processes

MDPI

Review

Data-Driven Perovskite Design via High-Throughput Simulation and Machine Learning

Yidi Wang, Dan Sun, Bei Zhao, Tianyu Zhu, Chengcheng Liu, Zixuan Xu, Tianhang Zhou * and Chunming Xu

State Key Laboratory of Heavy Oil Processing, China University of Petroleum-Beijing, Beijing 102249, China; 2023880028@cup.edu.cn (Y.W.); 2023010613@student.cup.edu.cn (D.S.); 2023010570@student.cup.edu.cn (B.Z.); 2023010616@student.cup.edu.cn (T.Z.); 2023210665@student.cup.edu.cn (C.L.); 2023010581@student.cup.edu.cn (Z.X.); xcm@cup.edu.cn (C.X.)
* Correspondence: zhouth@cup.edu.cn; Tel.: +86-8654621556

Abstract

Perovskites (ABX_3) exhibit remarkable potential in optoelectronic conversion, catalysis, and diverse energy-related fields. However, the tunability of A, B, and X-site compositions renders conventional screening methods labor-intensive and inefficient. This review systematically synthesizes the roles of physical simulations and machine learning (ML) in accelerating perovskite discovery. By harnessing existing experimental datasets and high-throughput computational results, ML models elucidate structure-property relationships and predict performance metrics for solar cells, (photo)electrocatalysts, oxygen carriers, and energy-storage materials, with experimental validation confirming their predictive reliability. While data scarcity and heterogeneity inherently limit ML-based prediction of material property, integrating high-throughput computational methods as external mechanistic constraints—supplementing standardized, large-scale training data and imposing loss penalties—can improve accuracy and efficiency in bandgap prediction and defect engineering. Moreover, although embedding high-throughput simulations into ML architectures remains nascent, physics-embedded approaches (e.g., symmetry-aware networks) show increasing promise for enhancing physical consistency. This dual-driven paradigm, integrating data and physics, provides a versatile framework for perovskite design, achieving both high predictive accuracy and interpretability—key milestones toward a rational design strategy for functional materials discovery.

Keywords: perovskite design; machine learning; high-throughput computational screening; bandgap prediction; defect engineering

check for updates

Academic Editor: Haiping Zhu

Received: 31 August 2025
Revised: 16 September 2025
Accepted: 21 September 2025
Published: 24 September 2025

Citation: Wang, Y.; Sun, D.; Zhao, B.; Zhu, T.; Liu, C.; Xu, Z.; Zhou, T.; Xu, C. Data-Driven Perovskite Design via High-Throughput Simulation and Machine Learning. *Processes* **2025**, *13*, 3049. https://doi.org/10.3390/pr13103049

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Perovskite (ABX_3) has both performance and cost advantages in optoelectronics, catalysis, and energy storage, due to the highly tunable crystal structure and high defect tolerance [1]. Component regulation of perovskite is achievable through multi-element doping strategies at A/B/X sites. Synergistic effects of multidimensional sites result in a discrete distribution of material properties, and multiple doping combinations lead to an increased computational complexity for optimal material screening due to explosive growth of combinations.

The performance optimization of perovskite is highly dependent on precise design of the composition and structure. The traditional experimental trial-and-error method manually adjusts the element composition of A/B/X sites, but the method is limited by low

efficiency and long period. First-principles calculations based on density functional theory (DFT) have revealed micro-mechanisms such as formation energy and defect states with the rise of computational materials science [2]. However, DFT optimization is time-consuming and makes it difficult to cover complex component spaces. Machine learning (ML) learns and identifies the complex relationship between material properties and structural features by training algorithms on partially known or simulation calculation data, and then predicts properties of unknown materials. The integration of machine learning and simulation calculation significantly reduces the demand for simulation calculation, and greatly expands the scope and depth of material searches [3].

This study reviewed a practical workflow that integrates machine learning and physical simulations for materials screening. The integration of ML with databases or high-throughput simulations has advanced rapidly in materials stability, optoelectronics, catalysis, and energy storage. Development trends are synthesized to offer a new route for predicting and discovering high performance perovskites.

2. Methods for Perovskite Material Screening

Beyond conventional experimental screening, perovskite materials can also be screened using materials databases, molecular simulations, and DFT. Furthermore, high-throughput simulations that integrate automated workflows and parallel computing enable batch evaluation and selection across vast candidate spaces.

2.1. Current Databases for Perovskite Materials

Perovskite databases integrate information from the literature and/or in-house experimental results, coupled with standardized processing and validation mechanisms, to build comprehensive platforms for materials properties and performance data. The Helmholtz-Zentrum Berlin Perovskite Database Project [4] systematically curates global perovskite data—key materials parameters (composition, crystal structure, optoelectronic properties) and detailed experimental conditions, fabrication processes, and measurement methods. Zhang et al. [5] analyzed 7419 PSC entries from the Perovskite Database Project and introduced TS80m—a stability metric defined as the tensile strength deviation after 80 m of coiling, incorporating acceleration factors for temperature, humidity, and irradiance to standardize test conditions. Their work identified synergistic optimization pathways via tolerance-factor control, 2D/3D device design, and carbon electrodes. However, existing materials databases remain non-application-specific and are predominantly used for early-stage structural screening or high-level trend analysis.

2.2. Applications of Molecular Simulation in Perovskites

Molecular simulation relies on classical mechanics and empirical potentials, with Molecular Dynamics and Monte Carlo being the most used. Molecular dynamics simulation provides critical insights into the dynamic behavior and stability of perovskite materials under varying temperatures and pressures [6]. Kholmurodov et al. [7] employed molecular dynamics simulations to investigate the dynamic behavior and structural properties of $CaTiO_3$ under ambient conditions. The influence of interatomic potential modifications on atomic pair correlations and structural reorganization was systematically elucidated. The Monte Carlo method is a numerical technique centered on random sampling, widely used in statistical physics and condensed matter research. It proves particularly effective for studying phase transitions, critical phenomena, and estimating thermodynamic properties. Zhang et al. [8] employed Monte Carlo simulations and other methods to investigate the lattice dynamics, ground-state structure, and magnetic properties of Sr_2RuO_4 under electric field modulation. This distorted structure hosts ferromagnetic half-metallic behavior with

dominant out-of-plane magnetization, while remaining tunable under external electric fields without phase transition. In summary, molecular simulation proves effective for investigating large-scale molecular systems and dynamic processes, particularly suitable for studying perovskite structural evolution, thermal stability, and interfacial behavior.

DFT calculations play an irreplaceable role in characterizing electronic defects and trap states, which is essential for a deep understanding of the performance-limiting mechanisms in perovskite optoelectronic devices. By calculating the formation energy of point defects (such as vacancies, interstitials, and substitutional impurities), one can determine their likelihood of formation and concentration under specific growth conditions (e.g., Pb-rich, I-rich environments). Further analysis of the position of defect energy levels within the bandgap can identify those "deep-level traps" located deep in the bandgap. These deep-level trap states act as non-radiative recombination centers, significantly shortening carrier lifetime and reducing the photoelectric conversion efficiency of solar cells or the radiative efficiency of light-emitting diodes. For instance, calculations on common defects in lead-based perovskites, such as lead vacancies (V_Pb), iodine vacancies (V_I), or interstitial iodine (I_i), have successfully revealed their key role as non-radiative recombination centers, corroborating experimentally observed carrier dynamics. Therefore, systematic defect calculations can provide direct theoretical guidance for tracing performance degradation to its microscopic origins and for designing material passivation strategies. By computing the interactions between different passivation molecules (e.g., passivators, additives) and defect states, it is possible to pre-evaluate their passivation effects (such as eliminating defect levels or increasing defect formation energy), thereby accelerating the rational design of high-performance and high-stability perovskite materials.

2.3. Application of DFT in Perovskite

DFT, rooted in quantum mechanics and electron density analysis, evaluates perovskite stability and microscopic properties by computing energy profiles, crystal structures, and phase transitions [9]. The methodology further characterizes band structures, point defects, interface properties, and photocarrier dynamics, elucidating fundamental mechanisms including electronic transitions, defect formation, and interfacial charge transfer. Yin et al. [10] employed DFT calculations to investigate defect formation energies in α-MAPbI3 under varying growth conditions. Their study revealed tunable conductivity from p-type to n-type behavior, demonstrating the exceptional defect property adjustability in perovskite materials. Through systematic growth condition optimization, the conductivity was precisely tuned across the p-type to n-type spectrum, revealing the exceptional defect property tunability inherent to perovskite systems. DFT calculations are particularly suitable for small to medium-scale systems. This method delivers high accuracy at the cost of substantial computational requirements. It has now become the predominant technique in perovskite material design.

3. ML–Assisted Screening of Perovskite Materials

A forementioned material screening methods still lack sufficient coverage and accuracy. Researchers are now building specialized datasets through computational modeling, combining active learning strategies to iteratively improve prediction models. This approach breaks through the limitations of traditional trial-and-error methods, enabling targeted design of high-performance perovskite materials. Figure 1 illustrates the fundamental ML workflow for perovskite material screening, with key components analyzed below.

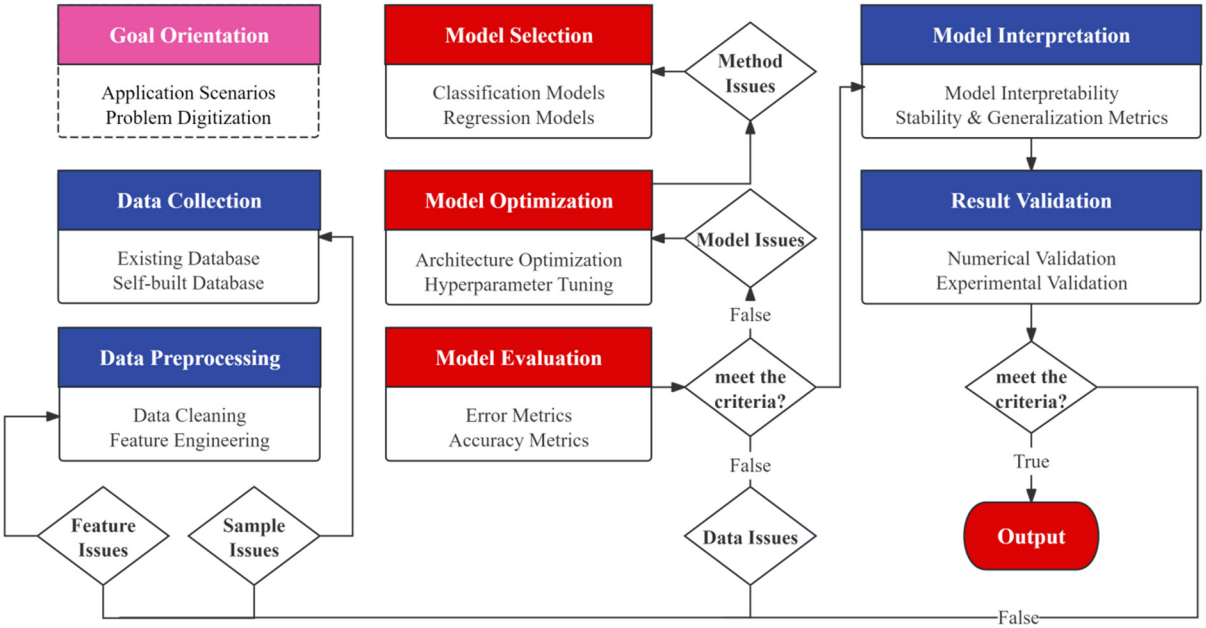


Figure 1. Fundamental ML workflow for perovskite material screening.

3.1. Data Collection and Preprocessing

High-dimensional feature modeling for perovskite is complicated by multi-scale coupling effects. Traditional semi-empirical models and feature-driven machine learning methods suffer from sparse feature spaces and insufficient valid samples due to high dimensionality and redundancy, leading to instability and overfitting [11]. Data preprocessing directly determines model performance and result reliability. Feature filtering and selection during modeling can effectively reduce dimensions, improve accuracy, and enhance generalization capability. Hu et al. [12] processed data collected from the Perovskite Database Project through missing-value treatment, stability screening, and standardization, enabling rapid high-accuracy prediction of halide double perovskite bandgaps and relative stability. Molecular feature importance rankings for passivation materials were generated and screening criteria were quantified, leading to the identification of perovskite solar cells with device efficiencies of 22.36% and 24.47%. On the other hand, molecular descriptors should be transferable, experimentally measurable, and quality-controllable, while balancing performance goals with physical interpretability. Iterative optimization through error analysis and importance evaluation is recommended.

3.2. Model Selection and Evaluation

Materials design requires differentiated use of classification (for categorical sorting) and regression (for continuous value prediction) algorithms. For classification tasks, Zhu et al. [13] developed a gradient boosting classification model that achieved 92.3% accuracy in predicting the stability of 2877 perovskites by utilizing key features like A/B-site ionic radius ratio and formation energy, significantly outperforming traditional tolerance factor empirical criteria. For regression tasks, Wan et al. [14] addressed the complex nonlinear relationships in bandgap prediction by designing a gradient boosting regression (GBR) model combined with a "last elimination" feature selection strategy. Their approach achieved a mean absolute error of 0.0287 eV across 5158 candidates while improving computational efficiency by three orders of magnitude compared to DFT calculations. The model also successfully identified the regulatory threshold of Pb^{2+}/I^{-} element ratio on bandgap. The key to improving model performance is a well-designed hyperparameter search combined with effective regularization, which together suppress overfitting and

enhance generalization. Evaluation and tuning should be guided by the task objective, using a mix of metrics—such as accuracy and recall—for comparative assessment. For regression, consider both MSE and RMSE rather than relying on a single metric. Applying cross-validation with multiple splits and repeated train–test cycles helps stabilize estimates, reduce randomness, and support more reliable choices of models and configurations.

3.3. Model Interpretation and Validation

Model interpretability is essential in data-driven materials design. While traditional "black-box" models can deliver high-accuracy predictions, they struggle to illuminate structure–property relationships and underlying physics, which undermines their contribution to knowledge building and, in turn, narrows their scope of application. Yang et al. [15] used symbolic regression to obtain an explicit formula that links the band gap to physical descriptors, directly quantifying how input features influence (or determine) Eg. The analysis points to the B–X electronegativity difference as the dominant factor, and the resulting compact expression achieves a test error (RMSE $\approx$ 0.08–0.09 eV) comparable to DFT-level uncertainty.

In the final stage of ML-based property prediction for perovskites, validate model-selected candidates via experiments or high-accuracy calculations. If results diverge from predictions, trace back to revise the model and features; if they align, confirm the findings and deliver the final conclusions and candidate list. Pritish et al. [16] trained a random-forest model on samples extracted from 182 papers to predict optical band gaps for 10,920 halide perovskite compositions, achieving RMSE $\approx$ 0.146 eV—on par with DFT accuracy. They synthesized Cs_2PbSnI_6 nanocrystals, whose measured photoluminescence peak at 794 nm matched the prediction, validating the model.

4. Applications of High-Throughput Simulation and ML in Perovskite Design

4.1. Structural Formability and Stability of Perovskite Design

The structural formability of perovskite lattice is the basis for material screening. Not all elements can occupy the A, B, or X sites in an ABX_3 perovskite, and the choice of ions directly determines whether the phase forms and what structure results. The tolerance factor (t) first proposed by Goldschmidt [17] and now the most widely used stability descriptor:

$$t = \frac{r_A + r_X}{\sqrt{2}(r_B + r_X)}$$

where r_A, r_B, and r_X are the ionic radii of the A-site cation, B-site cation, and X-site anion, respectively.

This t considers only ionic radii. By introducing the octahedral factor (μ) and the atomic packing ratio (ρ) to jointly refine t, the predictive accuracy for structural formability improves markedly. Sun et al. [18] built a stability–radius mapping from a high-throughput DFT database using kernel ridge regression, tuning regularization via five-fold cross-validation. They further proposed a composite descriptor $(\mu + t)^{\eta}$, where the octahedral factor (μ) enables the quantification of the intrinsic stability of BX_6 octahedra, while the exponent η—positively correlated with ρ—serves to balance the synergistic contributions of μ and t. This descriptor modestly improved the prediction accuracy (Figure 2).

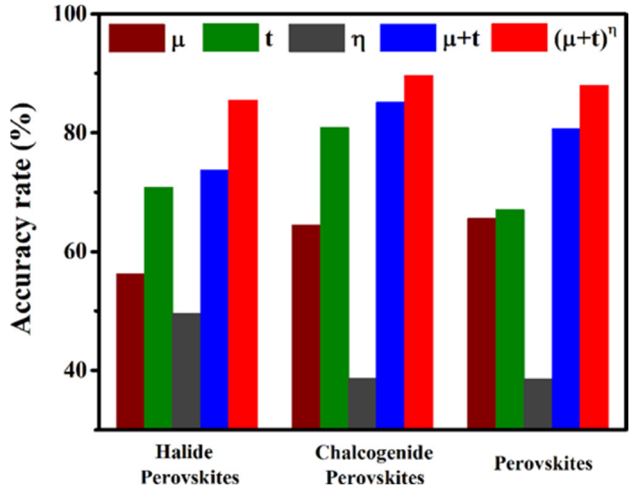


Figure 2. Accuracy rate for μ, t, η, (μ + t), and $(\mu + t)^{\eta}$ as a descriptor to predict the relative stabilities among two perovskites [18].

Perovskite stability depends on cation lattice parameters, bond strength and anisotropy, as well as anion selection and doping ratios. For multicomponent solid solutions, building high-dimensional compositional supercells makes DFT cost scale roughly cubically with atom count. Machine learning can effectively overcome this computational bottleneck. Shetty et al. [19] trained ML models on DFT data to predict hull energies E_{hull} for screening thermodynamically stable perovskites. Among six classifiers including RF, NN, and AdaBoost, AdaBoost delivered the best performance across four evaluation metrics for perovskite oxides. Krishnaraj et al. [20] noted that E_{hull} alone tends to overestimate stability and emphasized the need to include vibrational free energy. A polynomial–ML scheme (GPR/RF + SISSO) on 80 perovskites learned the coefficients of FH(T) and reduced the ZPE error from 18.9 to 8 meV/atom, which markedly improved thermodynamic stability assessment.

4.2. Applications of Perovskites in Photoelectric Effects

Perovskites, with high absorption, broad spectral response, excellent carrier transport, and low exciton binding energy, are valuable for solar cells, LEDs, and photodetectors. Even as lab efficiencies approach theoretical limits, screening materials for photoelectric performance remains a central focus in perovskite research.

4.2.1. Halide and Oxide Perovskite Solar Cells

The photovoltaic performance of perovskite solar cells is governed by light absorption, carrier transport and recombination, and interfacial stability [21]. In the ABX_3 lattice, corner-sharing $[BX_6]^{4-}$ octahedra formed by B-site cations (e.g., Pb^{2+}, Sn^{2+}) and halides (e.g., I^-, Br^-) give rise to direct bandgaps and large absorption coefficients, supporting effective spectral utilization [22]. The A-site cations regulate tolerance factors and lattice dynamics, stabilizing the perovskite framework through hydrogen-bonding interactions in hybrid compositions and predominantly ionic/steric effects in all-inorganic compositions. These structural and chemical controls lower defect densities, suppress nonradiative recombination [23], and extend carrier diffusion lengths, collectively improving device efficiency and operational robustness [24].

Given the wealth of data in the solar-cell field, key performance metrics for perovskite solar cells (e.g., bandgap, short-circuit current density, open-circuit voltage) can be taken directly from published literature. Notably, when the dataset is sufficiently large, missing values for some parameters have only a limited impact on training machine-learning

models [25]. Yan et al. [26] built a dataset from 248 papers and trained predictive models with Extreme Gradient Boosting and Random Forest (Figure 3). Descriptors included bandgap, short-circuit current density, open-circuit voltage, elemental ratios, and A-site cation type. The workflow predicted five previously unexplored low-bandgap perovskites of the form $(FAPbI_3)_X(MAPbBr_{2.8}Cl_{0.2})_{1-X}$ (X = 0.90, 0.92, 0.95, 0.97, 0.99). Experiments reported a best $(FAPbI_3)_{0.95}(MAPbBr_{2.8}Cl_{0.2})_{0.05}$ power conversion efficiency of 22.5% with a relative error below 2% versus the ML result. Li et al. [27] cleaned and curated bandgap data for 300 ($Cs_aFA_6MA_{(1-a-6)}Pb(Cl_XBr_YI_{(1-X-Y)})_3$, where the A-site is occupied by Cs^+, FA^+, or MA^+, and the X-site is occupied by Cl^-, Br^-, or I^-.) lead-halide perovskites and retained 109 valid entries. Three machine-learning models were trained on four input features (the mole fractions of Cs, FA, Cl, and Br) with five-fold cross-validation. The neural network achieved the lowest RMSE on the test set.

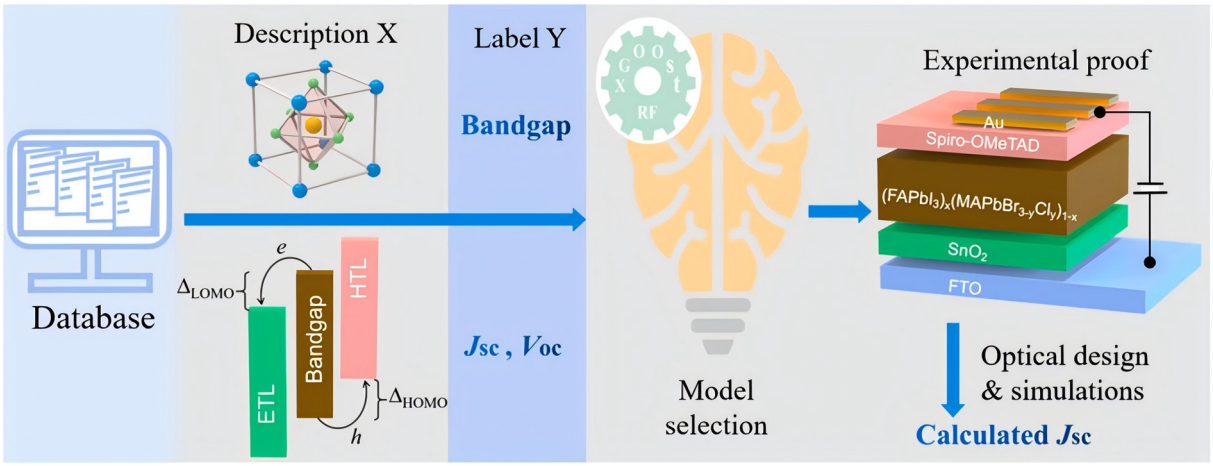


Figure 3. Flowchart for ML-assisted development of perovskite solar cells with high power conversion efficiency [26].

On the other hand, the literature data alone cannot support accurate prediction. First-principles calculations and related simulations are needed to re-evaluate the optoelectronic properties of perovskites and to supply more reliable training sets for machine-learning models [28]. Talapatra et al. [29] used high-throughput DFT to assemble bandgap data for 5152 perovskite oxides (ABO_3, $A_2BB'O_6$, $AA'B_2O_6$ and $AA'BB'O_6$) and built a staged pipeline with Random Forest classification and regression. The workflow first separates wide- and narrow-bandgap materials, then predicts bandgaps for the wide-bandgap set. The study flagged 13,589 potential wide-bandgap candidates (e.g., Ba_2TaInO_6, $LaCsTaDyO_6$), with 310 high-confidence new compounds (e.g., $DyCsTaInO_6$, Sr_2AuErO_6) showing a DFT-validated prediction error of 0.21 eV, and delivered a design map for how elemental substitution (e.g., alkaline earth or rare earth metals on the A/A′ sites and transition metals on the B/B′ sites) affects bandgaps. Zhu et al. [30] performed high-throughput DFT and computed 488 bandgap-related entries along with decomposition energy and spectroscopic-limit efficiency. Recursive feature elimination identified 10 key descriptors such as ionic radius and electronegativity. From 177,264 entries, the workflow retained 434 candidates and ultimately predicted four lead-free perovskites ($MABa_{0.125}Sn_{0.875}I_3$, $K_{0.25}FA_{0.75}Ba_{0.125}Sn_{0.875}I_3$, $FA_{0.25}MA_{0.75}Ba_{0.125}Sn_{0.875}I_3$ and $Cs_{0.375}FA_{0.625}Ba_{0.125}Sn_{0.875}I_3$) with high efficiency and strong stability (Figure 4). Xu et al. developed a machine learning-assisted pseudohalide anion screening method, selecting 168 candidate structures from 5 million molecules in the PubChem database through DFT—using $Cs_{0.05}FA_{0.9}MA_{0.05}Pb(I_{0.95}Br_{0.05})_3$ perovskite as a model system. Key descriptors such as topological polar surface area were used to construct a predictive model. Experimental

validation ultimately confirmed that sodium thioglycolate enhanced the perovskite solar cell efficiency to 24.56% and significantly improved device stability. Kim et al. significantly enhanced the performance of perovskite solar cells by developing a methoxyphenethylammonium iodide (M-PEAI)/cyclohexylammonium bromide (CHA)Br composite additive. This additive forms a quasi-2D intermediate layer on the perovskite surface, suppressing ion migration through strong intermolecular interactions and optimizing energy level alignment. Using a SnO_2/perovskite/additive/Spiro-OMeTAD device architecture, a power conversion efficiency of 26.28% was achieved, with 88% efficiency retention after 440 h of illumination. The study confirmed that the additive operates through a dual mechanism of interface modification and defect passivation, reducing trap density by an order of magnitude while endowing the device with 9000 h of environmental stability.

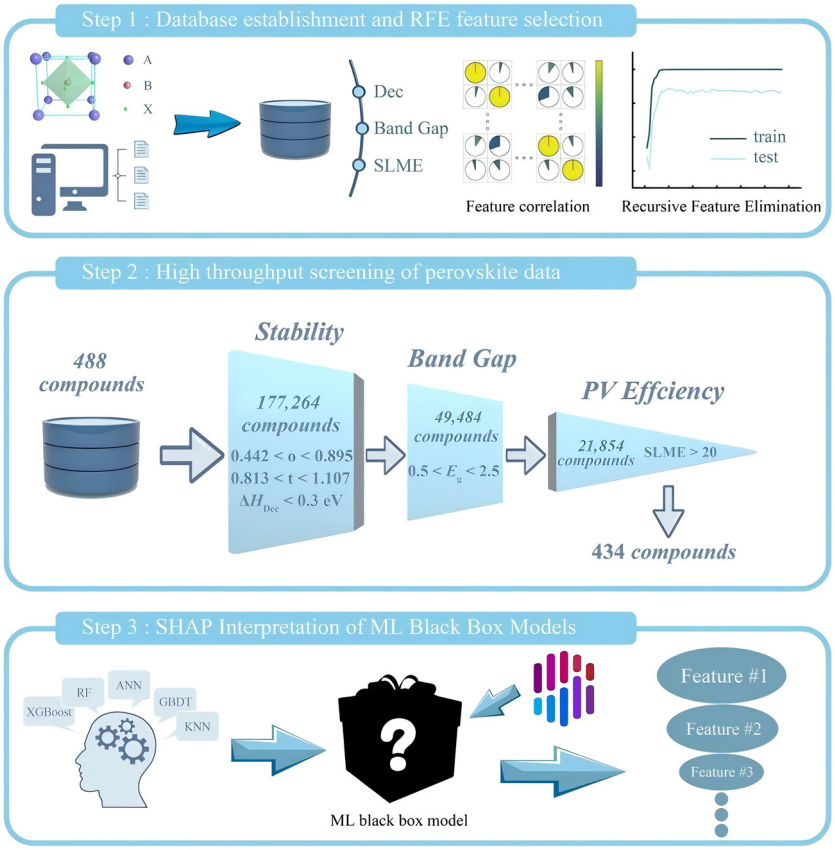


Figure 4. Step 1: database establishment and feature dimension reduction (by analyzing the correlations among all features through heat maps, it was found that there are strong correlations among many features, which may lead to overfitting of the model); Step 2: high-throughput screening; Step 3: ML model interpretation (different colors represent different input features. The length of the bar represents the magnitude and direction of the contribution of this feature to this prediction) [30].

4.2.2. Formamidinium Lead Triiodide Perovskite Light-Emitting Diodes

Perovskite light-emitting diode performance is governed by the interplay of exciton recombination, carrier balance, and defect control [31]. In the ABX_3 lattice, B–X quantum confinement enables narrowband emission with a very small full width at half maximum [32]. The A-site tunes carrier transport, yielding electron-to-hole mobility ratios on the same order of magnitude [33]. Zhang et al. [34] combined literature and in-house data to assemble a 132-additive database, merging MACCS, ECFP, and FP2 fingerprints with 208 RDKit descriptors to build an enriched molecular model. Joint feature integration with neural-network tuning mitigated small-sample bias and fingerprint redundancy, reaching

96% prediction accuracy. The workflow identified new additives, raising $FAPbI_3$ external quantum efficiency to 22.7%.

4.2.3. Two-Dimensional Metal Halide Perovskite Single-Crystal Photodetectors

Perovskite photodetectors benefit from the ABX_3 lattice, where strong B–X covalent hybridization yields a direct bandgap and high absorption [35]. Low defect density with strong dielectric screening suppresses nonradiative recombination [36]. Intrinsic ambipolar transport improves carrier dynamics [37]. Pandey et al. [38] compiled a 225-sample dataset (2D metal halide perovskites, e.g., BA_2PbI_4, $BA_2MAPb_2I_7$, $BA_2MA_3Pb_4I_{13}$, BA_2PbBr_4 and $BA_2MAPb_2Br_7$) from over 150 studies, covering nitrogen content, layer shift factor, bandgap, and related properties. Among four ML models, a decision tree gave the lowest errors for material responsivity and detectivity. Experiments on five photodetector types showed close agreement between predictions and measurements.

4.3. Perovskites as Catalytic Materials

Perovskites, with tunable crystal structures, strong redox activity, and intrinsic conductivity, have emerged as standout catalysts for green energy conversion and carbon neutrality [39]. A-site rare-earth/alkaline-earth cations (e.g., La^{3+}, Sr^{2+}) tune lattice distortion and oxygen-vacancy concentration, while B-site transition metals (e.g., Fe^{3+}, Co^{3+}) govern redox activity. Catalytic activity, selectivity, and stability of perovskite oxides are enhanced via A/B-site doping, oxygen-vacancy engineering, heterostructure construction, and double-perovskite design [40].

4.3.1. Perovskite Halide Photocatalyst

In perovskite halides photocatalysis, light generates electron–hole pairs that separate and migrate to the surface for redox reactions. Compared with conventional semiconductors, metal halide perovskites offer high absorption and strong tunability. Adjusting the halide composition tailors the bandgap to the visible range and improves photon-to-carrier conversion efficiency [41]. Typical applications include photocatalytic water splitting for H_2 [42], CO_2 photoreduction [43], and degradation of organic pollutants [44]. Zhai et al. [28] built a multi-parameter literature dataset and used support vector regression with an RBF kernel plus forward–backward stepwise selection to identify features governing perovskite halides specific surface area, yielding a high-accuracy model that speeds the design and screening of high-SSA photocatalysts. Biswas et al. [45] assembled a DFT-based dataset, added tree-structured regularization, and applied a fully corrected regularized greedy algorithm to predict decomposition energy and bandgap, outperforming ensemble baselines with broader generality. After hyperparameter tuning and evaluation using RMSE, the study selected 3043 halide perovskites from 151,140 compounds (Figure 5). Among these, the A-site contains inorganic cations (Cs, Rb) or organic cations (MA, FA), the B-site consists of alloyed divalent metals of Ge/Sn/Pb and Ca/Sr, the X-site is Br/I, and a typical example is $CsCa_{0.25}Ge_{0.75}Br_3$.

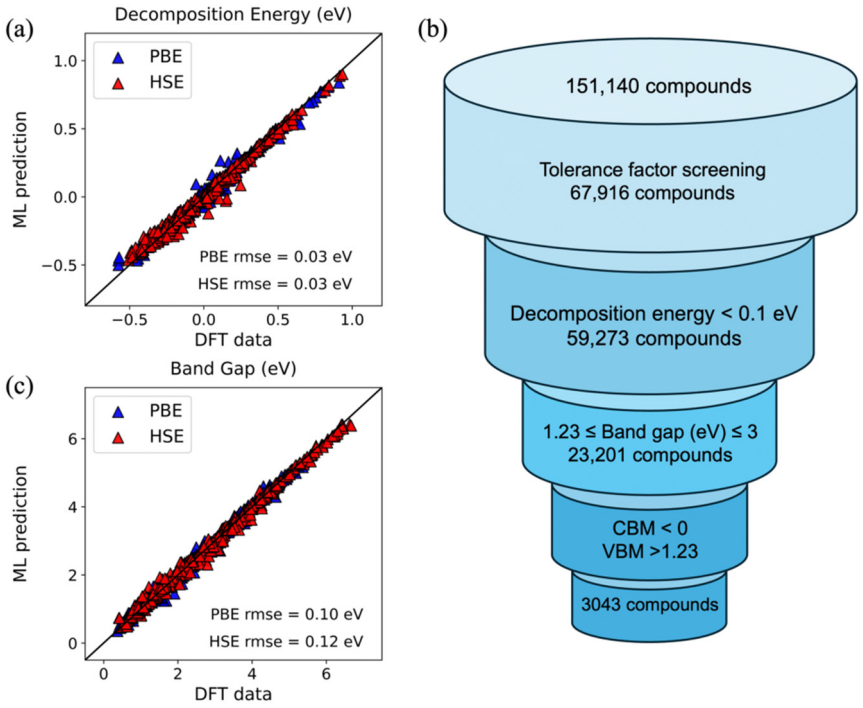


Figure 5. Parity plots for RGF models showing effective test predictions over 4000 runs plotted against ground truth DFT values for (**a**) decomposition energy per formula unit and (**c**) band gap. The screening procedure for identifying suitable perovskites for water splitting is pictured in (**b**) [46].

4.3.2. Perovskite Oxide Electrocatalyst

In electrocatalysis, perovskite oxides activity arises from the synergy of crystal structure, electronic states, and surface chemistry. Structural tuning enables precise optimization of the electronic structure, mainly via A/B-site doping and oxygen-vacancy engineering. This approach co-modulates intermediate adsorption, electron transfer efficiency, and reaction pathways. Perovskite oxide electrocatalysts are used in solid oxide fuel cells [46], water electrolysis for hydrogen [47], and CO_2 electroreduction [48].

The oxygen evolution reaction (OER) is a key anodic half-reaction in electrocatalysis. Perovskite oxides doped with selected transition metals, such as $LaCuO_3$ with B-site Co or Ni, have emerged as efficient OER catalysts. Li et al. [49] combined DFT-augmented datasets with Gaussian process regression to predict high-OER perovskite oxides, using composition and bulk-electronic descriptors with recursive feature elimination and an adaptive, uncertainty-aware learning strategy, yielding 10 stable and active double perovskite oxides such as $KRbCo_2O_6$ and $BaSrCo_2O_6$ from 4000 candidates. Chang et al. [50] introduced a transfer-learning framework for alkaline OER, building OER and non-OER datasets from literature, encoding cations with an autoencoder, clustering by K-Means and phase ranking, and using a tuned gradient boosting regressor with global ensembling and active learning, which identified high-performance compositions such as $Pr_{0.1}Sr_{0.9}Co_{0.5}Fe_{0.3}Mn_{0.3}O_3$ (Figure 6).

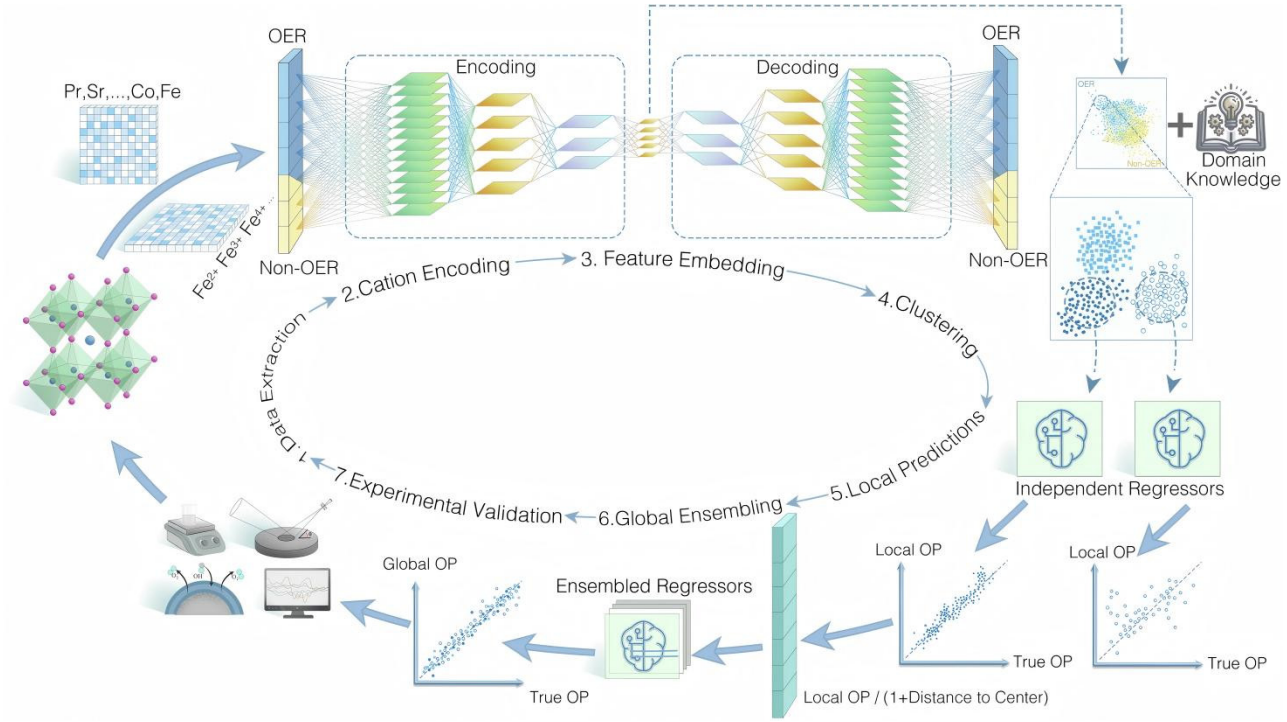


Figure 6. Transfer learning workflow to discover perovskite electrocatalysts for the oxygen evolution reaction [50].

Besides using intrinsic elemental properties as features, some studies take external conditions as the primary descriptors. Li et al. [51] used machine learning to design perovskite OER catalysts operable across temperatures. A literature-based dataset included oxygen-vacancy concentrations at different temperatures, with operating temperature as a key variable to predict equilibrium lattice vacancies. After benchmarking 36 models, a gradient boosting regressor performed best, then SHAP, DFT, and extensive experiments validated the model, yielding high-performance compositions at multiple temperatures, such as $SrCo_{0.8}Ta_{0.15}V_{0.05}O_{3-\delta}$ and $SrCo_{0.8}Ta_{0.16}W_{0.04}O_{3-\delta}$. Jacobs et al. [52] built a random-forest high-throughput framework to couple perovskite area-specific resistance with high-temperature operation, accelerating SOFC and electrolyzer catalyst discovery. A 749-entry database captured oxygen surface-exchange coefficients, diffusion, and area-specific resistance. Random-forest models for 500 °C and 800 °C mapped vacancy formation energy and active-site density to quantitative features, with SHAP and experiments confirming the approach, leading to low-ASR perovskite oxides at target temperatures.

4.3.3. Perovskite Oxide Chemical Looping Oxygen Carrier

Chemical looping splits complex reactions into multiple elementary steps via chemical intermediates. The oxygen carrier acts as the oxygen shuttle, typically a metal oxide, and reactions proceed through reversible lattice-oxygen release and uptake [53]. Perovskites offer high oxygen storage activity and synergistic catalysis, while precise A/B-site doping tunes equilibrium oxygen partial pressure and lattice-oxygen mechanisms, highlighting strong oxygen-carrier potential [54]. Perovskite carriers show promise in chemical looping air separation [55], CO_2 splitting [56], and oxygen uncoupling [57]. Oxygen-vacancy formation energy is the most common descriptor. Li et al. [58] used high-throughput DFT to build a perovskite oxide dataset, screened by oxygen-vacancy formation energy ΔG at fixed vacancy concentration, then trained a random forest with cation ratios as inputs and target ΔG under given conditions, yielding ~20,000 candidates from 227,273 high-entropy quinary-plus perovskites for chemical looping air separation and

CO_2 splitting, Such as $Sr_{0.875}Ba_{0.125}Fe_{0.5}Co_{0.5}O_{3-\delta}$ for chemical looping air separation, and $Sr_{0.5}Y_{0.5}Fe_{0.125}Ti_{0.875}O_{3-\delta}$ for CO_2 splitting. Ramazani et al. [59] compiled a DFT thermodynamic set for $SrFeO_{3-\delta}$ with single and co-doping on A/B sites, used ΔG as the key descriptor to assess oxygen release from 298–1200 K, and trained a k-nearest neighbors model with composition ratios as inputs and temperature-specific ΔG as targets, identifying two key families: $Sr_{1-x}Ca_xFe_{1-\gamma}Ni_\gamma O_{3-\delta}$ and the newly proposed $Sr_{1-x}Ba_xFe_{1-\gamma}Cu_\gamma O_{3-\delta}$.

To develop high-performance Chemical Looping with Oxygen Uncoupling materials, Duell et al. [60] constructed a DFT thermodynamic dataset for the $SrFeO_{3-\delta}$-based perovskite system. Using the Gibbs free energy of oxygen release (ΔG_{O2}) as the key descriptor, they evaluated the thermodynamic feasibility and kinetic potential of oxygen release for this system over the temperature range of 100–1200 K. In parallel, they trained a random forest regression model, with the characteristics of 39 substituent elements as inputs and the absolute value of ΔG_{O2} at a given temperature as the prediction objective. This model was then coupled with a genetic algorithm to iteratively optimize population composition, ultimately identifying two categories of key perovskite families: one is $Sr_{1-x}Ca_xFe_{1-\gamma}Ni_\gamma O_{3-\delta}$ (whose low-temperature performance advantages have been verified), and the other is the newly proposed $Sr_{1-x}A_xFe_{1-\gamma}B_\gamma O_{3-\delta}$ (where A = K; B = Mg, Bi, Co, Cu, Zn). To accelerate the development of perovskite oxides for thermochemical energy storage (TCES), Cai et al. [61] combined high-throughput DFT calculations with experimental methods. Based on a dataset of over 2000 $SrFeO_{3-\delta}$ materials doped at the A/B sites, they screened the materials using thermodynamic descriptors (e.g., formation energy and reaction enthalpy). Eventually, they demonstrated that $Sr_{0.875}Ba_{0.125}FeO_{3-\delta}$ and $Sr_{0.125}Ca_{0.875}Fe_{0.25}Mn_{0.75}O_{3-\delta}$ achieved energy densities of 85 $kJ \cdot kg^{-1}$ and 157 $kJ \cdot kg^{-1}$, respectively. This result confirms the effectiveness of high-throughput methods in accelerating the development of high-performance TCES materials.

4.4. Applications of Perovskites in Energy Storage

Perovskites offer high ionic conductivity and about 5% lattice strain tolerance [62]. Their strong defect tolerance broadens the electrochemical stability window from 0.5 to 4.5 V and suppresses side reactions [63]. Solution processing can form 2–50 nm hierarchical pores, increasing surface area and interfacial kinetics, which enables high specific capacity and durable cycling [64]. Current work largely focuses on antiperovskite solid electrolytes for solid-state batteries. Kim et al. [65] compiled over 600 DFT migration barriers for 36 alkali metal thiohalides based on X_3AZ antiperovskites, selected 20 key descriptors across five categories, and with AdaBoost plus extremely randomized trees (ERT) showed that channel width and hop distance govern vacancy migration, while interstitial formation energy strongly affects interstitial migration, highlighting synergy between anion polarizability and defect-formation features. Xiang et al. [66] aggregated room-temperature ionic conductivities for 403 electrolytes, built descriptors from Magpie and site-specific ion properties, used an extremely randomized trees model, and expanded virtual samples from 168 to 150,000 via ion substitution, identifying A-site electronegativity, density, and ionic radius as primary factors for conductivity. Zhang et al. [67] mined descriptors from 106 experimental samples, proposed the ratio of tolerance factor to atomic packing factor as a key descriptor, and combined classification with symbolic regression to pinpoint candidates with high ionic conductivity such as Li_6NClBr_2, informing solid-state electrolyte design. Table 1 summarizes the content of the previous paragraph.

Table 1. Summary of Machine Learning Models, Descriptors, and Datasets Used in Perovskite-Based Energy Storage Studies.

Descriptor	Model	Dataset	Metric	Found Candidates
CCD PW Anion polarizability	AdaBoost ERTR	603 DFT migration barriers	Vacancy RMSE (71 meV) Interstitial RMSE (46 meV)	Decreasing path distance Increasing width Highly polarizable anions
A-site electronegativity A-site density A-site ionic radius	ERT	403 sets of experimental SSE data 149,652 types of virtual AP materials	Classification accuracy (94%) MAE (0.57) R^2 (0.76) ROC_AUC (0.946)	16,807 superionic conductors
The ratio of tolerance factor to atomic packing factor (t/η)	Symbolic regression	106 experimental samples	RMSP (0.8)	Li_6NClBr_2 Li_6NBrBr_2 Li_6NBrI_2

5. Challenges and Future Directions for Applying ML to Perovskite Design

5.1. Bottlenecks and Challenges

Machine learning is emerging as a key driver for accelerating materials development, sharply shortening the discovery cycle for new materials. Current challenges center on three dimensions: data, process modeling, and model interpretability.

(1) Data Scarcity and Low Quality

Data standardization remains the main bottleneck for applying machine learning to materials development. Perovskite research lacks large, standardized experimental databases, and heterogeneity across groups causes feature distribution shift that degrades structure prediction and model training. Most studies still rely on few-shot or transfer learning, with data cleaning and standardization used to improve compatibility. The problem is acute in emerging areas such as perovskite LEDs and catalysis, where sparse data in high-dimensional spaces drives overfitting [68], while cross-lab variation in synthesis protocols and characterization methods introduces distribution shift and environmental noise that skew stability prediction. Even advanced models such as graph neural networks suffer when inputs are not standardized, showing poor cross-dataset generalization and reduced interpretability. Although data-standardization initiatives are underway worldwide, progress still trails algorithmic advances. Furthermore, perovskite materials and device performance are highly sensitive to micro-environmental conditions. Parameters often considered secondary or not systematically recorded—such as laboratory humidity during experiments, ambient temperature, and even the exposure time of precursor solutions to air—can significantly impact the performance and stability of the final product. This high dependency on hidden variables makes the standardization and comparability of datasets exceptionally difficult, greatly aggravating the issues of data scarcity and low quality.

(2) Model Generalization and Interpretability Gaps

Inherent limitations in model architecture, training strategies, and optimization methods can affect the accuracy of machine learning predictions. There is often a natural trade-off between the generalization capability and interpretability of AI algorithms. Traditional linear models or shallow neural networks struggle to capture the complex nonlinear relationships between perovskite composition, structure, and properties, limiting their generalization. Purely data-driven black-box models often lack explicit integration of

prior scientific knowledge, reducing interpretability and sometimes leading to unphysical predictions. Deep models handle complex process–structure relations yet still fall short on mechanisms: kGNNs cannot quantify how GBLD relates to device stability T_{80} [5]; SHAP showed I^- contributes 42% to bandgap yet explains interface recombination poorly [69]; lab-trained models lose about 30% on slot-die lines due to distribution shift [70]. These challenges stem from the unique ABX_3 octahedral structure of perovskites, and involve issues such as the data dependency of supervised learning (restricting generalization), the absence of physical constraints in unsupervised learning (affecting interpretability and consistency), and the large action space in reinforcement learning (hindering policy efficiency and generalization).

(3) Emphasis on Structure Prediction, Neglect of Fabrication and Application

Processing parameters (e.g., spin-coating speed, annealing curves) are often scattered in supplementary materials of research papers or kept as confidential industrial data, lacking standardized databases. Before 2020, most studies focused only on "predicting performance → recommending compositions" without addressing the full manufacturing process, leading to significant gaps between lab efficiency and industrial performance. This has pushed machine learning toward solving practical production challenges. As a representative example, Whitaker et al. [71] demonstrated that tuning drying/crystallization windows—via controlled wet film residence and coordinated flow rate, web speed, and die gap—enables slot die translation while preserving film morphology and device metrics during process transfer. On the other hand, fabrication processes involve multiphysics coupling (fluid dynamics, thermodynamics, crystallization kinetics), requiring temporal modeling and real-time feedback—far more complex than static structure prediction. Traditional models struggle to capture these dynamic correlations, necessitating deep learning to model nonlinear reactions. Examples include the nonlinear relationship between spin-coating speed (ω) and film thickness (d) [72], threshold effects in vapor-assisted crystallization [73], and the Arrhenius-type relationship between film crystallinity and annealing conditions [74]. These challenges require multimodal integration of process parameters for effective modeling.

5.2. Future Directions: Multimodal-Coupled Perovskite Design

In data generation, high-throughput computational methods have become an efficient approach for expanding datasets. Through high-throughput DFT and molecular dynamics simulations, vast amounts of material characteristic data can be systematically generated. High-throughput calculations enable continuous scanning of broader chemical spaces, increasing data volume by orders of magnitude while significantly accelerating exploration progress. Yao et al. [75] constructed the MatHub-2D database using high-throughput DFT, screening 1900 2D materials for carrier mobility and identifying 19 high-mobility semiconductors through deformation potential and elastic modulus analysis, achieving a data throughput 50 times greater than experimental synthesis. While experimental characterization provides localized information, high-throughput simulations can systematically compute all possible configurations to fill global data gaps. Thomas et al. [76] screened over 700 substitutional defect configurations in WS2 via high-throughput DFT, discovering that sulfur-site substitutions create ideal quantum defect energy levels, later confirmed by scanning tunneling microscopy experiments—covering defect chemical spaces far beyond experimental reach. Whereas experiments face limitations in multi-parameter control, high-throughput simulations can simultaneously scan multidimensional property spaces to establish comprehensive structure-property relationships. Li et al. [77] employed generative adversarial networks to create 50,000 virtual perovskites while calculating bandgap, defect tolerance, and carrier effective mass to guide targeted experimental validation. Mul-

timodal data fusion techniques enable effective integration of cross-scale, multi-source data through advanced machine learning methods. Data standardization serves as the bridge connecting experimental data with machine learning models, providing a stable and efficient foundation for intelligent full-process R&D. On the other hand, data standardization requires a minimal, machine-readable process metadata schema that the community can readily adopt, centered on universally accessible records; for halide perovskite solar cells this means recording formulation with precursor identities, purity and lot, solvent volume fraction and concentration, substrate and treatment, time-aligned coating and annealing profiles, clear XRD in 2θ with FWHM and film thickness, all provided in unified CSV or NetCDF with a JSON dictionary and SI units. For perovskite catalysts this means documenting synthesis or deposition history including route, stoichiometry and dopants, the thermal program with ramp, peak, dwell and cooling, gas partial pressure and flow, film deposition rate, clear XRD with scan parameters and optional XPS or BET, using the same unified CSV or NetCDF with JSON and SI units, and reporting electrochemical data in volts versus a stated reference and current density.

The integration of multiscale physical theories with data science enables high-performance predictions while preserving physical interpretability. Architecturally embedded physics approaches achieve this by coupling high-throughput screening with machine learning, where physical laws are mathematically transformed into model inductive biases to facilitate knowledge transfer in data-scarce scenarios. Li et al. [78] developed a multimodal material property prediction model incorporating physical constraints and graph neural networks (GNNs), along with a physics-guided generative optimization strategy, enabling efficient digital material design and performance prediction. Batzner et al. [79] created an equivariant GNN model for molecular dynamics simulations by embedding E(3) equivariance (including physical symmetries like rotation and reflection) into the network architecture, achieving efficient machine learning predictions for molecules and materials while supporting high-throughput screening. These architecturally embedded physics methods transform physical principles like molecular symmetry and topological properties into machine learning model biases. Some researchers have also embedded first-principles calculations as physical constraints. Lan et al. [80] proposed a framework combining high-throughput deep reinforcement learning with first-principles calculations, where potential energy surfaces derived from first principles were incorporated as internal constraints. This approach enabled thousands of parallel simulations on a single GPU for high-throughput screening, providing an efficient and general method for complex catalytic reaction pathway prediction. Current architecturally embedded physics methods face limitations in numerical stability and high implementation costs for scalable physical constraints. Although research remains limited in this area, the approach shows clear potential for deeply integrating physical priors with data-driven methods, representing a crucial development direction for intelligent material design.

Materials design is shifting from single-point trial-and-error to full-chain intelligent development. Most studies integrate physical constraints into simulation workflows. For instance, Lu et al. [81] incorporated the Goldschmidt tolerance factor (Tf = 0.8–1.2) and the octahedral factor (Of = 0.44–0.9) to assess structural stability, while also constraining the electronegativity of B-site metals (χ = 1.4–2.2 Pauling units) and the polarization of organic molecules (PA = 1–5 $\times$ 10^{-40} $C \cdot m^2 \cdot V^{-1}$) to optimize electronic properties. The application of these constraints reduced the rate of false-positive predictions in initial candidate screening from 35% to 12%, improved the mean squared error of bandgap predictions from 0.15 eV to 0.09 eV, and increased the coefficient of determination R^2 from 0.87 to 0.93. These results clearly demonstrate the substantial contribution of physical constraints in accelerating the discovery of stable lead-free perovskites. Researchers are building closed-loop optimization

that links high-throughput experiments with computation and real-time feedback, while coupling process modeling with dynamic, data-driven control to boost model performance, prediction accuracy, and process stability. MIT's platform [82,83] runs on the order of a hundred perovskite synthesis trials per day, adjusts annealing temperature and additive ratios on the fly, and tightly integrates an automated lab with machine learning for continuous process optimization. Ultrafast electron microscopy and other in situ probes feed degradation dynamics back into physics-informed neural networks to recalibrate diffusion and related parameters, which iteratively improves interface defect prediction. Li et al. [84] used LSTM models to parse in situ photoluminescence in real time, identified a critical threshold in vapor-assisted crystallization, and actively steered crystallization kinetics, cutting film defect density by three orders of magnitude. Graupner et al. [85] combined ML force fields with molecular dynamics in a multiscale framework to capture precursor clustering and film-scale morphology evolution, then designed graded 2D/3D architectures that enabled devices beyond 30% efficiency. Specifically, for single-junction solar cells based on $(BA)_2(MA)_3Pb_4I_{13}$, their work identified a complete energy funneling mechanism that suppresses non-radiative recombination and enhances exciton lifetime, suggesting a theoretical PCE exceeding 30% under ideal conditions. The USTC autonomous lab [86] integrates AI, robotics, and high-throughput experimentation into a self-driving pipeline that progresses from automated execution to closed-loop decide–experiment–optimize, and incorporates quantum computing with AI to accelerate materials discovery. This system shortens material discovery and optimization from years to weeks, improves candidate identification success rates by orders of magnitude, reduces R&D manpower by over 90%, and employs intelligent algorithms to screen stable material configurations. Similarly, the self-driving laboratory at MIT performs over 100 perovskite syntheses per day, reducing R&D duration from months to days. It achieves a three-order-of-magnitude reduction in defect density, a tenfold improvement in device stability, and enables solar cells with over 30% efficiency, while saving more than 95% of manual operational effort. Together these advances form an intelligent loop of real-time data capture, mechanism learning, and adaptive optimization, moving perovskite R&D from isolated automation toward fully integrated, end-to-end autonomy.

6. Conclusions

ML is becoming the core engine of data-driven perovskite design, especially in integration with high-throughput simulation. This pairing speeds stability evaluation, bandgap and emission tuning, defect and dopant optimization, and catalysis-relevant property discovery, moving candidates to validated leads more quickly. Open databases support rapid large-scale triage, and tight coupling with high-throughput DFT and molecular dynamics improves accuracy and robustness beyond the training domain. Future progress depends on standardized data and the explicit inclusion of physics in model architectures and in loss functions to enhance interpretability and reliability. With active learning and self-driving experimentation in the loop, the simulation-ML workflow closes the cycle from data acquisition and mechanism inference to process optimization, enabling trustworthy predictions and faster translation to devices.

A 24-month feasible roadmap

Months 0–6—Strengthen HT simulations as reliable ML data

Focus on scaling and validating high-throughput DFT/MD to provide consistent, high-quality labels for ML. Harmonize simulation protocols, units, and metadata; prioritize stability indicators (e.g., Ehull with finite-T corrections where available), bandgaps, and process-relevant properties.

Months 6–12—Embed physics for interpretability and reliability

Introduce symmetry-aware/equivariant architectures and lightweight physical constraints (charge neutrality, tolerance-factor bounds, defect thermodynamics) to limit unphysical extrapolation. Couple HT calculations with uncertainty-aware active learning and conduct small-batch validations.

Months 12–24—Multimodal, process-aware, end-to-end workflows

Develop multimodal models that fuse composition/structure descriptors, microstructural images, and key process variables to predict device metrics and stability under operating conditions. Deploy closed-loop optimization with self-driving or partner labs, integrate in-situ signals.

Author Contributions: Y.W.: Writing—original draft, Software, Data curation, Investigation. D.S.: Writing—original draft, Investigation, Formal analysis. B.Z.: Writing—original draft, Investigation, Formal analysis. T.Z. (Tianyu Zhu): Writing—review & editing, Software, Data curation. C.L.: Formal analysis, Validation, Data curation. Z.X.: Formal analysis, Validation, Funding acquisition. T.Z. (Tianhang Zhou): Writing—review & editing, Supervision, Funding acquisition, Conceptualization. C.X.: Writing—review & editing, Supervision, Funding acquisition. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China grant number No. 22408400 and Science Foundation of China University of Petroleum, Beijing grant number No. 2462023SZBH010.

Data Availability Statement: No new data were created or analyzed in this study.

Acknowledgments: We thank Mark Inglin (University of Basel) for his editorial assistance.

Conflicts of Interest: The author declares no conflict of interest.

References

1. Veldhuis, S.A.; Boix, P.P.; Yantara, N.; Li, M.; Sum, T.C.; Mathews, N.; Mhaisalkar, S.G. Perovskite Materials for Light-Emitting Diodes and Lasers. *Adv. Mater.* **2016**, *28*, 6804–6834. [CrossRef]
2. Emery, A.A.; Wolverton, C. High-throughput DFT calculations of formation energy, stability and oxygen vacancy formation energy of ABO_3 perovskites. *Sci. Data* **2017**, *4*, 170153. [CrossRef]
3. Du, X.; Damewood, J.K.; Lunger, J.R.; Millan, R.; Yildiz, B.; Li, L.; Gómez-Bombarelli, R. Machine-learning-accelerated simulations to enable automatic surface reconstruction. *Nat. Comput. Sci.* **2023**, *3*, 1034–1044. [CrossRef]
4. Jacobsson, T.J.; Hultqvist, A.; García-Fernández, A.; Anand, A.; Al-Ashouri, A.; Hagfeldt, A.; Crovetto, A.; Abate, A.; Ricciardulli, A.G.; Vijayan, A.; et al. An open-access database and analysis tool for perovskite solar cells based on the FAIR data principles. *Nat. Energy* **2022**, *7*, 107–115. [CrossRef]
5. Zhang, Z.; Wang, H.; Jacobsson, T.J.; Luo, J. Big data driven perovskite solar cell stability analysis. *Nat. Commun.* **2022**, *13*, 7639. [CrossRef]
6. Aladerah, B.; Alrousan, A.; Obeidat, A.; Al-Sharif, A. Properties of cubic $GdAlO_3$ perovskite under pressure: Density functional theory and Monte Carlo simulations. *J. Comput. Electron.* **2024**, *24*, 12. [CrossRef]
7. Kholmurodov, K.; Ibragimova, S.; Gladishev, P.; Vannikov, A.; Tameev, A.; Zelenyak, T. Molecular Dynamics Simulations of Perovskites: The Effect of Potential Function Representation on Equilibrium Structural Properties. *Open J. Phys. Chem.* **2015**, *5*, 110–121. [CrossRef]
8. Zhang, J.-T.; Ji, K.; Xie, Y.; Li, C. Perovskite-based two-dimensional ferromagnet Sr_2RuO_4 monolayer. *Acta Phys. Sin.* **2024**, *73*, 226101. [CrossRef]
9. Gu, F.; Li, Q.; Li, S.; Xiao, J. Micro-mechanism of interaction between components of perovskite energetic material ABX_3 on structure and sensitivity: DFT study. *Appl. Mater. Today* **2025**, *42*, 102593. [CrossRef]
10. Yin, W.-J.; Shi, T.; Yan, Y. Unusual Defect Physics in $CH_3NH_3PbI_3$ Perovskite Solar Cell Absorber. *Appl. Phys. Lett.* **2014**, *104*, 063903. [CrossRef]
11. Bessa, M.; Bostanabad, R.; Liu, Z.; Hu, A.; Apley, D.; Brinson, C.; Chen, W.; Liu, W. A framework for data-driven analysis of materials under uncertainty: Countering the curse of dimensionality. *Comput. Methods Appl. Mech. Eng.* **2017**, *320*, 633–667. [CrossRef]

12. Hu, T.; Ye, Z.; Wang, Y.; Gao, X.; Sun, Z.; Li, J.L.; Chen, S.; Lian, C.; Xu, Q.; Li, F. Synergistic Effect of H-bond Reconstruction and Interface Regulation for High-Voltage Aqueous Energy Storage. *Adv. Energy Mater.* **2023**, *13*, 2300567. [CrossRef]
13. Zhu, Y.; Zhang, J.; Qu, Z.; Jiang, S.; Liu, Y.; Wu, Z.; Yang, F.; Hu, W.; Xu, Z.; Dai, Y. Accelerating stability of ABX_3 perovskites analysis with machine learning. *Ceram. Int.* **2024**, *50*, 6250–6258. [CrossRef]
14. Wan, Z.; Ding, B.; Su, J.; Su, Z.; Li, Z.; Jia, C.; Jiang, Z.; Qin, Q.; Zhang, M.; Shi, J.; et al. Efficient perovskite solar modules with an ultra-long processing window enabled by cooling stabilized intermediate phases. *Energy Environ. Sci.* **2024**, *17*, 6302–6313. [CrossRef]
15. Yang, C.; Chong, X.; Hu, M.; Yu, W.; He, J.; Zhang, Y.; Feng, J.; Zhou, Y.; Wang, L.-W. Accelerating the Discovery of Hybrid Perovskites with Targeted Band Gaps via Interpretable Machine Learning. *ACS Appl. Mater. Interfaces* **2023**, *15*, 40419–40427. [CrossRef]
16. Mishra, P.; Zhang, M.; Kar, M.; Hellgren, M.; Casula, M.; Lenz, B.; Chen, A.; Recatala, J.; Padhy, S.; Trouche, M.; et al. Synthesis of Machine Learning-Predicted Cs_2PbSnI_6 Double Perovskite Nanocrystals. *ACS Nano* **2025**, *19*, 6107–6119. [CrossRef]
17. Goldschmidt, V.M. Die Gesetze der Krystallochemie. *Naturwissenschaften* **1926**, *14*, 477–485. [CrossRef]
18. Sun, Q.; Yin, W.-J. Thermodynamic Stability Trend of Cubic Perovskites. *JACS* **2017**, *139*, 14905–14908. [CrossRef]
19. Shetty, V.; Shedthi, B.S.; Kumaraswamy, J. Predicting the thermodynamic stability of perovskite oxides using multiple machine learning techniques. *Mater. Today* **2022**, *52*, 457–461. [CrossRef]
20. Kundavu, K.; Mondal, S.; Bhattacharya, A. Machine learning the vibrational free energy of perovskites. *Mater. Adv.* **2023**, *4*, 4238–4249. [CrossRef]
21. Kim, H.-S.; Lee, C.-R.; Im, J.-H.; Lee, K.-B.; Moehl, T.; Marchioro, A.; Moon, S.-J.; Humphry-Baker, R.; Yum, J.-H.; Moser, J.-E.; et al. Lead Iodide Perovskite Sensitized All-Solid-State Submicron Thin Film Mesoscopic Solar Cell with Efficiency Exceeding 9%. *Sci. Rep.* **2012**, *2*, 591. [CrossRef] [PubMed]
22. De Wolf, S.; Holovsky, J.; Moon, S.-J.; Löper, P.; Niesen, B.; Ledinský, M.; Haug, F.-J.; Yum, J.-H.; Ballif, C. Organometallic Halide Perovskites: Sharp Optical Absorption Edge and Its Relation to Photovoltaic Performance. *J. Phys. Chem. Lett.* **2014**, *5*, 1035–1039. [CrossRef]
23. Shi, D.; Adinolfi, V.; Comin, R.; Yuan, M.; Alarousu, E.; Buin, A.; Chen, Y.; Hoogland, S.; Rothenberger, A.; Katsiev, K.; et al. Solar cells. Low trap-state density and long carrier diffusion in organolead trihalide perovskite single crystals. *Science* **2015**, *347*, 519–522. [CrossRef] [PubMed]
24. Stranks, S.; Eperon, G.; Grancini, G.; Menelaou, C.; Alcocer, M.; Leijtens, T.; Herz, L.; Petrozza, A.; Snaith, H. Electron-Hole Diffusion Lengths Exceeding 1 Micrometer in an Organometal Trihalide Perovskite Absorber. *Science* **2013**, *342*, 341–344. [CrossRef]
25. Bull, K.; He, Y.-H.; Jejjala, V.; Challenger, M. Machine learning CICY threefolds. *Phys. Lett. B* **2018**, *785*, 65–72. [CrossRef]
26. Yan, W.; Liu, Y.; Zang, Y.; Cheng, J.; Wang, Y.; Chu, L.; Tan, X.; Liu, L.; Zhou, P.; Li, W.; et al. Machine learning enabled development of unexplored perovskite solar cells with high efficiency. *Nano Energy* **2022**, *99*, 107394. [CrossRef]
27. Li, Y.; Yao, L.; Huo, X.; Wei, D.; Meng, J.; Dong, J.; Qiao, B.; Zhao, S.; Xu, Z.; Song, D. Bandgap tuning strategy by cations and halide ions of lead halide perovskites learned from machine learning. *RSC Adv.* **2021**, *11*, 15688–15694. [CrossRef]
28. Zhai, X.; Chen, M. Accelerated Design for Perovskite-Oxide-Based Photocatalysts Using Machine Learning Techniques. *Materials* **2024**, *17*, 3026. [CrossRef]
29. Talapatra, A.; Uberuaga, B.; Stanek, C.; Pilania, G. Band gap predictions of double perovskite oxides using machine learning. *Commun. Mater.* **2023**, *4*, 46. [CrossRef]
30. Zhu, C.; Liu, Y.; Wang, D.; Zhu, Z.; Zhou, P.; Tu, Y.; Yang, G.; Chen, H.; Zang, Y.; Du, J.; et al. Exploration of highly stable and highly efficient new lead-free halide perovskite solar cells by machine learning. *Cell Rep. Phys. Sci.* **2024**, *5*, 102321. [CrossRef]
31. Cao, Y.; Wang, N.; Tian, H.; Guo, J.; Wei, Y.; Chen, H.; Miao, Y.; Zou, W.; Pan, K.; He, Y.; et al. Perovskite light-emitting diodes based on spontaneously formed submicrometre-scale structures. *Nature* **2018**, *562*, 249–253. [CrossRef]
32. Protesescu, L.; Yakunin, S.; Bodnarchuk, M.; Krieg, F.; Caputo, R.; Hendon, C.; Yang, R.; Walsh, A.; Kovalenko, M. Nanocrystals of Cesium Lead Halide Perovskites ($CsPbX_3$, X = Cl, Br, and I): Novel Optoelectronic Materials Showing Bright Emission with Wide Color Gamut. *Nano Lett.* **2015**, *15*, 3692–3696. [CrossRef] [PubMed]
33. Bretscher, H.; Li, Z.; Xiao, J.; Qiu, D.; Refaely-Abramson, S.; Alexander-Webber, J.; Tanoh, A.; Fan, Y.; Delport, G.; Williams, C.; et al. Rational Passivation of Sulfur Vacancy Defects in Two-Dimensional Transition Metal Dichalcogenides. *ACS Nano* **2021**, *15*, 8780–8789. [CrossRef]
34. Zhang, L.; Li, N.; Liu, D.; Tao, G.; Xu, W.; Li, M.; Chu, Y.; Cao, C.; Lu, F.; Hao, C.; et al. Deep Learning for Additive Screening in Perovskite Light-Emitting Diodes. *Angew. Chem. Int. Ed.* **2022**, *61*, e202209337. [CrossRef] [PubMed]
35. Rogalski, A.; Hu, W.; Wang, F.; Wang, Y.; Martyniuk, P. Perovskite versus Standard Photodetectors. *Materials* **2024**, *17*, 4029. [CrossRef]
36. Ghimire, S.; Klinke, C. Two-Dimensional halide perovskites: Synthesis, optoelectronic properties, stability, and applications. *Nanoscale* **2021**, *13*, 12394–12422. [CrossRef]

37. Miah, M.; Khandaker, M.; Islam, M.; Alam, M.; Osman, H.; Ullah, M.H. Perovskite materials in X-ray detection and imaging: Recent progress, challenges, and future prospects. *RSC Adv.* **2024**, *14*, 6656–6698. [CrossRef]
38. Pandey, S.V.; Parikh, N.; Kalam, A.; Prochowicz, D.; Satapathi, S.; Akin, S.; Tavakoli, M.M.; Yadav, P. A machine learning framework for predicting device performance in 2D metal halide perovskite photodetector. *Sol. Energy* **2024**, *270*, 112399. [CrossRef]
39. Humayun, M.; Li, Z.; Israr, M.; Khan, A.; Luo, W.; Wang, C.; Shao, Z. Perovskite Type ABO_3 Oxides in Photocatalysis, Electrocatalysis, and Solid Oxide Fuel Cells: State of the Art and Future Prospects. *Chem. Rev.* **2025**, *125*, 3165–3241. [CrossRef]
40. Labhasetwar, N.; Govindachetty, S.; Suresh, K.M.; Nilesh, M.; Rohini, K.; Pradeep, D.; Grasset, F. Perovskite-type catalytic materials for environmental applications. *Sci. Technol. Adv. Mater.* **2015**, *16*, 036002. [CrossRef]
41. Bresolin, B.-M.; Park, Y.; Bahnemann, D.W. Recent Progresses on Metal Halide Perovskite-Based Material as Potential Photocatalyst. *Catalysts* **2020**, *10*, 709. [CrossRef]
42. Subha, N.; Sankar, A.R.; Navaneethakrishnan, S.; Lavanya, J.; Aakash, M. Perovskite-based Z-scheme photocatalytic system for hydrogen production. *Catal. Commun.* **2024**, *187*, 106903. [CrossRef]
43. Wang, J.; Xiong, L.; Bai, Y.; Chen, Z.J.; Zheng, Q.; Shi, Y.Y.; Zhang, C.; Jiang, G.C.; Li, Z.Q. Mn-Doped Perovskite Nanocrystals for Photocatalytic CO_2 Reduction: Insight into the Role of the Charge Carriers with Prolonged Lifetime. *Sol. RRL* **2022**, *6*, 2200294. [CrossRef]
44. Masri, M.; Girisha, K.B.; Hezam, A.; Alkanad, K.; Prashantha, K.; Manjunath, S.H.; Udayabhanu; Masri, F.; Qahtan, T.F.; Byrappa, K. Metal halide perovskite-based photocatalysts for organic pollutants degradation: Advances, challenges, and future directions. *Colloid Surf. A* **2024**, *687*, 133387. [CrossRef]
45. Biswas, M.; Desai, R.; Mannodi-Kanakkithodi, A. Screening of novel halide perovskites for photocatalytic water splitting using multi-fidelity machine learning. *PCCP* **2024**, *26*, 23177–23188. [CrossRef] [PubMed]
46. Gazda, M.; Jasinski, P.; Kusz, B.; Bochentyn, B.; Gdula-Kasica, K.; Lendze, T.; Lewandowska-Iwaniak, W.; Mielewczyk-Gryn, A.; Molin, S. Perovskites in Solid Oxide Fuel Cells. *Solid State Phenom.* **2011**, *183*, 65–70. [CrossRef]
47. Kim, D.; Oh, L.S.; Park, J.H.; Kim, H.J.; Lee, S.G.Y.; Lim, E. Perovskite-based electrocatalysts for oxygen evolution reaction in alkaline media: A mini review. *Front. Chem.* **2022**, *10*, 1024865. [CrossRef]
48. Chen, J.Y.; Gao, X.; Chen, X.D.; Zhen, Z.; Chen, Y.; Zeng, X.T.; Cui, L.F. Recent advances of perovskite oxide-based cathodes in solid oxide electrolysis cells for CO_2 electroreduction. *Mater. Today Phys.* **2023**, *38*, 101237. [CrossRef]
49. Li, Z.; Achenie, L.E.K.; Xin, H. An Adaptive Machine Learning Strategy for Accelerating Discovery of Perovskite Electrocatalysts. *ACS Catal.* **2020**, *10*, 4377–4384. [CrossRef]
50. Jiang, C.; He, H.Y.; Guo, H.Q.; Zhang, X.X.; Han, Q.Y.; Weng, Y.H.; Fu, X.Z.; Zhu, Y.L.; Yan, N.; Tu, X.; et al. Transfer learning guided discovery of efficient perovskite oxide for alkaline water oxidation. *Nat. Commun.* **2024**, *15*, 6301. [CrossRef]
51. Li, Z.; Mao, X.; Feng, D.; Li, M.; Xu, X.; Luo, Y.; Zhuang, L.; Lin, R.; Zhu, T.; Liang, F.; et al. Prediction of perovskite oxygen vacancies for oxygen electrocatalysis at different temperatures. *Nat. Commun.* **2024**, *15*, 9318. [CrossRef]
52. Jacobs, R.; Liu, J.; Abernathy, H.; Morgan, D. Machine Learning Design of Perovskite Catalytic Properties. *Adv. Energy Mater.* **2024**, *14*, 2303684. [CrossRef]
53. Dawa, T.; Sajjadi, B. Exploring the potential of perovskite structures for chemical looping technology: A state-of-the-art review. *Fuel Process. Technol.* **2024**, *253*, 108022. [CrossRef]
54. Ahmad, A.; Al Mamun, M.A.; Al-Mamun, M.; Huque, S.; Ismail, M. LFO Perovskites as Oxygen Carriers for Chemical Looping Oxygen Uncoupling (CLOU). *J. Therm. Anal. Calorim.* **2022**, *147*, 6605–6613. [CrossRef]
55. Dou, J.; Krzystowczyk, E.; Mishra, A.; Liu, X.B.; Li, F.X. Perovskite Promoted Mixed Cobalt-Iron Oxides for Enhanced Chemical Looping Air Separation. *ACS Sustain. Chem. Eng.* **2018**, *6*, 15528–15540. [CrossRef]
56. Ramos, A.E.; Maiti, D.; Daza, Y.A.; Kuhn, J.N.; Bhethanabotla, V.R. Co, Fe, and Mn in La-perovskite oxides for low temperature thermochemical CO_2 conversion. *Catal. Today* **2019**, *338*, 52–59. [CrossRef]
57. Galinsky, N.; Mishra, A.; Zhang, J.; Li, F.X. $Ca_{1-x}A_xMnO_3$ (A = Sr and Ba) perovskite based oxygen carriers for chemical looping with oxygen uncoupling (CLOU). *Appl. Energy* **2015**, *157*, 358–367. [CrossRef]
58. Wang, X.J.; Gao, Y.F.; Krzystowczyk, E.; Iftikhar, S.; Dou, J.; Cai, R.X.; Wang, H.Y.; Ruan, C.Y.; Ye, S.; Li, F.X. High-throughput oxygen chemical potential engineering of perovskite oxides for chemical looping applications. *Energy Environ. Sci.* **2022**, *15*, 1512–1528. [CrossRef]
59. Ramazani, A.; Duell, B.A.; Popczun, E.J.; Natesakhawat, S.; Nandi, T.; Lekse, J.W.; Duan, Y.H. High-throughput *ab initio* calculations and machine learning to discover $SrFeO_{3-6}$-based perovskites for chemical-looping applications. *Cell Rep. Phys. Sci.* **2024**, *5*, 16. [CrossRef]
60. Duell, B.A.; Ramazani, A.; Natesakhawat, S.; Popczun, E.J.; Lekse, J.W.; Duan, Y. Targeted Chemical Looping Materials Discovery by an Inverse Design. *Adv. Intell. Syst.* **2025**, *7*, 2401118. [CrossRef]

61. Cai, R.; Bektas, H.; Wang, X.; McClintock, K.; Teague, L.; Yang, K.; Li, F. Accelerated Perovskite Oxide Development for Thermochemical Energy Storage by a High-Throughput Combinatorial Approach. *Adv. Energy Mater.* **2023**, *13*, 2203833. [CrossRef]
62. Sotoudeh, M.; Baumgart, S.; Dillenz, M.; Döhn, J.; Forster-Tonigold, K.; Helmbrecht, K.; Stottmeister, D.; Gross, A. Ion Mobility in Crystalline Battery Materials. *Adv. Energy Mater.* **2023**, *14*, 2302550. [CrossRef]
63. Dawson, J.; Famprikis, T.; Johnston, K. Anti-Perovskites for Solid-State Batteries: Recent Developments, Current Challenges and Future Prospects. *J. Mater. Chem. A* **2021**, *9*, 18746–18772. [CrossRef]
64. Ha, T.; Su, R.; Xing, J.; Zhang, Q.; Xiong, Q. Metal halide perovskite nanomaterials: Synthesis and applications. *Chem. Sci.* **2016**, *8*, 2522–2536. [CrossRef]
65. Kim, K.; Siegel, D. Machine Learning Reveals Factors that Control Ion Mobility in Anti-Perovskite Solid Electrolytes. *J. Mater. Chem. A* **2022**, *10*, 15169–15182. [CrossRef]
66. Xiang, S.; Lu, S.; Li, J.; Xie, K.; Zhu, R.; Wang, H.; Huang, K.; Li, C.; Wu, J.; Chen, S.; et al. Ionic Conductivity Study of Antiperovskite Solid-State Electrolytes Based on Interpretable Machine Learning. *ACS Appl. Energy Mater.* **2025**, *8*, 1620–1628. [CrossRef]
67. Zhang, Z.; Chu, J.; Zhang, H.; Liu, X.; He, M. Mining ionic conductivity descriptors of antiperovskite electrolytes for all-solid-state batteries via machine learning. *J. Energy Storage* **2024**, *75*, 109714. [CrossRef]
68. Jo, B.; Chen, W.; Jung, H. Comprehensive review of advances in machine-learning-driven optimization and characterization of perovskite materials for photovoltaic devices. *J. Energy Chem.* **2024**, *101*, 298–323. [CrossRef]
69. Zhang, Q.; Wang, H.; Qiangqiang, Z.; Ullah, A.; Zhong, X.; Wei, Y.; Zhang, C.; Xu, R.; De Wolf, S.; Wang, K. Machine-Learning-Assisted Design of Buried- Interface Engineering Materials for High- Efficiency and Stable Perovskite Solar Cells. *ACS Energy Lett.* **2024**, *9*, 5924–5934. [CrossRef]
70. Yang, C.; Hu, W.; Liu, J.; Han, C.; Gao, Q.; Mei, A.; Zhou, Y.; Guo, F.; Han, H. Achievements, challenges, and future prospects for industrialization of perovskite solar cells. *Light. Sci. Appl.* **2024**, *13*, 227. [CrossRef]
71. Whitaker, J.B.; Kim, D.H.; Larson, B.W.; Zhang, F.; Berry, J.J.; van Hest, M.F.A.M.; Zhu, K. Scalable slot-die coating of high performance perovskite solar cells. *Sustain. Energy Fuels* **2018**, *2*, 2442–2449. [CrossRef]
72. Milot, R.; Eperon, G.; Snaith, H.; Johnston, M.; Herz, L. Temperature-Dependent Charge-Carrier Dynamics in $CH_3NH_3PbI_3$ Perovskite Thin Films. *Adv. Funct. Mater.* **2015**, *25*, 6218–6227. [CrossRef]
73. Rong, Y.; Hou, X.; Hu, Y.; Mei, A.; Liu, L.; Wang, P.; Han, H. Synergy of ammonium chloride and moisture on perovskite crystallization for efficient printable mesoscopic solar cells. *Nat. Commun.* **2017**, *8*, 14555. [CrossRef] [PubMed]
74. Liu, H.; Antai, Y.; Zhong, C.; Zhu, X.; Meng, H.; Feng, Z.; Tang, J.; Yang, C.; Zhang, J.; Liu, J.; et al. Integration of Microstructural Image Data into Machine Learning Models for Advancing High-Performance Perovskite Solar Cell Design. *ACS Energy Lett.* **2025**, *10*, 1884–1891. [CrossRef]
75. Yao, M.; Ji, J.; Li, X.; Zhu, Z.; Ge, J.-Y.; Singh, D.J.; Xi, J.; Yang, J.; Zhang, W. MatHub-2d: A database for transport in 2D materials and a demonstration of high-throughput computational screening for high-mobility 2D semiconducting materials. *Sci. China Mater.* **2023**, *66*, 2768–2776. [CrossRef]
76. Thomas, J.C.; Chen, W.; Xiong, Y.; Barker, B.A.; Zhou, J.; Chen, W.; Rossi, A.; Kelly, N.; Yu, Z.; Zhou, D.; et al. A substitutional quantum defect in WS_2 discovered by high-throughput computational screening and fabricated by site-selective STM manipulation. *Nat. Commun.* **2024**, *15*, 3556. [CrossRef]
77. Li, J.-B.; Jiang, Z.-K.; Wang, R.; Zhao, J.-Z.; Wang, R. Ferroelectric order in hybrid organic-inorganic perovskite NH_4PbI_3 with non-polar molecules and small tolerance factor. *npj Comput. Mater.* **2023**, *9*, 62. [CrossRef]
78. Li, H.; Yang, J.; Yao, J.; Sheng, C. Digitized material design and performance prediction driven by high-throughput computing. *Front. Mater.* **2025**, *12*, 9439. [CrossRef]
79. Batzner, S.; Musaelian, A.; Sun, L.; Geiger, M.; Mailoa, J.P.; Kornbluth, M.; Molinari, N.; Smidt, T.E.; Kozinsky, B. E(3)-equivariant graph neural networks for data-efficient and accurate interatomic potentials. *Nat. Commun.* **2022**, *13*, 2453. [CrossRef]
80. Lan, T.; Wang, H.; An, Q. Enabling high throughput deep reinforcement learning with first principles to investigate catalytic reaction mechanisms. *Nat. Commun.* **2024**, *15*, 6281. [CrossRef]
81. Lu, S.; Zhou, Q.; Ouyang, Y.; Guo, Y.; Li, Q.; Wang, J. Accelerated discovery of stable lead-free hybrid organic-inorganic perovskites via machine learning. *Nat. Commun.* **2018**, *9*, 3405. [CrossRef] [PubMed]
82. Liu, Z.; Rolston, N.; Flick, A.C.; Colburn, T.W.; Ren, Z.; Dauskardt, R.H.; Buonassisi, T. Machine learning with knowledge constraints for process optimization of open-air perovskite solar cell manufacturing. *Joule* **2022**, *6*, 834–849. [CrossRef]
83. Yu, R.; Zeng, W.; Zhou, L.; Van Tendeloo, G.; Mai, L.; Yao, Z.; Wu, J. Layer-by-layer delithiation during lattice collapse as the origin of planar gliding and microcracking in Ni-rich cathodes. *Cell Rep. Phys. Sci.* **2023**, *4*, 101480. [CrossRef]
84. Li, R.; Yao, L.; Sun, J.; Sun, Z.; Zhang, K.; Xue, J.; Wang, R. Challenges and perspectives for the perovskite module research. *Chem* **2025**, *11*, 102542. [CrossRef]

85. Graupner, D.R.; Kilin, D.S. Nonadiabatic Dynamics in Two-Dimensional Perovskites Assisted by Machine Learned Force Fields. *J. Phys. Chem. C* **2024**, *128*, 3935–3944. [CrossRef]
86. Zhang, C.; Li, Y.; Hu, X.-M.; Xiang, Y.; Li, C.-F.; Guo, G.-C.; Tura Brugués, J.; Gong, Q.; He, Q.; Liu, B.-H. Randomness versus Nonlocality in Multiple-Input and Multiple-Output Quantum Scenario. *Phys. Rev. Lett.* **2025**, *134*, 090201. [CrossRef] [PubMed]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.