

Machine learning for halide perovskite materials

Lei Zhang^{a,b,*}, Mu He^a, Shaofeng Shao^a

^a School of Chemistry and Materials Science, Nanjing University of Information Science & Technology, Nanjing, 210044, China

^b Shanghai Ximi Information Technology Co., Ltd, Shanghai, 200001, China



ARTICLE INFO

Keywords:

Halide perovskites
Machine learning
Solar cells

ABSTRACT

Halide perovskite materials serve as excellent candidates for solar cell and optoelectronic devices. Recently, the design of the halide perovskite materials is greatly facilitated by machine learning techniques, which effectively identify suitable halide perovskite candidates and unveil hidden relationships by algorithms that mimic the human cognitive functions. In this manuscript, we review recent progresses on the machine learning studies of the halide perovskite materials, including the prediction and understanding of lead-free and stable halide perovskite materials. The structural descriptors to describe the property and performance of the halide perovskite materials are discussed. In addition, the design strategy of the additive species for the halide perovskite materials via the machine learning technique is provided. Suggestions to further develop the halide perovskite-based systems via the machine learning methods in the future are provided.

1. Introduction

Perovskite solar cells have received much attention in recently years owing to their outstanding power conversion efficiencies reaching 25.2% [1]. The exceptional power conversion efficiencies of the perovskite solar cells are attributed to the strong optical absorption, long carrier diffusion length, minimal recombination and defect tolerance of halide perovskite materials. The halide perovskite materials also enjoy low fabrication cost and facile synthesis, which are not available for the silicon-based traditional solar cells [2]. Apart from the solar cells, the halide perovskite materials have been applied to various other optoelectronic and energy devices, including light-emitting diodes, catalysts, batteries and photodetectors [3–6].

However, the halide perovskite materials suffer from severe environmental and stability issues, which prevent them from further industrial deployment [7,8]. For example, the B-site poisonous lead species may leak into the nature, causing environmental and health problems; as a result, the search for the lead-free halide perovskite materials is crucial. Several promising design strategies have been devised to solve the issues, such as the replacement of lead with tin and the employment of double halide perovskites. Besides, the halide perovskite materials are unstable in the presence of water and oxygen [9,10]; thus extensive works have been carried out to improve the stability of halide perovskite materials. Suitable halide perovskite candidates with tailored dimensions and chemical ingredients have been

demonstrated, and various interfacial engineering approaches have been proposed. For example, the molecular additives, such as those introducing Lewis-acid interactions, halogen bonds, carboxylate, phosphonate and metal halides are incorporated into the perovskite solar cells to passivate the photoactive materials and facilitate the interfacial charge transfer [11–13]. These additives also partially eliminate the hysteresis effects, which are essential to boost the device performance and improve the stability.

The prediction of new halide perovskite materials and the establishment of their hidden structure-property relationships are traditionally realized via the first-principles calculations, which obtain optimized the crystal structures, surfaces/interface structures and optoelectronic properties. However, the first-principles calculations are normally computationally expensive. Apart from the traditional trial and error experimental and computational routes, the machine learning techniques have been devised to accelerate the discovery process of the novel halide perovskite materials and unveil their hidden relationships. In addition, with the exponential expansion of data stored in publication, it is important to use the data-driven methods [14] such as the machine learning techniques to extract useful information from the databases and accelerate the materials design process [15–29]. In the machine learning process, the identification of accurate features that strongly correlate to the targeting properties are prerequisites. The features, or sometimes descriptors, can be selected via the feature engineering process. Once the input and output data of the halide

* Corresponding author. School of Chemistry and Materials Science, Nanjing University of Information Science & Technology, Nanjing, 210044, China.
E-mail address: 002699@nuist.edu.cn (L. Zhang).

perovskite materials are successfully represented numerically, the mapping between inputs and output data could be performed via the machine learning process (Fig. 1) [30].

In this manuscript, we review recent progresses of the design of halide perovskite materials via the machine learning techniques. The machine learning techniques are demonstrated to greatly accelerate the discovery process of the lead-free and stable halide perovskites for solar cells and unveil their structure-property relationships. The machine learning method also helps the synthetic design of dimensionally tailored halide perovskites. Last but not least, the machine learning technique facilitates the selection of the combination of materials such as the additives residing on the halide perovskite surfaces to further optimize the halide perovskite-based devices. The instability and the necessity of poisonous lead of halide perovskite materials are two notorious problems in the perovskite solar cell community; the insufficient candidates of stable and lead-free halide perovskites prevent their further industrial deployment. On the other hand, the machine learning techniques are especially effective to identify new stable and lead-free halide perovskite candidates and unveil the hidden relationships by algorithms that mimic the human cognitive functions. A comprehensive review of the halide perovskite material studies via the machine learning methods are provided, including the datasets, descriptors, lead-free halide perovskites, other halide perovskites, stability, dimensional tailoring, and additives. The “dataset” and “descriptors” sections describe the fundamental modules of the machine learning techniques tailored for the halide perovskite materials; the “lead-free halide perovskites”, “stability”, “dimensional tailoring”, and “additives” comprehensively address the successful researches of the machine learning-assisted halide perovskite materials. These contents are interconnected; for example, the molecular additive design can also improve stability of the halide perovskite materials via the proper selection of descriptors in the feature engineering stage. The dimensional tailoring method benefits both optoelectronic performance and stability.

2. Background

2.1. Halide perovskite materials

The prototypical halide perovskite materials share the common formula ABX_3 (A is a monovalent cation, B a divalent metal ion and X a

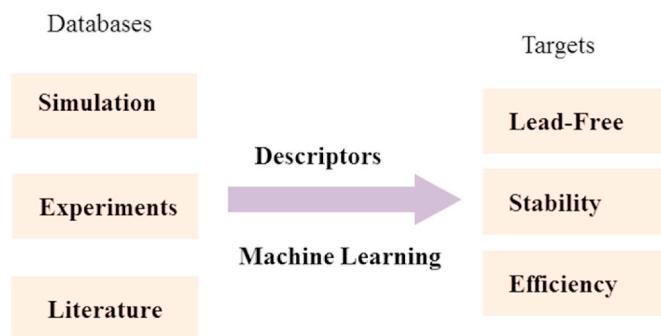


Fig. 1. General workflow of the halide perovskite design via the machine learning method. Firstly, the databases storing correct structural, property and synthetic information of halide perovskites are prepared as the input data. The databases can be obtained via first-principles calculations-based simulation, in-house experiments or information extraction directly from the literature. Secondly, the output data (targets) are conceptualized and quantified, such as the yes/no of stable lead-free material verified by experiments or first-principles calculations, the E_g values and the overall conversion efficiency values. Thirdly, the suitable machine learning algorithms are selected to correlate the input and output data, either unveiling the hidden relationships or predicting new candidates. Choosing the correct descriptors is critical in the machine-learning process; the features that efficiently describe the targets should be selected via the feature engineering process.

halogen counterion). They exhibit optimal band gap, superior light-harvesting properties, long carrier diffusion length and defect tolerance, which make them ideal candidates for solar cells, light-emitting diodes and photodetectors. The halide perovskites exhibit different phases, such as tetragonal, cubic and orthorhombic, depending on the temperature and pressure. The disordered A-site cations are suggested to contribute to the ferroelectric highways and the stabilization of the overall framework structures that are associated with the excellent electronic and optical properties [31]. Large polarons in the halide perovskite materials have been suggested to protect the charge carriers and contribute to the excellent optoelectronic properties [32], while excitons with the binding energies within millielectronvolt exist [33].

The poisonous lead species present in the halide perovskite materials can cause environmental issues. The lead-free halide perovskite materials have been fabricated, such as those with the B-site tin element. The lead-free halide double perovskites have been fabricated, such as Cs_2InAgX_6 and similar $A_2In(I)M(III)X_6$ -based compounds [34,35]. However, none of these halide perovskite alternatives have been surpassed the lead-based perovskite solar cell efficiencies so far. It is of critical importance to develop new lead-free and stable halide perovskite materials for the solar cell applications.

The degradation of the halide perovskite materials is one of the most notorious problems of the perovskite-based solar cells, which are accelerated by the environmental moisture, photon, thermal energy and external electric fields [36,37]. Many efforts have been made to improve the stability of halide perovskite materials. For example, the inorganic A-site species have been incorporated in the halide perovskite systems [38]. Additives have been introduced to improve the film quality and passivate the reactive perovskite surfaces [39,40]. Compositional engineering has been employed to tailor the halide perovskite structures and properties [40]. Elemental doping such as lithium, sodium, potassium and rubidium has been tried to improve the perovskite solar cell performance [41]. The light harvesting properties are engineered to broaden the absorption region and increase the absorptivity of halide perovskite materials [42]. The precursor solutions are found to affect the resulting morphology and performance [43]. In addition, the halide perovskites are coupled with various other materials to form nanocomposites to achieve tailored properties [44]. The ion migration in halide perovskites leads to the hysteresis phenomena, which contribute to the unstable photovoltaic performance with a forward or reverse bias. Both ion and electron migration inside halide perovskite materials contribute to the slow and fast ion transport signals [45]. Under different biases, the ions accumulate at the grain boundaries and interfaces, which are detrimental for the device performance. The incorporation of additives is demonstrated to effectively reduce the hysteresis. However, the good ion transportation in halide perovskite materials also a positive effect: the efficient ion transport makes this type of material ideal for lithium-ion batteries and super-capacitors, which are receiving attentions these days.

The halide perovskite materials have reached lower dimensions, with the dimensional tailoring methods offering improved stability compared with those of the traditional three-dimensional halide perovskites. For example, the stable two-dimensional halide perovskites are suggested to demonstrate better stability but suffer from inferior optoelectronic performance compared with the three dimensional (3D)-counterpart. As a result, the two-dimensional (2D)/3D dimensional mixing is proposed, which exhibits balanced stability and optoelectronic performance [46,47]. Apart from the 2D halide perovskites, the one-dimensional (1D) (nanowire and nanotube), zero-dimensional (0D) and the ultimate molecular counterparts (A_4BX_6) of halide perovskites have been designed, each offering unique and unprecedented phenomena that are not observed in other dimensions.

The halide perovskite materials are generally prepared via the one-step solution process or the two-step process. The former involves the mixing of the inorganic PbX_2 precursor and the organic CH_3NH_3I precursor in a single solution; the latter has the lead salt deposited first and

has the typical PbI_2 precursor added in a second step. Various solvent engineering methods have been developed to improve the film quality [48–51], since the chosen solvent molecules can strongly affect the crystallization process and reduce the defect states. In particular, the solvent molecules can coordinate with the precursor solution, forming complexes and preventing unnecessary reactions during the perovskite formation process [50,52]. In addition, the compositional engineering approach, such as the mixed cation method, is effective to further improve the perovskite solar cell performance [53,54]. The mixed cation method is probably one of the most effective way to modify the electronic properties and stability of the halide perovskite materials. For example, cations with disparate sizes can significantly alter the band gap of the halide perovskite materials, with the substitution of MA with HC (NH_2H_2^+) (FA) offering broader absorption and facile crystallization. In contrast, the routine single cation strategy leads to limited light absorption or unwanted yellow phase [55]. In addition to the binary system, the ternary cation perovskite including MA, FA and Cs has been developed to further improve the film quality and stability simultaneously.

2.2. Machine learning

The machine learning techniques include the computer algorithms that automatically evolve and adapt to the environment via experience, in order to make decisions and extract knowledge by the machine. A comprehensive database is necessary to help the machine learning model automatically predict candidates and hidden relationships without particular traditional programming by human. The machine learning techniques can be divided into supervised machine learning (with labeled or classified data to train the model), unsupervised machine learning (unlabeled or unclassified data to train the model), semi-supervised machine learning (bearing both supervised and unsupervised machine learning characters) and reinforcement machine learning (normally with reward and try-and-error feedback). A number of machine learning models have been developed, with the common ones such as decision trees, support vector machines, regression analysis, Bayesian networks and genetic algorithms and artificial neural networks. The artificial neural network probably serves as one of the most popular machine learning method used for the materials design process, with a brain-inspired system that consists of input layers, output layers and hidden layers that are “dark” for humans. Before the machine learning technique is applied to the materials science research, the high-throughput calculation is almost one of the most effective tools to obtain useful information and find suitable materials for particular applications [56]. Even now, the machine learning process is often coupled with the high-throughput calculations which generate a database for the machine learning method to extract useful information. New materials are explored via the data-mining method, which provides platforms of materials informatics and effectively unveil the hidden structure-property relationships [57–59]. Descriptors of the targeting properties/performances/applications are required to extract useful information from the large database [60]. The machine learning technique is found to rapidly identify candidates of solar cell materials [61,62]. For example, the machine-learning method is employed to predict the power conversion efficiencies of the organic solar cells [63,64]. The band gaps of the solids could be accurately predicted via the machine learning technique which enjoys better fidelity than the quantum mechanical calculations [65]. The machine learning method is employed to investigate the energy storage applications such as lithium-ion batteries, including their state-of-charge [66–70], the health diagnosis and the crack detection [71].

Several typical software and tools for the machine-learning studies have been well summarized in the literature [72], including Care (<http://topepo.github.io/caret>), DeepLearning4j (<https://deeplearning4j.org>), H2O.ai (<https://h2o.ai>), Keras (<https://keras.io>), Mlpack (<https://mlpack.org>), Scikit-learn (<http://scikit-learn.org>) and Weka (<https://cs.waikato.ac.nz/ml/weka>), while the machine-learning tools usually employed for the molecular and material sciences include Amp (<https://bitbucket.org/andrewpeterson/amp>), ANI (https://github.com/isayev/ASE_ANI), COMBO (<https://github.com/tsudalab/combo>), DeepChem (<https://deepchem.io>), GAP (<http://libatoms.org/Home/Software>), MatMiner (<https://hackingmaterials.github.io/matminer>), NOMAD (toolkit.nomad-coe.eu), PROPHet (<https://github.com/biklooost/PROPhet>) and TensorMol (<https://github.com/jparkhill/TensorMol>). In particular, the scikit-learn package acts one of the most versatile and accessible platform to perform the machine-learning investigations [73–77], while TensorFlow is another popular platform to provide relevant libraries for the deep learning methods.

The basic procedures for employing the machine learning approaches include data collection, feature engineering, modeling, validating and producing final results (Fig. 2) [22,78,79]. In the data collection stage, researchers extract useful information from experiments, computations and various available databases, and represent them in a unified format after the data optimization [80]. The feature engineering is required to remove redundant features and help establish proper structure-property relationships. The modeling procedure includes the selection of appropriate algorithms, the training and testing of data, and the subsequent accurate predictions. The machine learning algorithms can be divided into supervised learning, unsupervised learning, semi-supervised learning and reinforcement learning. The core of the machine learning techniques is associated with the training and testing of data, where the large amount of data are usually split into at least two parts. A series of criteria are selected to validate the model, such that the final results can be obtained.

3. Datasets

The Cambridge Structure Database (CSD) and the inorganic crystal structure database (ICSD) are the two major structure databases that store the experimental data of accurate crystal structures of the halide perovskite materials obtained via the single-crystal X-ray diffraction technique. The calculated structures and properties of ABX_3 -type halide perovskites obtained via the density-functional theory are also stored in several other databases. Kim et al. prepared an organic-inorganic perovskite dataset storing 1346 halide perovskite structures and their crystals, band gap, dielectric constant and relative energies [81]. Nevertheless, more types of A-site molecules and ions should be included in this database to further increase the database size. Materials projects store a number of halide perovskite materials with their crystal structure, band gap, formation energy, mechanical properties and surface/interface features obtained via the first-principles calculation software Vienna Ab-initio Simulation Package (VASP). Alternatively, researchers can directly retrieve useful halide perovskite information from the published papers; however, this is typically associated with the labor intensive manual editing process, and the data of the chemical names, properties, synthetic procedure and performance can be accumulated via the manual inspection of the literature. In addition, the information extraction can be accelerated by the text mining process such as those using ChemDataExtractor. Table 1 summarizes the several experimental, computational and published paper-based databases storing structures or properties of halide perovskite materials for future development.

4. Descriptors

Several important structural features to describe the halide perovskite materials for the solar cell applications have been identified (Table 2); these features dictate the formability, band gap (E_g) and device power conversion efficiency values. The Goldschmidt tolerance factor and the octahedral factor are two of the most widely used structural descriptors for the band gap prediction of the halide perovskite materials. In addition, the ionic radii, tolerance factor, and octahedral

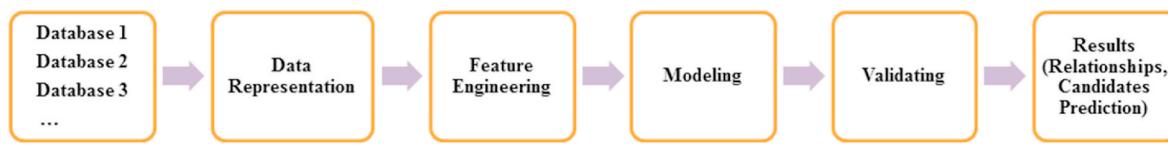


Fig. 2. Basic procedures to employ machine learning approaches.

Table 1

Exemplar databases storing information of halide perovskite materials for machine learning, where the data can be obtained from experiments (*Exp.*), calculations (*Cal.*) and literatures (*Lit.*).

Name	Type	Description	Website	Ref.
CSD, ICSD, COD	Exp.	Organic or inorganic crystal structures from single crystal X-ray diffraction experiments	https://www.ccdc.cam.ac.uk https://icsd.nist.gov/ crystallography.net	[82–84]
MP, AFLOW, OQMD	Cal.	DFT- or ML-based simulated structures and properties of different materials	https://materiasproject.org/ aflowlib.org http://oqmd.org/	[85–87]
Hybrid organic-inorganic perovskite dataset	Cal.	1346 HOIPs featuring 16 organic cations, 3 group-IV cations and 4 halide anions	https://www.nature.com/articles/sdata201757 http://khazana.uconn.edu/	[81]
JARVIS-DFT and JARVIS-ML	Cal.	DFT- and ML-based database for 2D materials, solar cells and thermoelectric	https://www.ctcms.nist.gov/jarvisml/ https://jarvis.nist.gov/ https://github.com/usnistgov/jarvis https://www.ctcms.nist.gov/~knc6/JVASP.html	[88]
Supporting information of published papers	Lit.	Synthetic methods, additives, V_{oc} , J_{sc} , FF and overall efficiency	https://doi.org/10.1016/j.nanoen.2018.11.069	[76]

factor values are found to be important quantities to classify the halide perovskite materials and describe the structural formability [89,90]. The Goldschmidt tolerance factor is expressed by

$$t = (r_A + r_x) / \sqrt{2(r_B + r_x)} \quad (1)$$

where r_A , r_B , r_x are the ionic radii values of the A, B and X species,

Table 2

Examples of features and the corresponding target (output) types that are described by the features.

Feature	Target (output)	References
Goldschmidt tolerance factor τ , octahedral factor μ and their revised versions	Formability	[36,91,97, 104–107]
Effective atomic radius and lone pair number on A-site	Structural stability	[100]
ΔH , ΔL , HOMO, LUMO and E_g	J_{sc} , V_{oc} , FF, and PCE	[98]
Atomistic/electronic structures, halogen compositions	Formation energies, ionic/electronic dynamics	[99]
Lead-halide bond strength, covalent bonds and phenyl-phenyl interactions	Young's modulus, Stiffness and hardness	[103]
Unified figure of merit	Light absorption	[101]

respectively ($0.9 < t \leq 1$ for cubic and $0.8 < t \leq 0.9$ for tetragonal or orthorhombic).

Another parameter, the octahedral factor, is expressed by Ref. [90]:

$$u = r_B / r_x \quad (2)$$

This is frequently employed to describe the perovskite structure and effectively predict the formability of perovskite structures. The Goldschmidt tolerance factors between 0.98 and 1 and the octahedral factors between 0.45 and 0.7 are used to screen the stable cubic perovskite structures [91]. For example, Park et al. employ the machine learning techniques to help establish the relationship between octahedral distortion and the perovskite band gap values [92].

The octahedral tilting is found to strongly affect the band gap values. The band gap increase, corresponding to a blue shift in the UV-vis absorption spectra, is suggested to be contributed by the octahedral tilting [93]. The “no-rattling” principle proposed by Goldschmidt successfully describes the perovskite structures with the prediction fidelity of 80% [94].

The tolerance factor sometimes fail to evaluate the perovskite formability, and many efforts have been tried to revise the tolerance factor for the perovskite structure prediction, including the anion-dependent ionic radii of the cations [95,96], and other new tolerance factor to predict the stability of the perovskite halide materials [97]. Bartel et al. propose a revised tolerance factor expressed by:

$$\tau = \frac{r_X}{r_B} - n_A \left(n_A - \frac{r_A/r_B}{\ln(r_A/r_B)} \right) \quad (3)$$

where n_A is the oxidation state of the A site species; this improves the prediction accuracy for the halide perovskite materials [97].

ΔH , defined by the energy difference between the highest occupied molecular orbitals (HOMO) of the neighboring two layers and ΔL , defined by the energy difference between the lowest unoccupied molecular orbital (LUMO) of the neighboring two layers, correlated well with the power conversion efficiencies, while a more robust descriptor involves the consideration of the HOMO, LUMO, E_g , ΔH , and ΔL for the overall conversion efficiencies [98].

The machine learning process identifies the relationships between the octahedral structures and the anharmonicity of the halogen species [99]. Park et al. suggest that the effective atomic radii and number of the lone pairs in the A-site are adequate to predict the perovskite stability [100]. The difference between the energy values of the cubic and the optimized perovskite structures is employed to describe the phase stability of perovskites [100]. The band gap value as the primary descriptor to predict the efficiencies is challenged, and multiple factors including the stability, optical absorption, and carrier lifetime should be accounted for [101]. In addition, 18 physical descriptors are recognized to explore the band gap values via random forest classification for the sodium- and lithium-based halide perovskites [102].

The possible features to accurately describe the mechanical properties of the halide perovskite materials are examined by Tu et al. The halogen species significantly influence the B-X bond length of the framework and the Young's modulus (Cl > Br > I). Introducing the ring-ring interactions to substitute the van der Waals cation interactions leads to the enhanced stiffness of the two-dimensional Ruddlesden-Popper phases [103].

5. Lead-free halide perovskites

The machine-learning technique expedites the lead-free halide perovskite discovery process [108,109]. Lu and Wang et al. employ the machine learning method to identify the lead-free halide perovskites, exploring a dataset consisting of 5158 potential perovskites that are trained using a DFT-based dataset of 212 halide perovskites (Fig. 3a–c). The feature engineering process selects and ranks 14 features via the Pearson correlation coefficient matrix. This leads to the identification of six lead-free halide perovskite materials, with two of them corresponding to the excellent light harvesting and stability performance. The 14 features selected from the feature engineering process confirm the tolerance factor, the total number of ionic charge and the octahedral factor as three of the most important features dictating the band gap values [78]. In a similar process, eight additional perovskites are selected with optimal machine-learned band gaps [107]. The titanium-incorporated halide perovskites have been identified to be suitable halide perovskite alternatives for perovskite solar cells [110, 111]. The titanium-incorporated halide perovskites with the A_2TiX_6 and $ATiX_3$ formula are investigated via the machine learning technique, because of their cuboctahedral stability and lessened interior stress; the bromine species are identified to outperform the other halogen counterparts for the structural stability [23]. High-throughput synthesis coupled with a deep neural network classifies the perovskites and the non-linear band gap identified in $Cs_3(Bi_{1-x}Sb_x)_2(I_{1-x}Br_x)_9$ [112]. Mannodi-Kanakkithodi et al. [113] employ the machine learning technique to identify that the energetics and energy levels of the defect introduced by the metal substitution can be described by a few bromide-related properties, and that the experimentally-verified halide perovskites with Bi, Sc, Ni, and Zr presents improved film morphology and enhanced in electron/hole lifetimes. Im et al. employ gradient-boosted regression trees to find the lead-free perovskite solar cell materials and estimate the possible formation and energy gap values [114]. Chen et al. employ $CsSn_{0.5}Ge_{0.5}I_3$ as the photoactive layer in a

perovskite solar cell, and achieve a PCE of 7.11% with decent stability; this is ascribed to the formation of a native-oxide layer that passivates the perovskite surface. The employment of the $CsSn_{0.5}Ge_{0.5}I_3$ halide perovskite materials is predicted by the electron/hole mobility and the electron/hole trap density calculations [115,116]. Wu et al. predict the lead-free halide perovskite materials from a database consisting of 230808 entries of halide perovskite structures, resulting in 686 machine-learned (gradient boosting regression, support vector regression and kernel ridge regression) outputs and 132 of them validated by the density functional theory calculations [117].

The double halide perovskites have the lead species replaced by the mixed monovalent and trivalent cations. A revised tolerance factor τ is employed to identify 23,314 double perovskites, with 91% accuracy for 1034 experimentally verified perovskites and the probability of the formation of the new double perovskite are determined [97]. Li and Yin et al. employ a revised machine learning model to investigate the stability of halide double perovskites based on the decomposition energies obtained by the first-principles calculations (Fig. 3d), which outperforms the tolerance factor-based method [74].

6. Machine learning other halide perovskites

The machine learning technique is coupled with the density functional theory to evaluate the factors that determine the band gaps of the halide perovskite materials [118,119]. Park et al. identify 151 ferroelectric perovskites possessing optimal band gap values with eight for the single-junction solar cell [100]. Saidi et al. employ a hierarchical convolutional neural network to investigate the electronic properties of the halide perovskite materials, using neutral network element to calculate the proper structural and electronic features [120]. The perovskites with the formula ABX_3 and $(A'A'')(B'B'')X_6$ that are available in experiments and the Materials Projects database are analyzed, with 16 machine learning algorithms identifying several suspicious formability values [121]. The power conversion efficiencies could be predicted via

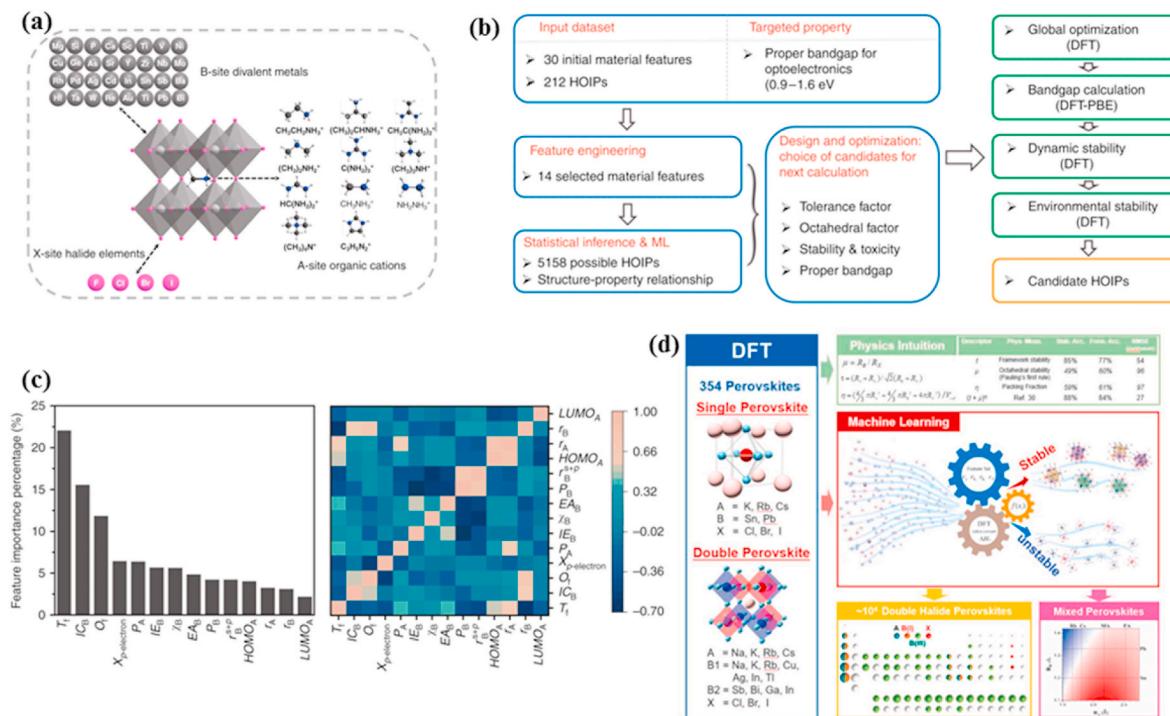


Fig. 3. (a) Representation of ABX_3 with different moieties. (b) The flowchart of the machine-learning-accelerated process to discover the lead-free halide perovskites. (c) The feature engineering process selects and ranks 14 features via the Pearson correlation coefficient matrix [78]. Reproduced with permission from Refs. [78]. Copyright Macmillan Publishers Limited (2018). (d) Combined machine learning and density functional theory-based method to screen halide double perovskites, based on the decomposition energies obtained by the first-principles calculations [74]. Reproduced with permission from Ref. [74]. Copyright Wiley (2019).

the band gap values; however, the band-gap method rather crude and DFT-based methods is the spectroscopic limited maximum efficiency (SLME) is employed to better predict the solar cell overall conversion efficiency. Based on this, Li et al. employ a machine learning method to accelerate the prediction of SLME that is time-consuming to obtain via the density functional theory [88]. The Bayesian optimization is employed to design the halide perovskite materials with the optimal combination of cations, halides and solvents [122]. Choubisa and Sargent et al. propose a crystal site feature embedding with convolutional neural networks (CNNs) and extensive deep neural networks (EDNNs) to accelerate the prediction of the band gaps and formation energies of the halide perovskites with low mean absolute errors (Fig. 4a) [123]. The band gap values of the halide perovskites are predicted via the machine learning techniques based on the atomic arrangement [124]. Wang et al. employ Bayesian optimization to discover the potential halide perovskite light absorbers via a revised figure of merit rather than the biased band gap values [101]. The random forest classification is employed to estimate the band gaps of the halide perovskite materials, with eleven lithium- and sodium-based perovskite materials identified to be proper for the solar cell applications [102]. The support vector machine-based machine learning method is used to identify new halide perovskites materials based on the steric features [89].

Apart from the band gap values, various other structure and property quantities have been employed to expedite the machine learning process [125,126]. Li and Thomas et al. employ the machine learning technique to optimize the compositions and rationalize the design strategies to achieve the high-performance perovskite solar cells. A database obtained from 2000 publications is constructed to perform the machine learning process, while the ΔH and ΔL values for higher power conversion efficiencies are found to differ in the high- and low-band gap perovskites [98]. A neural network is employed to facilitate the optimization of Ag core-shell structures inside the halide perovskite film with improved optical absorption properties [127]. The genetic algorithm with Nelder-Mead method is employed to estimate the suitable parameters for the perovskite solar cell performance, which quantifies the ideality factors that dictate the efficiencies [128].

Based on the published papers, Odabaşı and Yıldırım [76] employ machine learning to analyze the experimental details and device performance of perovskite solar cells that are stored in the literature from 2013 to 2018, and confirm that the machine learning techniques not

only help develop the new halide perovskite materials, but also facilitate the design of the compatible neighboring materials and the fabrication routes. This facilitates the selection of suitable HTL, ETL, solvents, anti-solvents and spinning times. Suitable processing conditions of the perovskite solar cells are optimized via the machine learning model [129].

The machine learning technique is combined with various experimental techniques to dynamically evaluate the functions of the halide perovskite materials. Ali et al. employ the machine learning technique, X-ray diffraction and scanning electron microscope to investigate the perovskite structure recovery of the double-cation perovskites in the cubic phase and optimize the cation mixing via the addition of <10 mol % cesium additive [130]. The machine learning method is employed to detect and optimize the suitable conditions of the dynamic performance (Reap-Rest-Recovery Cycle) of the halide perovskite materials (Fig. 4b) [131]. The crystallization of the halide perovskite materials is critical to realize the improved film quality and the reduced recombination. The machine learning method with the convolutional neural network is employed to analyze the crystal images and optimize the synthetic conditions to guide the crystallization process; this results in the acquirement of a new single crystal halide perovskite ($(3\text{-PLA})_2\text{PbCl}_4$) [132]. The logistic regression algorithm is applied to analyze the size distributions of the perovskite crystals and classifies the pixels of perovskite materials [133].

Jinnouchi and Bokdam et al. develop the machine-learning force fields to facilitate the simulation of the phase transition of the hybrid perovskite materials (Fig. 4c) [134]. The machine-learned force fields expedite the materials simulation process since they intelligently decide whether the first-principles calculations during the simulation stage should be neglected via the prediction of the Bayesian error.

7. Stability

In general, the stability is expressed by the formability values, which can be calculated by the high-throughput calculations and expressed as the output values for the machine learning process. The machine learning method helps understand the association between decomposition energy and the structural factors such as ionic radius [88]. Based on the extracted information from the published papers, the optimal additives, the electron transporting layer (ETL)/hole transporting layer

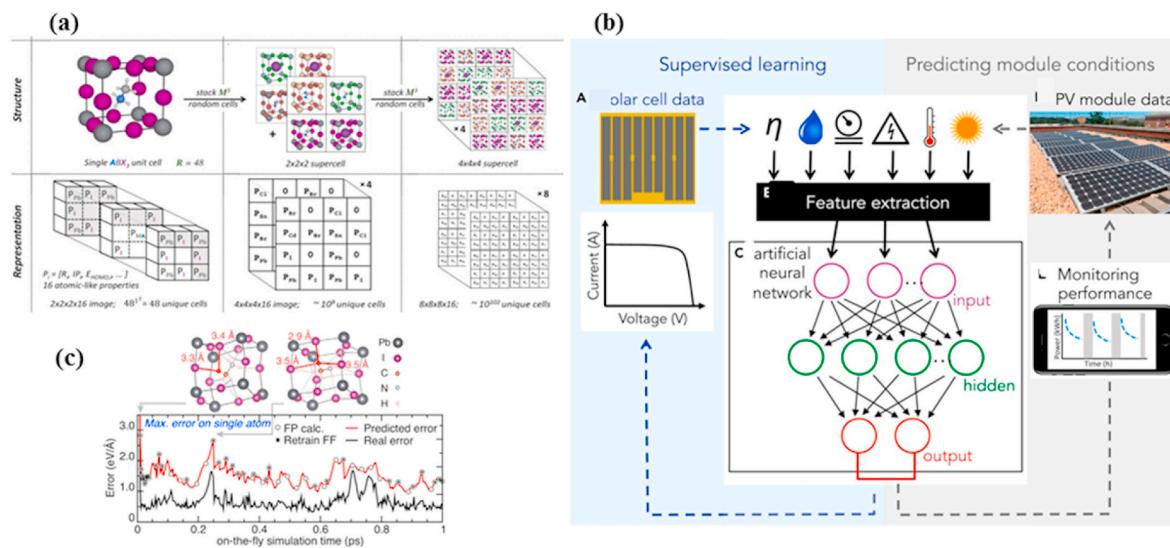


Fig. 4. (a) Crystal site feature embedding is applied to predict possible halide perovskite candidates with suitable band gaps for light absorption [123]. Reproduced with permission from Ref. [123]. Copyright Elsevier Inc. (2020). (b) An artificial neural network to facilitate the detection and optimization of the dynamic performance (Reap-Rest-Recovery Cycle) of halide perovskite materials [131]. Reproduced with permission from Refs. [131]. Copyright Elsevier Inc. (2019). Development of machine-learning force fields to decide whether first-principles calculations are necessary via the Bayesian error. The machine-learning force fields identify the phase transition temperatures of the halide perovskites [134]. Reproduced with permission from Refs. [134]. Copyright APS (2019).

(HTL) materials and the fabrication methods that improve the stability and impede the degradation are revealed by the association rule mining and decision trees [79]. The formability of the titanium-based halide perovskites are investigated based on the tolerance and octahedral factors, with the dataset prepared via the decision tree classifier [23]. The association rule mining is employed to evaluate the hysteresis data, while the optimal precursor solution, spin coating steps and HTL dopants are found to strongly influence the hysteresis [135].

The moisture and oxygen degradations are significant in halide perovskite solar cells and various surface and interface strategies have been proposed [77,136–139]; this can be accelerated via the data-driven and machine learning methods. For example, the passivating molecules are incorporated onto the halide perovskite layer, where different machine learning models are employed to identify critical structural parameters such as hydrogen bond donors and acceptors, steric bulk, secondary, tertiary amines, and pyridine derivatives that are compatible with the halide perovskite film and possibly alleviate moisture and oxygen degradation [77]. The employment of the two-dimensional halide perovskite materials, which can be identified efficiently via the machine learning and data mining methods [140], is suggested to improve the overall stability performance including the water and oxygen stability. The A-site long alkyl chain in the halide perovskite materials is suggested to repel the water molecules via the hydrophobic chain and improve the water stability [141], and the machine learning techniques are recommended to be used in the future to identify suitable long alkyl chains with proper geometrical factors via the calculation of their thermodynamic stabilities. In addition, the machine-learning-assisted design of suitable hole transport materials and electron transport materials is equally important to improve the moisture and oxygen stability [142]. The machine learning techniques are employed to monitor the impact of environmental stressors on perovskite modules, in which the intrinsic and extrinsic parameters (H_2O , O_2 , bias, temperature, and light) are effectively tracked via the artificial neural networks [131, 143]. Mining the perovskite solar cell literature has given valuable insights on the perovskite stabilities. Different solvent engineering process and different types of charge transport layer materials and additives can be associated with the moisture and oxygen stability of the perovskite solar cells via the machine learning methods, which predict excellent materials ingredients to achieve the honorable target [76,131]. Certain charge transport materials such as PEDOT:PSS is found to be especially unstable in the precursor solution in the presence of the water molecules [76]. More accurate relationships between the input values (descriptors) and the output values (such as specific oxygen and moisture stabilities) should be established; this requires the establishment of more comprehensive and updated databases of perovskite solar cells, while the machine learning studies should specifically accumulate the quantified data of the oxygen and moisture stabilities in the literature [74,88]. The thermodynamic stability in terms of the decomposition energy is another important parameter to identify the appropriate halide perovskite materials via the machine learning methods [144]. Modifying the charge transport layer can help improve the heat stability of perovskite solar cells [142]. The decomposition energy can be calculated via the first-principles calculations, and the octahedral factor and the atomic packing fraction are well related with the thermodynamic stability [145, 146].

8. Dimensional tailoring

The design of the dimensionally tailored halide perovskites is realized via the combination of the high-throughput experiments and the machine learning technique. With the experimentally obtained 75 halide perovskite materials, the deep neural network is employed to classify 0D, 2D, and 3D halide perovskites, reaching 90% accuracy (Fig. 5a). The input of the algorithms is the simulated and experimental powder X-ray diffraction patterns along with a list of precursor elements and expected crystal symmetry for the target compounds while the

output is the crystallographic dimensionality of the experimentally synthesized compounds. Four $A_3B_2Br_9$ inorganic perovskites are obtained while a new $Cs_3(Bi_{1-x}Sb_x)_2(I_{1-x}Br_x)_9$ candidate is fabricated [112]. Braham et al. use the machine learning method to develop the synthetic routes to obtain the two-dimensional $CsPbBr_3$ nanoplatelets and the corresponding crystal growth process mediated by the ligands via SVM (Fig. 5b) [147]. It would be interesting to envisage a robotic synthetic setup to provide various precursors and conditions to make new perovskites with tailored dimensions, chirality and performances that can be accelerated by the machine learning process.

9. Additives

The machine learning techniques have been employed to design the molecular capping agents to passivate the halide perovskite materials. Yu et al. use different machine learning algorithms to find the optimal molecular species for post-treating the halide perovskite material and unveil important structural descriptors of the molecules compatible with the halide perovskite substrates. The outputs label to quantify the film damage is determined by the threshold value of the absorptivity in the UV-vis absorption spectra of the molecule/perovskite composite systems. This confirms the important compatibility factors such as the fewer hydrogen bond donors/acceptors, the availability of the secondary/tertiary amines and the pyridine moieties (Fig. 6a–b) [77]. Hartono et al. employ the random forest regression to screen the capping layer types to prevent the degradation of the perovskite films. When 21 organic halide salts on $CH_3NH_3PbI_3$ are considered, the small number of hydrogen-bond donors and the small topological polar surface area of the capping layer are associated with the prerequisites of the capping layer structures (Fig. 6c). This results in the employment of an capping agent phenyltriethylammonium iodide that leads to the improved stability than that offered by the traditional octylammonium bromide [138]. In addition, based on the numerical and textual data stored the published papers, the stability of the perovskite solar cells is analyzed by the association rule mining and the decision trees; the optimal additives, ETL/HTL materials and fabrication methods that improve the stability and impede the degradation are revealed [76,79].

10. Suggestions

More comprehensive databases storing accurate structures and properties of the halide perovskites are critical for the machine learning process to generate ideal halide perovskite candidates for solar cell applications and unveil the structure-property relationships. The databases should be made public or stored in the supporting information in a readable format. This should be particularly addressed by the high-throughput calculation and high-throughput experimental studies where a readable source data format (such as the csv format) is often missing [148]. The machine learning techniques can also identify suspicious data that are stored in a databases [121]. More efforts should be made to employ the machine learning techniques to remove the incorrect data from the perovskite databases. In addition, the control device should be emphasized to obtain the reasonable machine learned models, since the different efficiencies of the control devices from different research groups can introduce significant biases on the effectiveness of the new materials. More experimental studies should incorporate the machine learning techniques, such that the high-throughput experimental data could be timely analyzed via the data-driven method to expedite the materials design process. The combined work of the high-throughput calculation, machine-learning, data mining and experiments facilitates the establishment of a “perovskite genome project”, which encompasses the comprehensive structural, mechanical, electronic and optical properties of the halide perovskite materials.

There have been a limited number of machine learning papers predicting the combinations of halide perovskite materials. The selection of different material combinations dictates the device performance, since

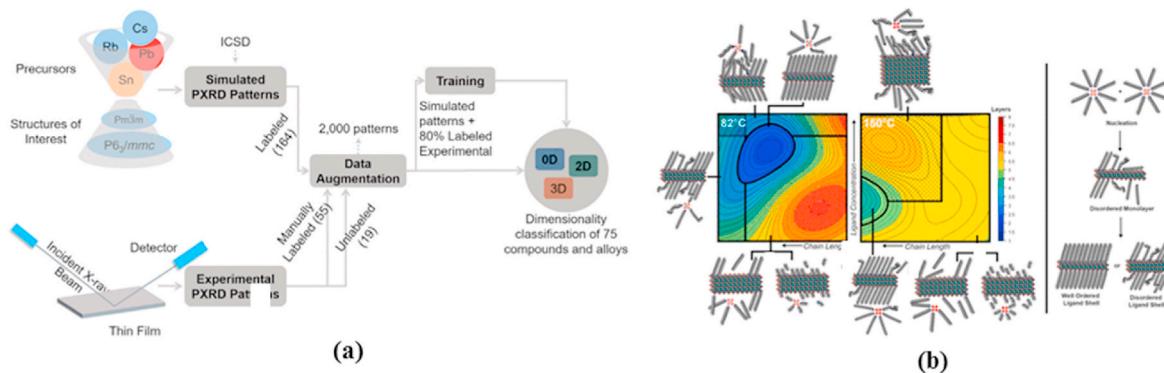


Fig. 5a. (a) A schematic workflow to employ the machine-learning algorithms to assist the structural characterization for the perovskite-inspired materials and guide the analysis of the diagnosis results. The input of the algorithms is the simulated and experimental powder X-ray diffraction patterns along with a list of precursor elements and expected crystal symmetry for the target compounds [112]. Reproduced with permission from Ref. [112]. Copyright Elsevier Inc. (2019). (b) Machine-learning assisted design strategy to obtain CsPbBr_3 nanoplatelets and the corresponding crystal growth process mediated by the ligands. Copyright ACS [147]. Reproduced with permission from Ref. [147]. Copyright ACS (2019).

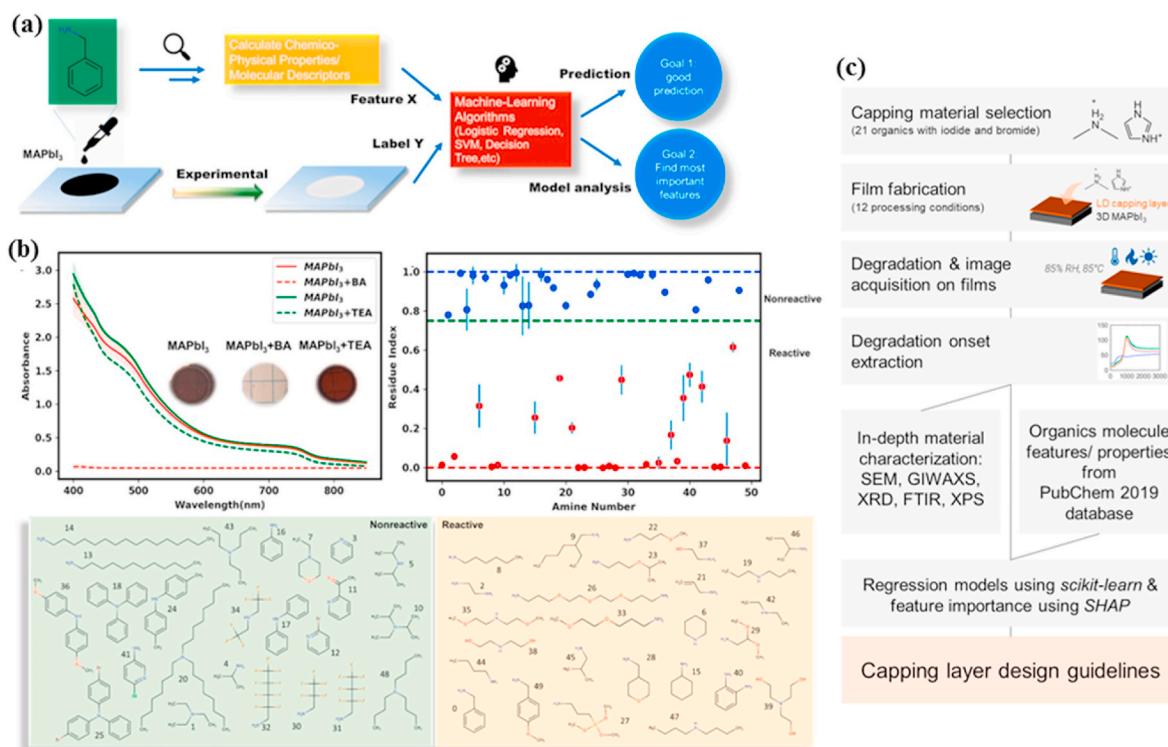


Fig. 6. (a) Machine learning-assisted investigations of the post-treatment amine molecules for the passivation of the halide perovskite materials. The outputs label to quantify the film damage is expressed by the evaluation of the absorptivity in the UV-vis absorption spectra. (b) The UV-vis absorption spectra of the molecule/perovskite systems, the categories of residue index to define the labels for the machine learning process, and the input parameters with amine terminal groups for the post-treatment amine molecules [77]. Reproduced with permission from Refs. [77]. Copyright ACS (2019). (c) A general machine learning design guideline of the capping layers, which helps the selection of the optimal capping layer types to prevent the degradation of the perovskite films [138]. Reproduced with permission from Ref. [138]. Copyright chemrxiv.org (2020).

the interfaces and the energy offsets are important descriptors for the halide perovskite solar cell performance. For example, the choices of the hole transporting materials, electron transporting materials, additives, passivating layers, interfacial modifiers, nearby electrode materials and deposition sequence are equally important for the high-performance perovskite solar cell.

The majority of the machine learning-based halide perovskite studies are associated with the formability and band gap calculations. A series of other parameters and output values should be considered, such as the light harvesting performance, overall conversion efficiency and water stability. For example, the SLME-determined values could be employed

to predict the power conversion efficiencies for the machine learning purposes. The modeling of the excited state structures and properties of halide perovskites could be optimized via the machine learning method. The excited state rather than the ground state is the de-facto working state of the halide perovskite materials in perovskite solar cells and light-emitting diodes. For example, the machine learning technique has been applied to the non-adiabatic molecular dynamic study of MAPbI_3 and the I-I-I angle is found to be critical parameter that determines the electron-phonon coupling and the band gap values [149]. Apart from the halide perovskite materials, other types of new perovskite structures have been investigated via the machine learning technique, such as

ABO_3 [127,150–156], ABN_3 [157] and charcogenide perovskites [158]. These new types of perovskite structures identified by the machine learning methods are suggested to exhibit improved stability and environmental friendliness that remain as open issues for the halide perovskites.

Since the halide perovskites demonstrate promising prospects for the new energy conversion and storage applications, new descriptors should be developed to design the halide perovskite materials for new energy applications, such as lithium-ion batteries, photo-batteries and catalysts [159–162]. For example, the lithium-ion batteries require strong interactions between the halide perovskites and the lithium ions while the photo-rechargeable batteries require the light-driven lithium ion movement and the photo-charging capability, which should be described by more accurate structural descriptors to proceed with the machine learning process.

The text mining techniques that incorporate the machine learning method to deal with the large number of data could be employed to predict new halide perovskite materials and unveil latent knowledge from the literature. For example, the text mining process coupled with the machine learning technique has been demonstrated to successfully predict new functional materials and hidden relationships via the text vectorization [163]. The image recognition is another technique to effectively understand and predict new materials via a series of knowledge-containing figures [164–167]. A larger number of data could be retrieved from the paper images from the experimental (X-ray diffraction, Fourier transform infrared spectrometer, scanning electron microscope, transmission electron microscopy and scanning tunneling microscope) and the computational results (such as highest occupied molecular orbital, lowest unoccupied molecular orbital distribution, band structures and density of states) of halide perovskite materials. These are particularly pertinent since there are extensive text- and figure-based contents in the halide perovskite and materials science literature.

Interestingly, not only the artificial intelligence can benefit the halide perovskite development, but the advance in halide perovskite materials can also benefit the artificial intelligence community in reward. For instance, the incorporation of the halide perovskite materials has been successfully implemented to design the neuromorphic learning and artificial synapse, in which the artificial neuromorphic system mimics the biological synapses [168–177]. This is because the metal halide perovskite-based synapse has achieved sufficiently low energy consumption similar to the energy consumption of the real biological synapses [40]. The efficient ion migration in halide perovskite materials is suggested to contribute to the low energy consumption [178]. It is important to find new halide perovskite alternatives for artificial synapses for neuromorphic learning and intelligence.

11. Conclusion

The machine learning techniques have been successfully employed to identify suitable halide perovskite candidates for the solar cell applications, addressing the lead-free and stability issues. The machine learning techniques efficiently deal with the large number of experimental and computational data of halide perovskite materials. In addition, the hidden relationships such as the structure-property relationships are effectively established via the machine learning by selecting suitable structural descriptors. The machine learning technique also extends to various other compatible materials such as the capping additives to further improve the performance of the halide perovskite materials. A “halide perovskite materials genome” project is called for, which requires the preparation of a more comprehensive database storing the structures, properties and device performance of the halide perovskite materials obtained from the high-throughput experiments and the high-throughput calculations.

Author contributions

The author has given approval to the final version of the manuscript.

Declaration of competing interest

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

Acknowledgements

This work was supported by the National Natural Science Foundation of China (No. 51702165). The authors acknowledge computational support from NSCCSZ Shenzhen, China.

References

- [1] F. Sahli, J. Werner, B.A. Kamino, M. Bräuninger, R. Monnard, B. Paviet-Salomon, L. Barraud, L. Ding, J.J. Diaz Leon, D. Sacchetto, G. Cattaneo, M. Despesse, M. Boccard, S. Nicolay, Q. Jeangros, B. Niesen, C. Ballif, Fully textured monolithic perovskite/silicon tandem solar cells with 25.2% power conversion efficiency, *Nat. Mater.* 17 (2018) 820–826, <https://doi.org/10.1038/s41563-018-0154>.
- [2] M. Saliba, Polyelemental, multicomponent perovskite semiconductor libraries through combinatorial screening, *Adv. Energy Mater.* 9 (2019), 1803754, <https://doi.org/10.1002/aenm.201803754>.
- [3] A. Bayon, A. de la Calle, K.K. Ghose, A. Page, R. McNaughton, Experimental, computational and thermodynamic studies in perovskites metal oxides for thermochemical fuel production: a review, *Int. J. Hydrogen Energy* 45 (2020) 12653–12679, <https://doi.org/10.1016/j.ijhydene.2020.02.126>.
- [4] M. Crespo-Quesada, L.M. Pazos-Outón, J. Waranan, M.F. Kuehnle, R.H. Friend, E. Reisner, Metal-encapsulated organolead halide perovskite photocathode for solar-driven hydrogen evolution in water, *Nat. Commun.* 7 (2016) 12555, <https://doi.org/10.1038/ncomms12555>.
- [5] Y. Wu, P. Wang, Z. Guan, J. Liu, Z. Wang, Z. Zheng, S. Jin, Y. Dai, M.-H. H. Wangbo, B. Huang, Enhancing the photocatalytic hydrogen evolution activity of mixed-halide perovskite $\text{CH}_3\text{NH}_3\text{PbBr}_{3-x}\text{I}_x$ achieved by bandgap funneling of charge carriers, *ACS Catal.* 8 (2018) 10349–10357, <https://doi.org/10.1021/acs.catal.8b02374>.
- [6] Q. Fu, C. Draxl, Hybrid organic-inorganic perovskites as promising substrates for Pt single-atom catalysts, *Phys. Rev. Lett.* 122 (2019), 046101, <https://doi.org/10.1103/PhysRevLett.122.046101>.
- [7] J. Xu, O. Vožny, R. Comin, X. Gong, G. Walters, M. Liu, P. Kanjanaboons, X. Lan, E.H. Sargent, Crosslinked remote-doped hole-extracting contacts enhance stability under accelerated lifetime testing in perovskite solar cells, *Adv. Mater.* 1–9 (2016), <https://doi.org/10.1002/adma.201505630>.
- [8] Y. Bai, Q. Dong, Y. Shao, Y. Deng, Q. Wang, L. Shen, D. Wang, W. Wei, J. Huang, Enhancing stability and efficiency of perovskite solar cells with crosslinkable silane-functionalized and doped fullerene, *Nat. Commun.* 7 (2016) 12806, <https://doi.org/10.1038/ncomms12806>.
- [9] N.Z. Koocher, D. Saldana-Greco, F. Wang, S. Liu, A.M. Rappe, Polarization dependence of water adsorption to $\text{CH}_3\text{NH}_3\text{PbI}_3$ (001) surfaces, *J. Phys. Chem. Lett.* 6 (2015) 4371–4378, <https://doi.org/10.1021/acs.jpclett.5b01797>.
- [10] C. Müller, T. Glaser, M. Plogmeyer, M. Sendner, S. Döring, A.a. Bakulin, C. Brzuska, R. Scheer, M.S. Pshenichnikov, W. Kowalsky, A. Pucci, R. Lovrincic, Water infiltration in methylammonium lead iodide perovskite: fast and inconspicuous, *Chem. Mater.* 27 (2015) 7835–7841, <https://doi.org/10.1021/acs.chemmater.5b03883>.
- [11] T. Wu, Y. Wang, X. Li, Y. Wu, X. Meng, D. Cui, X. Yang, L. Han, Efficient defect passivation for perovskite solar cells by controlling the electron density distribution of donor-π-acceptor molecules, *Adv. Energy Mater.* 9 (2019), 1803766, <https://doi.org/10.1002/aenm.201803766>.
- [12] Y. Chen, L. Zhang, J. Miao, J. Li, Dye-sensitized halide perovskite: a case study on calcein dye, *Dyes Pigments* (2020), 108608, <https://doi.org/10.1016/j.dyepig.2020.108608>.
- [13] Y. Zhou, H. Zhong, J. Han, M. Tai, X. Yin, M. Zhang, Z. Wu, H. Lin, Synergistic effect of charge separation and defect passivation using zinc porphyrin dye incorporation for efficient and stable perovskite solar cells, *J. Mater. Chem. A.* 7 (2019) 26334–26341, <https://doi.org/10.1039/c9ta09369a>.
- [14] L. Himanen, A. Geurts, A.S. Foster, P. Rinke, Data-driven materials science: status, challenges, and perspectives, *Adv. Sci.* 6 (2019), 1900808, <https://doi.org/10.1002/advs.201900808>.
- [15] T. Wang, C. Zhang, H. Snoussi, G. Zhang, Machine learning approaches for thermoelectric materials research, *Adv. Funct. Mater.* 30 (2020), 1906041, <https://doi.org/10.1002/adfm.201906041>.
- [16] K.A. Brown, S. Britzman, N. Maccaferri, D. Jariwala, U. Celano, Machine learning in nanoscience: big data at small scales, *Nano Lett.* 20 (2020) 2–10, <https://doi.org/10.1021/acs.nanolett.9b04090>.

- [17] L.D. Williams, P. Ghanshyam, Machine Learning using local environment descriptors to predict new scintillator materials, *Nat. Commun.* 5 (2018) 4377, <https://doi.org/10.2172/1469502>.
- [18] G.R. Schleider, A.C.M. Padilha, C.M. Acosta, M. Costa, A. Fazzio, From DFT to machine learning: recent approaches to materials science—a review, *J. Phys. Mater.* 2 (2019), 032001, <https://doi.org/10.1088/2515-7639/ab084b>.
- [19] P. Raccuglia, K.C. Elbert, P.D.F. Adler, C. Falk, M.B. Wenny, A. Mollo, M. Zeller, S.A. Friedler, J. Schrier, A.J. Norquist, Machine-learning-assisted materials discovery using failed experiments, *Nature* 533 (2016) 73–76, <https://doi.org/10.1038/nature17439>.
- [20] J. Lee, A. Seko, K. Shitara, K. Nakayama, I. Tanaka, Prediction model of band gap for inorganic compounds by combination of density functional theory calculations and machine learning techniques, *Phys. Rev. B* 93 (2016), 115104, <https://doi.org/10.1103/PhysRevB.93.115104>.
- [21] C. Kim, G. Pilania, R. Ramprasad, From organized high-throughput data to phenomenological theory using machine learning: the example of dielectric breakdown, *Chem. Mater.* 28 (2016) 1304–1311, <https://doi.org/10.1021/acs.chemmater.5b04109>.
- [22] G.H. Gu, J. Noh, I. Kim, Y. Jung, Machine learning for renewable energy materials, *J. Mater. Chem. A* 7 (2019) 17096–17117, <https://doi.org/10.1039/C9TA02356A>.
- [23] H.A. Maddah, V. Berry, S.K. Behura, Cuboctahedral stability in Titanium halide perovskites via machine learning, *Comput. Mater. Sci.* 173 (2020), 109415, <https://doi.org/10.1016/j.commatsci.2019.109415>.
- [24] J.-P. Sun, G.C. McKeown Wessler, T. Wang, T. Zhu, V. Blum, D.B. Mitzi, Structural tolerance factor Approach to defect-resistant I 2 -II-IV-X 4 semiconductor design, *Chem. Mater.* 32 (2020) 1636–1649, <https://doi.org/10.1021/acs.chemmater.9b05107>.
- [25] J. Schmidt, L. Chen, S. Botti, M.A.L. Marques, Predicting the stability of ternary intermetallics with density functional theory and machine learning, *J. Chem. Phys.* 148 (2018), 241728, <https://doi.org/10.1063/1.5020223>.
- [26] T.D. Huan, R. Batra, J. Chapman, S. Krishnan, L. Chen, R. Ramprasad, A universal strategy for the creation of machine learning-based atomistic force fields, *Npj Comput. Mater.* 3 (2017) 37, <https://doi.org/10.1038/s41524-017-0042-y>.
- [27] A. Jha, A. Chandrasekaran, C. Kim, R. Ramprasad, Impact of dataset uncertainties on machine learning model predictions: the example of polymer glass transition temperatures, *Model. Simulat. Mater. Sci. Eng.* 27 (2019), 024002, <https://doi.org/10.1088/1361-651X/aaf8ca>.
- [28] X.-Y. Ma, H.-Y. Lyu, K.-R. Hao, X. Qian, Q.-B. Yan, G. Su, Large Family of Two-Dimensional Ferroelectric Metals Discovered via Machine Learning, 2020. <http://arxiv.org/abs/2004.08527>.
- [29] A.O. Oliynyk, A. Mar, Discovery of intermetallic compounds from traditional to machine-learning approaches, *Acc. Chem. Res.* 51 (2018) 59–68, <https://doi.org/10.1021/acs.accounts.7b00490>.
- [30] R. Ramprasad, R. Batra, G. Pilania, A. Mannodi-Kanakkithodi, C. Kim, Machine learning in materials informatics: recent applications and prospects, *Npj Comput. Mater.* 3 (2017) 54, <https://doi.org/10.1038/s41524-017-0056-5>.
- [31] F. El-Melouhi, A. Marzouk, E.T. Bentria, S.N. Rashkeev, S. Kais, F.H. Alharbi, Hydrogen bonding and stability of hybrid organic-inorganic perovskites, *ChemSusChem* 9 (2016) 2648–2655, <https://doi.org/10.1002/cssc.201600864>.
- [32] D. Ghosh, E. Welch, A.J. Neukirch, A. Zakhidov, S. Tretiak, Polarons in halide perovskites: a perspective, *J. Phys. Chem. Lett.* 11 (2020) 3271–3286, <https://doi.org/10.1021/acs.jpclett.0c00018>.
- [33] M. Baranowski, P. Plochocka, Excitons in metal-halide perovskites, *Adv. Energy Mater.* (2020), 1903659, <https://doi.org/10.1002/aenm.201903659>.
- [34] G. Volonakis, A.A. Haghighirad, R.L. Milot, W.H. Sio, M.R. Filip, B. Wenger, M. B. Johnston, L.M. Herz, H.J. Snaith, F. Giustino, $\text{Cs}_2\text{InAgCl}_6$: a new lead-free halide double perovskite with direct band gap, *J. Phys. Chem. Lett.* 8 (2017) 772–778, <https://doi.org/10.1021/acs.jpclett.6b02682>.
- [35] Z. Xiao, K.-Z. Du, W. Meng, J. Wang, D.B. Mitzi, Y. Yan, Intrinsic instability of $\text{Cs}_2\text{In(I)M(III)X}_6$ ($M = \text{Bi, Sb}$; $X = \text{halogen}$) double perovskites: a combined density functional theory and experimental study, *J. Am. Chem. Soc.* 139 (2017) 6054–6057, <https://doi.org/10.1021/jacs.7b02227>.
- [36] C.-X. Zhang, T. Shen, D. Guo, L.-M. Tang, K. Yang, H.-X. Deng, Reviewing and Understanding the Stability Mechanism of Halide Perovskite Solar Cells, *InfoMat*, 2020, <https://doi.org/10.1002/inf2.12104>.
- [37] M.-G. Ju, M. Chen, Y. Zhou, J. Dai, L. Ma, N.P. Padture, X.C. Zeng, Toward eco-friendly and stable perovskite materials for photovoltaics, *Joule* 2 (2018) 1231–1241, <https://doi.org/10.1016/j.joule.2018.04.026>.
- [38] I. Lignos, S. Stavraklis, G. Nedelcu, L. Protesescu, A.J. DeMello, M.V. Kovalenko, Synthesis of cesium lead halide perovskite nanocrystals in a droplet-based microfluidic platform: fast parametric space mapping, *Nano Lett.* 16 (2016) 1869–1877, <https://doi.org/10.1021/acs.nanolett.5b04981>.
- [39] C.H. Hendon, R.X. Yang, L.A. Burton, A. Walsh, Assessment of polyanion (BF_4^- and PF_6^-) substitutions in hybrid halide perovskites, *J. Mater. Chem. A* 3 (2015) 9067–9070, <https://doi.org/10.1039/C4TA05284F>.
- [40] J. Wu, X. Li, Y. Zhao, L. Liu, W. Qu, R. Luo, R. Chen, Y. Li, Q. Chen, Interface engineering in solid state Li metal batteries by quasi-2D hybrid perovskites, *J. Mater. Chem. A* 6 (2018) 20896–20903, <https://doi.org/10.1039/C8TA07643J>.
- [41] Z. Fang, H. He, L. Gan, J. Li, Z. Ye, Understanding the role of lithium doping in reducing nonradiative loss in lead halide perovskites, *Adv. Sci.* 5 (2018), 1800736, <https://doi.org/10.1002/advs.201800736>.
- [42] C. Monblona, O. Malinkiewicz, C. Roldán-Carmona, A. Soriano, L. Gil-Escríg, E. Bandiello, M. Scheepers, E. Edri, H.J. Bolink, Efficient methylammonium lead iodide perovskite solar cells with active layers from 300 to 900 nm, *Apl. Mater.* 2 (2014), 081504, <https://doi.org/10.1063/1.4890056>.
- [43] D.-N. Jeong, D.-K. Lee, S. Seo, S.Y. Lim, Y. Zhang, H. Shin, H. Cheong, N.-G. Park, Perovskite cluster-containing solution for scalable D-bar coating toward high-throughput perovskite solar cells, *ACS Energy Lett.* 4 (2019) 1189–1195, <https://doi.org/10.1021/acsenergylett.9b00042>.
- [44] M. Righetto, D. Meggiolaro, A. Rizzo, R. Sorrentino, Z. He, G. Meneghesso, T. C. Sum, T. Gatti, F. Lamberti, Coupling halide perovskites with different materials: from doping to nanocomposites, beyond photovoltaics, *Prog. Mater. Sci.* 110 (2020), 100639, <https://doi.org/10.1016/j.jpmatsci.2020.100639>.
- [45] O.S. Game, G.J. Buchsbaum, Y. Zhou, N.P. Padture, A.I. Kingon, Ionic matter: description of the anomalous electronic behavior in methylammonium lead halide perovskite devices, *Adv. Funct. Mater.* 27 (2017), 1606584, <https://doi.org/10.1002/adfm.201606584>.
- [46] A. Krishna, S. Gottis, M.K. Nazeeruddin, F. Sauvage, Mixed dimensional 2D/3D hybrid perovskite absorbers: the future of perovskite solar cells? *Adv. Funct. Mater.* 29 (2019), 1806482, <https://doi.org/10.1002/adfm.201806482>.
- [47] G. Grancini, C. Roldán-Carmona, I. Zimmermann, E. Mosconi, X. Lee, D. Martineau, S. Narbey, F. Oswald, F. De Angelis, M. Graetzel, M. K. Nazeeruddin, One-Year stable perovskite solar cells by 2D/3D interface engineering, *Nat. Commun.* 8 (2017) 15684, <https://doi.org/10.1038/ncomms15684>.
- [48] Z. wen Kwang, C.W. Chang, T.Y. Hsieh, T.C. Wei, S.Y. Lu, Solvent-modulated reaction between mesoporous PbI_2 film and $\text{CH}_3\text{NH}_3\text{I}$ for enhancement of photovoltaic performances of perovskite solar cells, *Electrochim. Acta* 266 (2018) 118–129, <https://doi.org/10.1016/j.electacta.2018.02.026>.
- [49] T. Li, Y. Pan, Z. Wang, Y. Xia, Y. Chen, W. Huang, Additive engineering for highly efficient organic–inorganic halide perovskite solar cells: recent advances and perspectives, *J. Mater. Chem. A* 5 (2017) 12602–12652, <https://doi.org/10.1039/C7TA01798G>.
- [50] Y. Li, L. Ji, R. Liu, C. Zhang, C.H. Mak, X. Zou, H.-H.H. Shen, S.-Y.Y. Leu, H.-Y. Y. Hsu, A review on morphology engineering for highly efficient and stable hybrid perovskite solar cells, *J. Mater. Chem. A* 6 (2018) 12842–12875, <https://doi.org/10.1039/c8ta04120b>.
- [51] Y. Zhao, K. Zhu, Solution chemistry engineering toward high-efficiency perovskite solar cells, *J. Phys. Chem. Lett.* 5 (2014) 4175–4186.
- [52] N.J. Jeon, J.H. Noh, Y.C. Kim, W.S. Yang, S. Ryu, S. Il Seok, Solvent engineering for high-performance inorganic–organic hybrid perovskite solar cells, *Nat. Mater.* 13 (2014) 897–903, <https://doi.org/10.1038/nmat4014>.
- [53] N.J. Jeon, J.H. Noh, W.S. Yang, Y.C. Kim, S. Ryu, J. Seo, S. Il Seok, Compositional engineering of perovskite materials for high-performance solar cells, *Nature* 517 (2015) 476–480, <https://doi.org/10.1038/nature14133>.
- [54] D. Bi, W. Tress, M.I. Dar, P. Gao, J. Luo, C. Renévier, K. Schenk, A. Abate, F. Giordano, J.-P. Correa Baena, J.-D. Decoppet, S.M. Zakeeruddin, M. K. Nazeeruddin, M. Graetzel, A. Hagfeldt, Efficient luminescent solar cells based on tailored mixed-cation perovskites, *Sci. Adv.* 2 (2016), e1501170, <https://doi.org/10.1126/sciadv.1501170>.
- [55] F. Xu, T. Zhang, G. Li, Y. Zhao, Mixed cation hybrid lead halide perovskites with enhanced performance and stability, *J. Mater. Chem. A* 5 (2017) 11450–11461, <https://doi.org/10.1039/C7TA00042A>.
- [56] Y. Li, D. Maldonado-Lopez, V. Ríos Vargas, J. Zhang, K. Yang, Stability diagrams, defect tolerance, and absorption coefficients of hybrid halide semiconductors: high-throughput first-principles characterization, *J. Chem. Phys.* 152 (2020), 084106, <https://doi.org/10.1063/1.5127929>.
- [57] A. Jain, G. Hautier, S.P. Ong, K. Persson, New opportunities for materials informatics: resources and data mining techniques for uncovering hidden relationships, *J. Mater. Res.* 31 (2016) 977–994, <https://doi.org/10.1557/jmr.2016.80>.
- [58] F. Oviedo, Z. Ren, X. Hansong, S.I.P. Tian, K. Zhang, M. Layurova, T. Heumueller, N. Li, E. Birgersson, S. Sun, B. Mayurama, I.M. Peters, C.J. Brabec, J. Fisher, T. Buonassisi, Bridging the Gap between Photovoltaics R&D and Manufacturing with Data-Driven Optimization, 2020. <http://arxiv.org/abs/2004.13599>.
- [59] E.P. Shapera, A. Schleife, Database-driven materials selection for semiconductor heterojunction design, *Adv. Theory Simulations*. 1 (2018), 1800075, <https://doi.org/10.1002/adts.201800075>.
- [60] Z. Song, Q. Liu, Tolerance factor, phase stability and order-disorder of the pyrochlore structure, *Inorg. Chem. Front.* 7 (2020) 1583–1590, <https://doi.org/10.1039/DQQI0016G>.
- [61] C. Chen, Y. Zuo, W. Ye, X. Li, Z. Deng, S.P. Ong, A critical review of machine learning of energy materials, *Adv. Energy Mater.* 10 (2020), 1903242, <https://doi.org/10.1002/aenm.201903242>.
- [62] O. Voznyy, L. Levina, J.Z. Fan, M. Askerka, A. Jain, M.-J. Choi, O. Ouellette, P. Todorović, L.K. Sagar, E.H. Sargent, Machine learning accelerates discovery of optimal colloidal quantum dot synthesis, *ACS Nano* 13 (2019) 11122–11128, <https://doi.org/10.1021/acsnano.9b03864>.
- [63] M. Lee, Insights from machine learning techniques for predicting the efficiency of fullerene derivatives-based ternary organic solar cells at ternary blend design, *Adv. Energy Mater.* (2019), 1900891, <https://doi.org/10.1002/aenm.201900891>.
- [64] M.-H. Lee, A machine learning-based design rule for improved open-circuit voltage in ternary organic solar cells, *Adv. Intell. Syst.* 2 (2020), 1900108, <https://doi.org/10.1002/aisy.201900108>.
- [65] G. Pilania, J.E. Gubernatis, T. Lookman, Multi-fidelity machine learning models for accurate bandgap predictions of solids, *Comput. Mater. Sci.* 129 (2017) 156–163, <https://doi.org/10.1016/j.commatsci.2016.12.004>.
- [66] A. Nuhic, T. Terzimehic, T. Soczka-Guth, M. Buchholz, K. Dietmayer, Health diagnosis and remaining useful life prognostics of lithium-ion batteries using

- data-driven methods, *J. Power Sources* 239 (2013) 680–688, <https://doi.org/10.1016/j.jpowsour.2012.11.146>.
- [67] M.-F. Ng, J. Zhao, Q. Yan, G.J. Conduit, Z.W. Seh, Predicting the state of charge and health of batteries using data-driven machine learning, *Nat. Mach. Intell.* 2 (2020) 161–170, <https://doi.org/10.1038/s42256-020-0156-7>.
- [68] E. Chemali, P.J. Kollmeyer, M. Preindl, A. Emadi, State-of-charge estimation of Li-ion batteries using deep neural networks: a machine learning approach, *J. Power Sources* 400 (2018) 242–255, <https://doi.org/10.1016/j.jpowsour.2018.06.104>.
- [69] H. Sheng, J. Xiao, Electric vehicle state of charge estimation: nonlinear correlation and fuzzy support vector machine, *J. Power Sources* 281 (2015) 131–137, <https://doi.org/10.1016/j.jpowsour.2015.01.145>.
- [70] X. Hu, S.E. Li, Y. Yang, Advanced machine learning approach for lithium-ion battery state estimation in electric vehicles, *IEEE Trans. Transp. Electrif.* 2 (2016) 140–149, <https://doi.org/10.1109/TTE.2015.2512237>.
- [71] L. Petrich, D. Westhoff, J. Feinauer, D.P. Finegan, S.R. Daemi, P.R. Shearing, V. Schmidt, Crack detection in lithium-ion cells using machine learning, *Comput. Mater. Sci.* 136 (2017) 297–305, <https://doi.org/10.1016/j.commatsci.2017.05.012>.
- [72] K.T. Butler, D.W. Davies, H. Cartwright, O. Isayev, A. Walsh, Machine learning for molecular and materials science, *Nature* 559 (2018) 547–555, <https://doi.org/10.1038/s41586-018-0337-2>.
- [73] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, É. Duchesnay, Scikit-learn: machine learning in Python, *J. Mach. Learn. Res.* 12 (2011) 1532–4435, <https://doi.org/10.5555/1953048.2078195>.
- [74] Z. Li, Q. Xu, Q. Sun, Z. Hou, W.-J. Yin, Thermodynamic stability landscape of halide double perovskites via high-throughput computing and machine learning, *Adv. Funct. Mater.* 29 (2019), 1807280, <https://doi.org/10.1002/adfm.201807280>.
- [75] Rs. Team, RStudio, Integrated Development for R, 2020. <http://www.rstudio.com/>.
- [76] Ç. Odabaşı, R. Yıldırım, Performance analysis of perovskite solar cells in 2013–2018 using machine-learning tools, *Nanomater. Energy* 56 (2019) 770–791, <https://doi.org/10.1016/j.nanoen.2018.11.069>.
- [77] Y. Yu, X. Tan, S. Ning, Y. Wu, Machine learning for understanding compatibility of organic–inorganic hybrid perovskites with post-treatment amines, *ACS Energy Lett.* 4 (2019) 397–404, <https://doi.org/10.1021/acsenergylett.8b02451>.
- [78] S. Lu, Q. Zhou, Y. Ouyang, Y. Guo, Q. Li, J. Wang, Accelerated discovery of stable lead-free hybrid organic–inorganic perovskites via machine learning, *Nat. Commun.* 9 (2018) 3405, <https://doi.org/10.1038/s41467-018-05761-w>.
- [79] Ç. Odabaşı, R. Yıldırım, Machine learning analysis on stability of perovskite solar cells, *Sol. Energy Mater. Sol. Cells* 205 (2020), 110284, <https://doi.org/10.1016/j.solmat.2019.110284>.
- [80] J. Wei, X. Chu, X. Sun, K. Xu, H. Deng, J. Chen, Z. Wei, M. Lei, Machine learning in materials science, *Info* 1 (2019) 338–358, <https://doi.org/10.1002/inf.212028>.
- [81] C. Kim, T.D. Huan, S. Krishnan, R. Ramprasad, A hybrid organic–inorganic perovskite dataset, *Sci. Data* 4 (2017), 170057, <https://doi.org/10.1038/sdata.2017.57>.
- [82] C.R. Groom, I.J. Bruno, M.P. Lightfoot, S.C. Ward, The Cambridge structural database, *Acta Crystallogr. Sect. B Struct. Sci. Cryst. Eng. Mater.* 72 (2016) 171–179, <https://doi.org/10.1107/S2052520616003954>.
- [83] M. Hellenbrandt, The inorganic crystal structure database (ICSD)—present and future, *Crystallogr. Rev.* 10 (2004) 17–22, <https://doi.org/10.1080/08893110410001664882>.
- [84] S. Gräßl, A. Daskević, A. Merkys, D. Chateigner, L. Lutterotti, M. Quiros, N. R. Serebryanaya, P. Mocek, R.T. Downs, A. Le Bail, Crystallography Open Database (COD): an open-access collection of crystal structures and platform for world-wide collaboration, *Nucleic Acids Res.* 40 (2012) D420–D427, <https://doi.org/10.1093/nar/gkr900>.
- [85] A. Jain, S.P. Ong, G. Hautier, W. Chen, W.D. Richards, S. Dacek, S. Cholia, D. Gunter, D. Skinner, G. Ceder, K.A. Persson, Commentary: the Materials Project: a materials genome approach to accelerating materials innovation, *Apl. Mater.* 1 (2013), 011002, <https://doi.org/10.1063/1.4812323>.
- [86] S. Curtarolo, W. Setyawan, G.L.W. Hart, M. Jahnake, R.V. Chepulskii, R. H. Taylor, S. Wang, J. Xue, K. Yang, O. Levy, M.J. Mehl, H.T. Stokes, D. O. Demchenko, D. Morgan, AFLOW: an automatic framework for high-throughput materials discovery, *Comput. Mater. Sci.* 58 (2012) 218–226, <https://doi.org/10.1016/j.commatsci.2012.02.005>.
- [87] S. Kirklin, J.E. Saal, B. Meredig, A. Thompson, J.W. Doak, M. Aykol, S. Rühl, C. Wolverton, The Open Quantum Materials Database (OQMD): assessing the accuracy of DFT formation energies, *Npj Comput. Mater.* 1 (2015) 15010, <https://doi.org/10.1038/npjcomput.2015.10>.
- [88] K. Choudhary, M. Bercx, J. Jiang, R. Pachter, D. Lamoen, F. Tavazza, Accelerated discovery of efficient solar cell materials using quantum and machine-learning methods, *Chem. Mater.* 31 (2019) 5900–5908, <https://doi.org/10.1021/acs.chemmater.9b02166>.
- [89] G. Pilania, P.V. Balachandran, C. Kim, T. Lookman, Finding new perovskite halides via machine learning, *Front. Mater.* 3 (2016), <https://doi.org/10.3389/fmats.2016.00019>.
- [90] R. Ouyang, Exploiting ionic radii for rational design of halide perovskites, *Chem. Mater.* 32 (2020) 595–604, <https://doi.org/10.1021/acs.chemmater.9b04472>.
- [91] O. Allam, C. Holmes, Z. Greenberg, K.C. Kim, S.S. Jang, Density functional theory - machine learning approach to analyze the bandgap of elemental halide perovskites and ruddlesden-popper phases, *ChemPhysChem* 19 (2018) 2559–2565, <https://doi.org/10.1002/cphc.201800382>.
- [92] H. Park, R. Mall, F.H. Alharbi, S. Sanvito, N. Tabet, H. Bensmail, F. El-Mellohi, Exploring new approaches towards the formability of mixed-ion perovskites by DFT and machine learning, *Phys. Chem. Chem. Phys.* 21 (2019) 1078–1088, <https://doi.org/10.1039/C8CP06528D>.
- [93] R.X. Yang, J.M. Skelton, E.L. da Silva, J.M. Frost, A. Walsh, Spontaneous octahedral tilting in the cubic inorganic cesium halide perovskites CsSn_3 and CsPb_3X_3 ($\text{X} = \text{F}, \text{Cl}, \text{Br}, \text{I}$), *J. Phys. Chem. Lett.* 8 (2017) 4720–4726, <https://doi.org/10.1021/acs.jpclett.7b02423>.
- [94] M.R. Filip, F. Giustino, The geometric blueprint of perovskites, *Proc. Natl. Acad. Sci. Unit. States Am.* 115 (2018) 5397–5402, <https://doi.org/10.1073/pnas.1719179115>.
- [95] W. Travis, E.N.K. Glover, H. Bronstein, D.O. Scanlon, R.G. Palgrave, On the application of the tolerance factor to inorganic and hybrid halide perovskites: a revised system, *Chem. Sci.* 7 (2016) 4548–4556, <https://doi.org/10.1039/C5SC04845A>.
- [96] R.D. Shannon, Revised effective ionic radii and systematic studies of interatomic distances in halides and chalcogenides, *Acta Crystallogr. A* 32 (1976) 751–767, <https://doi.org/10.1107/S0567739476001551>.
- [97] C.J. Bartel, C. Sutton, B.R. Goldsmith, R. Ouyang, C.B. Musgrave, L. M. Ghiringhelli, M. Scheffler, New tolerance factor to predict the stability of perovskite oxides and halides, *Sci. Adv.* 5 (2019) 1–10, <https://doi.org/10.1126/sciadv.aav0693>.
- [98] J. Li, B. Pradhan, S. Gaur, J. Thomas, Predictions and strategies learned from machine learning to develop high-performing perovskite solar cells, *Adv. Energy Mater.* 9 (2019) 1–10, <https://doi.org/10.1002/aenm.201901891>.
- [99] J. Yang, Y. Wang, T. Wu, S. Li, Correlating the composition-dependent structural and electronic dynamics of inorganic mixed halide perovskites, *Chem. Mater.* 32 (2020) 2470–2481, <https://doi.org/10.1021/acs.chemmater.9b04995>.
- [100] H. Park, R. Mall, F.H. Alharbi, S. Sanvito, N. Tabet, H. Bensmail, F. El-Mellohi, Learn-and-Match molecular cations for perovskites, *J. Phys. Chem. C* 123 (2019) 7323–7334, <https://doi.org/10.1021/acs.jpcc.9b06208>.
- [101] A.Y.-T. Wang, R.J. Murdoch, S.K. Kauwe, A.O. Oliynyk, A. Gurlo, J. Brugge, K. A. Persson, T.D. Sparks, Machine Learning for Materials Scientists: an introductory guide towards best practices, *Chem. Mater.* 32 (2020) 4954–4965, <https://doi.org/10.1021/acs.chemmater.0c01907>.
- [102] K. Takahashi, L. Takahashi, I. Miyazato, Y. Tanaka, Searching for hidden perovskite materials for photovoltaic systems by combining data science and first principle calculations, *ACS Photonics* 5 (2018) 771–775, <https://doi.org/10.1021/acspophotonics.7b01479>.
- [103] Q. Tu, I. Spanopoulos, E.S. Vasileiadou, X. Li, M.G. Kanatzidis, G.S. Shekhawat, V. P. Dravid, Exploring the factors affecting the mechanical properties of 2D hybrid organic–inorganic perovskites, *ACS Appl. Mater. Interfaces* 12 (2020) 20440–20447, <https://doi.org/10.1021/acsami.0c02313>.
- [104] V.M. Goldschmidt, Die gesetze der Krystallochemie, *Naturwissenschaften* 14 (1926) 477–485, <https://doi.org/10.1007/BF01507527>.
- [105] C. Li, K.C.K. Soh, P. Wu, Formability of ABO_3 perovskites, *J. Alloys Compd.* 372 (2004) 40–48, <https://doi.org/10.1016/j.jallcom.2003.10.017>.
- [106] C. Li, X. Lu, W. Ding, L. Feng, Y. Gao, Z. Guo, Formability of ABX_3 ($\text{X} = \text{F}, \text{Cl}, \text{Br}, \text{I}$) halide perovskites, *Acta Crystallogr. Sect. B Struct. Sci.* 64 (2008) 702–707, <https://doi.org/10.1107/S0108768108032734>.
- [107] S. Lu, Q. Zhou, L. Ma, Y. Guo, J. Wang, Rapid discovery of ferroelectric photovoltaic perovskites and material descriptors via machine learning, *Small Methods* 3 (2019), 1900360, <https://doi.org/10.1002/smtd.201900360>.
- [108] S. Chakraborty, W. Xie, N. Mathews, M. Sherburne, R. Ahuja, M. Asta, S. G. Mhaisalkar, Rational design: a high-throughput computational screening and experimental validation methodology for lead-free and emergent hybrid perovskites, *ACS Energy Lett.* 2 (2017) 837–845, <https://doi.org/10.1021/acsenergylett.7b00035>.
- [109] J.C. Stanley, F. Mayr, A. Gagliardi, Machine learning stability and bandgaps of lead-free perovskites for photovoltaics, *Adv. Theory Simulations*. 3 (2020), 1900178, <https://doi.org/10.1002/adts.201900178>.
- [110] M.G. Ju, M. Chen, Y. Zhou, H.F. Garces, J. Dai, L. Ma, N.P. Padture, X.C. Zeng, Earth-abundant nontoxic titanium(IV)-based vacancy-ordered double perovskite halides with tunable 1.0 to 1.8 eV bandgaps for photovoltaic applications, *ACS Energy Lett.* 3 (2018) 297–304, <https://doi.org/10.1021/acsenergylett.7b01167>.
- [111] M. Chen, M.G. Ju, A.D. Carl, Y. Zong, R.L. Grimm, J. Gu, X.C. Zeng, Y. Zhou, N. P. Padture, Cesium titanium(IV) bromide thin films based stable lead-free perovskite solar cells, *Joule* 2 (2018) 558–570, <https://doi.org/10.1016/j.joule.2018.01.009>.
- [112] S. Sun, N.T.P. Hartono, Z.D. Ren, F. Oviedo, A.M. Buscemi, M. Layurova, D. X. Chen, T. Ogunfunmi, J. Thapa, S. Ramasamy, C. Settens, B.L. DeCost, A. G. Kusne, Z. Liu, S.I.P. Tian, I.M. Peters, J.-P. Correa-Baena, T. Buonassisi, Accelerated development of perovskite-inspired materials via high-throughput synthesis and machine-learning diagnosis, *Joule* 3 (2019) 1437–1451, <https://doi.org/10.1016/j.joule.2019.05.014>.
- [113] A. Mannodi-Kanakkithodi, J.-S. Park, N. Jeon, D.H. Cao, D.J. Gosztola, A.B. F. Martinson, M.K.Y. Chan, Comprehensive computational study of partial lead substitution in methylammonium lead bromide, *Chem. Mater.* 31 (2019) 3599–3612, <https://doi.org/10.1021/acs.chemmater.8b04017>.
- [114] J. Im, S. Lee, T.-W. Ko, H.W. Kim, Y. Hyon, H. Chang, Identifying Pb-free perovskites for solar cells by machine learning, *Npj Comput. Mater.* 5 (2019) 37, <https://doi.org/10.1038/s41524-019-0177-0>.
- [115] M. Chen, M.-G. Ju, H.F. Garces, A.D. Carl, L.K. Ono, Z. Hawash, Y. Zhang, T. Shen, Y. Qi, R.L. Grimm, D. Pacifici, X.C. Zeng, Y. Zhou, N.P. Padture, Highly stable and

- efficient all-inorganic lead-free perovskite solar cells with native-oxide passivation, *Nat. Commun.* 10 (2019) 16, <https://doi.org/10.1038/s41467-018-07951-y>.
- [116] W. Ke, M.G. Kanatzidis, Prospects for low-toxicity lead-free perovskite solar cells, *Nat. Commun.* 10 (2019) 965, <https://doi.org/10.1038/s41467-019-08918-3>.
- [117] T. Wu, J. Wang, Global discovery of stable and non-toxic hybrid organic-inorganic perovskites for photovoltaic systems by combining machine learning method with first principle calculations, *Nanomater. Energy* 66 (2019), 104070, <https://doi.org/10.1016/j.nanoen.2019.104070>.
- [118] R. Jacobs, G. Luo, D. Morgan, Materials discovery of stable and nontoxic halide perovskite materials for high-efficiency solar cells, *Adv. Funct. Mater.* 29 (2019), 1804354, <https://doi.org/10.1002/adfm.201804354>.
- [119] B.R. Sutherland, Solar materials find their band gap, *Joule* 4 (2020) 984–985, <https://doi.org/10.1016/j.joule.2020.05.001>.
- [120] W.A. Saidi, W. Shadid, I.E. Castelli, Machine-learning structural and electronic properties of metal halide perovskites using a hierarchical convolutional neural network, *Npj Comput. Mater.* 6 (2020) 36, <https://doi.org/10.1038/s41524-020-0307-8>.
- [121] Q. Xu, Z. Li, M. Liu, W.-J. Yin, Rationalizing perovskite data for machine learning and materials design, *J. Phys. Chem. Lett.* 9 (2018) 6948–6954, <https://doi.org/10.1021/acs.jpclett.8b03232>.
- [122] H.C. Herbol, W. Hu, P. Frazier, P. Clancy, M. Poloczek, Efficient search of compositional space for hybrid organic–inorganic perovskites via Bayesian optimization, *Npj Comput. Mater.* 4 (2018) 51, <https://doi.org/10.1038/s41524-018-0106-7>.
- [123] H. Choubisa, M. Askerka, K. Ryczko, O. Voznyy, K. Mills, I. Tamblyn, E. H. Sargent, Crystal site feature embedding enables exploration of large chemical spaces, *Matter* (2020), <https://doi.org/10.1016/j.matt.2020.04.016>.
- [124] J. Stanley, A. Gagliardi, Machine learning bandgaps of inorganic mixed halide perovskites, in: *IEEE 18th Int. Conf. Nanotechnol.*, IEEE, 2018, pp. 1–4, <https://doi.org/10.1109/NANO.2018.8626420>.
- [125] H.-A. Chen, C.-W. Pao, Fast and accurate artificial neural network potential model for MAPbI₃ perovskite materials, *ACS Omega* 4 (2019) 10950–10959, <https://doi.org/10.1021/acsomega.9b00378>.
- [126] J.C. Thomas, J.S. Bechtel, A.R. Natarajan, A. Van der Ven, Machine learning the density functional theory potential energy surface for the inorganic halide perovskite CsPbBr₃, *Phys. Rev. B* 100 (2019), 134101, <https://doi.org/10.1103/PhysRevB.100.134101>.
- [127] M. Nelson, M. Di Vece, Using a neural network to improve the optical absorption in halide perovskite layers containing core-shells silver nanoparticles, *Nanomaterials* 9 (2019) 437, <https://doi.org/10.3390/nano9030437>.
- [128] E. Velilla, J. Cano, K. Jimenez, J. Valencia, D. Ramirez, F. Jaramillo, Numerical analysis to determine reliable one-diode model parameters for perovskite solar cells, *Energies* 11 (2018) 1963, <https://doi.org/10.3390/en11081963>.
- [129] Z. Ren, S. Tian, T. Heumueller, E. Bergstrand, F. Lin, A. Aberle, S. Sun, I.M. Peters, R. Stangl, C.J. Brabec, T. Buonassisi, F. Oviedo, H. Xue, M. Thway, K. Zhang, N. Li, J.D. Perea, M. Layuрова, Y. Wang, Physics-guided characterization and optimization of solar cells using surrogate machine learning model, in: *2019 IEEE 46th Photovolt. Spec. Conf.*, IEEE, 2019, pp. 3054–3058, <https://doi.org/10.1109/PVSC40753.2019.8890715>.
- [130] A. Ali, H. Park, R. Mall, B. Aissa, S. Sanvito, H. Bensmail, A. Belaïdi, F. El-Mellouhi, Machine learning accelerated recovery of the cubic structure in mixed-cation perovskite thin films, *Chem. Mater.* 32 (2020) 2998–3006, <https://doi.org/10.1021/acs.chemmater.9b05342>.
- [131] J.M. Howard, E.M. Tenneyson, B.R.A. Neves, M.S. Leite, Machine learning for perovskites’ re-ap-rest-recovery Cycle, *Joule* 3 (2019) 325–337, <https://doi.org/10.1016/j.joule.2018.11.010>.
- [132] J. Kirman, A. Johnston, D.A. Kuntz, M. Askerka, Y. Gao, P. Todorović, D. Ma, G. G. Privé, E.H. Sargent, Machine-learning-accelerated perovskite crystallization, *Matter* 2 (2020) 938–947, <https://doi.org/10.1016/j.matt.2020.02.012>.
- [133] T. Chen, Y. Zhou, M. Rafailovich, Application of machine learning in perovskite solar cell crystal size distribution analysis, *MRS Adv* 4 (2019) 793–800, <https://doi.org/10.1557/adv.2019.145>.
- [134] R. Jinnouchi, J. Lahnsteiner, F. Karsai, G. Kresse, M. Bokdam, Phase transitions of hybrid perovskites simulated by machine-learning force fields trained on the fly with bayesian inference, *Phys. Rev. Lett.* 122 (2019), 225701, <https://doi.org/10.1103/PhysRevLett.122.225701>.
- [135] Ç. Odabaşı, R. Yıldırım, Assessment of reproducibility, hysteresis, and stability relations in perovskite solar cells using machine learning, *Energy Technol.* (2020), 1901449, <https://doi.org/10.1002/ente.201901449>.
- [136] M.J. Hong, R.Y. Johnson, J.G. Labram, Impact of moisture on mobility in methylammonium lead iodide and formamidinium lead iodide, *J. Phys. Chem. Lett.* 11 (2020) 4976–4983, <https://doi.org/10.1021/acs.jpclett.0c01369>.
- [137] Z. Hawash, L.K. Ono, Y. Qi, Moisture and oxygen enhance conductivity of LiTFSI-doped spiro-MeOTAD hole transport layer in perovskite solar cells, *Adv. Mater. Interfaces* 3 (2016), 1600117, <https://doi.org/10.1002/admi.201600117>.
- [138] N.T.P. Hartono, J. Thapa, A. Tiihonen, F. Oviedo, C. Batalí, J.J. Yoo, Z. Liu, R. Li, D.F. Marrón, M.G. Bawendi, T. Buonassisi, S. Sun, How machine learning can help select capping layers to suppress perovskite degradation, *Nat. Commun.* 11 (2020) 4172, <https://doi.org/10.1038/s41467-020-17945-4>.
- [139] S. Yang, Y. Wang, P. Liu, Y.-B. Cheng, H.J. Zhao, H.G. Yang, Functionalization of perovskite thin films with moisture-tolerant molecules, *Nat. Energy* 1 (2016) 15016, <https://doi.org/10.1038/nenergy.2015.16>.
- [140] L. Zhang, Z. Chen, J. Su, J. Li, Data mining new energy materials from structure databases, *Renew. Sustain. Energy Rev.* 107 (2019) 554–567, <https://doi.org/10.1016/j.rser.2019.03.036>.
- [141] J. Kim, S. Cho, F. Dinic, J. Choi, C. Choi, S.M. Jeong, J.-S. Lee, O. Voznyy, M.J. Ko, Y. Kim, Hydrophobic stabilizer-anchored fully inorganic perovskite quantum dots enhance moisture resistance and photovoltaic performance, *Nanomater. Energy* 75 (2020), 104985, <https://doi.org/10.1016/j.nanoen.2020.104985>.
- [142] R. Prasanna, T. Leijtens, S.P. Dunfield, J.A. Raiford, E.J. Wolf, S.A. Swifter, J. Werner, G.E. Eperon, C. de Paula, A.F. Palmstrom, C.C. Boyd, M.F.A.M. van Hest, S.F. Bent, G. Teeter, J.J. Berry, M.D. McGehee, Design of low bandgap tin–lead halide perovskite solar cells to achieve thermal, atmospheric and operational stability, *Nat. Energy* 4 (2019) 939–947, <https://doi.org/10.1038/s41560-019-0471-6>.
- [143] A. Bhowmik, T. Vegge, AI fast track to battery fast charge, *Joule* 4 (2020) 717–719, <https://doi.org/10.1016/j.joule.2020.03.016>.
- [144] W. Li, R. Jacobs, D. Morgan, Predicting the thermodynamic stability of perovskite oxides using machine learning models, *Comput. Mater. Sci.* 150 (2018) 454–463, <https://doi.org/10.1016/j.commatsci.2018.04.033>.
- [145] Q. Sun, W.-J. Yin, Thermodynamic stability trend of cubic perovskites, *J. Am. Chem. Soc.* 139 (2017) 14905–14908, <https://doi.org/10.1021/jacs.7b09379>.
- [146] M.A. Haque, S. Kee, D.R. Villalva, W. Ong, D. Baran, Halide perovskites: thermal transport and prospects for thermoelectricity, *Adv. Sci.* 7 (2020), 1903389, <https://doi.org/10.1002/advs.201903389>.
- [147] E.J. Graham, J. Cho, K.M. Forlano, D.F. Watson, R. Arròyave, S. Banerjee, Machine learning-directed navigation of synthetic design space: a statistical learning approach to controlling the synthesis of perovskite halide nanoplatelets in the quantum-confined regime, *Chem. Mater.* 31 (2019) 3281–3292, <https://doi.org/10.1021/acs.chemmater.9b00212>.
- [148] P. De Luna, J. Wei, Y. Bengio, A. Aspuru-Guzik, E. Sargent, Use machine learning to find energy materials, *Nature* 552 (2017) 23–27, <https://doi.org/10.1038/d41586-017-07820-6>.
- [149] G. Zhou, W. Chu, O.V. Prezhdo, Structural deformation controls charge losses in MAPbI₃: unsupervised machine learning of nonadiabatic molecular dynamics, *ACS Energy Lett.* (2020) 1930–1938, <https://doi.org/10.1021/acsenergylett.0c00899>.
- [150] H. Liu, J. Cheng, H. Dong, J. Feng, B. Pang, Z. Tian, S. Ma, F. Xia, C. Zhang, L. Dong, Screening stable and metastable ABO₃ perovskites using machine learning and the materials project, *Comput. Mater. Sci.* 177 (2020), 109614, <https://doi.org/10.1016/j.commatsci.2020.109614>.
- [151] C. Li, H. Hao, B. Xu, G. Zhao, L. Chen, S. Zhang, H. Liu, A progressive learning method for predicting the band gap of ABO₃ perovskites using an instrumental variable, *J. Mater. Chem. C* 8 (2020) 3127–3136, <https://doi.org/10.1039/C9TC06632B>.
- [152] S.G. Javed, A. Khan, A. Majid, A.M. Mirza, J. Bashir, Lattice constant prediction of orthorhombic AB₃ perovskites using support vector machines, *Comput. Mater. Sci.* 39 (2007) 627–634, <https://doi.org/10.1016/j.commatsci.2006.08.015>.
- [153] Z. Li, L.E.K. Achenie, H. Xin, An adaptive machine learning strategy for accelerating discovery of perovskite electrocatalysts, *ACS Catal.* 10 (2020) 4377–4384, <https://doi.org/10.1021/acscatal.9b05248>.
- [154] G. Pilania, A. Mannodi-Kanakkithodi, B.P. Uberuaga, R. Ramprasad, J. E. Gubernatis, T. Lookman, Machine learning bandgaps of double perovskites, *Sci. Rep.* 6 (2016) 19375, <https://doi.org/10.1038/srep19375>.
- [155] P.V. Balachandran, B. Kowalski, A. Sehirlioglu, T. Lookman, Experimental search for high-temperature ferroelectric perovskites guided by two-step machine learning, *Nat. Commun.* 9 (2018) 1668, <https://doi.org/10.1038/s41467-018-03821-9>.
- [156] A. Majid, A. Khan, G. Javed, A.M. Mirza, Lattice constant prediction of cubic and monoclinic perovskites using neural networks and support vector regression, *Comput. Mater. Sci.* 50 (2010) 363–372, <https://doi.org/10.1016/j.commatsci.2010.08.028>.
- [157] R. Sarmiento-Pérez, T.F.T. Cerqueira, S. Körbel, S. Botti, M.A.L. Marques, Prediction of stable nitride perovskites, *Chem. Mater.* 27 (2015) 5957–5963, <https://doi.org/10.1021/acs.chemmater.5b02026>.
- [158] M.L. Agiorgousis, Y. Sun, D. Choe, D. West, S. Zhang, Machine learning augmented discovery of chalcogenide double perovskites for photovoltaics, *Adv. Theory Simulations*. 2 (2019), 1800173, <https://doi.org/10.1002/adts.201800173>.
- [159] Y. Fujii, D. Ramirez, N.C. Rosero-Navarro, D. Julian, A. Miura, F. Jaramillo, K. Tadanaga, Two-dimensional hybrid halide perovskite as electrode materials for all-solid-state lithium secondary batteries based on sulfide solid electrolytes, *ACS Appl. Energy Mater.* 2 (2019) 6569–6576, <https://doi.org/10.1021/acsaem.9b01118>.
- [160] W. Zhang, G.E. Eperon, H.J. Snaith, Metal halide perovskites for energy applications, *Nat. Energy* 1 (2016) 16048, <https://doi.org/10.1038/nenergy.2016.48>.
- [161] A. Kostopoulou, D. Vernardou, K. Savva, E. Stratakis, All-inorganic lead halide perovskite nanohexagons for high performance air-stable lithium batteries, *Nanoscale* 11 (2019) 882–889, <https://doi.org/10.1039/C8NR10009H>.
- [162] Q. Jiang, X. Zeng, N. Wang, Z. Xiao, Z. Guo, J. Lu, Electrochemical lithium doping induced property changes in halide perovskite CsPbBr₃ crystal, *ACS Energy Lett.* 3 (2018) 264–269, <https://doi.org/10.1021/acsenergylett.7b01230>.
- [163] V. Tshitoyan, J. Dagdelen, L. Weston, A. Dunn, Z. Rong, O. Kononova, K. Persson, G. Ceder, A. Jain, Unsupervised word embeddings capture latent knowledge from materials science literature, *Nature* 571 (2019) 95–98, <https://doi.org/10.1038/s41586-019-1335-8>.
- [164] J. Madsen, P. Liu, J. Kling, J.B. Wagner, T.W. Hansen, O. Winther, J. Schiøtz, A deep learning approach to identify local structures in atomic-resolution transmission electron microscopy images, *Adv. Theory Simulations* 1 (2018), 1800037, <https://doi.org/10.1002/adts.201800037>.

- [165] Y. Liu, Q. Sun, W. Lu, H. Wang, Y. Sun, Z. Wang, X. Lu, K. Zeng, General resolution enhancement method in atomic force microscopy using deep learning, *Adv. Theory Simulations*. 2 (2019), 1800137, <https://doi.org/10.1002/adts.201800137>.
- [166] Y. Zhang, D. Lu, M. Gao, M. Lai, J. Lin, T. Lei, Z. Lin, L.N. Quan, P. Yang, Quantitative imaging of anion exchange kinetics in halide perovskites, *Proc. Natl. Acad. Sci. Unit. States Am.* 116 (2019) 12648–12653, <https://doi.org/10.1073/pnas.1903448116>.
- [167] B.R. Jany, A. Janas, F. Krok, Retrieving the quantitative chemical information at nanoscale from scanning electron microscope energy dispersive X-ray measurements by machine learning, *Nano Lett.* 17 (2017) 6520–6525, <https://doi.org/10.1021/acs.nanolett.7b01789>.
- [168] Y. Wang, Z. Lv, L. Zhou, X. Chen, J. Chen, Y. Zhou, V.A.L. Roy, S.-T. Han, Emerging perovskite materials for high density data storage and artificial synapses, *J. Mater. Chem. C*. 6 (2018) 1600–1617, <https://doi.org/10.1039/C7TC05326F>.
- [169] N.K. Upadhyay, H. Jiang, Z. Wang, S. Asapu, Q. Xia, J. Joshua Yang, Emerging memory devices for neuromorphic computing, *Adv. Mater. Technol.* 4 (2019), 1800589, <https://doi.org/10.1002/admt.201800589>.
- [170] T. Hasegawa, T. Ohno, K. Terabe, T. Tsuruoka, T. Nakayama, J.K. Gimzewski, M. Aono, Learning abilities achieved by a single solid-state atomic switch, *Adv. Mater.* 22 (2010) 1831–1834, <https://doi.org/10.1002/adma.200903680>.
- [171] H. Yu, J. Gong, H. Wei, W. Huang, W. Xu, Mixed-halide perovskite for ultrasensitive two-terminal artificial synaptic devices, *Mater. Chem. Front.* 3 (2019) 941–947, <https://doi.org/10.1039/C9QM00061E>.
- [172] R.A. John, N. Yantara, Y.F. Ng, G. Narasimman, E. Mosconi, D. Meggiolaro, M. R. Kulkarni, P.K. Gopalakrishnan, C.A. Nguyen, F. De Angelis, S.G. Mhaisalkar, A. Basu, N. Mathews, Ionotronic halide perovskite drift-diffusive synapses for low-power neuromorphic computation, *Adv. Mater.* 30 (2018), 1805454, <https://doi.org/10.1002/adma.201805454>.
- [173] S. Ham, S. Choi, H. Cho, S.-I. Na, G. Wang, Photonic organolead halide perovskite artificial synapse capable of accelerated learning at low power inspired by dopamine-facilitated synaptic activity, *Adv. Funct. Mater.* 29 (2019), 1806646, <https://doi.org/10.1002/adfm.201806646>.
- [174] A.M. Shen, C.-L. Chen, K. Kim, B. Cho, A. Tudor, Y. Chen, Analog neuromorphic module based on carbon nanotube synapses, *ACS Nano* 7 (2013) 6117–6122, <https://doi.org/10.1021/nn401946s>.
- [175] X. Zhao, H. Xu, Z. Wang, Y. Lin, Y. Liu, Memristors with Organic-inorganic Halide Perovskites, *InfoMat*, 2019, <https://doi.org/10.1002/inf2.12012> inf2.12012.
- [176] E. Wlaźiak, D. Przyczyna, R. Gutiérrez, G. Cuniberti, K. Szaciłowski, Towards synthetic neural networks: can artificial electrochemical neurons be coupled with artificial memristive synapses? *Jpn. J. Appl. Phys.* 59 (2020), S10801 <https://doi.org/10.35848/1347-4065/ab7e11>.
- [177] B. Li, W. Hui, X. Ran, Y. Xia, F. Xia, L. Chao, Y. Chen, W. Huang, Metal halide perovskites for resistive switching memory devices and artificial synapses, *J. Mater. Chem. C*. 7 (2019) 7476–7493, <https://doi.org/10.1039/C9TC02233C>.
- [178] W. Xu, H. Cho, Y.-H. Kim, Y.-T. Kim, C. Wolf, C.-G. Park, T.-W. Lee, Organometal halide perovskite artificial synapses, *Adv. Mater.* 28 (2016) 5916–5922, <https://doi.org/10.1002/adma.201506363>.