Screening Perovskites from $\text{ABO}_3$ Combinations Generated by Constraint Satisfaction Techniques Using Machine Learning

Jie Zhao* and Xiaoyan Wang*

**Cite This:** ACS Omega 2022, 7, 10483−10491

**ACCESS**

**ABSTRACT:** Perovskite oxides are attractive candidates for various scientific applications because of their outstanding structure flexibilities and attractive physical and chemical properties. However, labor-intensive and high-cost experimental and density functional theory calculation approaches are normally used to screen candidate perovskites. Herein, a machine learning method is employed to identify perovskites from $\text{ABO}_3$ combinations formulated as constraint satisfaction problems based on the restrictions of charge neutrality and Goldschmidt tolerance factor. By eliminating five features based on their correlation and importance, 16 features refined from 21 features are employed to describe 343 known $\text{ABO}_3$ compounds for perovskite formability and stability model training. It is found that the top three features for predicting formability are structural features of the $\text{A}−\text{O}$ bond length, tolerance, and octahedral factors, whereas the top nine features for predicting the stability are elemental and structural features related to the $\text{B}$-site elements. The precision and recall of the two models are 0.983, 1.00 and 0.971, 0.943, respectively. The formability prediction model categorizes 2229 $\text{ABO}_3$ combinations into 1373 perovskites and 856 nonperovskites, whereas the stability prediction model distinguishes 430 stable perovskites from 1799 unstable ones. Three hundred thirty-eight combinations are recognized as both formable and stable perovskites for future investigation.

1. **INTRODUCTION**

In the past few decades, substantial attention has been devoted to perovskite oxides because of their fascinating electrical, optical, magnetic, dielectric, thermal properties, etc.1-6 For example, their mixed electronic and ionic conductivities enable them to be used as electrode materials of solid oxides fuel cells.6 Due to their visible light response capability and intrinsic activity for oxygen evolution reaction, some of the perovskites have been used as photo-electrochemical and electrochemical catalysts for water splitting.7 Their multiferroics responses also make them promising actuators and sensors.8

Another reason for the enormous application of perovskite oxides is their structure flexibility, through which their physical and chemical properties can be easily tuned. Although their chemical formula $\text{ABO}_3$ is simple, the 12-fold coordinated $\text{A}$ site can be occupied with low-valent and large-sized alkali, alkali earth, and rare earth metal cations, while the $\text{B}$ site coordinated with 6-fold oxygen anions can be filled with high-valent and small-sized transition metal cations.9 Therefore, tens of elements can be accommodated at the $\text{A}$ and $\text{B}$ sites because of the structural flexibility of perovskite oxides.10 More $\text{ABO}_3$ perovskites can be acquired by doping the $\text{A}$ and/or $\text{B}$ sites with more than one cation based on charge balance.5

However, not all of the stable $\text{ABO}_3$ compounds are perovskite oxides. For example, $\text{CsNbO}_3$ is a stable but not a perovskite oxide due to its larger $\text{A}−\text{O}$ and smaller $\text{B}−\text{O}$ bond lengths.11 Therefore, it is indispensable to investigate the formability and stability of $\text{ABO}_3$ perovskite oxides before their utilization. Traditionally, trial-and-error approaches are used for that. The oxides are initially synthesized using their corresponding stoichiometric precursors and then characterized to check their formability and stability.9 However, those methods depend heavily on researchers’ knowledge and instinct, and they are time-consuming and labor-intensive.12

It is impractical for researchers to check all possible $\text{ABO}_3$ combinations using experimental methods.

Density functional theory (DFT) calculations are also potential approaches to predict the formability and stability of perovskite oxides, achieved by solving the Kohn−Sham equations.12 Jacobs et al.13 screened 2145 perovskites using high-throughput DFT calculations to confirm their stability and catalytic activity for the oxygen reduction reaction. Fifty-two potential candidates were singled out with good stability on par with high activity. Emery et al.14 investigated the thermodynamic stability and oxygen vacancy formation energy

**Received:** January 1, 2022  
**Accepted:** March 7, 2022  
**Published:** March 16, 2022
A-HOMO  A-IE  A-LUMO  Zr  A-ZR  A-EA  B-HOMO  B-IE  B-LUMO  Zr  B-ZR  B-EA  μ (μ)  r_A/ r_O  d_AO  d_BO  M_B  \delta E_{AO} r_A / r_O  \delta E_{BO} r_B / r_O  Figure 1. Pearson correlation coefficients between the 21 features.

of 5329 perovskites using high-throughput DFT calculations. They discerned 139 favorable perovskites for thermochemical water splitting. Tezsevin et al.\textsuperscript{15} used high-throughput DFT calculation to screen cubic perovskites for solid oxide fuel cell cathode materials. Thirty-one candidates were picked out from 270 ABO\textsubscript{3} compounds. Although DFT calculations are effective measures in screening potential perovskites for various applications, a large number of calculations that have to be carried out result in high computational expense.\textsuperscript{16,17} Therefore, it is highly desirable to screen potential perovskite oxides with a more economical and practical method.

Machine learning (ML) approaches use both successful and failed experimental and computational data to train models. The models are then used to forecast whether ABO\textsubscript{3} compounds are perovskites or not.\textsuperscript{18} For example, Talapatra et al.\textsuperscript{10} used available experimental and computational data of 1505 single perovskites and 3469 double perovskites to train classification models for predicting new formable and stable perovskites. They found that 414 compounds are promising candidates for future evaluation. Li et al.\textsuperscript{16} trained ML models with 1929 perovskite oxide energies calculated with DFT to predict the energy above the convex hull (\(E_{\text{hull}}\)) and phase stability of perovskite oxides. Liu et al.\textsuperscript{19} trained an ML model for predicting the formability of perovskite with known 397 ABO\textsubscript{3} compounds. The model was then used to classify 891 ABO\textsubscript{3} compounds from Materials Project (MP) database to perovskites and nonperovskites. The stability of the predicted perovskites was described with \(E_{\text{hull}}\). Sharma et al.\textsuperscript{20} investigated the feasibility and stability of defect formation in perovskite oxides using ML methods, anticipating to determine the factors that affect the defect formation energy of perovskite oxides during substitution.

For the ML methods, appropriate features related to the formability and stability of perovskite oxides should be carefully selected for model training. Up till now, many features have been employed to be indicators for that.\textsuperscript{10,16,19,20} Goldschmidt tolerance factor \(t\) was initially proposed to geometrically describe the likelihood of perovskite formation.\textsuperscript{21} It is defined as \(t = (r_A + r_O)/\sqrt{2(r_B + r_O)}\), where \(r_A, r_B, \text{and } r_O\) are the Shannon ionic radii of the A, B, and O ions\textsuperscript{22} respectively. For the ideal cubic perovskite structure, \(t\) is close to 1. Experimental evidence suggests that the tolerance factor of the cubic perovskite structure is in the range of 0.9–1.0\textsuperscript{10} and the range of \(t\) can be extended through structure distortion.\textsuperscript{10,23} Nevertheless, some ABO\textsubscript{3} compounds with 0.8 < \(t\) < 0.9 do not have perovskite structures, which indicates that additional features should be used in addition to tolerance factor.\textsuperscript{24} The octahedron BO\textsubscript{6} is a basic component for perovskite structure.\textsuperscript{24} As the ratio of \(r_B\) and \(r_O\) is limited in a certain range for BO\textsubscript{6} octahedron, it is appropriate to use their ratio (\(\mu = r_B/r_O\)) to describe the formability and stability of perovskites. Another tolerance factor defined as \(t_{BV} = d_{A-O}/\sqrt{2d_{B-O}}\) was also proposed.\textsuperscript{25} Rather than using Shannon ionic radii of the A, B, and O ions, it utilized A–O and B–O bond lengths (\(d_{A-O}\) and \(d_{B-O}\)) based on the bond-valence model to calculate \(t_{BV}\). Features with respect to elemental properties of A- and B-site atoms such as highest occupied molecular orbital (HOMO), lowest unoccupied molecular orbital (LUMO) energies, ionization energy (1E), electronegativity (X), and Mendeleev numbers are also applied for predicting the formability and stability of perovskites.\textsuperscript{10,19} Instead of single features, some complex features were also effective in describing the structural formability of ABO\textsubscript{3} perovskites.\textsuperscript{19,25}

In this work, formable and stable perovskite oxides are screened from unexplored ABO\textsubscript{3} combinations using ML approaches. The unexplored combinations are generated by a constraint satisfaction problem (CSP) technique from 73 elements based on the restrictions of charge neutrality and the Goldschmidt tolerance factor. An input data set composed of 343 known ABO\textsubscript{3} compounds, which are, respectively, described by 21, 16, and 17 features, is employed to train ML models for predicting the formability and stability of perovskites. The 16 and 17 features are refined from the 21 features on the basis of their correlations and importance. The
performance of the six models is then evaluated using a confusion matrix. The formability prediction model trained with 16 features classifies the unexplored ABO$_3$ combinations into 1373 perovskites and 856 nonperovskites, while the stability prediction model trained with those features categorizes them into 430 stable and 1799 unstable perovskites.

2. RESULTS AND DISCUSSION

2.1. Feature Correlation and Importance. In this work, 21 features were employed to describe the input data set for ML model training. To refine them and eliminate those with high correlation and less importance, the pairwise Pearson correlation coefficients of the 21 features were calculated, and their importance was evaluated using the random forest (RF) algorithm and the recursive feature elimination (RFE) feature selection approach, respectively.

Figure 1 shows the Pearson correlation coefficient ($r$, $r$ ∈ $[-1, 1]$) for the 21 features. The number in each square represents the coefficient of each pair of features. The positive values mean that the two features are positively correlated; on the contrary, the negative values signify that the two features have negative correlations. In addition, the bigger the absolute value of the coefficient, the stronger the correlation between the features. Normally, $|r| > 0.8$ indicates a very strong relationship. Therefore, the feature pairs of ($X_A$, $M_d$), ($X_A$, $\Delta X_{AO}$*$r_A$/$r_O$), ($B$-IE, $X_B$), ($X_B$, $\Delta X_{BO}$*$r_B$/$r_O$), ($B$-ZR, $\Delta X_{BO}$*$r_B$/$r_O$), ($\mu$, $d_{BO}$-O), ($\mu$, $\Delta X_{BO}$*$r_B$/$r_O$), and ($M_d$, $\Delta X_{AO}$*$r_A$/$r_O$) have strong correlation because their coefficients are 0.91, 0.88, 0.85, 0.86, −0.86, 0.86, −0.81, and 0.83, respectively. Therefore, some of these features can be eliminated since they have strong linear correlations.

To explore the importance of the 21 features, we collect their importance values from the results of constructing 100 ML models. Figure 2a shows the Pearson correlation coefficients of the 21 features. The heat map portrays the receiver operating characteristic (ROC) curves of those formability and stability models. The accuracy and precision for models 1−3 are 0.977, 0.967, 0.988, 0.983, and 0.942 and 0.950, respectively. Their recalls and F1 scores are 1.00, 0.983, 1.00, 0.992, and 0.966 and 0.958, respectively. The ROC curves show that all of the models are effective in predicting the formability and stability of perovskites.

Figure 2b displays the feature importance for predicting the stability of perovskites. The top five features are $d_{BO}$-O, $t$, $B$-HOMO, and $B$-ZM. The top nine features are all elemental and structural features with respect to the B-site elements, indicating that the B-site element properties are determining factors for predicting the stability of perovskites. Most of the features related to the A-site element rank at the bottom.

Therefore, based on the feature correlation and importance investigation, the 21 features were refined to 16 features for model training for the perovskite formability and stability prediction model training by removing the feature sets of [$X_A$, $X_B$, $\Delta X_{AO}$*$r_A$/$r_O$, $\Delta X_{BO}$*$r_B$/$r_O$, $d_{BO}$-O] and [$d_{BO}$-O, $\Delta X_{BO}$*$r_B$/$r_O$, $A$-IE, $X_A$, $M_d$] respectively.

For comparison, the RF feature selection approach was also used to investigate the importance of the 21 features. It was found that A-HOMO, B-EA, B-LUMO, and A-EA were the least important features for the perovskite formability model training, while A-HOMO, A-EA, A-LUMO, and A-ZR were eliminated for the most times for the perovskite stability model training; therefore, they were removed from the 21 features, leaving 17 features for perovskite formability and stability model training.

2.2. Predicting the Formability of Perovskites. The refined 16 and 17 features (FG-2 and FG-3) were used for the formability model training, and the models were named model 2 and model 3, respectively (Figure 7b). For comparison, the 21 features (FG-1) were also adopted for model training, and this model was named model 1 (Figure 7b). Confusion matrices for the three models are shown in Figure 3a−c. The accuracy and precision for models 1−3 are 0.977, 0.967, 0.988, 0.983, and 0.942 and 0.950, respectively. Their recalls and F1 scores are 1.00, 0.983, 1.00, 0.992, and 0.966 and 0.958, respectively. The 100% recall signifies that all of the perovskites are differentiated from the test data set. The almost 100% F1 score implies the outstanding reliability of our RF classifier models to distinguish perovskites from nonperovskite compounds. Figure 3d−f portrays the receiver operating characteristic (ROC) curves of those formability
prediction models. The area under the ROC curve (AUC) denotes a measure of separability. The higher the AUC, the better the classification model. The AUCs of the three models are 0.995, 0.999, and 0.9856, respectively. The performance of the models for predicting the formability of perovskites in this study is better or comparable to that in other studies. For example, the accuracy, precision, recall, and AUC of a perovskite formability prediction model were 0.9401, 0.9344, 0.9913, and 0.96, respectively.10 The accuracy of another RF model for predicting the formability of the ABX3 perovskite was 0.9655.26

Moreover, it can be seen that, compared with the other two models, the ML model trained with FG-2 displayed the highest accuracy, precision, F1 scores, and AUC, indicating its excellent performance in distinguishing formable perovskite oxides from ABO3 compounds, probably owing to the elimination of the features with less importance and strong correlations.

Due to its better performance, model 2 was used to predict the formability of the 2229 ABO3 combinations. It was found that 1373 and 856 combinations were classified as perovskites and nonperovskites, respectively. The probability heat map for the classified perovskites is shown in Figure 4. Each small rectangle in it represents an ABO3 combination. The deeper the red of the rectangles, the higher the probability. It can be seen that the heat map is divided into several regions by A = [Re, Li, Ca, Ag, Al, Zn, W, Mg, I] and B = [Re, Rb, Cs, Ba, Sr, I]. This should be attributed to the exclusion of ABO3 compounds in the MP database from our ABO3 combinations and the low possibility of them to be A- and B-site elements. Furthermore, Nb, Bi, Tl, In, and Zr are more likely to be the A-site elements, while Lu, Y, Ce, Zr, and Sc tend to be the B-site elements.

2.3. Predicting the Stability of Perovskites. Models 4–6 (Figure 7c) for predicting the stability of perovskites were trained with 305 known compounds from the input data set. Each compound was described with 21 features (FG-1), refined 16 (FG-4), and 17 (FG-5) features, respectively. Apart from those features, $E_{\text{hull}}$ was also used to describe the thermodynamic stability of the compounds. The confusion matrixes and ROC curves for the three models are shown in Figure 5. In general, our models exhibited excellent performance in distinguishing stable perovskites from ABO3 compounds, especially for model 5. The accuracy, precision, recall, F1 score, and AUC of model 4 were 0.948, 0.970, 0.914, 0.941, and 0.985, respectively, while those for model 5 were 0.961, 0.971, 0.943, 0.956, and 0.983, respectively. Hence, model 5 was more robust in differentiating stable and unstable ABO3 perovskites than model 4, possibly owing to the elimination of redundant features with high correlations. The accuracy, precision, recall, F1 score, and AUC of model 6 were 0.961, 1.00, 0.914, 0.955, and 0.966, respectively. Apparently, model 5 also showed better quality than model 6, although it showed a little bit lower precision than model 6. The models for forecasting the perovskite stability in this work demonstrated better or comparable performance compared to the models in...
other studies. For instance, the accuracy, precision, recall, and AUC of an RF model for predicting the perovskite stability were 0.941, 0.933, 0.936, and 0.98, respectively. Li et al. developed several ML models for predicting perovskite thermodynamic stability using several machine learning algorithms, and the best model showed an accuracy of 0.93 and a precision of 0.89. As model 5 showed better performance than the other two models, it was used to predict the thermodynamic stability of the 2229 ABO₃ combinations. It was found that 430 of them were classified as stable perovskites. The probability heat map for the predicted stable combinations is displayed in Figure 6. It could be seen that ABO₃ combinations with A = [Mg, Mn, Pb, Zr, Sm] and B = [Zr, Hf, Nb] were stable with high probabilities.

On the basis of the prediction results, we found that 338 ABO₃ combinations were predicted to be both formable and stable perovskites. Of these, 17 combinations were had a probability of higher than 0.8, as shown in Table 1. These compounds would be promising candidates for further evaluation.

3. METHODS

3.1. Design Workflow. The overarching ML workflow for the prediction of the formability and stability of ABO₃ perovskites in this study is shown in Figure 7. Initially, 129 cations and oxygen ions were used to generate ABO₃ combinations by a CSP technique based on the restrictions of charge neutrality (m + n = 6) and tolerance factor (0.76 ≤ t ≤ 1.44), and 3057 ABO₃ combinations were therefore obtained (Figure 7a). The combinations were then compared with the ABO₃ compounds existing in the Materials Project (MP) database. Surprisingly, 2229 ABO₃ combinations were not included in the database, and they are used for further investigation.

An input data set composed of 343 ABO₃ compounds was employed to train machine learning models for predicting the formability of perovskites (Figure 7b), which is an intersection of the data sets gathered by Talapatra et al., Liu et al., and Zhang et al. Of these, 218 compounds were perovskites, while 125 compounds were nonperovskites. Each compound was described with three groups of features. Feature group 1 (FG-1) consisted of 21 features, while feature group 2 (FG-2) and feature group 3 (FG-3) were composed of 16 and 17 features, respectively, which were defined from FG-1 based on their correlation and importance. The models trained with FG-1, FG-2, and FG-3 were named model 1, model 2, and model 3, respectively. The performance of the three models was then evaluated and compared. Afterward, model 2 was used to navigate through a data set of 2229 ABO₃ combinations (Figure 7a).

Three hundred fifty ABO₃ compounds from the input data set, which were described with FG-1, feature group 4 (FG-4), and feature group 5 (FG-5), were selected to train stability classification models (Figure 7c). FG-4 and FG-5 were acquired from FG-1 by getting rid of five and four features from it on the basis of their correlation and importance, respectively. Apart from these features groups, the compounds...
were also described with $E_{\text{hull}}$. An $E_{\text{hull}}$ threshold of 50 meV/atom was used to distinguish thermodynamically stable and unstable compounds. The models trained with FG-1, FG-4, and FG-5 were named model 4, model 5, and model 6, respectively, and model 5 was then utilized to sieve stable perovskites from the 2229 ABO$_3$ combinations.

3.2. Generation of ABO$_3$ Combinations. Due to the structural flexibility of ABO$_3$ perovskite oxides, A and B sites can accommodate various cations. Generating ABO$_3$ combinations and screening potential perovskite oxides from them with high accuracy and efficiency is therefore of great importance. In this study, the generation of new ABO$_3$ combinations was formulated as constraint satisfaction problems (CSPs): an outstanding problem-solving paragon in artificial intelligence. For a CSP, a set of variables are included, each variable is correlated with a finite domain, and a number of limitations related to these variables are built. The CSP is solved by satisfying all of the limitations. Formally, the combination problem is defined as a tuple $(V, D, C)$

$$\text{ABO}_3\_\text{Gen} = (V, D, C)$$

where $V$ is a set of variables, $A_D$ ($B_D$) is the possible A-site (B-site) cations, and $C$ is a set of constraints: the sum of the A and B valence values equals 6. The tolerance factor is in the range of 0.76 and 1.44. The backtracking technique was used to search for satisfactory solutions (new ABO$_3$ combinations). Seventy-three elements forming 129 cations due to the multivalence of some elements were used for generating the ABO$_3$ combination (Figure 8). With the restrictions of charge neutrality and tolerance factor, 3057 ABO$_3$ combinations were generated. After eliminating the combinations that exist in the MP database, 2229 combinations were left, which were further used for predicting the formability and stability of perovskites.

3.3. Feature Selection and Reduction. Until now, many features have been used to indicate the formability and stability of perovskite oxides. For example, Zhang et al.\textsuperscript{23} applied bond lengths of A–O and B–O ($d_{A-O}$ and $d_{B-O}$) to fundamentally show the formability and stability of perovskite oxides. Liu et al.\textsuperscript{19} used a group of nine features to train a Gradient Boosting Decision Tree model for predicting the formability of perovskites, which were the Goldschmidt tolerance factor ($t$), octahedron factor ($\mu = r_B/r_O$), radius ratio of A to O ($r_A/r_O$), $d_{A-O}$ and $d_{B-O}$, electronegativity difference between A(B) and O ($\Delta X_{AO}$ and $\Delta X_{BO}$) multiplied by $r_A/r_O$ ($\mu r_A/r_O$), and the Mendeleev numbers of A and B ($M_A$ and $M_B$). In another work, Talapatra et al.\textsuperscript{10} adopted a group of 14 features (reduced from 28 features due to the elimination of mismatch factors for ABO$_3$ combinations) to train ML models for predicting the formability and stability of perovskites. The structural features include the Goldschmidt tolerance factor ($t$)
and octahedral factor ($\mu$), while the elemental features contain the highest occupied molecular orbital (HOMO) energy, lowest unoccupied molecular orbital (LUMO) energy, ionization energy (IE), electronegativity ($X$), electron affinity (EA), and Zunger’s pseudopotential radius (ZR) for A- and B-site cations.

Although the features in refs 10, 19, and 23 were employed to predict the formability and/or stability of perovskite oxides, they were different dimensional features. This indicated that the formability and stability of perovskite oxides could be described with various dimensional features. Therefore, in this study, the intersection of the features (21 features) from those references was adopted, and their correlations and importance were investigated to get rid of the redundant and less important features. The correlations of the features were investigated by calculating the Pearson correlation coefficient between each pair of them, while the feature importance was evaluated using the RF algorithm and the RFE approach.

3.4. Classification Model Training. Classification is a supervised learning notion that categorizes a set of data into classes. In our study, the input data set was employed to train the formability and stability classification models using the RF classifier implemented in the Scikit-learn package. An RF algorithm consists of many decision trees and merges them together to improve its classification accuracy and stability. It uses bootstrapping to select training data and construct a classifier. To tune the models for better fit and prediction, a grid search combined with a 10-fold cross-validation method is normally used for parameter selection and evaluation. Grid search optimizes the model by traversing parameter combinations. Cross-validation uses the training data set repeatedly, splits the data set, and combines them into different training and test data sets. The training data set is used to train the RFC model, while the test data set is used for model evaluation. Our formability and stability classification models were both trained on 80% of the input data set and then tested on the residual 20%. The number of trees and the maximum tree depth in the forest for the formability and stability classification models were 11, 12, 10, and 7, respectively.

3.5. Model Evaluation. Confusion matrixes were used to evaluate the performance of the RFC models. There are two rows and two columns for each matrix. Each row signifies the perovskites in an actual class, while each column denotes the perovskites in a predicted class. Therefore, four blocks are included for each confusion matrix, which are true positive (TP), false positive (FP), true negative (TN), and false negative (FN).

Four metrics, accuracy, precision, recall, and F1 score, calculated from the confusion matrixes were used for classification model evaluation. Accuracy is a ratio of correct predictions to the total number of samples in the test data set (eq 1). Precision is the proportion of positive predictions that is actually correct (eq 2). It reflects the model’s ability to distinguish the positive class. A recall is defined as the ratio of TP to the sum of TP and FN (eq 3). The higher the recall, the stronger the model’s ability to recognize positive samples. F1 score is the harmonic mean of the precision and recall, as shown in eq 4. It tries to find the balance between precision and recall. The higher the F1 score, the more robust the model.

Figure 6. Probability heat map of classified 430 stable perovskites with a probability of more than 50%.

Table 1. Formable and Stable Perovskite Oxides with Probability Higher than 0.8

| ABO₃     | formable probability | stable probability | ABO₃     | formable probability | stable probability |
|----------|---------------------|--------------------|----------|---------------------|--------------------|
| HfVO₃    | 1                   | 1                  | ErCeO₃   | 1                   | 0.818              |
| OsVO₃    | 0.909               | 1                  | VZrO₃    | 0.909               | 0.818              |
| TaCrO₃   | 0.909               | 0.909              | CoNbO₃   | 0.909               | 0.818              |
| CdZrO₃   | 1                   | 0.909              | PdHfO₃   | 0.909               | 0.818              |
| PtCrO₃   | 1                   | 0.909              | TcNbO₃   | 0.909               | 0.818              |
| PdVVO₃   | 1                   | 0.909              | AnNbO₃   | 0.818               | 0.818              |
| NiNbO₃   | 0.818               | 0.909              | MoScO₃   | 0.909               | 0.818              |
| TcMnO₃   | 1                   | 0.818              | RbScO₃   | 0.909               | 0.818              |
| TaMnO₃   | 0.909               | 0.818              |          |                     |                    |
**Figure 7.** Workflow for the prediction of the formability and stability of perovskites in this study. (a) Three thousand fifty-seven ABO3 combinations are generated from 129 A and B cations and oxygen ions using a CSP method. The combinations are then compared with the ABO3 compounds in the Materials Project database. Two thousand two hundred twenty-nine ABO3 combinations are found to be not included in the database. (b) Threeiability classification models are trained with the input data set of 343 known ABO3 compounds, and each compound is described with three groups of features (FG-1, FG-2, and FG-3), respectively. Model 2 is then used to investigate the formability of the 2229 ABO3 combinations. (c) Three hundred five ABO3 compounds from the input data set are selected for stability prediction model training. The stability status for them is described with Eh and the feature groups of FG-1, FG-5, and FG-S are also used to describe the compounds. Model S is then used to predict the stability of the 2229 ABO3 combinations.

**Figure 8.** Elements used for generating the ABO3 combination using a CSP technique.

\[
\text{accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)
\]
\[
\text{precision} = \frac{TP}{TP + FP} \quad (2)
\]
\[
\text{recall} = \frac{TP}{TP + FN} \quad (3)
\]
\[
\text{F1 - score} = \frac{2 \times \text{precision} \times \text{recall}}{\text{precision} + \text{recall}} \quad (4)
\]

### 4. CONCLUSIONS

In summary, a machine learning method was employed to screen formable and thermodynamically stable perovskite oxides. We introduced a constraint satisfaction technique in artificial intelligence to generate ABO3 combinations that meet the constraints of charge neutrality and tolerance factor. An input data set of 343 known compounds was used to train machine learning models using a random forest classifier. The models were then adopted to screen formable and thermodynamically stable perovskite oxides from the ABO3 combinations. Based on the findings in this study, the following conclusions can be drawn: The constraint satisfaction technique is efficient in generating a new ABO3 combination. The refined 16 features were enough to predict the perovskite formability and thermodynamic stability. The perovskite formability was dominantly determined by the structural features regarding the A- and B-site elements, whereas the perovskite stability was primarily governed by the elemental features related to the B-site element. Three hundred thirty-eight combinations that do not exist in the Materials Project database were recognized as promising candidate perovskites for future evaluation.

**AUTHOR INFORMATION**

Corresponding Authors

Jie Zhao – College of Chemical Engineering, Nanjing Tech University, Nanjing, Jiangsu 211816, People’s Republic of China; orcid.org/0000-0002-7712-6635; Email: j.zhao1@njtech.edu.cn

Xiaoyan Wang – School of Information Engineering, Nanjing Audit University, Nanjing, Jiangsu 211815, People’s Republic of China; orcid.org/0000-0003-1939-3839; Email: xywang@nau.edu.cn
Acknowledgments

This work was financially supported by the Natural Science Foundation of Jiangsu Province (Grant No. BK202000690), the Natural Science Foundation of Jiangsu Higher Education Institutions (Grant Nos. 20KJB530010 and 17KJB520014), and the Innovation and Entrepreneurship Program of Jiangsu Province.

References

(1) Grabowska, E. Selected perovskite oxides: Characterization, preparation and photocatalytic properties—A review. Appl. Catal., B 2016, 186, 97–126.
(2) Kubicek, M.; Bork, A. H.; Rupp, J. L. M. Perovskite oxides—a review on a versatile material class for solar-to-fuel conversion processes. J. Mater. Chem. A 2017, 5, 11983–12000.
(3) Gao, W.; Zhu, Y.; Wang, Y.; Yuan, G.; Liu, J.-M. A review of flexible perovskite oxide ferroelectric films and their application. J. Materiomics 2020, 6, 1–16.
(4) Riemer, L. M.; Jin, L.; Uršič, H.; Otonicar, M.; Rojac, T.; Damjanovic, D. Dielectric and electro-mechanic nonlinearities in perovskite oxide ferroelectrics, relaxors, and relaxor ferroelectrics. J. Appl. Phys. 2021, 129, No. 054101.
(5) Yin, W.-J.; Weng, B.; Ge, J.; Sun, Q.; Li, Z.; Yan, Y. Oxide perovskites, double perovskites and derivatives for electrocatalysis, photocatalysis, and photovoltaics. Energy Environ. Sci. 2019, 12, 442–462.
(6) Kan, W. H.; Samson, A. J.; Thangadurai, V. Trends in electrode development for next generation solid oxide fuel cells. J. Mater. Chem. A 2016, 4, 17913–17932.
(7) Fabbi, E.; Nachtegaal, M.; Binninger, T.; Cheng, X.; Kim, B.-J.; Durst, J.; Bozza, F.; Graule, T.; Schäublin, R.; Wiles, L.; et al. Dynamic surface self-reconstruction is the key of highly active perovskite nanoelectrocatalysts for water splitting. Nat. Mater. 2017, 16, 925–931.
(8) Srivastava, A.; Singh, A. K.; Srivastava, O. N.; Tewari, H. S.; Massoud, K. B.; Singh, J. Magnetic and Dielectric Properties of La and Ni Co-substituted BiFeO3 Nanoceramics. Front. Phys. 2020, 8, No. 282.
(9) Zhu, J.; Li, H.; Zhong, L.; Xiao, P.; Xu, X.; Yang, X.; Zhao, Z.; Li, J. Perovskite Oxides: Preparation, Characterizations, and Applications in Heterogeneous Catalysis. ACS Catal. 2014, 4, 2917–2940.
(10) Talapatra, A.; Uberuaga, 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.
(11) Pilania, G.; Balachandran, P.; Gubernatis, J. E.; Lookman, T. Classification of ABO3 perovskite solids: a machine learning study. Acta Crystallogr., B: Struct. Sci., Cryst. Eng. Mater. 2015, 71, 507–513.
(12) Tao, Q.; Xu, P.; Li, M.; Lu, W. Machine learning for perovskite materials design and discovery. npj Comput. Mater. 2021, 7, No. 23.
(13) Jacobs, R.; Mayeshiba, T.; Booske, J.; Morgan, D. Material Discovery and Design Principles for Stable, High Activity Perovskite Cathodes for Solid Oxide Fuel Cells. Adv. Energy Mater. 2018, 8, No. 1707208.
(14) Emery, A. A.; Saal, J. E.; Kirklin, S.; Hegde, V. I.; Wolverton, C. High-Throughput Computational Screening of Perovskites for Thermochemical Water Splitting Applications. Chem. Mater. 2016, 28, 5621–5634.
(15) Tezsevin, I.; van de Sanden, M. C. M.; Er, S. High-Throughput Computational Screening of Cubic Perovskites for Solid Oxide Fuel Cell Cathodes. J. Phys. Chem. Lett. 2021, 12, 4160−4165.

Notes
The authors declare no competing financial interest.