, M.; Friedler, S. A.; Schrier, J.; Norquist, A. J. Machine-Learning-Assisted Materials Discovery Using Failed Experiments. *Nature* **2016**, *533*, 73−77.
(44) Topol, E. J. High-Performance Medicine: The Convergence of Human and Artificial Intelligence. *Nat. Med.* **2019**, *25*, 44−56.
(45) Takahashi, K.; Takahashi, L.; Miyazato, I.; Tanaka, Y. Searching for Hidden Perovskite Materials for Photovoltaic Systems by Combining Data Science and First Principle Calculations. *ACS Photonics* **2018**, *5*, 771−775.
(46) Pilania, G.; Mannodi-Kanakkithodi, A.; Uberuaga, B. P.; Ramprasad, R.; Gubernatis, J. E.; Lookman, T. Machine Learning Bandgaps of Double Perovskites. *Sci. Rep.* **2016**, *6*, 19375.
(47) Li, J.; Pradhan, B.; Gaur, S.; Thomas, J. Predictions and Strategies Learned from Machine Learning to Develop High-Performing Perovskite Solar Cells. *Adv. Energy Mater.* **2019**, *9*, 1901891.
(48) Xu, Q.; Li, Z.; Liu, M.; Yin, W. J. Rationalizing Perovskites Data for Machine Learning and Materials Design. *J. Phys. Chem. Lett.* **2018**, *9*, 6948−6954.
(49) Javed, S. G.; Khan, A.; Majid, A.; Mirza, A. M.; Bashir, J. Lattice Constant Prediction of Orthorhombic ABO_3 Perovskites Using Support Vector Machines. *Comput. Mater. Sci.* **2007**, *39*, 627−634.
(50) Chen, S.; Hou, Y.; Chen, H.; Tang, X.; Langner, S.; Li, N.; Stubhan, T.; Levchuk, I.; Gu, E.; Osvet, A.; et al. Exploring the Stability of Novel Wide Bandgap Perovskites by a Robot Based High Throughput Approach. *Adv. Energy Mater.* **2018**, *8*, 1701543.
(51) Epps, R. W.; Bowen, M. S.; Volk, A. A.; Abdel-Latif, K.; Han, S.; Reyes, K. G.; Amassian, A.; Abolhasani, M. Artificial Chemist: An Autonomous Quantum Dot Synthesis Bot. *Adv. Mater.* **2020**, *32*, No. 2001626.
(52) Higgins, K.; Valleti, S. M.; Ziatdinov, M.; Kalinin, S. V.; Ahmadi, M. Chemical Robotics Enabled Exploration of Stability in Multicomponent Lead Halide Perovskites via Machine Learning. *ACS Energy Lett.* **2020**, *5*, 3426−3436.
(53) Correa-Baena, J.-P.; Hippalgaonkar, K.; van Duren, J.; Jaffer, S.; Chandrasekhar, V. R.; Stevanovic, V.; Wadia, C.; Guha, S.; Buonassisi, T. Accelerating Materials Development via Automation, Machine Learning, and High-Performance Computing. *Joule* **2018**, *2*, 1410−1420.
(54) Aristidou, N.; Sanchez-Molina, I.; Chotchuangchutchaval, T.; Brown, M.; Martinez, L.; Rath, T.; Haque, S. A. The Role of Oxygen in the Degradation of Methylammonium Lead Trihalide Perovskite Photoactive Layers. *Angew. Chem., Int. Ed.* **2015**, *54*, 8208−8212.
(55) Abdelsamie, M.; Xu, J.; Bruening, K.; Tassone, C. J.; Steinrück, H. G.; Toney, M. F. Impact of Processing on Structural and Compositional Evolution in Mixed Metal Halide Perovskites During Film Formation. *Adv. Funct. Mater.* **2020**, *30*, 2001752.
(56) Phung, N.; Al-Ashouri, A.; Meloni, S.; Mattoni, A.; Albrecht, S.; Unger, E. L.; Merdasa, A.; Abate, A. The Role of Grain Boundaries on Ionic Defect Migration in Metal Halide Perovskites. *Adv. Energy Mater.* **2020**, *10*, 1903735.
(57) Leguy, A. M. A.; Hu, Y.; Campoy-Quiles, M.; Alonso, M. I.; Weber, O. J.; Azarhoosh, P.; van Schilfgaarde, M.; Weller, M. T.; Bein, T.; Nelson, J.; et al. Reversible Hydration of $CH_3NH_3PbI_3$ in Films, Single Crystals, and Solar Cells. *Chem. Mater.* **2015**, *27*, 3397−3407.
(58) Milot, R. L.; Eperon, G. E.; Snaith, H. J.; Johnston, M. B.; Herz, L. M. Temperature-Dependent Charge-Carrier Dynamics in $CH_3NH_3PbI_3$ Perovskite Thin Films. *Adv. Funct. Mater.* **2015**, *25*, 6218−6227.
(59) Beal, R. E.; Hagström, N. Z.; Barrier, J.; Gold-Parker, A.; Prasanna, R.; Bush, K. A.; Passarello, D.; Schelhas, L. T.; Brüning, K.; Tassone, C. J.; et al. Structural Origins of Light-Induced Phase Segregation in Organic-Inorganic Halide Perovskite Photovoltaic Materials. *Matter* **2020**, *2*, 207−219.
(60) Leijtens, T.; Bush, K.; Cheacharoen, R.; Beal, R.; Bowring, A.; McGehee, M. D. Towards Enabling Stable Lead Halide Perovskite Solar Cells; Interplay Between Structural, Environmental, and Thermal Stability. *J. Mater. Chem. A* **2017**, *5*, 11483−11500.
(61) Berhe, T. A.; Su, W.-N.; Chen, C.-H.; Pan, C.-J.; Cheng, J.-H.; Chen, H.-M.; Tsai, M.-C.; Chen, L.-Y.; Dubale, A. A.; Hwang, B.-J. Organometal Halide Perovskite Solar Cells: Degradation and Stability. *Energy Environ. Sci.* **2016**, *9*, 323−356.
(62) Domanski, K.; Alharbi, E. A.; Hagfeldt, A.; Grätzel, M.; Tress, W. Systematic Investigation of the Impact of Operation Conditions on the Degradation Behaviour of Perovskite Solar Cells. *Nat. Energy* **2018**, *3*, 61−67.
(63) Stoddard, R. J.; Dunlap-Shohl, W. A.; Qiao, H.; Meng, Y.; Kau, W. F.; Hillhouse, H. W. Forecasting the Decay of Hybrid Perovskite Performance Using Optical Transmittance or Reflected Dark-Field Imaging. *ACS Energy Lett.* **2020**, *5*, 946−954.
(64) Howard, J. M.; Tennyson, E. M.; Barik, S.; Szostak, R.; Waks, E.; Toney, M. F.; Nogueira, A. F.; Neves, B. R. A.; Leite, M. S. Humidity-Induced Photoluminescence Hysteresis in Variable Cs/Br Ratio Hybrid Perovskites. *J. Phys. Chem. Lett.* **2018**, *9*, 3463−3469.
(65) Jaeger, H.; Haas, H. Harnessing Nonlinearity: Predicting Chaotic Systems and Saving Energy in Wireless Communication. *Science* **2004**, *304*, 78−80.
(66) Forestier, G.; Petitjean, F.; Dau, H. A.; Webb, G. I.; Keogh, E. Generating Synthetic Time Series to Augment Sparse Datasets. *IEEE Int. Conf. Data Min. ICDM* **2017**, 865−870.
(67) Oh, C.; Han, S.; Jeong, J. Time-Series Data Augmentation Based on Interpolation. *Procedia Comput. Sci.* **2020**, *175*, 64−71.
(68) Eslami, T.; Mirjalili, V.; Fong, A.; Laird, A. R.; Saeed, F. ASD-DiagNet: A Hybrid Learning Approach for Detection of Autism Spectrum Disorder Using fMRI Data. *Front. Neuroinform.* **2019**, *13*, 70.
(69) Iwana, B. K.; Uchida, S. *An Empirical Survey of Data Augmentation for Time Series Classification with Neural Networks*; arXiv, 2020.
(70) Grancini, G.; Roldán-Carmona, C.; Zimmermann, I.; Mosconi, E.; Lee, X.; Martineau, D.; Narbey, S.; Oswald, F.; De Angelis, F.; Graetzel, M.; et al. One-Year Stable Perovskite Solar Cells by 2D/3D Interface Engineering. *Nat. Commun.* **2017**, *8*, 15684.
(71) Christians, J. A.; Schulz, P.; Tinkham, J. S.; Schloemer, T. H.; Harvey, S. P.; Tremolet de Villers, B. J.; Sellinger, A.; Berry, J. J.; Luther, J. M. Tailored Interfaces of Unencapsulated Perovskite Solar Cells for > 1,000 h Operational Stability. *Nat. Energy* **2018**, *3*, 68−74.
(72) Tao, C.; Neutzner, S.; Colella, L.; Marras, S.; Srimath Kandada, A. R.; Gandini, M.; Bastiani, M. D.; Pace, G.; Manna, L.; Caironi, M.; et al. 17.6% Stabilized Efficiency in Low-Temperature Processed Planar Perovskite Solar Cells. *Energy Environ. Sci.* **2015**, *8*, 2365−2370.
(73) Chen, M.; Ju, M. G.; Garces, H. F.; Carl, A. D.; Ono, L. K.; Hawash, Z.; Zhang, Y.; Shen, T.; Qi, Y.; Grimm, R. L.; et al. Highly Stable and Efficient All-Inorganic Lead-Free Perovskite Solar Cells with Native-Oxide Passivation. *Nat. Commun.* **2019**, *10*, 16.
(74) Tress, W.; Domanski, K.; Carlsen, B.; Agarwalla, A.; Alharbi, E. A.; Graetzel, M.; Hagfeldt, A. Performance of Perovskite Solar Cells Under Simulated Temperature-Illumination Real-World Operating Conditions. *Nat. Energy* **2019**, *4*, 568−574.
(75) Guo, R.; Khenkin, M. V.; Arnaoutakis, G. E.; Samoylova, N. A.; Barbé, J.; Lee, H. K. H.; Tsoi, W. C.; Katz, E. A. Initial Stages of Photodegradation of $MAPbI_3$ Perovskite: Accelerated Aging with Concentrated Sunlight. *Solar RRL* **2020**, *4*, 1900270.
(76) Divitini, G.; Cacovich, S.; Matteocci, F.; Cinà, L.; Di Carlo, A.; Ducati, C. In situ Observation of Heat-Induced Degradation of Perovskite Solar Cells. *Nat. Energy* **2016**, *1*, 15012.
(77) Holzhey, P.; Yadav, P.; Turren-Cruz, S.-H.; Ummadisingu, A.; Grätzel, M.; Hagfeldt, A.; Saliba, M. A Chain Is As Strong As Its Weakest Link − Stability Study of $MAPbI_3$ Under Light and Temperature. *Mater. Mater. Today* **2019**, *29*, 10−19.
(78) Bryant, D.; Aristidou, N.; Pont, S.; Sanchez-Molina, I.; Chotchunangatchaval, T.; Wheeler, S.; Durrant, J. R.; Haque, S. A. Light and Oxygen Induced Degradation Limits the Operational Stability of Methylammonium Lead Triiodide Perovskite Solar Cells. *Energy Environ. Sci.* **2016**, *9*, 1655−1660.

https://doi.org/10.1021/acs.jpclett.1c01961
J. Phys. Chem. Lett. 2021, 12, 7866−7877

(79) Huang, J.; Tan, S.; Lund, P. D.; Zhou, H. Impact of H_2O on Organic−Inorganic Hybrid Perovskite Solar Cells. *Energy Environ. Sci.* **2017**, *10*, 2284−2311.
(80) Hochreiter, S.; Schmidhuber, J. Long Short-Term Memory. *Neural Comput* **1997**, *9*, 1735−1780.
(81) Schubert, M. C.; Mundt, L. E.; Walter, D.; Fell, A.; Glunz, S. W. Spatially Resolved Performance Analysis for Perovskite Solar Cells. *Adv. Energy Mater.* **2020**, *10*, 1904001.
(82) Collins, L.; Ahmadi, M.; Qin, J.; Liu, Y.; Ovchinnikova, O. S.; Hu, B.; Jesse, S.; Kalinin, S. V. Time Resolved Surface Photovoltage Measurements Using a Big Data Capture Approach to KPFM. *Nanotechnology* **2018**, *29*, 445703.
(83) Kirman, J.; Johnston, A.; Kuntz, D. A.; Askerka, M.; Gao, Y.; Todorović, P.; Ma, D.; Privé, G. G.; Sargent, E. H. Machine-Learning-Accelerated Perovskite Crystallization. *Matter* **2020**, *2*, 938−947.
(84) Taherimakhsousi, N.; MacLeod, B. P.; Parlane, F. G. L.; Morrissey, T. D.; Booker, E. P.; Dettelbach, K. E.; Berlinguette, C. P. Quantifying Defects in Thin Films Using Machine Vision. *npj Comput. Mater.* **2020**, *6*, 111.
(85) Shahcheraghi, N.; Keast, V. J.; Gentle, A. R.; Arnold, M. D.; Cortie, M. B. Anomalously Strong Plasmon Resonances in Aluminium Bronze by Modification of the Electronic Density-of-States. *J. Phys.: Condens. Matter* **2016**, *28*, 405501.
(86) Suram, S. K.; Xue, Y.; Bai, J.; Le Bras, R.; Rappazzo, B.; Bernstein, R.; Bjorck, J.; Zhou, L.; van Dover, R. B.; Gomes, C. P.; et al. Automated Phase Mapping with AgileFD and Its Application to Light Absorber Discovery in the V-Mn-Nb Oxide System. *ACS Comb. Sci.* **2017**, *19*, 37−46.
(87) Bai, J. W.; Xue, Y. X.; Bjorck, J.; Le Bras, R.; Rappazzo, B.; Bernstein, R.; Suram, S. K.; van Dover, R. B.; Gregoire, J. M.; Gomes, C. P. Phase-Mapper: Accelerating Materials Discovery with AI. *AI Mag* **2018**, *39*, 15−26.
(88) Voznyy, O.; Levina, L.; Fan, J. Z.; Askerka, M.; Jain, A.; Choi, M.-J.; Ouellette, O.; Todorović, P.; Sagar, L. K.; Sargent, E. H. Machine Learning Accelerates Discovery of Optimal Colloidal Quantum Dot Synthesis. *ACS Nano* **2019**, *13*, 11122−11128.
(89) Ren, F.; Ward, L.; Williams, T.; Laws, K. J.; Wolverton, C.; Hattrick-Simpers, J.; Mehta, A. Accelerated Discovery of Metallic Glasses Through Iteration of Machine Learning and High-Throughput Experiments. *Sci. Adv.* **2018**, *4*, No. eaaq1566.
(90) Yao, K.; Unni, R.; Zheng, Y. Intelligent Nanophotonics: Merging Photonics and Artificial Intelligence at the Nanoscale. *Nanophotonics* **2019**, *8*, 339−366.
(91) Ma, W.; Cheng, F.; Liu, Y. Deep-Learning-Enabled On-Demand Design of Chiral Metamaterials. *ACS Nano* **2018**, *12*, 6326−6334.

https://doi.org/10.1021/acs.jpclett.1c01961
J. Phys. Chem. Lett. 2021, 12, 7866−7877