



MIT Open Access Articles

Fast and interpretable classification of small X-ray diffraction datasets using data augmentation and deep neural networks

The MIT Faculty has made this article openly available. **Please share** how this access benefits you. Your story matters.

As Published	10.1038/S41524-019-0196-X
Publisher	Springer Science and Business Media LLC
Version	Final published version
Citable link	https://hdl.handle.net/1721.1/136192
Terms of Use	Creative Commons Attribution 4.0 International license
Detailed Terms	https://creativecommons.org/licenses/by/4.0/

ARTICLE OPEN

Fast and interpretable classification of small X-ray diffraction datasets using data augmentation and deep neural networks

Felipe Oviedo¹, Zekun Ren², Shijing Sun¹, Charles Settens¹, Zhe Liu¹, Noor Titan Putri Hartono¹, Savitha Ramasamy³, Brian L. DeCost⁴, Siyu I. P. Tian¹, Giuseppe Romano¹, Aaron Gilad Kusne⁴ and Tonio Buonassisi^{1,2}

X-ray diffraction (XRD) data acquisition and analysis is among the most time-consuming steps in the development cycle of novel thin-film materials. We propose a machine learning-enabled approach to predict crystallographic dimensionality and space group from a limited number of thin-film XRD patterns. We overcome the scarce data problem intrinsic to novel materials development by coupling a supervised machine learning approach with a model-agnostic, physics-informed data augmentation strategy using simulated data from the Inorganic Crystal Structure Database (ICSD) and experimental data. As a test case, 115 thin-film metal-halides spanning three dimensionalities and seven space groups are synthesized and classified. After testing various algorithms, we develop and implement an all convolutional neural network, with cross-validated accuracies for dimensionality and space group classification of 93 and 89%, respectively. We propose average class activation maps, computed from a global average pooling layer, to allow high model interpretability by human experimentalists, elucidating the root causes of misclassification. Finally, we systematically evaluate the maximum XRD pattern step size (data acquisition rate) before loss of predictive accuracy occurs, and determine it to be $0.16^\circ 2\theta$, which enables an XRD pattern to be obtained and classified in 5.5 min or less.

npj Computational Materials (2019)5:60; <https://doi.org/10.1038/s41524-019-0196-x>

INTRODUCTION

High-throughput material synthesis and rapid characterization are necessary ingredients for inverse design and accelerated material discovery.^{1,2} X-ray diffraction (XRD) is a workhorse technique to determine crystallography and phase information, including lattice parameters, crystal symmetry, phase composition, density, space group, and dimensionality.³ This is achieved by comparing XRD patterns of candidate materials with the measured or simulated XRD patterns of known materials.⁴ Despite its indispensable utility, XRD is a common bottleneck in materials characterization and screening loops: 1 hour is typically required for XRD data acquisition with high angular resolution, and another 1–2 hours are typically required for Rietveld refinement by an expert crystallographer, assuming the possible crystalline phases are known. It is widely recognized that machine learning methods have potential to accelerate this process; however, practical implementations have thus far focused on well-established materials,^{5–7} require combinatorial datasets spanning among various phases,^{8,9} or require large datasets,^{5,10} whereas material screening using the inverse design paradigm often involves less-studied materials, spanning multiple classes of different material/phase compositions, and smaller prototype datasets.

Typically, experimental XRD pattern data are analyzed by obtaining descriptors such as peak shape, height, and position. Matching descriptors of the test pattern to known XRD patterns in crystalline databases allows the identification of the compound of interest.⁴ Refinement methods such as Rietveld refinement and Pawley refinement have been used for decades to analyze

experimental XRD patterns.⁴ For novel compounds in thin-film form, however, the use of Rietveld refinement is limited due to the lack of reference patterns in material databases, as well as unknown film textures. The direct-space method, statistical methods, and the growth of single crystals have been used to obtain crystal symmetry information for novel materials,^{7,11–14} but the significant iteration time, feature engineering, human expertise, and knowledge of specific material required makes these methods impractical for high-throughput experimentation, where sample characterization rates are of the order of one material per minute or faster,^{2,15} explored over various material families.

An alternative approach consists in using machine learning methods to obtain more robust spectral descriptors and quickly classify crystalline structure based on peak location and shape in the XRD pattern. Breakthrough methods have been developed for the similar problem of phase attribution in combinatorial alloys,^{16,17} but only few studies have been developed for solution-processed material screening, such as perovskite screening, where phase attribution is usually not as important as correct classification of materials into groups according to crystal parameters. The most successful methods^{10,18} for material screening use convolutional neural networks (CNNs) trained with hundreds of thousands of XRD powder patterns simulated with data from the Inorganic Crystalline Structure Database (ICSD). Further CNN and other deep learning algorithms have been employed to obtain crystalline information for other kinds of diffraction data.^{19–21} In a couple studies, noise-based data

¹Massachusetts Institute of Technology, Cambridge, MA 02139, USA; ²Singapore-MIT Alliance for Research and Technology, Singapore 138602, Singapore; ³Institute for Infocomm Research (I2R), Agency for Science, Technology and Research (A*STAR), Singapore 138632, Singapore and ⁴National Institute of Standards and Technology, MS 8520, Gaithersburg, MD 20899, USA

Correspondence: Felipe Oviedo (foviedo@mit.edu) or Zekun Ren (zekun@smart.mit.edu) or Tonio Buonassisi (buonassisi@mit.edu)
These authors contributed equally: Felipe Oviedo, Zekun Ren

Received: 1 December 2018 Accepted: 25 April 2019

Published online: 17 May 2019

augmentation, a common technique of image preprocessing for machine learning, has been used to avoid overfitting in a broader kind of X-ray characterization problems^{20,22,23} and more broadly in other fields such as Transmission Electron Microscopy imaging;²¹ however, the augmentation procedure has not been based in physical knowledge of actual experimental samples. Furthermore, the best-performing machine learning methods developed up-to-date for XRD analysis do not allow any kind of interpretation by the experimentalists,¹⁰ hindering improvements of experimental design.

While similar approaches produce good results for crystal structure classification, we have found that applying them to high-throughput characterization of novel solution-processed compounds is generally not practical, given the limited access to large datasets of clean, preprocessed, relevant, XRD spectra. Furthermore, most materials of interest developed in high-throughput synthesis loops are thin-film materials. The preferred orientation of the crystalline planes in thin-films causes their experimental XRD patterns to differ from the thousands of simulated XRD powder patterns available in most databases.^{24,25} Thin-film compounds usually will present spectrum shifting and periodic scaling of peaks in preferred orientations, reducing the accuracy of machine learning models trained with powder data,^{8,9,26} even in the cases when noise-based data augmentation techniques are used.¹⁸

Considering these challenges, we propose a supervised machine learning framework for rapid crystal structure identification of novel materials from thin-film XRD measurements. For this work, we created a library of 164 XRD patterns of thin-film halide materials extracted from the >100,000 compounds available in the ICSD²⁷; these 164 XRD patterns include lead halide perovskite^{28,29} and lead-free perovskites-inspired materials.³⁰ These XRD patterns were manually classified among different crystal dimensionalities using ICSD information. Based on this small dataset of relevant XRD powder patterns extracted from the ICSD and an additional 115 experimental XRD patterns, we propose a model-agnostic, physics-informed data augmentation to generate a suitable and robust training dataset for thin-film materials, and subsequently test the space group and dimensionality classification accuracy of multiple machine learning algorithms. A one-dimensional (1D) implementation of an “*all convolutional neural network*” (a-CNN)³¹ is proposed, implemented and identified as the most accurate and interpretable classifier for this problem. We propose a new way to use class activation maps (CAMs),³² computed from the weight distribution of a global average pooling layer and adapted to the context of our problem, to provide interpretability of root causes of classification success or failure to the experimentalist. Subsequently, the effect of the augmented dataset size and the XRD pattern granularity is investigated. Our proposed methodology could be applied to other crystal descriptors of thin-film materials, such as lattice parameters or atomic coordinates, as long as labeled information is available.

Our contributions can be summarized as: (a) development of physics-informed data augmentation for thin-film XRD, which successfully addresses the sparse/scarce data problem, breaching the gap between the thousands of XRD patterns in crystalline databases and real thin-film materials, (b) development of highly interpretable, highly accurate, a-CNN for XRD material screening, (c) proposal of average class activation maps as a feasible interpretability tool in CNNs trained on spectral data.

RESULTS AND DISCUSSION

Framework for rapid XRD classification

The framework developed for rapid classification of XRD thin-film patterns according to crystal descriptors is shown in Fig. 1a. The methodology makes use of both experimental and simulated XRD patterns to train a machine learning classification algorithm. A

simulated dataset is defined by extracting crystal structure information from the ICSD. The experimental dataset consists of a set of synthesized samples, which are manually labeled for training and testing purposes. The datasets are subjected to data augmentation based on the three spectral transformations shown in Fig. 2.

The crystal descriptors of interest, *space group and crystal dimensionality*, are chosen because of their importance for material screening in accelerated material development. In many inorganic material systems, the crystalline dimensionality—i.e., a generalization of the crystalline symmetry into 0-dimensional (0D), 1D, 2D, or 3D symmetry—constitutes a figure of merit for experimental material screening as it correlates with observed charge-transport properties.³³ In perovskites and perovskites-inspired materials, for instance, 3D crystalline structures have been shown to have good carrier-transport properties for solar cells and light-emitting diode applications,^{33,34} whereas 3D–2D mixtures have been found to have greater stability in lead halide perovskites than pure-phase 3D crystals.³⁵ With further detail, the space group number describes the standardized symmetry group of a configuration in space, classifying crystal symmetries into 230 groups. Identifying the space group number of a sample provides crystal information beyond dimensionality, including atomic bonding angles and relative distances, which are believed to be of importance for predicting material properties.³⁶ In this specific study, the framework relies on the relation between XRD patterns and the crystal descriptors of interest. For example, among perovskite-inspired materials for photovoltaic applications, 3D cubic lead halide perovskites of multiple compositions show distinct features in their XRD patterns compared with 2D layered bismuth perovskites.^{37,38}

Typically, the powder XRD pattern is used to identify space group through Rietveld refinement, but the compression of crystalline 3D crystallographic information into a 1D diffraction pattern causes the space group to be impossible to determine unambiguously in certain low-symmetry phases, independently of the measurement technique.³⁹ In this work, the space groups of interest are able to be determined from XRD information only.

To better account for noise measurement and the physical difference between randomly oriented powder patterns and experimental thin-film patterns, the patterns were subjected to a process of data augmentation based on domain knowledge. Subsequently, both augmented experimental and simulated XRD pattern datasets are used for testing, training and cross-validation of machine learning algorithms.

Figure 1b shows the architecture of the final a-CNN, which is proposed and identified as the best-performing machine learning algorithm in subsequent sections.

Experimental measurement and labeling of XRD patterns

The experimental dataset consists of 75 XRD patterns for dimensionality classification, summarized in Supplementary Table S1, and 88 XRD patterns for space group classification, summarized in Supplementary Table S2. A total of 115 unique labeled XRD patterns are considered among both datasets. For this work, perovskites-inspired 3D materials based on lead halide perovskites (space group $Pm\bar{3}m$), tin halide perovskite ($I4/mcm$), cesium silver bismuth bromide double perovskite ($Fm\bar{3}m$), bismuth and antimony halide 2D ($P\bar{3}m1$, Pc , $P2_1/a$), and 0D ($P6_3/mmc$) perovskite-inspired materials are synthesized and used as training and testing dataset. The details of the synthesis and characterization methodology are described in great detail in our experimental study.⁴⁰ Supplementary Fig. S1 shows the t-Distributed Stochastic Neighbor Embedding representation of the XRD patterns labeled with dimensionality and space group, providing evidence of the complexity of the classification problem.

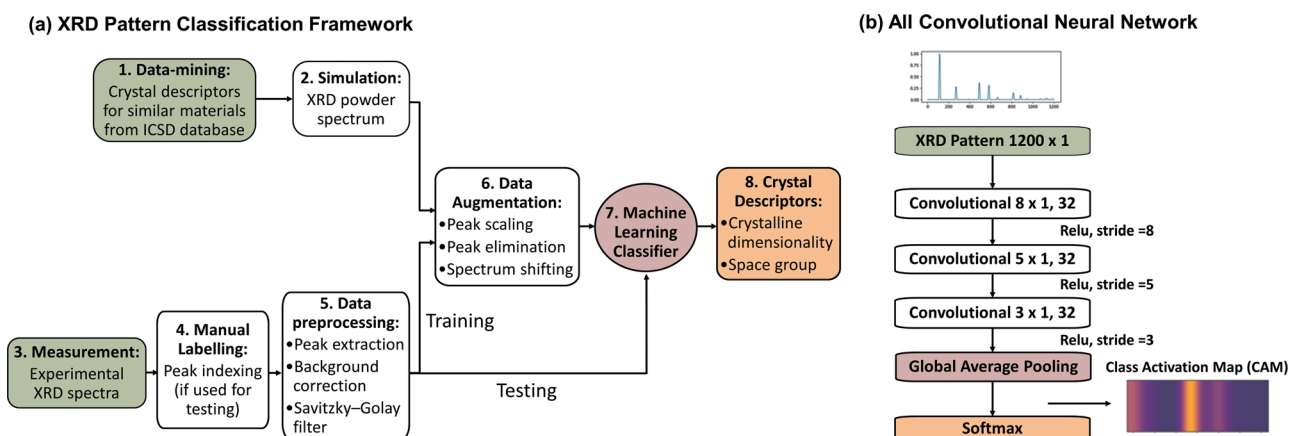


Fig. 1 General framework and a-CNN architecture. **a** Schematic of our X-ray diffraction data classification framework, with physics-informed data augmentation. **b** Schematic of the best-performing algorithm, our all convolutional neural network

In this case, Rietveld refinement is restricted due to the unknown preferred orientation and texture of the thin-film samples. In consequence, the XRD patterns are subjected to peak indexing and the dimensionality and space group are labeled based on ICSD data.

Data mining and simulation of XRD patterns

The simulated training dataset consists of 164 compounds extracted from ICSD with a similar composition, expected crystal symmetry, and space group as the synthesized materials of interest. All the possible single, double, ternary, and quaternary combinations of the elements of interest were extracted during database mining. The compositions of all the materials of interest along with labeled dimensionality and space group information is available in Supplementary Table S3. The fundamental crystal descriptors extracted from the material database are used to simulate XRD powder patterns with random crystalline orientations, as explained in the Methods section.

Data augmentation based on domain knowledge

Experimental thin-film XRD patterns vary greatly compared to simulated, idealized, randomly oriented XRD patterns. Due to expansions and contractions in the crystalline lattice, XRD peaks shift along the 2θ axis according to the specific size and location of the different elements present in a compound, while maintaining similar periodicity based on crystal space group.^{3,8,26,41} In addition, for thin-film samples, the XRD pattern can be shifted due to strain in the film induced during the fabrication process.⁴² Polycrystalline thin-films are also known to have preferred orientations along certain crystallographic planes. The preferred orientation is influenced by the crystal growth process and the growth substrate,²⁴ and is common for most solution-processing and vapor-deposition fabrication methods. Ideal random powders contain multiple grains without any preferred global orientations, thus all crystallographic orientations are represented evenly in the peak intensity and periodicity of the XRD pattern. As a consequence of their preferred orientations along crystallographic planes, thin-film XRD relative peak intensities are scaled up periodically in the preferred plane orientation, and scaled down periodically or even eliminated in the non-preferred orientations.

To increase the size and robustness of the limited training dataset and to account for these fundamental differences between real thin-films and simulated XRD powder patterns, we perform a three-step data augmentation procedure based on physical domain knowledge:

(1) Peak scaling, (2) Peak elimination, and (3) Pattern shifting. These transformations are described in detail in Methods. Figure 2 summarizes the data augmentation steps and its effects on a representative pattern. Given the hyperparameters S and ϵ in Eqs. 1–3 in the Methods section, 2000 patterns are augmented from the simulated dataset, and 2000 patterns are augmented from experimentally measured spectra.

We choose to perform physics-informed data augmentation instead of explicit regularization for the following reasons: (1) data augmentation has been found to be more robust at avoiding overfitting than explicit regularization when using convolutional neural networks,⁴³ (2) data augmentation is model-agnostic, allowing our approach to successfully bridge the gap between experimental XRD patterns and thousands of XRD patterns available in databases without depending on a specific model that might not generalize well in all cases (*no-free-lunch theorem*), and (3) physics-informed data augmentation allows high interpretability, and is found to be more robust than traditional noise-based data augmentation approaches (Supplementary Table S6).

Classification results and a-CNN

Preprocessed, augmented experimental data, and augmented simulated data are fed into various supervised machine learning algorithms for training and testing purposes. The best-performing algorithm is evaluated. The XRD patterns are classified into three crystal dimensionalities (0D, 2D, and 3D) and seven space groups ($Pm\bar{3}m$, $I4/mcm$, $Fm\bar{3}m$, $P\bar{3}m1$, Pc , $P2_1/a$, and $P6_3/mmc$).

For this purpose, we represent the XRD pattern as either a vector or a time series. For each kind of data representation, different classification algorithms are considered. Using a vector representation of the XRD pattern, the following classification methods are tested: Naive Bayes, k -Nearest Neighbors, Logistic Regression, Random Forest, Decision Trees, Support Vector Machine, Gradient Boosting Decision Trees, a Fully Connected Deep Neural Network, and an a-CNN with a global pooling layer.^{44–46} The XRD patterns are also analyzed as a time series with a normalized Dynamic Time Warping (DTW) distance metric⁴⁷ combined with a k -Nearest Neighbors classification algorithm, which was found in literature as the most adequate metric for measuring similarity among metal-alloy XRD spectra.^{7,26}

The problem of novel material development is inherently a multi-class classification problem, in which the classes for training and testing purposes can often be imbalanced as some material families are better characterized than others (e.g., lead-based perovskites are better represented in material databases than newer lead-free perovskites).³⁸ Common metrics for binary classification such as accuracy might not be the most adequate

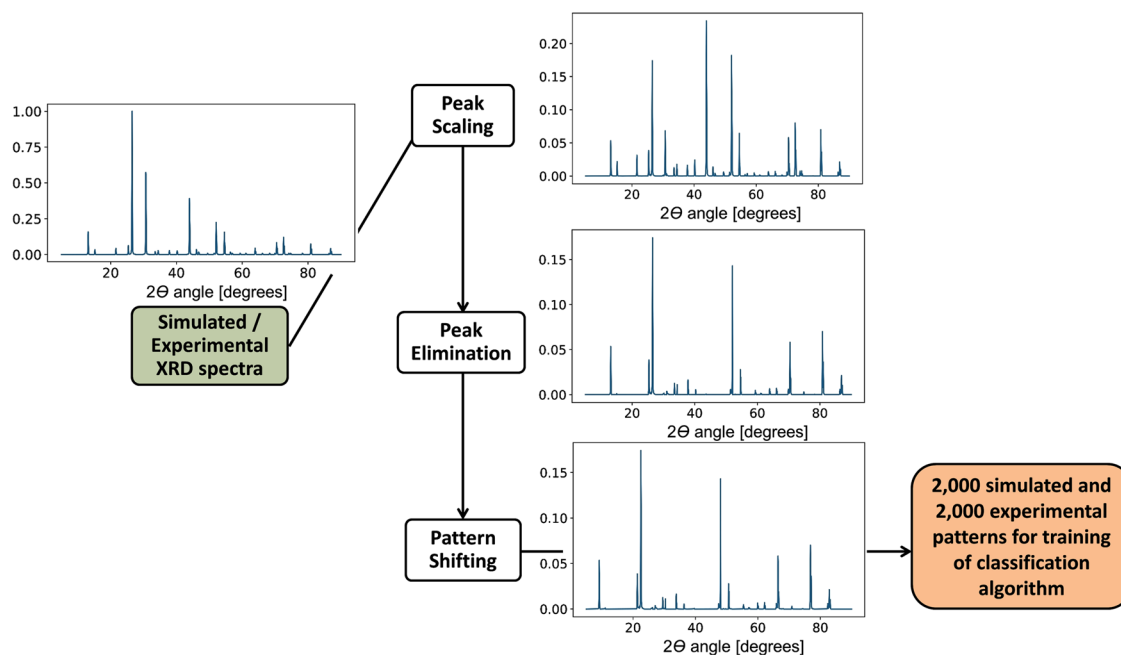


Fig. 2 Physics-informed data augmentation: schematic of the physics-informed data augmentation algorithm, which accounts for the particularities of thin-film XRD spectra, as described by the subsequent transformations stated in Eqs. 1–3

in this context.⁴⁵ The final choice for adequate metrics depends on the relative importance of false positive and false negatives in minority and majority classes, according to the goals of the experimentalist. For method development in this work, we consider the following metrics: *subset accuracy*, defined as the number of correctly classified patterns among all test patterns, and *F₁ score*. *F₁ score* in this problem can be interpreted as the weighted harmonic mean of precision and recall; the closer it is to 1.0, the higher the classifier's precision and recall.⁴⁵ Intuitively, precision is the ability of the classifier not to produce a false positive, whereas recall is the ability of a classifier to find all the true positives. An *F₁* metric is calculated for each class label, and it is combined into an overall score by taking either the *micro* or the *macro* average of the individual scores. The *macro* average calculates the mean of the metrics of all the individual classes, hence treating all classes equally. The *micro* average adds the individual contribution of all samples to compute the overall metric.

When there is class imbalance, accuracy and *F₁ micro* score characterize the classifier's performance over all classes, whereas *F₁ macro* emphasizes the accuracy on infrequent classes.⁴⁵ Thus, a natural choice for high-throughput experiments across multiple material classes seems to be accuracy/*F₁ micro* score, except in those cases when we are especially interested in analyzing an infrequent material class, being *F₁ macro* a more representative metric in that case. In this work, we choose to report the classification accuracy and both *F₁ micro* and *F₁ macro*, whereas recall and precision results are included in the Supplementary Information (Supplementary Tables S4 and S5).

We measure the performance of the dimensionality and space group classification methods based on three different approaches of splitting the training and testing datasets:

Case 1: Exclusively simulated XRD patterns are used for testing and training. Fivefold cross-validation is performed.

Case 2: The simulated XRD patterns are used for training, and the experimental patterns for known materials are used for testing.

Case 3: All of the simulated data and 80% of the experimental data are used for training, and 20% of the experimental data are used for testing. Fivefold cross-validation is performed.

Each one of the training/testing cases mentioned earlier are tested for crystal dimensionality and space group prediction accuracy and *micro/macro F₁* score. The results are reported in Table 1. In each cell, the crystal dimensionality classification metric is reported first, followed by the metric for space group classification. Case 1, presenting fivefold cross-validation results of the simulated dataset, has the highest accuracy as it does not predict any experimental data and thus is free of experimental errors for both crystal descriptors. Case 2 performs the experimental prediction solely based on simulated patterns, thus having the lowest accuracy. Finally, Case 3 has a significant higher accuracy than Case 2 for both crystal dimensionality and space group prediction. *F₁* scores follow these trends as well.

In general, the model's accuracy and *F₁* score is lower for space group classification than that for crystal dimensionality classification. This discrepancy is caused by the lower number of per-class labeled examples for space group classification compared to crystal dimensionality classes. Class imbalance can also systematically affect the training performance of the classifier, to avoid this issue, we performed an oversampling test with synthetic training data according to,^{48,49} and observed little discrepancy of accuracy between the balanced and imbalanced datasets after fivefold cross-validation.

The use of experimental data as part of the training set increases the model's accuracy and robustness. This fact can be explained by the high variability of experimental thin-film XRD patterns, even after data preprocessing. The relatively high accuracy with the relatively small number of experimental samples (on the order of 10–10²) confirms the potential of our data augmentation strategy to yield high predictive accuracies even with small datasets. Supplementary Table S6 in the Supplementary Information compares our strategy with traditional noise-based augmentation approaches, and shows an average increase of classification accuracy of >12% absolute.

Table 1. Results of the ML algorithms under consideration

Method	Case 1 accuracy [%]	Case 2 accuracy [%]	Case 3 accuracy [%]	Case 3 micro F ₁ score	Case 3 macro F ₁ score
Naive Bayes	84.9/81.2	66.7/25.5	83.3/60.5	0.833/0.605	0.830/0.433
K-nearest neighbors	95.1/93.2	68.3/21.2	83.3/74.5	0.833/0.745	0.820/0.621
K-nearest neighbors with dynamic time warping	96.0/94.1	75.3/78.5	86.6/80.1	0.866/0.801	0.836/0.645
Logistic regression	92/82.3	68.3/13.9	91.7/76.6	0.917/0.766	0.847/0.601
Random forest	93.8/81.6	63.3/56.2	83.3/82.1	0.833/0.821	0.827/0.597
Decision tree	87.1/80.7	45.2/31.2	81.7/69.1	0.817/0.691	0.793/0.550
Support vector machine	94.4/93.7	67.1/67.3	88.1/78.5	0.881/0.785	0.805/0.659
Gradient boosting regression trees	96.2/92.1	62.1/62.1	89.7/81.1	0.897/0.811	0.866/0.653
Dense neural network	99.2/98.1	84/75.4	92.1/84.1	0.921/0.841	0.858/0.695
All convolutional neural network (a-CNN)	99.1/99.1	83.7/80.2	92.9/89.3	0.929/0.893	0.859/0.714

Reported average accuracy, *micro* and *macro* F₁ score for dimensionality/space group classification after fivefold cross-validation for each classification algorithm and case. In Case 1, exclusively simulated XRD patterns are used for testing and training. In Case 2, the simulated XRD patterns are used for training, and the experimental patterns for known materials are used for testing. In Case 3, all of the simulated data and 80% of the experimental data are used for training, and 20% of the experimental data are used for testing. The best performing algorithm is our a-CNN architecture

Naturally, the F₁ *macro* score is systematically lower than the F₁ *micro* score, reflecting the impact of misclassification of those dimensionality and space group classes with less training examples. However, the F₁ *macro* score is still fairly high for most classifiers. This fact reflects the importance of adequate experimental design to achieve good generalization among classes.

For all three test cases, the a-CNN classifier performs better than any other classification technique. The 1D a-CNN architecture implemented is composed of three 1D convolutional layers, with 32 filters each, and strides and kernel sizes of 8, 5, and 3 units, respectively. The activation function between layers is *ReLU*. A global average pooling layer⁵⁰ (acting as a weak regularizer) and a final dense layer with *softmax* activation is used. The loss function minimized is binary cross entropy. We use early stopping with a batch size of 128 during training, and use the *Adam* optimizer algorithm to minimize the loss function. The CNN is implemented in Keras 2.2.1 with the Tensorflow background. Figure 1b contains a schematic of the proposed a-CNN architecture.

Our a-CNN architecture, in contrast with other CNNs, does not have max pooling layers between convolutional layers, and also lacks a set of dense layers in the final *softmax* classification layer. These modifications, in contrast with the architectures used in,^{10,18} significantly reduces the number of parameters in the neural network, allows faster and simpler training, and are less prone to overfitting. Another advantage of our implementation is the possibility to extract CAMs from the weight distribution of the global average pooling layer. This characteristic of the network, properly adapted to our problem, allows us to visualize how the classified XRD patterns are mapped to the weight distribution at the end of the a-CNN. The results are further discussed in the interpretability subsection.

The a-CNN trained after data augmentation has an accuracy of >93 and 89% for crystal dimensionality and space group classifications, respectively. As far as we know, the accuracy is among the highest described in literature for space group classification algorithms, comparable to those trained with thousands of ICSD patterns and manual labeling by human experts,^{10,51} and is also comparable to similar approaches in other kinds of diffraction data.^{8,19} The neural network seems to be the most adequate method for high-throughput synthesis and characterization loops, as it also performs relatively well in terms of algorithm speed and in conditions of class imbalance. In the future, our methodology can be extended to other materials

systems, and may include other crystal descriptors as predicted outputs, such as lattice parameters and atomic coordinates.

Furthermore, the a-CNN performs better than the traditional *k*-nearest neighbors method using DTW. In our test case and dataset, the differences between thin-film and powder spectra seem not to be captured properly by DTW alone. Arguably, DTW could be more useful if a larger XRD thin-film pattern dataset is available for *k*-Nearest Neighbors classification, or if it exists greater similarity between XRD patterns of the same class, allowing the DTW warping path to be better captured within the DTW window under consideration.²⁶ CNNs have been found to perform better than DTW for classification of time series, which is consistent with our results.⁵²

Effect of augmented dataset size. The size of the dataset is critical for obtaining a high accuracy and F₁ score. To explore the effect of augmented dataset size, the a-CNN accuracy was computed for various combinations of augmented experimental patterns (i.e., number of augmented XRD spectra originated from the 88 measured spectra, varying *S* and ϵ in Eqs. 1–3) and augmented simulated dataset sizes (i.e., number of augmented XRD spectra originating from the 164 simulated ICSD spectra). Figure 3a, b summarizes this sensitivity analysis for Case 3 training/testing conditions for dimensionality and space group classification. Twenty different fivefold cross-validation runs were performed to calculate the 1-standard deviation error bars for each data point.

In general, as the size of the experimental and augmented datasets increase, the mean accuracy quickly approaches the asymptotic accuracy reported in Table 1. This trend reaffirms and quantifies the importance of data augmentation for the predictive accuracy of our model.

The critical augmented dataset size seems to be around 700 augmented spectra. The model's accuracy is more sensitive to the augmented experimental dataset size, likely because most of the dataset variance comes from the experimental XRD patterns. The data augmentation of the simulated dataset causes the accuracy to grow monotonically in Fig. 3b; however, this trend is not satisfied in the case of Fig. 3a, where no augmented simulated data seems on average to perform the best. We hypothesize that augmenting simulated data could actually introduce excessive noise to the model, hampering classification when the number of possible classes is small.

Figure 3 illustrates that if no data augmentation is used (i.e., the origin, 0, 0), the predictive accuracy could be below 50% for space

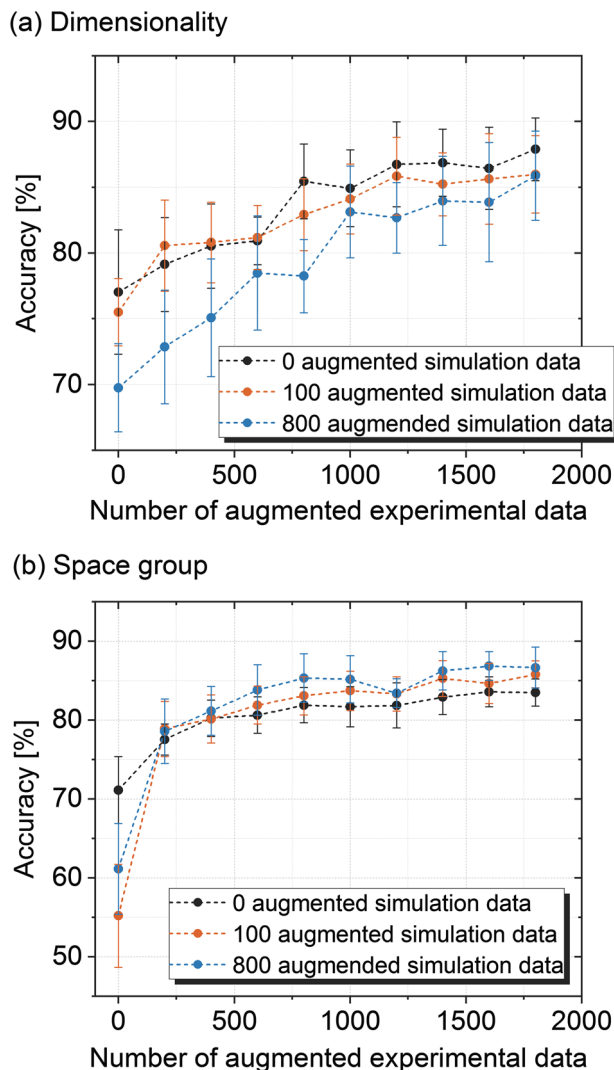


Fig. 3 Accuracy of the a-CNN model with augmented data: line plot showing mean Case 3 a-CNN accuracy, as a function of the number of augmented patterns (Methods, Eqs. 1–3) included in the training set, for (a) dimensionality classification and (b) space group classification. The x axis shows augmented experimental data (based on the original experimental XRD patterns), and the legend shows simulated data (based on the 164 simulated powder diffraction patterns obtained from the ICSD). The error bars correspond to one standard deviation from the mean

group and below 70% for dimensionality. Our physics-informed data augmentation directly increase accuracy by up to 23% in the case of dimensionality and 19% in the case of space group classification. This result reinforces the need for data augmentation for sparse/scarce datasets, as is typical with early-stage material development.

Impact of data coarsening. To evaluate the trade-offs between accuracy and XRD acquisition speed, we investigate how data coarsening of the XRD pattern impacts the accuracy of ML algorithm prediction. In Fig. 4, we report Case 3 accuracy with increasing 2θ angle step size. The baseline step size of the 2θ scan in our XRD patterns is 0.04° . Data coarsening is performed by selectively removing the data with increasing step sizes and rerunning the augmentation and classification algorithms. For crystal dimensionality and space group classification, the highest accuracies are achieved at 0.04 – 0.08° , whereas 85%+ accuracy is achieved when the 2θ step size is 0.16° or less for both cases.

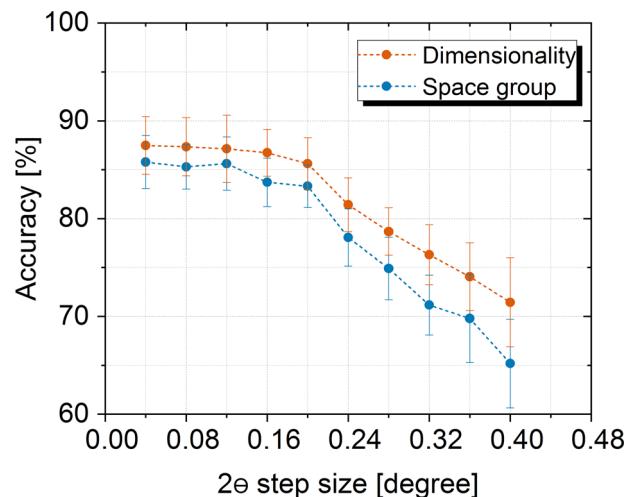


Fig. 4 Reducing XRD pattern acquisition time: simulation of the trade-off between XRD pattern acquisition rate and predictive accuracy. Accuracies for crystal dimensionality and space group predictions are estimated by coarsening the XRD spectrum 2θ step size for Case 3 conditions. The error bars correspond to one standard deviation from the mean

Using the larger step size, the XRD pattern acquisition time can be reduced by 75%, allowing the full spectra to be measured and classified in <5.5 min with our tool setup.

Interpretability using CAMs. Class activation maps (CAM) are representations of the weights in the last layer of a CNN, before performing classification. A CAM for a certain class and pattern indicates the main discriminative regions (in our case, peak and series of peaks in the XRD pattern) that the network uses to identify that class.³² Details of the CAM computation are included in the Methods section.

A similar approach has been followed to interpret and improve object recognition in images and videos using 2D CNNs.³² A single pattern or image produces a unique CAM revealing the main discriminative features. In addition, in the context of our problem, we propose to generalize CAMs to all training samples within a class by averaging over each of the computed CAM weights for all training samples within a class, as explained in the Methods section. This averaging procedure is justified as the location and periodicity of discriminative peaks within a class varies only slightly along all labeled samples in the training set, and can be extended to many spectral measurement techniques. This *average CAM* allows to visualize the main discriminative features of an XRD pattern that were used to classify all the training data belonging to a certain class.

By comparing the CAM for a single pattern with the average CAM for given class, we can identify the root causes of correct and incorrect classification by the a-CNN. Figure 5 illustrates this procedure for XRD patterns of space group classes Class 2 ($P2_1/a$) and Class 6 ($Pm\bar{3}m$). Figure 5a, b shows the average CAM maps of Class 6 and Class 2 space groups. Figure 5c shows the CAM of an individual, correctly classified XRD pattern. If we compare the individual CAM with Class 6 Average CAM, we can see that the neural network is identifying the same reference pattern as most of Class 6 samples, which translates into classification as Class 6. In contrast, Fig. 5d shows an incorrectly classified XRD pattern, which was determined to belong to Class 6 by the CNN, when in reality it belongs to Class 2. A comparison of the individual CAM with the average CAMs of classes 6 and 2 reveals that the misclassified CAM is more similar to Fig. 5a than Fig. 5b. A closer look to the misclassified patterns, shows that the periodicity of peaks before 30° and the relative lack of peaks between 30° and 50° (likely caused by mixed phases), are causing the misclassification.

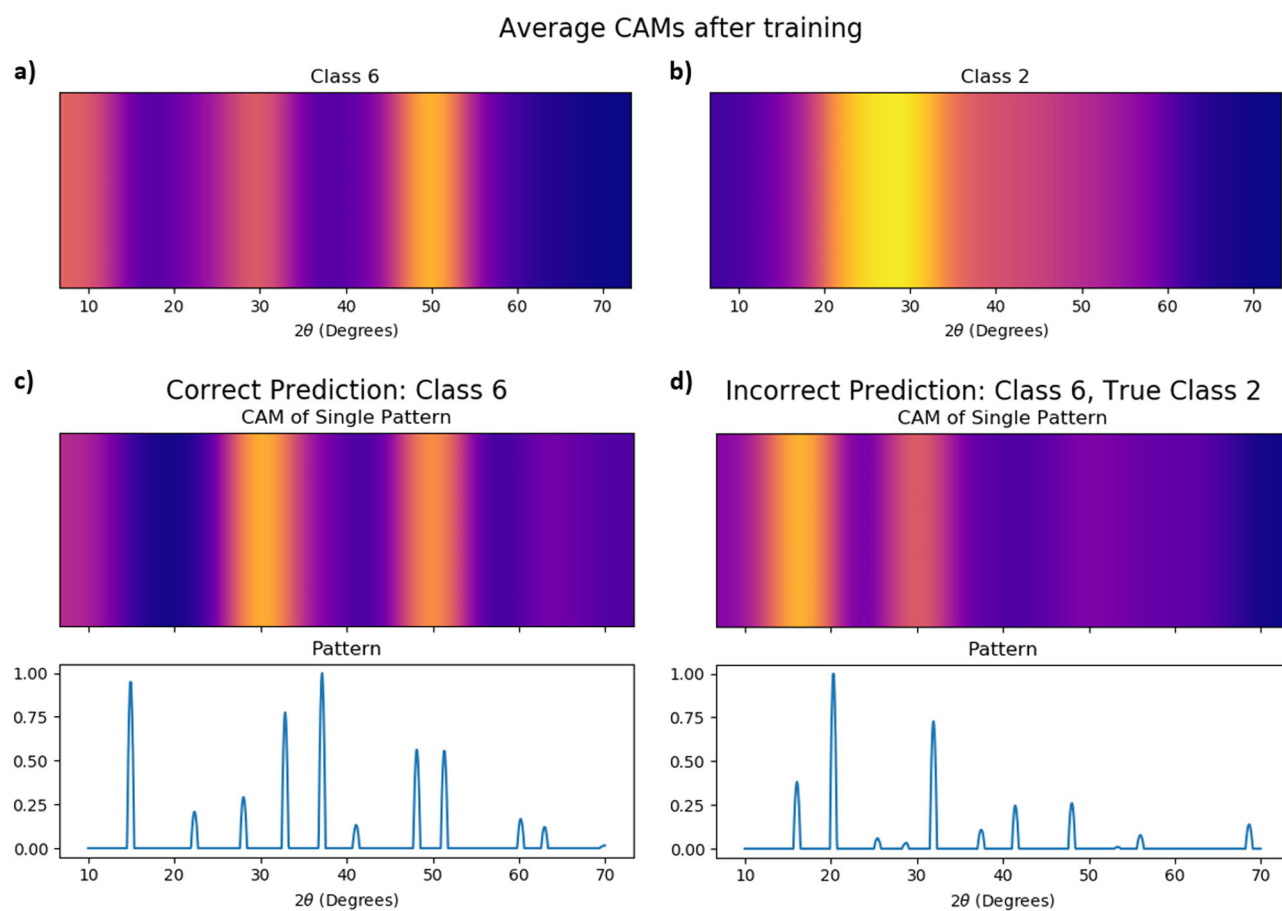


Fig. 5 Class activation maps for misclassification interpretability: class activation maps (CAMs) generated by the a-CNN architecture, representing space group classification of Class 2 ($P2_1/a$) and Class 6 ($Pm\bar{3}m$). **(a, b)** correspond to maps generated by averaging all training samples in a certain class, whereas **(c, d)** correspond to the CAMs of correctly classified and incorrectly classified individual patterns, respectively. The correctly predicted pattern of **(c)** is explained by the similarity of its CAM to the average CAM of Class 6; whereas, the incorrect prediction **(d)** can be explained by comparing its CAM with the average CAMs of classes 2 and 6

Several representative misclassification cases are identified and included in the Section VII of Supplementary Information. In our work, by comparing the CAM of wrongly classified cases with the average CAM of all training data within certain class, we conclude there are three main causes for failed classification. The first cause is the mixture of phases in the sample, which increases the chance of the CAM of a particular pattern differing from the average class CAM. The second cause is lack of XRD patterns in the *per-class* training data. We only have <5 patterns in certain space group classes, thus the average CAM of minority classes has only a limited number of discriminative features compared to majority classes, increasing the testing error. The third cause of misclassification is missing peaks, or too few peaks, present in the XRD pattern; because of the lack of enough discriminative features, the CAM of the incorrectly labeled pattern could be ambiguously similar to two or more average class CAMs, causing the misclassification.

After the experimentalist identifies these root causes of error, they can be mitigated by increasing the number of experimental training points for certain classes or increasing the phase purity of the material coming from the synthesis and solution-processing procedure.

Summary of contributions. In this work, we develop a supervised machine learning framework to screen novel materials based on the analysis of their XRD spectra. The framework is designed specifically for cases when only sparse datasets are available, e.g., early-stage high-throughput material development and discovery loops. Specifically, we propose a physics-informed data

augmentation method that extends small, targeted experimental and simulated datasets, and captures the possible differences between simulated XRD powder patterns and experimental thin-film XRD patterns. A few thousand accuracy augmented spectra are found to increase our classification accuracy from <60 to 93% for dimensionality and 89% for space group. The F_1 *macro* score is also over 0.85 for various algorithms, reflecting the model's capacity to deal with significantly imbalanced classes.

When trained with both augmented simulated and experimental XRD spectra, a-CNNs are found to have the highest accuracy/ F_1 *macro* score among the many supervised machine learning methods studied. Our proposed a-CNN architecture allows high performance and interpretably through CAMs. The use of our proposed average CAMs, allow to identify the root cause of misclassification, and allow the design of a robust experiment. Furthermore, we find that the neural network model tolerates coarsening of the training data, providing future opportunities for online learning, i.e., the on-the-fly adaptive adjustment of XRD measurement parameters by taking feedback from machine learning algorithms.¹⁵

Our approach can be extended to XRD-based high-throughput screening of any kind of thin-film materials, beyond perovskite and perovskite-inspired materials. Since most material databases lack information for various kinds of novel materials, the data augmentation approach to tackle data scarcity can be broadly applied. The underlying difference between XRD patterns of thin-films and powders is common to a broad range of materials and is commonly seen in most thin-film characterization experiments.

The high interpretability of our approach could allow future work in semi-supervised or active learning, allowing the CAM maps to guide manual XRD refinement or actual XRD experiments. The framework may also be extended beyond XRD classification, to any spectrum containing information-rich features that require classification (XPS, XRF, PL, mass spectroscopy, etc.). The advantages of our approach include: **(1)** interpretable error analysis; **(2)** a data augmentation strategy that enables fast and accurate classification, even with small and imbalanced datasets.

METHODS

Measurement and preprocessing of experimental XRD patterns

The XRD patterns for each sample are obtained by using a parallel beam, X-ray powder diffraction Rigaku SmartLab system⁵³ with 2θ angle from 5 to 60° with a step size of 0.04°. The tool is configured in a symmetric setup. We preprocess the raw XRD patterns to reduce the experimental noise and the background signal. For this purpose, the background signal is estimated and subtracted along the 2θ axis, and the spectrum is smoothed conserving the peak width and relative peak size applying the Savitzky-Golay filter.⁵⁴ A representative example of a preprocessed and raw XRD spectra is included in Supplementary Fig. S2.

Simulation of powder XRD patterns

The powder XRD simulations are carried out with Panalytical Highscore v4.7 software based on the Rietveld algorithm implementation by Hill and Howard.^{55,56} The XRD crystal is assumed to have fully random grain orientations. The unit cell lattice parameters, atomic coordinates, atomic displacement parameters, and space group information are considered for the structure factor calculation in the Rietveld model.

Physics-informed data augmentation

Suppose we describe the series of peaks in an XRD pattern by a discrete function $(2\theta) : I \rightarrow \mathbb{R}^+$, which maps a set of discrete angles I to positive real numbers \mathbb{R}^+ corresponding to peak intensities. We augment the data through the following sequential process of transformations $\Phi_{S,c}^1$, Φ_S^2 , and Φ_ε^3 :

Random peak scaling is applied periodically along the 2θ axis to account for different thin-film preferred orientations. A subset of random peaks at periodic angles S is scaled by factor c , such that

$$\Phi_{S,c}^1 = c \cdot f|_S + f|_{\setminus S} \quad (1)$$

Random peak elimination (with a different randomly selected S than Eq. 1) is applied periodically along the 2θ axis, to account for different thin-film preferred orientations, such that

$$\Phi_S^2 = 0 \cdot f|_S + f|_{\setminus S} \quad (2)$$

Pattern shifting by small random value ε along the 2θ direction to allow for different material compositions and film strain conditions, such that

$$\Phi_\varepsilon^3 = f(2\theta - \varepsilon) \quad (3)$$

CAMs

We perform global average pooling in the last convolutional layer. The pooling results are defined as $\sum f_k(x)$ where $f_k(x)$ is the activation of unit k at convolution location x . The pooling results $\sum f_k(x)$ are subsequently fed into the *softmax* classifier. The class confidence score, thus can be computed as:

$$Sc = \sum_k w_k^c \sum_x f_k(x) = \sum_x \sum_k w_k^c f_k(x) \quad (4)$$

We define $M_C(x) = \sum_k w_k^c f_k(x)$, as the CAM for class C , where w_k^c is the weight of unit k of the last convolution layer of class C . Therefore, Sc can be rewritten as:

$$Sc = \sum_x M_C(x) \quad (5)$$

$M_C(x)$ directly indicates the importance of the activation at the discrete location x for a given XRD pattern in class C .

In the context of XRD and other spectral measurements, the location of relevant intensity peaks is likely to be displaced only little from sample to sample within a certain class. Thus, the CAMs for all samples within a class

are similar, and could be averaged to obtain the discriminative features during activation. Thus, we can define *average CAMs* for class C as $\bar{M}_C(x)$:

$$\bar{M}_C(x) = \frac{1}{n} \sum_{i=0 \in C}^n M_C(x) \quad (6)$$

where n corresponds to the training patterns labeled as class C .

DATA AVAILABILITY

The experimental dataset analyzed during the current study is available in the following GitHub repository: [<https://github.com/PV-Lab/AUTO-XRD/tree/master/Datasets/Experimental>]. The simulated and labeled data that support the findings of this study is available from the Inorganic Crystalline Structure Database (ICSD), but restrictions apply to the availability of these data, which were used under license for the current study, and so are not publicly available. Data are, however, available from the authors upon reasonable request and with permission of ICSD.

CODE AVAILABILITY

The codes used for preprocessing, data augmentation, and classification are available in the GitHub repository [<https://github.com/PV-Lab/AUTO-XRD/>]. The classification algorithms are implemented using *scikit-learn* 0.20⁴⁵ and the *fastdtw* python library. The a-CNN is implemented using Keras with the Tensorflow background.

ACKNOWLEDGEMENTS

We thank FengXia Wei (A*STAR), Jason Hattrick-Simpers (NIST), Max Hutchinson (Citricine Informatics), and Zachary T. Trautt (NIST) for helpful suggestions on our analysis. Prof. Juan Pablo Correa Baena at Georgia Tech for sharing various experimental XRD patterns. This work was supported by a TOTAL SA research grant funded through MITeI (supporting the experimental XRD), the National Research Foundation (NRF), Singapore through the Singapore Massachusetts Institute of Technology (MIT) Alliance for Research and Technology's Low Energy Electronic Systems research program (supporting the machine learning algorithm development), the Accelerated Materials Development for Manufacturing Program at A*STAR via the AME Programmatic Fund by the Agency for Science, Technology and Research under Grant No. A1898b0043 (for ML algorithm error analysis), and by the U.S. Department of Energy under the Photovoltaic Research and Development program under Award DE-EE0007535 (for code framework development). This work made use of the CMSE at MIT, which is supported by NSF award DMR-0819762.

AUTHOR CONTRIBUTIONS

F.O. and Z.R. contributed equally for this work. F.O., Z.R., S.S., and T.B. conceived of the research. C.S., F.O., and S.S. prepared the simulated XRD data. Z.R., F.O., C.S., B.L.D.C., and A.G.K. developed data augmentation. S.S., N.T.P.H., and C.S. collected the XRD data. F.O., Z.R., and G.R. developed and tested ML algorithms, with key intellectual contributions from R.S., Z.L., B.L.D.C., and A.G.K. F.O. and Z.R. implemented and tuned the final a-CNN architecture and adapted class activation maps to the context of the problem. F.O., Z.R., S.I.P.T., and T.B. wrote the manuscript, with input from all co-authors.

ADDITIONAL INFORMATION

Supplementary Information accompanies the paper on the *npj Computational Materials* website (<https://doi.org/10.1038/s41524-019-0196-x>).

Competing interests: The authors declare no competing interests.

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

REFERENCES

1. Tabor, A., Roch, D. & Saikin, L. Lawrence Berkeley National Laboratory recent work title accelerating the discovery of materials for clean energy in the era of smart automation. *Nat Rev Mater.* <https://doi.org/10.1038/s41578-018-0005-z> (2018).
2. Correa-Baena, J.-P. et al. Accelerating materials development via automation, machine learning, and high-performance computing. *Joule* **2**, 1410–1420 (2018).
3. Dinnebier, R. E. *Powder Diffraction: Theory and Practice*. (RSC Publ, Cambridge, 2009).
4. Rietveld, H. M. A profile refinement method for nuclear and magnetic structures. *J. Appl. Crystallogr.* **2**, 65–71 (1969).

5. Carr, D. A., Lach-hab, M., Yang, S., Vaisman, I. I. & Blaisten-Barojas, E. Machine learning approach for structure-based zeolite classification. *Microporous Mesoporous Mater.* **117**, 339–349 (2009).
6. Baumes, L. A., Moliner, M., Nicoloyannis, N. & Corma, A. A reliable methodology for high throughput identification of a mixture of crystallographic phases from powder X-ray diffraction data. *CrystEngComm* **10**, 1321–1324 (2008).
7. Baumes, L. A., Moliner, M. & Corma, A. Design of a full-profile-matching solution for high-throughput analysis of multiphase samples through powder X-ray diffraction. *Chem. - A Eur. J.* **15**, 4258–4269 (2009).
8. Stanev, V. et al. Unsupervised phase mapping of X-ray diffraction data by non-negative matrix factorization integrated with custom clustering. *npj Comput. Mater.* **4**, 43 (2018).
9. Kusne, A. G., Keller, D., Anderson, A., Zaban, A. & Takeuchi, I. High-throughput determination of structural phase diagram and constituent phases using GRENDEL. *Nanotechnology* **26**, 444002 (2015).
10. Park, W. B. et al. Classification of crystal structure using a convolutional neural network. *IUCrJ.* **4**, 486–494 (2017).
11. Park, W. B., Singh, S. P., Yoon, C. & Sohn, K. S. Combinatorial chemistry of oxynitride phosphors and discovery of a novel phosphor for use in light emitting diodes, $\text{Ca}_{1.5}\text{Ba}_{0.5}\text{Si}_5\text{N}_6\text{O}_3:\text{Eu}^{2+}$. *J. Mater. Chem. C* **1**, 1832–1839 (2013).
12. Rybakov, V. B., Babaev, E. V., Pasichnichenko, K. Y. & Sonneveld, E. J. X-ray mapping in heterocyclic design: VI. X-ray diffraction study of 3-(isonicotinoyl)-2-oxooxazolol[3,2-*a*]pyridine and the product of its hydrolysis. *Crystallogr. Rep.* **47**, 473–477 (2002).
13. Hirotsaki, N., Takeda, T., Funahashi, S. & Xie, R. J. Discovery of new nitridosilicate phosphors for solid state lighting by the single-particle-diagnosis approach. *Chem. Mater.* **26**, 4280–4288 (2014).
14. Suram, S. K. et al. Automated phase mapping with AgileFD and its application to light absorber discovery in the V-Mn-Nb oxide system. *ACS Comb. Sci.* **19**, 37–46 (2017).
15. Ren, F. et al. Accelerated discovery of metallic glasses through iteration of machine learning and high-throughput experiments. *Sci. Adv.* **4**, eaaq1566 (2018).
16. Bunn, J. K. et al. Generalized machine learning technique for automatic phase attribution in time variant high-throughput experimental studies. *J. Mater. Res.* **30**, 879–889 (2015).
17. Kusne, A. G. et al. On-the-fly machine-learning for high-throughput experiments: search for rare-earth-free permanent magnets. *Sci. Rep.* **4**, 1–7 (2014).
18. Vecsei, P. M. et al. Neural network-based classification of crystal symmetries from X-ray diffraction patterns. arXiv preprint arXiv:1812.05625 (2018).
19. Ziletti, A., Kumar, D., Scheffler, M. & Ghiringhelli, L. M. Insightful classification of crystal structures using deep learning. *Nat. Commun.* **9**, 1–10 (2018).
20. Ke, T. W. et al. A convolutional neural network-based screening tool for X-ray serial crystallography. *J. Synchrotron Radiat.* **25**, 655–670 (2018).
21. Ziatdinov, M. et al. Deep learning of atomically resolved scanning transmission electron microscopy images: chemical identification and tracking local transformations. *ACS Nano* **11**, 12742–12752 (2017).
22. Le Bras, R. et al. A computational challenge problem in materials discovery: synthetic problem generator and real-world datasets. *Proceedings of the Twenty-Eighth AAAI Conference of Artificial Intelligence System*, pp. 438–443 (2014).
23. LeBras, R. et al. Constraint reasoning and kernel clustering for pattern decomposition with scaling. in *Principles and Practice of Constraint Programming – CP 2011* (ed. Lee, J.) 508–522 (Springer, Berlin, Heidelberg, 2011).
24. Järvinen, M. Application of symmetrized harmonics expansion to correction of the preferred orientation effect. *J. Appl. Crystallogr.* **26**, 525–531 (1993).
25. Fewster, P. F., Langford, J. I. & Fewster, P. F. Reports on progress in physics related content X-ray analysis of thin films and multilayers. **59**, 11 (1996).
26. Iwasaki, Y., Kusne, A. G. & Takeuchi, I. Comparison of dissimilarity measures for cluster analysis of X-ray diffraction data from combinatorial libraries. *npj Comput. Mater.* **3**, 1–8 (2017).
27. Belkly, A., Helderman, M., Karen, V. L. & Ulkch, P. New developments in the Inorganic Crystal Structure Database (ICSD): accessibility in support of materials research and design. *Acta Crystallogr. Sect. B Struct. Sci.* **58**, 364–369 (2002).
28. Eperon, G. E. et al. Formamidinium lead trihalide: a broadly tunable perovskite for efficient planar heterojunction solar cells. *Energy Environ. Sci.* **7**, 982–988 (2014).
29. Lee, M. M., Teuscher, J., Miyasaka, T., Murakami, T. N. & Snaith, H. J. Efficient hybrid solar cells based on meso-superstructured organometal halide perovskites. *Sci. (80-)* **338**, 643–647 (2012).
30. Hoyer, R. L. Z. et al. Perovskite-inspired photovoltaic materials: toward best practices in materials characterization and calculations. *Chem. Mater.* **29**, 1964–1988 (2017).
31. Springenberg, J. T. et al. Striving for simplicity: the all convolutional net. arXiv preprint arXiv:1412.6806. (2014).
32. Zhou, B. et al. Learning deep features for discriminative localization. *Proceedings of the IEEE conference on computer vision and pattern recognition.* <https://doi.org/10.1109/CVPR.2016.319> (2016).
33. Etgar, L. The merit of perovskite's dimensionality; Can this replace the 3D halide perovskite? *Energy Environ. Sci.* **11**, 234–242 (2018).
34. Xiao, Z., Meng, W., Wang, J., Mitzi, D. B. & Yan, Y. Searching for promising new perovskite-based photovoltaic absorbers: the importance of electronic dimensionality. *Mater. Horiz.* <https://doi.org/10.1039/C6MH00519E> (2017).
35. Zhang, T., Long, M., Liu, P., Xie, W. & Xu, J.-B. Stable and efficient 3D-2D perovskite-perovskite planar heterojunction solar cell without organic hole transport layer. *Joule* <https://doi.org/10.1016/j.joule.2018.09.022> (2018).
36. Kurchin, R. C., Gorai, P., Buonassisi, T. & Stevanović, V. Structural and chemical features giving rise to defect tolerance of binary semiconductors. *Chem. Mater.* **30**, 5583–5592 (2018).
37. Baikie, T. et al. Synthesis and crystal chemistry of the hybrid perovskite (CH₃NH₃)PbI₃ for solid-state sensitised solar cell applications. *J. Mater. Chem. A* **1**, 5628–5641 (2013).
38. Sun, S. et al. Synthesis, crystal structure, and properties of a perovskite-related bismuth phase, (NH₄)₃Bi₂I₉. *APL Mater.* **4**, 031101 (2016).
39. Coelho, A. A. TOPAS-Academic, Version 6: technical reference. (Coelho Software, Brisbane, Australia, 2016).
40. Shijing Sun, et al. Accelerating photovoltaic materials development via high-throughput experiments and machine-learning-assisted diagnosis. *Joule* **2**, 1410–1420 (2018).
41. Ermon, S. et al. Pattern decomposition with complex combinatorial constraints: application to materials discovery. *AAAI'15 Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence.* (2014).
42. Zhao, J. et al. Strained hybrid perovskite thin films and their impact on the intrinsic stability of perovskite solar cells. *Sci. Adv.* **3**, eaa05616 (2017).
43. Hernández-García, A. & König, P. Data augmentation instead of explicit regularization. arXiv preprint arXiv:1806.03852 (2018).
44. Goodfellow, I., Bengio, Y., Courville, A. & Bengio, Y. *Deep Learning*. 1 (MIT Press, Cambridge, 2016).
45. Pedregosa, F. et al. Scikit-learn: machine learning in {Python}. *J. Mach. Learn. Res.* **12**, 2825–2830 (2011).
46. Friedman, J., Hastie, T. & Tibshirani, R. *The Elements of Statistical Learning*. 1 (Springer Series in Statistics, New York, NY, USA, 2001).
47. Salvador, S. & Chan, P. FastDTW: toward accurate dynamic time warping in linear time and space. *Time* **11**, 70–80 (2004).
48. Haixiang, G. et al. Learning from class-imbalanced data: review of methods and applications. *Expert Syst. Appl.* **73**, 220–239 (2017).
49. Chawla, N. V., Bowyer, K. W., Hall, L. O. & Kegelmeyer, W. P. SMOTE: synthetic minority over-sampling technique. *J. Artif. Intell. Res.* **16**, 321–357 (2002).
50. Lin, M., Chen, Q. & Yan, S. Network in network. arXiv preprint arXiv:1312.4400. (2013).
51. Yoon, C. H. et al. Unsupervised classification of single-particle X-ray diffraction snapshots by spectral clustering. *Opt. Express* **19**, 16542 (2011).
52. Wang, Z., Yan, W. & Oates, T. Time series classification from scratch with deep neural networks: a strong baseline. *Proc. Int. Jt. Conf. Neural Netw.* **2017**, 1578–1585 (2017).
53. Kobayashi, S. & Inaba, K. X-ray thin-film measurement techniques. *Mass Spectrosc. Equipped a Ski. Interface* **28**, 8 (2012).
54. Press, W. H. & Teukolsky, S. A. Savitzky-Golay smoothing filters. *Comput. Phys.* **4**, 669 (1990).
55. Hill, R. J. & Howard, C. J. Quantitative phase analysis from neutron powder diffraction data using the Rietveld method. *J. Appl. Crystallogr.* **20**, 467–474 (1987).
56. Degen, T., Sadki, M., Bron, E., König, U. & Nénért, G. The HighScore suite. *Powder Diff.* **29**, S13–S18 (2014).



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.