

Predicting the Formability of Hybrid Organic–Inorganic Perovskites via an Interpretable Machine Learning Strategy

Shilin Zhang,[§] Tian Lu,[§] Pengcheng Xu, Qiuling Tao, Minjie Li,* and Wencong Lu*



Cite This: *J. Phys. Chem. Lett.* 2021, 12, 7423–7430



Read Online

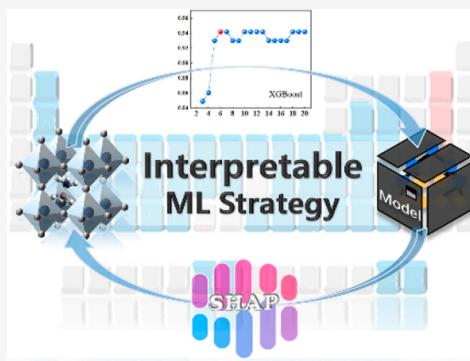
ACCESS |

Metrics & More

Article Recommendations

Supporting Information

ABSTRACT: Predicting the formability of perovskite structure for hybrid organic–inorganic perovskites (HOIPs) is a prominent challenge in the search for the required materials from a huge search space. Here, we propose an interpretable strategy combining machine learning with a shapley additive explanations (SHAP) approach to accelerate the discovery of potential HOIPs. According to the prediction of the best classification model, top-198 nontoxic candidates with a probability of formability (P_f) of >0.99 are screened from 18560 virtual samples. The SHAP analysis reveals that the radius and lattice constant of the B site (r_B and LC_B) are positively related to formability, while the ionic radius of the A site (r_A), the tolerant factor (t), and the first ionization energy of the B site (I_{1B}) have negative relations. The significant finding is that stricter ranges of t (0.84–1.12) and improved tolerant factor τ (critical value of 6.20) do exist for HOIPs, which are different from inorganic perovskites, providing a simple and fast assessment in the design of materials with an HOIP structure.



Perovskites, some of the most promising functional materials, have been widely applied in many fields, such as photovoltaics, ferroelectrics, and electrocatalysts.^{1–3} To explore perovskite compounds that meet specific application requirements, various new types of perovskites have emerged. Among them, hybrid organic–inorganic perovskites (HOIPs) have attracted tremendous attention due to the advantages of inorganic and organic materials.^{4–8} In its simplest form, the cubic crystal structure is composed of a three-dimensional framework of corner-sharing BX_6 octahedra, where the B site is a divalent cation and the X site is a halogen element (X = F, Cl, Br, or I), while A-site organic cation lies in a void surrounded by the octahedron. Several unique properties of HOIPs are largely dependent on the perovskite structure, but the distortion of the cubic structure caused by the mismatch of the cation and anion may lead to a nonperovskite structure. Tolerance factor t , proposed by Goldschmidt a century ago,⁹ is generally used to determine whether oxide perovskites can form a perovskite structure, but it cannot always be applied to other types of perovskites.¹⁰ Therefore, improving the accuracy of formability of HOIPs has been a major research challenge.

The traditional trial-and-error method requires extensive synthesis and characterization of materials to determine the crystal structure, resulting in long and costly time scales. To accelerate material design, density functional theory (DFT) can calculate electronic structures for the material composition to avoid the extra expense of experimental synthesis. However, the complexity of the electronic structure of a set of materials leads to a long calculation time and a low accuracy to a certain extent.^{11,12} On the contrary, machine learning (ML), a more

efficient material design technology in comparison with conventional methods, can build models to make corresponding predictions and explore hidden rules based on known data.^{13–15} Recently, ML has been employed to study the formability of perovskites.^{16–23} Though previous studies have made outstanding contributions to the formability of perovskites, they focused on only inorganic perovskites. Due to the complex structures and excessive radii of organic cations, the models and geometric factors proposed would be greatly restricted in the formability prediction of HOIPs. Different from the comprehensive applications in the field of inorganic perovskite design, the application of ML techniques in the study of the formability of HOIPs has received less attention.

In this work, we propose an interpretable strategy through the machine learning-based shapley additive explanations (SHAP) approach to accurately predict the formability of HOIPs and screen many potential HOIPs. First, we collect 44 HOIP and 58 non-HOIP samples from the publications for modeling. Then, optimal feature subsets are filtered out from the original feature pool for modeling with four algorithms in feature engineering. After model selection from four different algorithms, the well-trained extreme gradient boosting

Received: June 21, 2021

Accepted: July 29, 2021

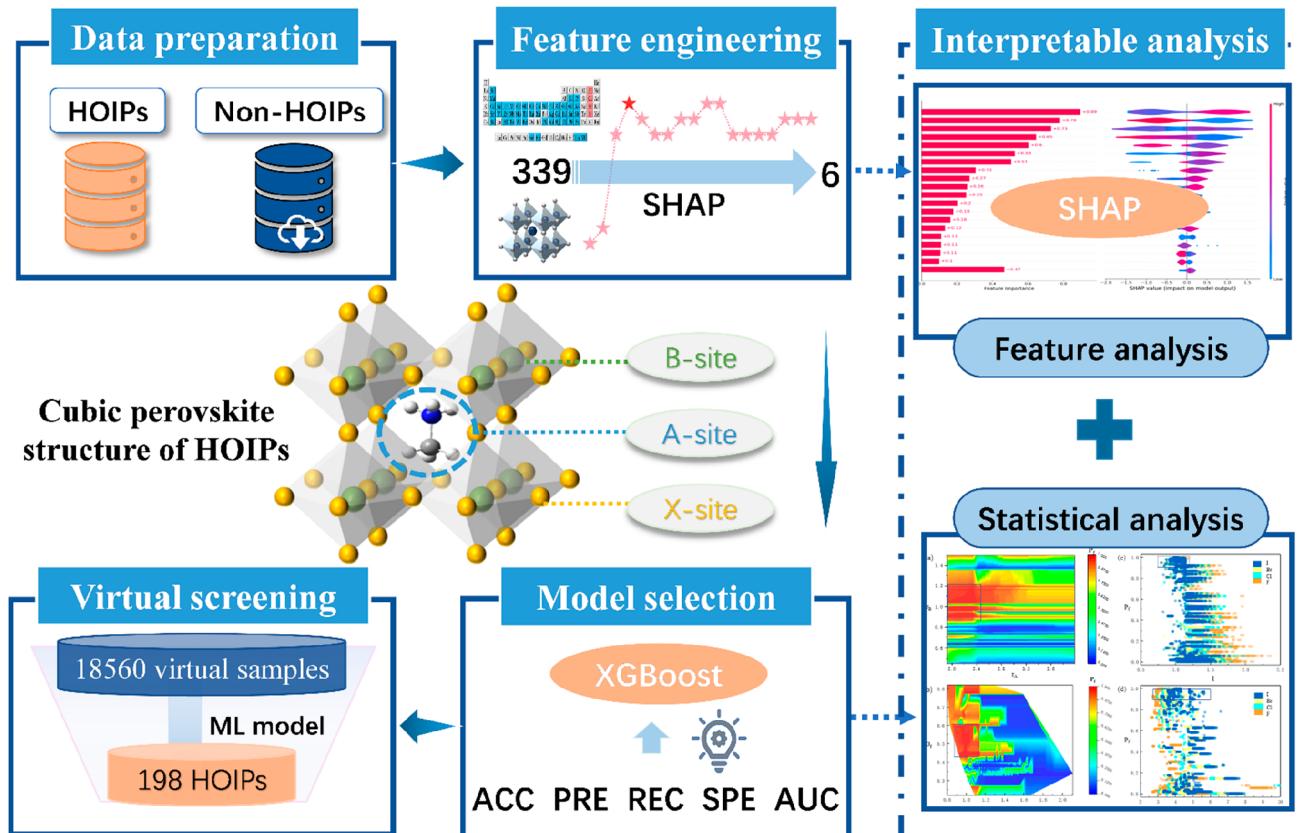


Figure 1. Flowchart of HOIP machine learning in this work.

(XGBoost) model is employed to classify unexplored materials into HOIPs and non-HOIPs. On the basis of the optimal model, the chemical space is gradually narrowed in the high-throughput screening process. Finally, 198 nontoxic compounds with a probability of formability (P_f) of >0.99 are screened. In addition, the optimal features are deeply discussed by combining feature analysis and statistical analysis. The complete workflow is shown in Figure 1.

Training the model requires gaining information from both the positive samples (HOIPs) and the negative samples (non-HOIPs). In this work, we employed 102 samples calculated by the same Perdew–Burke–Ernzerh (PBE) functional from one study to ensure the consistency of the data, including 44 undoped HOIP and 58 non-HOIP samples.²⁴ The total of 102 samples contain two different A-site cations [CH_3NH_3^+ and $\text{HC}(\text{NH}_2)_2^+$] and 14 B-site cations (Be^{2+} , Mg^{2+} , Ca^{2+} , Sr^{2+} , Ba^{2+} , Cr^{2+} , Mn^{2+} , Fe^{2+} , Co^{2+} , Ni^{2+} , Cu^{2+} , Hg^{2+} , Sn^{2+} , and Pb^{2+}). The X site has four possible choices (F^- , Cl^- , Br^- , and I^-).

Many features affect the targeted property. We establish an initial pool containing 339 features, far exceeding the number of features in previous work. Among them, three structural parameters [tolerance factor (t), improved tolerance factor (τ),¹⁰ and octahedral factor (O_f)¹⁶] are adopted as features, and the remaining features are obtained from the Mendeleev database²⁵ and the Villars database.^{26,27}

To avoid the curse of dimensionality, the best strategy is to carefully select features that not only uniquely characterize the material in the data set but also are relevant to the properties of desired material.²⁸ The two-step method is employed to select optimal feature subsets in feature engineering. First, the whole feature set is preprocessed by pruning the redundant features

that have missing values, or the high Pearson correlation values (0.95 is set as the criterion) to the other features, stemming from the subset of only 45 features (listed in Table S1). In the second step, recursive feature addition (RFA)²⁹ is used to perform feature selection to shortlist the key features, in which we have evaluated the models that are built up on the basis of the top 2–20 features by adding one more n th feature under the specified feature importance order. Depending on the various ML algorithms, the specified feature importance is extracted from the SHAP method for XGBoost³⁰ and the gradient boosting classifier (GBC),³¹ while the maximum relative minimum redundancy (mRMR)³² method for the support vector classifier (SVC)³³ and the K nearest neighbor classifier (KNN).³⁴ To evaluate the performance of the model, the accuracy of leaving-one-out cross-validation (LOOCV) is selected as the evaluation indicator. The changing trend of accuracy and the number of features corresponding to each algorithm are shown in Figure 2. The models of four algorithms, including XGBoost, GBC, SVC, and KNN, reach the optimal values of LOOCV accuracy of 0.94, 0.91, 0.90, and 0.83 with six, four, four, and three features, respectively. The optimal feature subsets of each algorithm according to the maximum accuracy are selected as shown in Table S2. In addition, the Pearson correlation coefficient matrix is calculated to identify positive and negative correlations between paired features, as shown in Figure S1. The linear correlation of most features is weak, indicating that redundant and irrelevant features have been successfully deleted.

In ML technology, the crucial step is choosing an appropriate algorithm. The GBC, XGBoost, KNN, and SVC algorithms are used for model selection. In addition, five

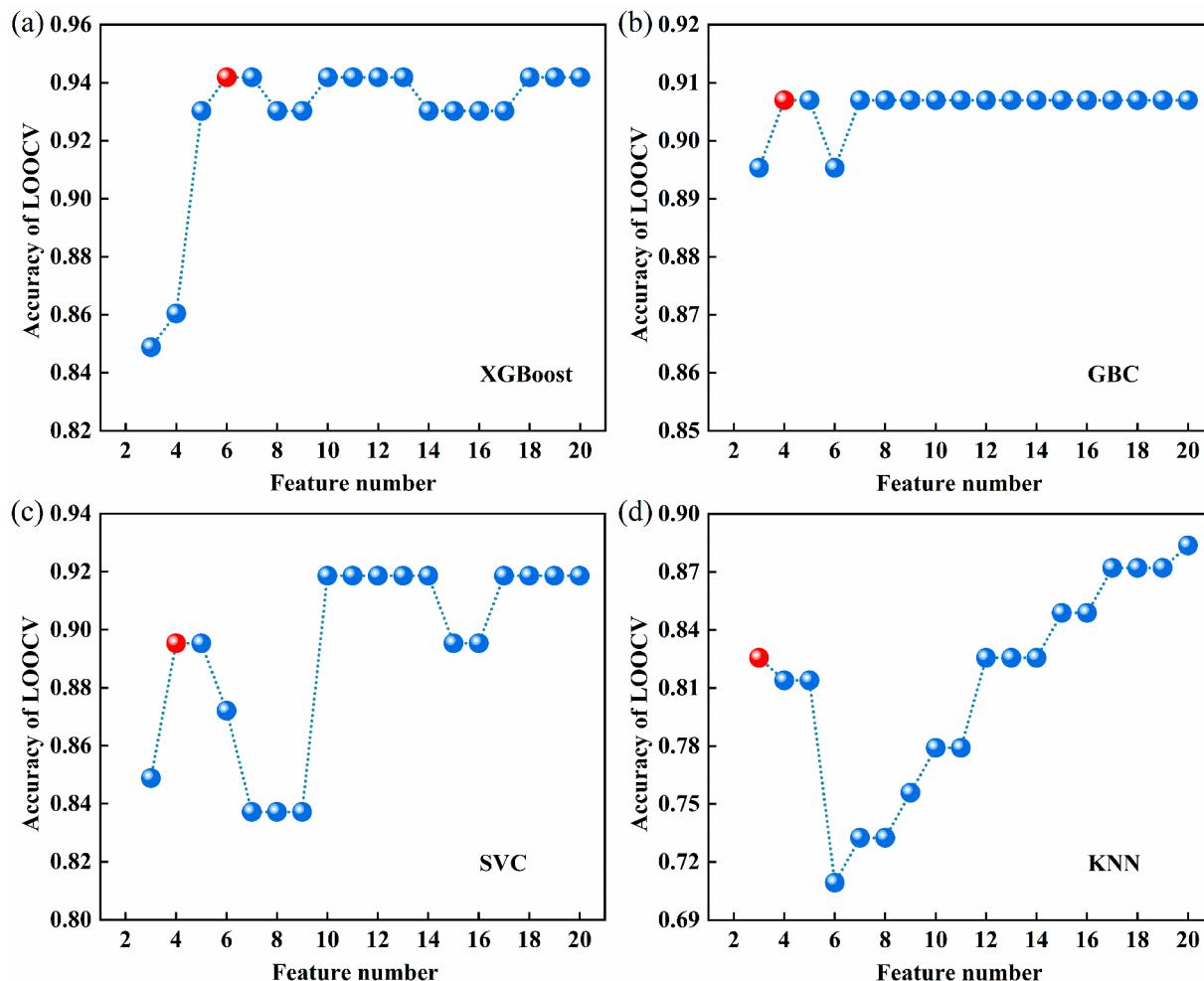


Figure 2. Accuracy of LOOCV of ML models in each selection process. The red points represent the optimal values of accuracies in each model: (a) XGBoost model, (b) GBC model, (c) SVC model, and (d) KNN model.

important validation metrics are used to evaluate the model performance, including the LOOCV classification accuracy, precision, recall, specificity, and area under the curve (AUC).¹³ The first four metrics are defined in Note S1. The AUC is calculated on the basis of the receiver operating characteristic curve (ROC), which reveals the degree of separability. In general, the larger the evaluation metrics, the better the model performance. As shown in Figure 3a, the XGBoost model has great robustness. After hyperparameter optimization (Note S2 and Figure 3b), the accuracy of LOOCV, precision, recall, specificity, and AUC increased by ~1%, reaching 0.95, 0.91, 0.96, 0.97, and 0.97, respectively. In Figure 3c, the three-dimensional confusion matrix more intuitively shows the excellent classification capabilities of the XGBoost model.

To validate the generalization ability of the model, the divided independent test set (Note S2) is used to predict the selected model, and the accuracy of the test set reaches 0.88 (the confusion matrix as shown in Figure 3d). It is worth mentioning that the prediction accuracy values of t and τ in the data set are only 0.50 and 0.71, respectively, while the accuracies of the LOOCV and test set of the selected model are 0.95 and 0.88, respectively. In addition, we collect 22 HOIPs (Table S3) from the experimental literature^{35–50} as an external independent test set. The accuracy is 0.91, and the probability of formability of all 17 samples is above 0.94. Thus,

both results prove the good prediction capability and the satisfactory generalization of our model. In addition, the data set is randomly divided 100 times and modeled on the basis of the 85:15 ratio of the train set and test set. The average accuracies of the LOOCV and test set reach 0.89 and 0.84, respectively, demonstrating the stability of the selected model.

The constructed model has great prediction ability and can be employed for high-throughput screening. Currently, thousands of candidates are waiting to be discovered. In previous work, up to 21 A-site cations were considered for high-throughput screening.⁵ We collected 80 organic cations (all of the A-site cation structures can be seen in Figure S2) that have been reported by experimental synthesis or theoretical calculations as A-site cations. Meanwhile, 58 ions are considered as B sites, covering most of the elements in the periodic table except for radioactive elements, and the X sites are four halogen elements contained in the input set. Consequently, 18560 possible virtual samples have been generated. Excluding non-electrically neutral compounds that are difficult to synthesize, we selected 14298 electrically neutral samples. Through further model prediction screening, 6343 potential samples are predicted to form HOIP structures with P_f values of >0.50 , and the predicted probability distribution of these compounds is as shown in Figure 4a. Subsequently, 258 compounds with P_f values of >0.99 are screened. Among them,

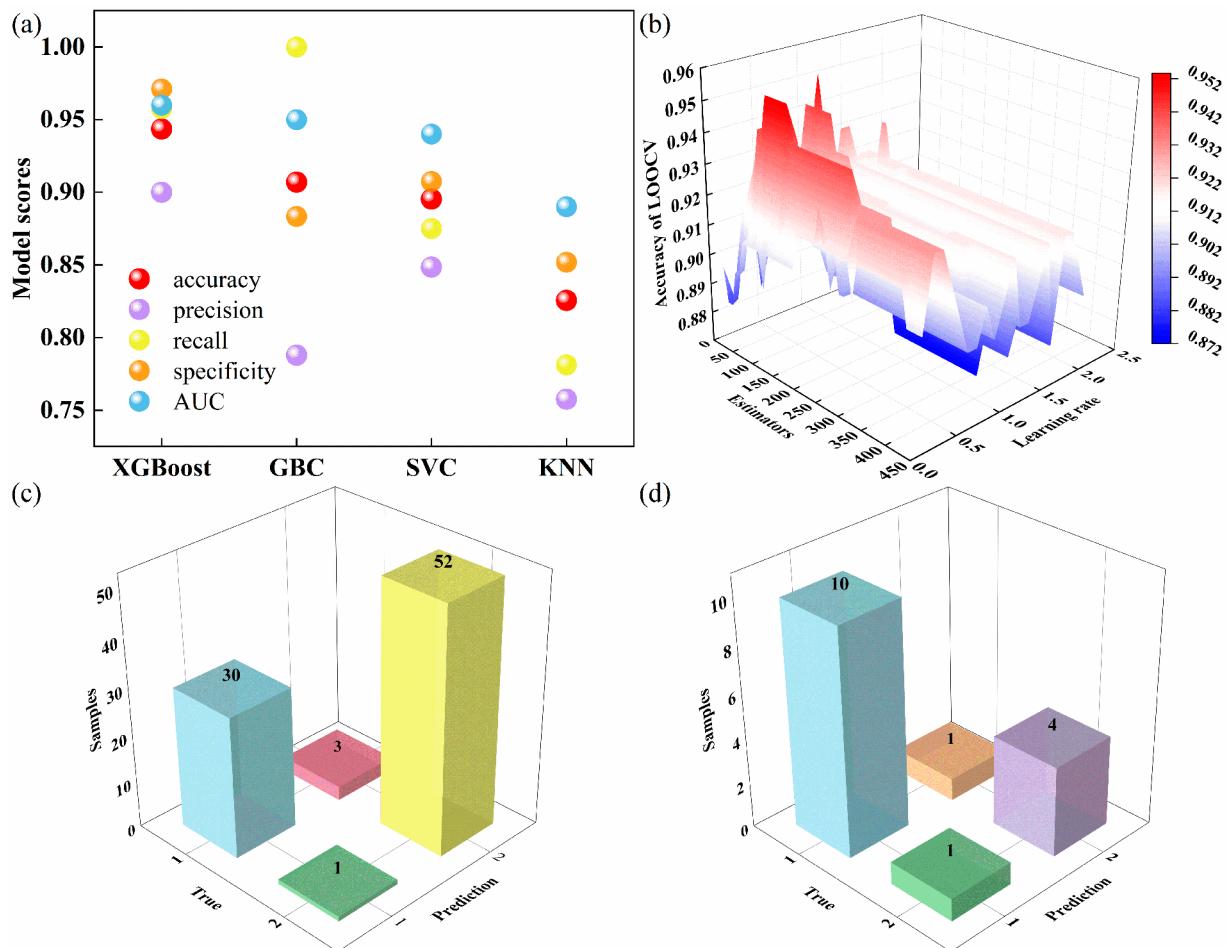


Figure 3. (a) Model performance of different ML algorithms. (b) Accuracy of LOOCV vs estimators and learning rate. Three-dimensional confusion matrices for classification based on the accuracy of (c) LOOCV and (d) the test set.

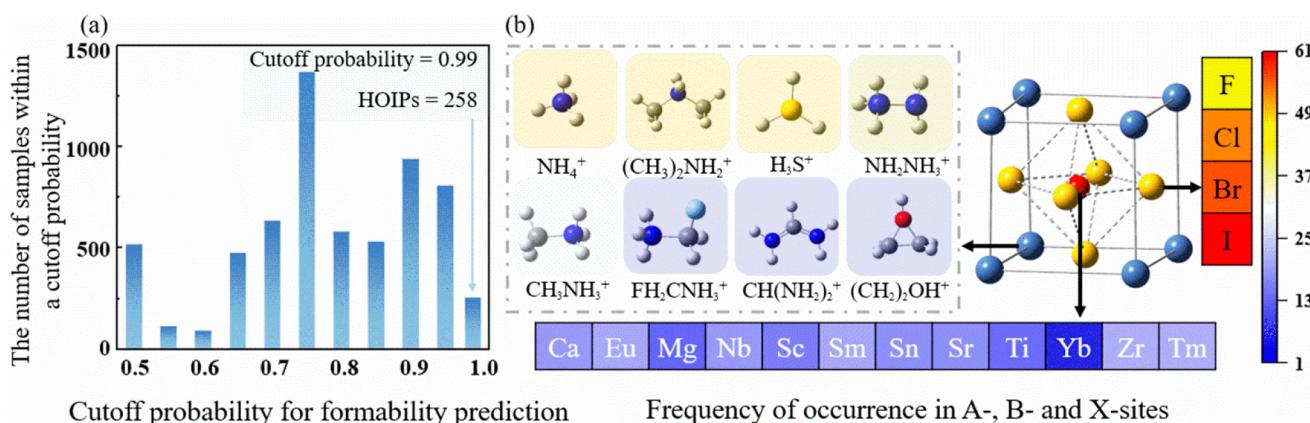


Figure 4. (a) Distribution plot for prediction probabilities of HOIPs. (b) Numbers of occurrences of each cation in the A site and each element of the B site and X site in the prediction data set with a cutoff probability of 0.99.

MAPbCl₃, MAPbBr₃, MAPbI₃, MASnCl₃, MASnBr₃, and MASnI₃ have been synthesized experimentally, and they are labeled as HOIP structures, which are expected to have good application prospects in photovoltaics, further validating the reliability and generalization ability of our model. Once compounds containing toxic elements had been removed to ensure environmental friendliness, 198 compounds remained and were finally selected (Table S4). The frequency of organic cations at the A site and the elements at the B site and X site of

198 candidates are shown in Figure 4b. Here, candidates for the A site are dominated by NH₄⁺, (CH₃)₂NH₂⁺, H₃S⁺, NH₂NH₃⁺, and CH₃NH₃⁺. The distribution of the B site is relatively balanced, with Ca²⁺, Eu²⁺, Nb²⁺, Sm²⁺, Sn²⁺, Sr²⁺, Zr²⁺, and Tm³⁺ cations occupying the majority, while the number of iodide ions is the largest at the X site.

To further analyze the contributions of features to the classification model and to reveal the relationship between the features and the target variable, the importance of features is

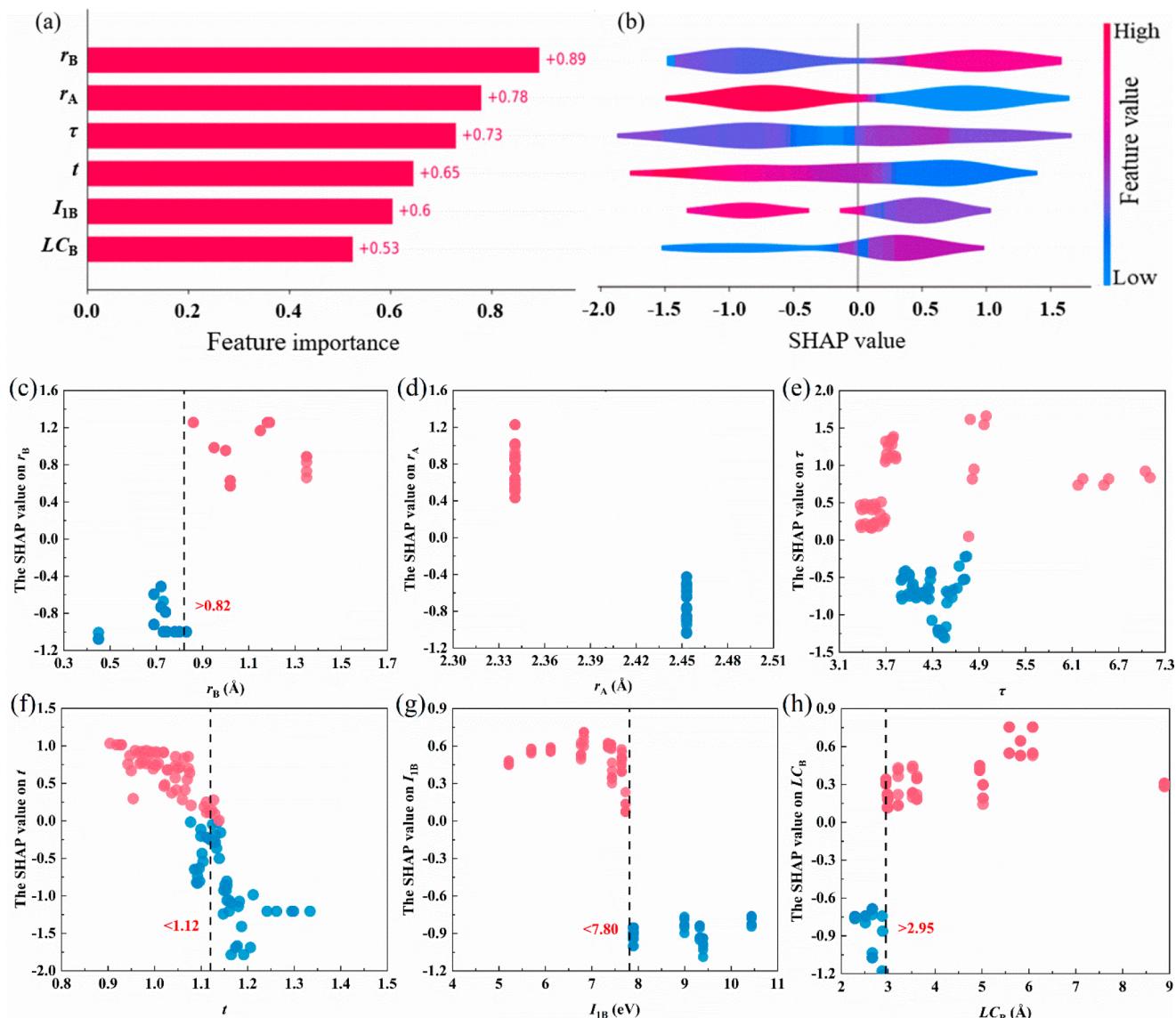


Figure 5. Optimal selected subset of (a) the importance rankings and (b) the SHAP value distribution. SHAP values of (c) r_B , (d) r_A , (e) τ , (f) t , (g) I_{1B} , and (h) LC_B for every one of the data. Note that blue represents negative correlation and red represents positive correlation.

ranked by SHAP. As shown in Figure 5a, the ionic radius of the B site (r_B) is the most important for the formability of HOIPs, and the following features are the ionic radii of the A site (r_A), τ , t , the first ionization energy of the B site (I_{1B}), and the lattice constant of the B site (LC_B). Moreover, the features related to the B site show a greater correlation than the A-site and X-site features, which is consistent with previous studies of the formation of inorganic perovskites.^{21–23}

Compared with traditional feature interpretation methods, the advantages of the SHAP method are not only ranking the importance of features but also reflecting the positive and negative influence of each feature in each sample and the important law of the feature on the prediction result, which effectively unifies the global and local explanatory properties of the model. The SHAP violin diagram is shown in Figure 5b. The gray straight line in the middle represents the baseline, which is the prediction value. Each row represents a feature, and the abscissa represents the specific output SHAP value, with reddish color representing the larger feature value and bluish color representing the smaller feature value. If the SHAP

value is more positive, this means that the feature pushes the prediction to the direction where the perovskite structure is more likely to be formed. Obviously, r_B and LC_B are positively correlated with formability, while r_A , t , and I_{1B} are negatively correlated with formability. After further SHAP analysis, we found that a critical value exists for these features to distinguish between positive and negative regions. The SHAP values are positive when the $r_B > 0.82 \text{ \AA}$, $t < 1.12$, $I_{1B} < 7.80 \text{ eV}$, and $LC_B > 2.95 \text{ \AA}$ in Figure 5c–h.

Furthermore, the visualization graphs can intuitively reveal the hidden trends from the data set predicted by the ML model. The relationship between the selected features of the training modeling and P_f is shown in Figure 6a–f. The four X-site elements (F, Cl, Br, and I) are distinguished by different colors, and the areas with P_f values of >0.90 are marked with boxes. The radii of the A site and B site not only show high importance for the model but also show a strong correlation to the formability in the prediction data set. As shown in panels a and b of Figure 6, though almost all A-site and B-site ions have the potential to form HOIP structure, the smaller the A site,

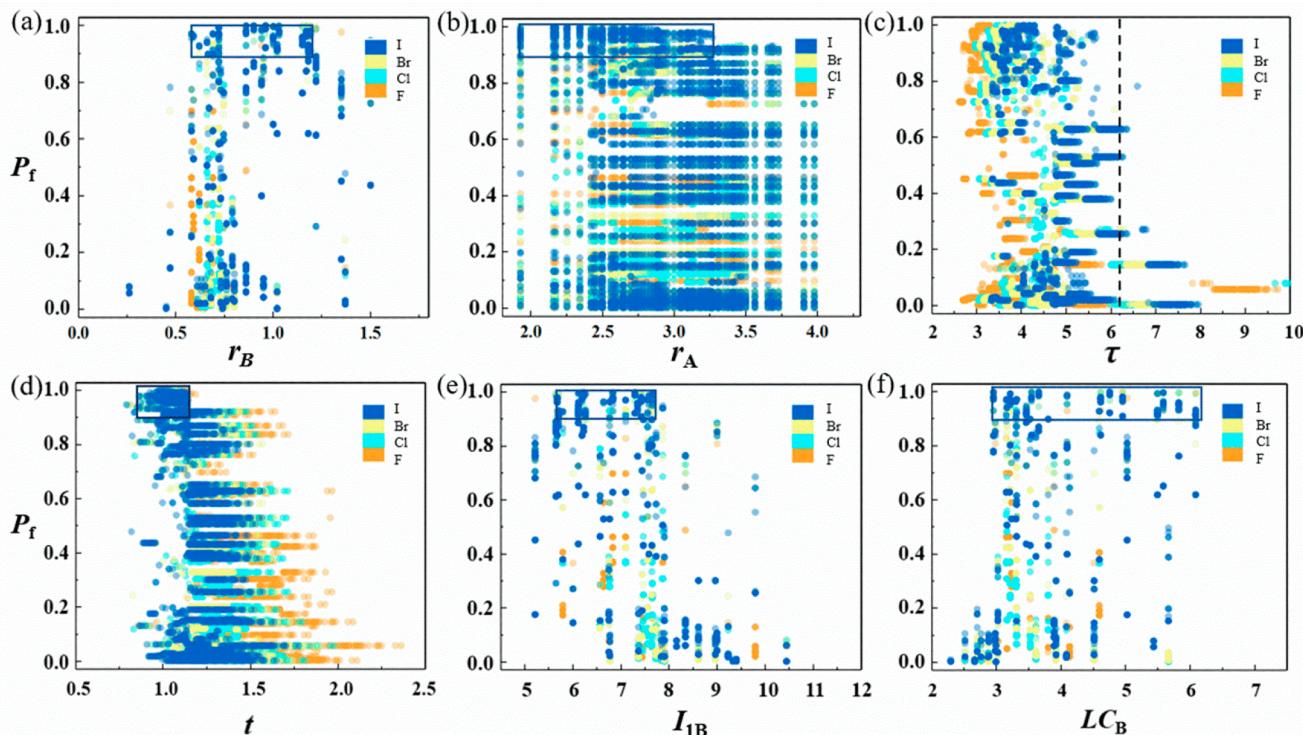


Figure 6. Scatter plots of P_f for the prediction data set with (a) r_B , (b) r_A , (c) τ , (d) t , (e) I_{1B} , and (f) LC_B . The box indicates the most appropriate range of each feature.

the better the formability of the HOIP structure. More importantly, when r_A is in the range of 1.95–3.25 Å and r_B is in the range 0.60–1.20 Å, the P_f of the compound is close to 1. Bartal et al. reported that τ achieved a classification accuracy of 0.92 on a data set of 576 oxide and halide perovskites based on nonperovskite classification for τ values of >4.18 .¹⁰ However, as shown in Figure 6c, when $\tau > 6.20$, the compound could be labeled as a non-HOIP structure. In Figure 6d, the P_f values of the compounds are maximal when t is in the range of 0.84–1.12. Compared to the range for inorganic perovskites,⁹ the stricter range for HOIPs could ensure a higher probability of formability with a value of >0.90 . The optimal I_{1B} and LC_B in the ranges of 5.60–7.65 eV and 2.95–6.10 Å, respectively, tend to yield a formability probability of >0.90 in panels e and f of Figure 6. Overall, the iodine-based compounds (ABI_3) account for the most candidates with high P_f values of >0.90 . Because limiting the scope of a single feature cannot distinguish the two categories well, the two-dimensional structure diagram is further used for more rigorous classification. The tolerance factor octahedral factor map is commonly used to determine whether inorganic perovskites can be formed. In Figure S3a, there is a greater probability of forming HOIP structures when t is in the range of 0.88–1.12 and O_f is in the range of 0.42–0.61. Figure S3b shows that the radii of the selected A-site and B-site ions are in the optimal ranges of 1.95–2.42 and 0.85–1.20 Å, respectively, providing guidance for researchers to design new HOIPs.

Interestingly, we find that the conclusions obtained by SHAP analysis are in good agreement with the results of statistical analysis. In other words, the larger the r_B and LC_B , the better the structure formation of HOIPs. On the contrary, the larger the r_A , t , and I_{1B} , the more unlikely they are to form HOIP structure. Meanwhile, the rules discovered show a consistent trend with the properties of experimentally

synthesized HOIPs. For example, the B-site ions of Sn^{2+} and Pb^{2+} have larger radii of 1.15 and 1.19 Å, respectively, along with their smaller ionization energies of 7.34 and 7.41 eV and larger lattice constants of 4.95 and 5.82 Å, respectively. The A-site cations (MA^+ , FA^+ , $NH_2NH_3^+$, $CH_3CH_2NH_3^+$, $OH-(CH_2)_2NH_3^+$, and $C_4H_{10}N^+$) have smaller radii in the range of 1.92–2.95 Å. For the most synthesized compounds, t is in the range of 0.91–1.12. In summary, the predictive results of the interpretable ML model are mutually confirmed with experiments, indicating the reliability of the model. We believe that the selected candidates are most likely to form the structure of HOIPs.

In summary, this work proposes an interpretable strategy combining machine learning technology with the SHAP method to accelerate the discovery of potential HOIPs. The results of LOOCV demonstrate that the XGBoost model has stronger predictive performance than the others, with a much higher accuracy of 0.95, while the test set accuracy reaches 0.88. In addition, the classification accuracy of the external test set composed of the experimental samples reaches 0.91, indicating the superior generalization ability of the model. Through high-throughput screening, 198 nontoxic HOIP candidates with a probability of formability of >0.99 are screened from 18560 virtual samples. In particular, it is more likely to form perovskite structure when the main cations are NH_4^+ , $(CH_3)_2NH_2^+$, H_3S^+ , $NH_2NH_3^+$, and $CH_3NH_3^+$ at the A site and Ca^{2+} , Eu^{2+} , Nb^{2+} , Sm^{2+} , Sn^{2+} , Sr^{2+} , Zr^{2+} , and Tm^{2+} at the B site. On the basis of SHAP analysis and statistical analysis, we find that the values of r_B and LC_B are positively related to formability while r_A , t , and I_{1B} are negatively related. Furthermore, in terms of ionic properties, it is more likely to form HOIP structure when the r_A of the compound is in the range of 1.95–3.25 Å and r_B is in the range of 0.60–1.20 Å. The optimal I_{1B} and LC_B in the ranges of 5.60–7.65 eV and

2.95–6.10 Å, respectively, tend to yield a formability probability of >0.90. In traditional evaluation factors, the range of t for HOIPs is adjusted to 0.84–1.12, and the critical value of τ is tuned to 6.20 to distinguish whether the compounds can form HOIP structure. Given the high predicted probabilities, the candidates are promising for the synthesis of perovskite materials and we look forward to experimental collaborations for further studies. It is also expected that the explainable approach can illuminate the acceleration of the discovery of HOIP materials and be extended to the discovery of other materials.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acs.jpcllett.1c01939>.

Evaluation indicators, hyperparameter optimization process, correlation matrix of variables, A-site cation structures, heat map of features and target variables, preprocessed feature set, optimal feature set selected by different algorithms, experimental test set, and 198 candidate materials ([PDF](#))

AUTHOR INFORMATION

Corresponding Authors

Minjie Li – Department of Chemistry, College of Sciences, Shanghai University, Shanghai 200444, China;
✉ orcid.org/0000-0001-5048-6211; Email: minjieli@shu.edu.cn

Wencong Lu – Department of Chemistry, College of Sciences, Shanghai University, Shanghai 200444, China; Materials Genome Institute, Shanghai University, Shanghai 200444, China; ✉ orcid.org/0000-0001-5361-6122; Email: wclu@shu.edu.cn

Authors

Shilin Zhang – Department of Chemistry, College of Sciences, Shanghai University, Shanghai 200444, China

Tian Lu – Materials Genome Institute, Shanghai University, Shanghai 200444, China

Pengcheng Xu – Materials Genome Institute, Shanghai University, Shanghai 200444, China

Qiuming Tao – Department of Chemistry, College of Sciences, Shanghai University, Shanghai 200444, China

Complete contact information is available at:
<https://pubs.acs.org/10.1021/acs.jpcllett.1c01939>

Author Contributions

✉ S.Z. and T.L. contributed equally to this work.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

Financial support to this work from the National Key Research and Development Program of China (2018YFB0704400) and the Science and Technology Commission of Shanghai Municipality (18520723500) is gratefully acknowledged.

REFERENCES

- (1) Yin, W. J.; Weng, B.; Ge, J.; Sun, Q.; Li, Z.; Yan, Y. Oxide Perovskites, Double Perovskites and Derivatives for Electrocatalysis, and Photovoltaics. *Energy Environ. Sci.* **2019**, *12*, 442–462.
- (2) Tao, Q.; Xu, P.; Li, M.; Lu, W. Machine Learning for Perovskite Materials Design and Discovery. *npj Comput. Mater.* **2021**, *7*, 23.
- (3) Tao, Q.; Lu, T.; Sheng, Y.; Li, L.; Lu, W.; Li, M. Machine Learning Aided Design of Perovskite Oxide Materials for Photocatalytic Water Splitting. *J. Energy Chem.* **2021**, *60*, 351–359.
- (4) Wang, Y.; Tang, Y.; Jiang, J.; Zhang, Q.; Sun, J.; Hu, Y.; Cui, Q.; Teng, F.; Lou, Z.; Hou, Y. Mixed-Dimensional Self-Assembly Organic-Inorganic Perovskite Microcrystals for Stable and Efficient Photodetectors. *J. Mater. Chem. C* **2020**, *8*, 5399–5408.
- (5) 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.
- (6) Wu, T.; Wang, J. Global Discovery of Stable and Non-Toxic Hybrid Organic-Inorganic Perovskites for Photovoltaic Systems by Combining Machine Learning Method with First Principle Calculations. *Nano Energy* **2019**, *66*, 104070.
- (7) Chen, H. A.; Tang, P. H.; Chen, G. J.; Chang, C. C.; Pao, C. W. Microstructure Maps of Complex Perovskite Materials from Extensive Monte Carlo Sampling Using Machine Learning Enabled Energy Model. *J. Phys. Chem. Lett.* **2021**, *12*, 3591–3599.
- (8) Zhang, C.; Wang, Y.; Lin, X.; Wu, T.; Han, Q.; Zhang, Y.; Han, L. Effects of A Site Doping on the Crystallization of Perovskite Films. *J. Mater. Chem. A* **2021**, *9*, 1372–1394.
- (9) Goldschmidt, V. M. Die Gesetze Der Krystallochemie. *Naturwissenschaften* **1926**, *14*, 477–485.
- (10) Bartel, C. J.; Sutton, C.; Goldsmith, B. R.; Ouyang, R.; Musgrave, C. B.; Ghiringhelli, L. M.; Scheffler, M. New Tolerance Factor to Predict the Stability of Perovskite Oxides and Halides. *Sci. Adv.* **2019**, *5*, No. eaav0693.
- (11) Mounet, N.; Gibertini, M.; Schwaller, P.; Campi, D.; Merkys, A.; Marrazzo, A.; Sohier, T.; Castelli, I. E.; Cepellotti, A.; Pizzi, G.; et al. Two-Dimensional Materials from High-Throughput Computational Exfoliation of Experimentally Known Compounds. *Nat. Nanotechnol.* **2018**, *13*, 246–252.
- (12) Kim, C.; Huan, T. D.; Krishnan, S.; Ramprasad, R. A Hybrid Organic-Inorganic Perovskite Dataset. *Sci. Data* **2017**, *4*, 170057.
- (13) Im, J.; Lee, S.; Ko, T.-W.; Kim, H. W.; Hyon, Y.; Chang, H. Identifying Pb-Free Perovskites for Solar Cells by Machine Learning. *npj Comput. Mater.* **2019**, *5*, 37.
- (14) Ma, X. Y.; Lewis, J. P.; Yan, Q. B.; Su, G. Accelerated Discovery of Two-Dimensional Optoelectronic Octahedral Oxyhalides via High-Throughput Ab Initio Calculations and Machine Learning. *J. Phys. Chem. Lett.* **2019**, *10*, 6734–6740.
- (15) Jin, H.; Zhang, H.; Li, J.; Wang, T.; Wan, L.; Guo, H.; Wei, Y. Discovery of Novel Two-Dimensional Photovoltaic Materials Accelerated by Machine Learning. *J. Phys. Chem. Lett.* **2020**, *11*, 3075–3081.
- (16) Li, C.; Soh, K. C. K.; Wu, P. Formability of ABO_3 Perovskites. *J. Alloys Compd.* **2004**, *372*, 40–48.
- (17) Li, C.; Lu, X.; Ding, W.; Feng, L.; Gao, Y.; Guo, Z. Formability of ABX_3 ($X = F, Cl, Br, I$) Halide Perovskites. *Acta Crystallogr., Sect. B: Struct. Sci.* **2008**, *64*, 702–709.
- (18) Pilania, G.; Balachandran, P. V.; Gubernatis, J. E.; Lookman, T. Classification of ABO_3 Perovskite Solids: A Machine Learning Study. *Acta Crystallogr., Sect. B: Struct. Sci., Cryst. Eng. Mater.* **2015**, *71*, 507–520.
- (19) Pilania, G.; Balachandran, P. V.; Kim, C.; Lookman, T. Finding New Perovskite Halides via Machine Learning. *Front. Mater.* **2016**, *3*, 19.
- (20) Xu, Q.; Li, Z.; Liu, M.; Yin, W. J. Rationalizing Perovskite Data for Machine Learning and Materials Design. *J. Phys. Chem. Lett.* **2018**, *9*, 6948–6954.
- (21) Yang, X.; Li, L.; Tao, Q.; Lu, W.; Li, M. Rapid Discovery of Narrow Bandgap Oxide Double Perovskites using Machine Learning. *Comput. Mater. Sci.* **2021**, *196*, 110528.
- (22) Liu, H.; Cheng, J.; Dong, H.; Feng, J.; Pang, B.; Tian, Z.; Ma, S.; Xia, F.; Zhang, C.; Dong, L. Screening Stable and Metastable

- ABO₃ Perovskites using Machine Learning and the Materials Project. *Comput. Mater. Sci.* **2020**, *177*, 109614.
- (23) Talapatra, A.; Überuaga, B. P.; Stanek, C. R.; Pilania, G. A Machine Learning Approach for the Prediction of Formability and Thermodynamic Stability of Single and Double Perovskite Oxides. *Chem. Mater.* **2021**, *33*, 845–858.
- (24) Nakajima, T.; Sawada, K. Discovery of Pb-Free Perovskite Solar Cells via High-Throughput Simulation on the K Computer. *J. Phys. Chem. Lett.* **2017**, *8*, 4826–4831.
- (25) Mentel, L. M. A Python Resource for Properties of Chemical Elements, Ions and Isotopes. 2014. <https://github.com/lmentel/mendeleev>.
- (26) Villars, P.; Brandenburg, K.; Berndt, M.; LeClair, S.; Jackson, A.; Paod, Y.-H.; Igelnik, B.; Oxley, M.; Bakshi, B.; Chen, P.; et al. Interplay of Large Materials Databases, Semi-Empirical Methods, Neuro-Computing and First Principle Calculations for Ternary Compound Former/Nonformer Prediction. *Eng. Appl. Artif. Intel.* **2000**, *13*, 497–505.
- (27) Pierre, V.; Karin, C.; Roman, G.; Iwata, S. *Materials Informatics: Methods, Tools and Applications*; Wiley-VCH Verlag GmbH & Co. KgaA: Weinheim, Germany, 2019.
- (28) Nasrabadi, N. M. *Pattern Recognition and Machine Learning*; Springer Press: New York, 2016.
- (29) Jablonka, K. M.; Ongari, D.; Moosavi, S. M.; Smit, B. Big-Data Science in Porous Materials: Materials Genomics and Machine Learning. *Chem. Rev.* **2020**, *120*, 8066–8129.
- (30) Lu, Z.; Chen, X.; Liu, X.; Lin, D.; Wu, Y.; Zhang, Y.; Wang, H.; Jiang, S.; Li, H.; Wang, X.; Lu, Z. Interpretable Machine-Learning Strategy for Soft-Magnetic Property and Thermal Stability in Fe-Based Metallic Glasses. *npj Comput. Mater.* **2020**, *6*, 187.
- (31) Friedman, J. H. Greedy Function Approximation: A Gradient Boosting Machine. *Annals of Statistics* **2001**, *29*, 1189–1232.
- (32) Berrendero, J. R.; Cuevas, A.; Torrecilla, J. L. The mRMR Variable Selection Method: A Comparative Study for Functional Data. *J. Stat. Comput. Sim.* **2016**, *86*, 891–907.
- (33) Shi, L.; Chang, D.; Ji, X.; Lu, W. Using Data Mining to Search for Perovskite Materials with Higher Specific Surface Area. *J. Chem. Inf. Model.* **2018**, *58*, 2420–2427.
- (34) Lapins, M.; Wikberg, J. E. Kinome-Wide Interaction Modelling using Alignment-Based and Alignment-Independent Approaches for Kinase Description and Linear and Non-Linear Data Analysis Techniques. *BMC Bioinf.* **2010**, *11*, 339.
- (35) Binek, A.; Hanusch, F. C.; Docampo, P.; Bein, T. Stabilization of the Trigonal High-Temperature Phase of Formamidinium Lead Iodide. *J. Phys. Chem. Lett.* **2015**, *6*, 1249–1253.
- (36) Campbell, E. V.; Dick, B.; Rheingold, A. L.; Zhang, C.; Liu, X.; Vardeny, Z. V.; Miller, J. S. Structures of a Complex Hydrazinium Lead Iodide, (N₂H₅)₁₅Pb₃I₂₁, Possessing [Pb₂I₉]⁵⁻, [PbI₆]⁴⁻, and I⁻ Ions and α - and β -(N₂H₅)PbI₃. *Chem. - Eur. J.* **2018**, *24*, 222–229.
- (37) Chiarella, F.; Zappettini, A.; Licci, F.; Borriello, I.; Cantele, G.; Ninno, D.; Cassinese, A.; Vaglio, R. Combined Experimental and Theoretical Investigation of Optical, Structural, and Electronic Properties of CH₃NH₃SnX₃ thin films (X = Cl, Br). *Phys. Rev. B: Condens. Matter Mater. Phys.* **2008**, *77*, 045129.
- (38) Huang, S.; Huang, P.; Wang, L.; Han, J.; Chen, Y.; Zhong, H. Halogenated-Methylammonium Based 3D Halide Perovskites. *Adv. Mater.* **2019**, *31*, No. 1903830.
- (39) Jokar, E.; Chien, C. H.; Tsai, C. M.; Fathi, A.; Diau, E. W. Robust Tin-Based Perovskite Solar Cells with Hybrid Organic Cations to Attain Efficiency Approaching 10%. *Adv. Mater.* **2019**, *31*, No. 1804835.
- (40) Kalthoum, R.; Ben Bechir, M.; Ben Rhaiem, A. CH₃NH₃CdCl₃: A Promising New Lead-Free Hybrid Organic-Inorganic Perovskite for Photovoltaic Applications. *Phys. E (Amsterdam, Neth.)* **2020**, *124*, 114235.
- (41) Krishnamoorthy, T.; Ding, H.; Yan, C.; Leong, W. L.; Baikie, T.; Zhang, Z.; Sherburne, M.; Li, S.; Asta, M.; Mathews, N.; et al. Lead-Free Germanium Iodide Perovskite Materials for Photovoltaic Applications. *J. Mater. Chem. A* **2015**, *3*, 23829–23832.
- (42) Kumawat, N. K.; Tripathi, M. N.; Waghmare, U.; Kabra, D. Structural, Optical, and Electronic Properties of Wide Bandgap Perovskites: Experimental and Theoretical Investigations. *J. Phys. Chem. A* **2016**, *120*, 3917–3923.
- (43) Mittal, M.; Jana, A.; Sarkar, S.; Mahadevan, P.; Sapra, S. Size of the Organic Cation Tuned the Band Gap of Colloidal Organolead Bromide Perovskite Nanocrystals. *J. Phys. Chem. Lett.* **2016**, *7*, 3270–3277.
- (44) Pisani, A.; Speltini, A.; Quadrelli, P.; Drera, G.; Sangaletti, L.; Malavasi, L. Enhanced Air-Stability of Sn-Based Hybrid Perovskites Induced by Dimethylammonium (DMA): Synthesis, Characterization, Aging and Hydrogen Photogeneration of the MA_{1-x}DMA_xSnBr₃ System. *J. Mater. Chem. C* **2019**, *7*, 7020–7026.
- (45) Tsai, C.-M.; Lin, Y. P.; Pola, M. K.; Narra, S.; Jokar, E.; Yang, Y.-W.; Diau, E. W. G. Control of Crystal Structures and Optical Properties with Hybrid Formamidinium and 2-Hydroxyethylammonium Cations for Mesoscopic Carbon-Electrode Tin-Based Perovskite Solar Cells. *ACS Energy Lett.* **2018**, *3*, 2077–2085.
- (46) Vega, E.; Mollar, M.; Marí, B. Optoelectronic Properties and the Stability of Mixed MA_{1-x}IA_xPbI₃ Perovskites. *Energy Technol.* **2020**, *8*, 1900743.
- (47) Xu, A. F.; Wang, R. T.; Yang, L. W.; Jarvis, V.; Britten, J. F.; Xu, G. Pyrrolidinium Lead Iodide from Crystallography: A New Perovskite with Low Bandgap and Good Water Resistance. *Chem. Commun.* **2019**, *55*, 3251–3253.
- (48) Yu, S.; Liu, H.; Wang, S.; Zhu, H.; Dong, X.; Li, X. Hydrazinium Cation Mixed FAPbI₃-Based Perovskite with 1D/3D Hybrid Dimension Structure for Efficient and Stable Solar Cells. *Chem. Eng. J.* **2021**, *403*, 125724.
- (49) Zhang, H. Y.; Chen, X. G.; Zhang, Z. X.; Song, X. J.; Zhang, T.; Pan, Q.; Zhang, Y.; Xiong, R. G. Methylphosphonium Tin Bromide: A 3D Perovskite Molecular Ferroelectric Semiconductor. *Adv. Mater.* **2020**, *32*, No. 2005213.
- (50) Zhou, H.; Cui, X.; Yuan, C.; Cui, J.; Shi, S.; He, G.; Wang, Y.; Wei, J.; Pu, X.; Li, W.; et al. Band-Gap Tuning of Organic-Inorganic Hybrid Palladium Perovskite Materials for a Near-Infrared Optoelectronics Response. *ACS Omega* **2018**, *3*, 13960–13966.

■ NOTE ADDED AFTER ASAP PUBLICATION

This paper was published ASAP on July 31, 2021, with an error in an author name due to a production error. The corrected version was reposted on August 3, 2021.