

Machine learning of octahedral tilting in oxide perovskites by symbolic classification with compressed sensing

Stephen R. Xie^{a,b,1}, Parker Kotlarz^{a,1}, Richard G. Hennig^{a,b,*}, Juan C. Nino^a

^a Department of Materials Science and Engineering, University of Florida, Gainesville, FL 32611, United States

^b Quantum Theory Project, University of Florida, Gainesville, FL 32611, United States

ARTICLE INFO

Keywords:

Machine learning
Symbolic regression
Classification
Oxide perovskites
Crystal structure
Octahedral tilting

ABSTRACT

The steady growth of online materials databases, coupled with efforts in materials informatics, has invited the reexamination of existing empirical models through the lens of modern machine learning techniques. Inspired by recent efforts to improve on the Goldschmidt tolerance factor for perovskite formation, we apply the symbolic regression to the problem of predicting octahedral tilting. In addition to its impact on the crystal structure, octahedral tilting is related to functional properties, including dielectric permittivity, ferroelectricity, magnetic properties, and metal-insulator transitions. By relating a selection of physical parameters (e.g., atomic radii, electronegativity) with mathematical operations (e.g., addition, exponentiation), we identify an analytical equation that correctly predicts the octahedral tilting classification for 49 perovskite oxides in a dataset of 60 materials. Using the same training dataset, we additionally fit and compare seven models generated by other common machine learning methods. Despite the increased complexity afforded by support vector machines, decision trees/random forests, and artificial neural networks, we find that our equation outperforms the other models as well as the original tolerance factor in predicting octahedral tilting.

1. Introduction

The use of machine learning in the accelerated and cost-efficient discovery of materials has been driven by the Materials Genome Initiative and the availability of extensive materials databases such as AFLOW, Materials Project, and OQMD [1–4]. Machine learning has been applied to materials discovery and characterization, including applications from empirical potential development to the prediction of observable properties like bandgap and superconducting critical temperature [5–15]. Arising from the emergence of such machine learning methods is the opportunity to revisit past explorations in materials science for the prospect of overlooked discoveries or trends (Fig. 1).

One recent noteworthy implementation of machine learning in materials science that can enable this line of research is the symbolic regression SISO (Sure Independence Screening and Sparsifying Operator) framework, which generates analytical expressions for regression and classification using a relatively small number of physical quantities [16,17]. Using a compressed-sensing approach, SISO selects the optimum descriptor from the massive space of analytical expressions. SISO's versatility has been demonstrated in several applications, from Anderson *et al.* employing SISO to find predictions for adsorption

energy for catalytic materials [18] to Bartel *et al.* using SISO to identify a descriptor to predict Gibbs energy for inorganic compounds [19], and Mera Acosta *et al.* using SISO to predict quantum spin Hall insulators while providing insights into the mechanisms engendering topological transitions [20]. In one notable example of SISO's computational prowess, Bartel *et al.* used SISO to predict a new tolerance factor for perovskite oxides and halides [21].

Originating in 1926, the Goldschmidt tolerance factor has been widely used for structural analysis of the perovskite crystal structure (ABX₃) [22]. The perovskite crystal structure consists of A-site cations in twelve-fold coordination with corner-sharing BX₆ octahedra. Goldschmidt developed the tolerance factor for the stability of the perovskite structure from the tolerance factor of a hard-sphere model of the cubic perovskite structure:

$$t = \frac{1}{\sqrt{2}} \frac{r_A + r_X}{r_B + r_X}, \quad (1)$$

where r_A , r_B , r_X are the ionic radii of the A, B, and X site, respectively. The Goldschmidt parameter is $t = 1$ for the ideal perovskite structure. The motivation for understanding the stability of compounds in the perovskite structures stems from their importance in applications

* Corresponding author.

E-mail address: rhennig@ufl.edu (R.G. Hennig).

¹ These authors contributed equally to this work.

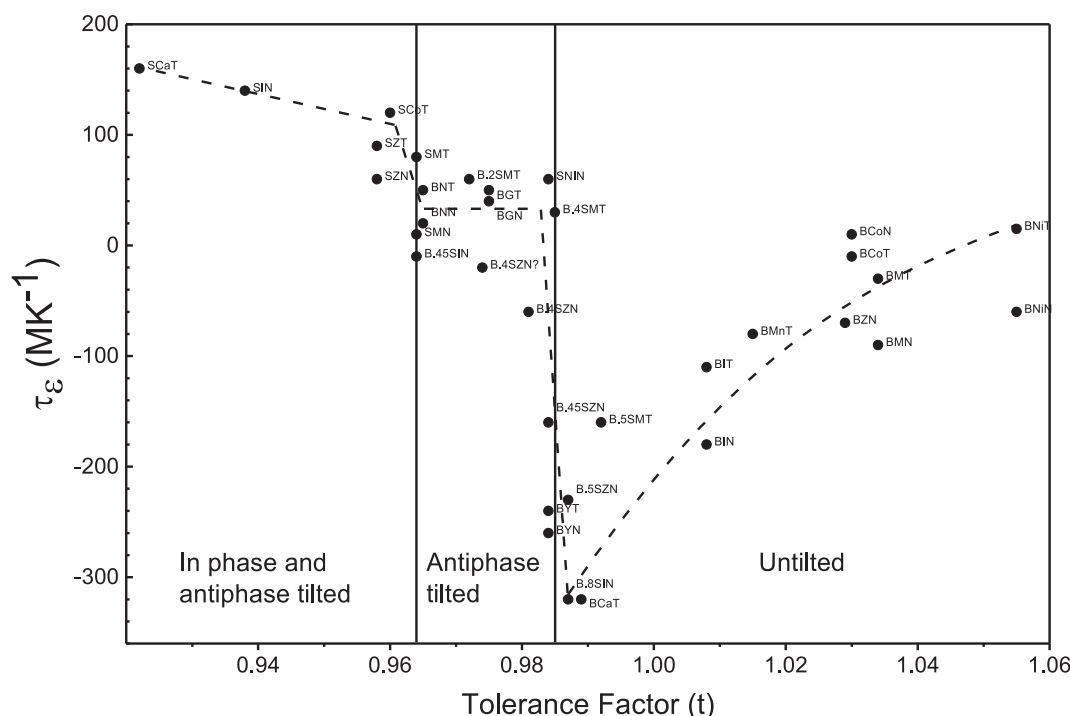


Fig. 1. Reaney *et al.* graph of tolerance factor (t) versus the temperature coefficient of the dielectric permittivity (τ_ϵ) at room temperature [29]. The line drawn through the points is a guide to the eye and shows a relative relationship between τ_ϵ and Goldschmidt's tolerance factor, t . Based on this work, the concept of applying Goldschmidt's tolerance factor spurred our investigation of a specific descriptor to predict octahedral tilting. The following is a key for the shorthand of perovskites and references are extensive in Reaney *et al.* original work: $\text{Ba}_x\text{Sr}_{1-x}(\text{Zn}_{1/3}\text{Nb}_{2/3})\text{O}_3$ (B_xSZN), $\text{Ba}_x\text{Sr}_{1-x}(\text{Mg}_{1/3}\text{Ta}_{2/3})\text{O}_3$ (B_xSMT), $\text{Ba}_x\text{Sr}_{1-x}(\text{In}_{1/2}\text{Nb}_{1/2})\text{O}_3$ (B_xSIN), $\text{Ba}(\text{Nd}_{1/2}\text{Ta}_{1/2})\text{O}_3$ (BNT), $\text{Ba}(\text{Gd}_{1/2}\text{Ta}_{1/2})\text{O}_3$ (BGT), $\text{Ba}(\text{Y}_{1/2}\text{Ta}_{1/2})\text{O}_3$ (BYT), $\text{Ba}(\text{Ca}_{1/3}\text{Ta}_{2/3})\text{O}_3$ (BCaT), $\text{Sr}(\text{Ca}_{1/3}\text{Ta}_{2/3})\text{O}_3$ (SCaT), $\text{Ba}(\text{Co}_{1/3}\text{Ta}_{2/3})\text{O}_3$ (BCoT), $\text{Sr}(\text{Co}_{1/3}\text{Ta}_{2/3})\text{O}_3$ (SCoT), $\text{Sr}(\text{Zn}_{1/3}\text{Ta}_{2/3})\text{O}_3$ (SZT), $\text{Sr}(\text{Ni}_{1/3}\text{Ta}_{2/3})\text{O}_3$ (SNTi), $\text{Ba}(\text{Ni}_{1/3}\text{Ta}_{2/3})\text{O}_3$ (BNiTi), $\text{Ba}(\text{Mn}_{1/3}\text{Ta}_{2/3})\text{O}_3$ (BMnTi), $\text{Ba}(\text{Mn}_{1/3}\text{Nb}_{2/3})\text{O}_3$ (BMnN), $\text{Ba}(\text{Mg}_{1/3}\text{Nb}_{2/3})\text{O}_3$ (BMN), $\text{Sr}(\text{Mg}_{1/3}\text{Nb}_{2/3})\text{O}_3$ (SMN), $\text{Ba}(\text{Zn}_{1/3}\text{Ta}_{2/3})\text{O}_3$ (BZTi), $\text{Ba}(\text{Ni}_{1/3}\text{Nb}_{2/3})\text{O}_3$ (BNiN), $\text{Sr}(\text{Ni}_{1/3}\text{Nb}_{2/3})\text{O}_3$ (SNiN), $\text{Ba}(\text{Co}_{1/3}\text{Nb}_{2/3})\text{O}_3$ (BCoN).

such as ferroelectrics [23], high-temperature superconductors [24], ferromagnets [25], spin filters [26], and electrocatalysis [27]. Although the Goldschmidt tolerance factor has been applied to perovskite analysis, ranging from Suarez *et al.* correlating tolerance factor increase to a decrease in paraelectric-to-ferroelectric phase transition temperature [28] to Reaney *et al.* demonstrating the relationship between tolerance factor and the temperature coefficient of dielectric permittivity and octahedral tilting [29], its limitations have prompted the explorations for a more accurate model of perovskite stability and structure [30].

Expanding on Goldschmidt's tolerance factor, Bartel *et al.* applied the SISO method to a database composed of 576 ABX₃ perovskite and non-perovskite materials to develop a new tolerance factor:

$$\tau = \frac{r_X}{r_B} - n_A \left(n_A - \frac{r_A/r_B}{\log(r_A/r_B)} \right), \quad (2)$$

where the additional parameter n_A is the oxidation state of the A-cation. Bartel’s tolerance factor distinguished between perovskites and non-perovskites with an accuracy of 92%, compared to only 74% accuracy for the Goldschmidt tolerance factor, with significant improvements for chlorides (90% to 51%), bromides (93% to 56%), and iodides (91% to 33%) [21]. This discovery of a more accurate tolerance factor for assessing the stability of the perovskite structure in a plethora of compounds warrants an exploration of previous trends determined by Goldschmidt’s tolerance factor for other perovskite properties.

One notable trend determined by Reaney *et al.* is the correlation of octahedral tilting with Goldschmidt’s tolerance factor [29]. Octahedral tilting involves the rotation of the BX octahedra about its three principal axes, resulting in different perovskite structures with lower overall symmetry compared to the conventional structure [31]. Glazer enumerated 23 unique tilt systems describing each unit cell length and the directions of octahedral tilting. A later group-theoretical analysis by Howard *et al.* suggested only 15 possible tilt structures [32]. In Glazer’s

notation, the letters a , b , and c represent unique principle axes lengths with superscripts denoting either a tilt in-phase along the axis (+), antiphase with the axis (-), or untilted (0). Octahedral tilting affects the unit-cell length, angles, and X-ray reflections. Octahedral tilting determines the space-group symmetry of perovskites [33]. Tilting also influences various characteristics of the material, including the temperature coefficient of the relative permittivity, τ_{ϵ} , important for the development of microwave dielectric materials [29,34,35], the ferroelectricity in perovskite heterostructures [36], the magnetic properties, and metal-insulator transitions [37].

Due to the vital role that octahedral tilting plays in determining perovskite structure and characteristics, it is pertinent to have a predictor for the onset of octahedral tilting in perovskite development. The difficulty of experiments that characterize perovskite tilt structures, and therefore correlate tilting with properties, results from the difficulty in identifying X-ray diffraction peaks, which are sometimes exceedingly small and difficult to classify correctly [31]. It is therefore widely understood that, in order to accurately detect and classify tilting experimentally, transmission electron microscopy is almost always necessary.

Based on the work by Reaney *et al.*, Goldschmidt's tolerance factor has been used to predict the onset of tilting in perovskites [29,38]. In addition to predicting tilt structures through Goldschmidt's tolerance factor, Lufaso and Woodward developed the Structure Prediction Diagnostic Software (SPuDS) based on the bond valence sum concept [39] and incrementally adjusting the tilt angle to minimize the global instability index (GII) [40] to predict the relative stability among the various perovskite structures and 10 tilt systems [41]. SPuDS calculates the bond valence s_{ji} between each cation-anion interaction using:

$$s_{ij} = \exp((R_0 - R_{ij})/B) \quad (3.1)$$

$$V_{i(\text{calc})} = \sum_j s_{ij} \quad (3.2)$$

$$d_i = V_{i(\text{ox})} - V_{i(\text{calc})} \quad (3.3)$$

where both B and R_0 are empirically determined parameters, and R_{ij} is the distance between the cation and anion. B can be considered a universal constant with a value of 0.37, and R_0 depends on the ions forming the bond. The discrepancy factor d_i is the difference between the formal valence $V_{i(\text{ox})}$ and the sum $V_{i(\text{calc})}$ of the atomic valences around the A, B, and X-site, which should be minimized to lessen the strain on the compound [42]. The GII measures the overall structural stability through:

$$GII = \sqrt{\frac{\sum_{i=1}^N d_i^2}{N}}, \quad (4)$$

where N is the number of atoms in the asymmetric unit cell, and d_i is the discrepancy factor mentioned above. Through incremental changes in the tilt angle, SPuDS provides each tilt structure with its associated GII. In tilt systems where the distance between A and X is not solely determined by tilt angle, SPuDS adjusts the tilt angle to first minimize the GII with the A-site ions in their highest symmetry positions. Each A and X bond valence is considered a vector parallel to the bond and magnitude equal to the valence. All 12 A and X vector bonds are then summed with the A-site iteratively changed to minimize the magnitude of the resultant vector, and then, based on the new A-site position, the tilt angle is iteratively changed to minimize the GII. This process of adjusting the A-site position and tilt angle is continued until both the resultant A-site vector sum and GII are minimized. SPuDS has been used to provide initial structural data [43], calculate tolerance factors with its bond valence parameters [44], and create structural models that take into account tilting [45].

In this work, we use SISO to search for a new descriptor to predict octahedral tilting in perovskites. We identify multiple equations that improve on the Goldschmidt tolerance factor and SPuDS in classifying octahedral tilting, with functional forms that rely primarily on atomic radii and suggest a geometric basis for tilting.

2. Data collection

The general steps to employ SISO in material sciences are: a) to determine a single target factor to be predicted by SISO; b) acquire a dataset that includes numerous properties that relate to the target factor; c) implement SISO in determining the most accurate equation describing said target factor from a training set of data; and d) analyze the newfound descriptor's accuracy using a testing set. The first two steps may be interchangeable depending on whether readily available data leads to the determination of a material's property that may be worth exploring using SISO. Nonetheless, properties must still be chosen that relate to the target property.

Another important consideration is whether the target property will be a quantitative or qualitative measure, which will alter the way SISO utilizes the property. In a quantitative target property, SISO will find a descriptor using the available data to best fit the numerical value of the target property. In a qualitative target property, categories must be created to differentiate each property, which will prompt SISO to find a descriptor that functions as a classifier rather than an equation attempting to find a specific property value. Due to tilting being classified as a qualitative target property, three categories of antiphase tilting, both in-phase and antiphase tilting (along separate axes), and untilted were chosen. These groups were modeled after Reaney's notable discovery of the relationship between Goldschmidt's tolerance factor with octahedral tilting and τ_c , which subsequently was separated into the same three groups [29]. The reasoning behind excluding solely in-phase tilting perovskites was further reinforced due to the scarcity of in-phase oxide perovskites found during the literature review. Most perovskites favor both in-phase and antiphase or solely antiphase tilt structures over solely in-phase tilt structures [46], which subsequently resulted in

only one oxide perovskite found with solely in-phase tilting during our literature review [47].

After we learn descriptors for the prediction of tilting in oxide perovskites, we assemble an experimental dataset through an extensive series of literature reviews of known perovskites. First, Reaney's work was adapted into a database to provide an initial foundation of perovskites of known tilt structures [29]. Next, two further review articles augment the database [38,48]. After noticing a trend in an underrepresentation of perovskites with either antiphase or untilted tilt structure, a more direct attempt to include a broader range of tilt structures was conducted using Bartel's database of perovskites and Goldschmidt's tolerance factor to identify a predicted set of perovskites with the desired tilting [21,29]. For this predicted set of perovskites, a further literature review followed to find the experimentally derived tilt structures for these compounds [49–56].

Since octahedral tilting is inherently difficult to characterize experimentally [31], the emphasis was placed on compiling a dataset of perovskites with accurately known tilting structures and properties while ensuring a sufficiently large database to perform symbolic regression with SISO.

Effective crystal radii and the effective valence of A and B cations were chosen as the associated characteristics or properties due to their importance in atomic bonding and in determining structure as relayed by Goldschmidt's tolerance factor and Bartel's SISO tolerance factor [21,22]. Crystal radii were chosen as opposed to ionic radii as recommended by Shannon due to their closer relationship to the physical size of ions in a solid [57]. Electronegativity and polarizability of A and B cations were also added to the list of associated properties due to their inherent influence on bonding and crystal structure stability, as well as their empirical prevalence in other machine learning classification techniques [58].

3. Finding low-dimensional descriptors for classification using machine learning

To identify predictive models for classification, the symbolic regression SISO framework searches for analytical formulas relating a small number of physical quantities, or "features," with the desired property. Beginning with a set of scalar-valued features Φ_0 , SISO generates additional features by recursively applying unary (e.g., exponentiation, logarithm) and binary (e.g., addition, multiplication) operations. The final feature space after i iterations, Φ_i , consists of various analytical expressions that each map the input parameters to a low-dimensional space. The domain of each class is defined by the convex hull formed by the corresponding training data. During the sure independence screening (SIS) step, SISO ranks these expressions by the sum of the overlapping one-dimensional region (or distance between regions, in the negative case) between class domains. For the one-dimensional solution ($\Omega = 1$), the sparsifying operator (SO) step returns the analytical expression with the least overlap between categories. To determine Ω -dimensional solutions, the SO step selects tuples of Ω equations that minimize the Ω -dimensional overlap between domains (area or volume, in the 2D and 3D cases, respectively).

As previously mentioned, we selected eight primary features to construct Φ_0 : atomic radii (r_a, r_b), valence numbers (v_a, v_b), electronegativities (χ_a, χ_b), and dielectric polarizabilities (α_b, α_b). We chose four binary operators (+, −, ×, /) and five unary operators ($^{-1}, ^2, ^3, \sqrt{}, \ln()$) to recursively apply to the feature space, resulting in over 182 million features in Φ_3 after three iterations. It is important to note that limitations in computational resources prevent the generation of Φ_4 using the SISO framework, meaning that equations with many nested terms, such as the GII, will not be generated or explored by SISO when Φ_0 consists of basic scalar properties.

We applied six-fold stratified cross-validation to benchmark the performance of the predicted classifiers framework, as depicted in

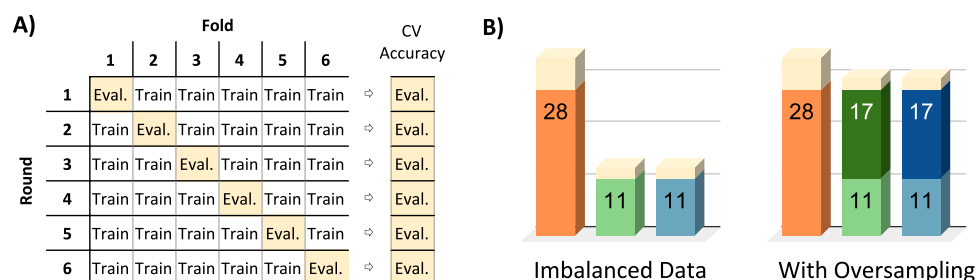


Fig. 2a. During N -fold cross-validation (CV), all samples are partitioned into N subsets, called folds. The stratified variant of CV involves preserving the distribution of classes within each fold, yielding stratified subsets that are each representative of the overall data. During each round of cross-validation, a different combination of $N - 1$ folds is used as training data, while the excluded fold is used to evaluate performance. The accuracy across all N folds, where each sample is used for validation exactly once, is a measure of how well a machine learning method performs on unseen data. In this work, we partitioned the data into six folds and used the SISSO framework to analyze each subset of fifty samples.

To address the imbalance between classes in the data set, where the ratio of in-phase and antiphase tilted materials to either antiphase tilted or untilted is greater than 5:2, we employed the oversampling technique as illustrated in Fig. 2b. Within each fold, we augmented both minority classes by randomly sampling materials with replacement until all three classes were equal. We then repeated the cross-validation process for comparison.

While the two tolerance factors presented by Goldschmidt and Bartel feature discrete boundaries between categories, we additionally consider the transition between categories using the probability estimates offered by modern machine learning methods. Based on the 1D SISSO function evaluations of the training data and the corresponding experimental tiltings, we use the k nearest-neighbor method to generate a continuous probability estimate for all 1D coordinates. We select $k = 10$ and additionally weigh the probability contributions from each neighbor by its distance to the sampled coordinate. In practice, any testing sample can be assigned a percent probability for each tilting class, and a prediction may be made based on the most-likely class or classes. We may also obtain discrete decision boundaries based on coordinates where the highest probability designations change.

4. Benchmarking the classification accuracy of competing approaches

We additionally optimized seven alternative machine learning models using both the imbalanced dataset and the oversampling scheme, to compare the SISSO framework with other modern methods for classification. Support vector machine (SVM), artificial neural network (NN), decision tree (DT), and random forest (RF) models were optimized using the Scikit-learn Python package [59].

The SVM is, traditionally, a deterministic classifier that separates classes by calculating the optimum decision boundaries between sets of samples after applying a transformation to a higher dimensional space using a kernel function [60,61]. We considered both the linear kernel and the nonlinear Gaussian radial basis function. The DT is a deterministic classifier that yields predictions (leaves) based on a flow-chart-like series of decisions (branches) based on the input variables and optimized thresholds [62]. The RF is an ensemble classification method that makes predictions based on the majority vote of a multitude of DTs, each trained on a randomized subset of the available data [63,64]. In this work, we constructed each RF with ten DTs. The NN, specifically the multi-layer perceptron in this work, is an algorithm that can learn highly nonlinear function approximators for mapping inputs

to their corresponding outputs [65]. The first layer of the NN consists of one neuron for each input feature, while the last layer of the NN consists of one neuron for each class. Between the input and output are one or more hidden layers of neurons. The value at each hidden neuron or output neuron is a weighted linear summation of all values in the preceding layer, plus a constant bias term, followed by an activation function. These weights and biases are trained using the back-propagation of error and stochastic gradient descent. We considered the linear, rectified linear unit (ReLU) and hyperbolic tangent (tanh) activation functions [66].

We tuned hyperparameters for each model using six-fold cross-validation rather than a holdout set due to the relatively small amount of available data. Probability estimates in SVMs were calculated by the pairwise coupling method for multi-class classification described by Wu et al. [67]. Probability estimates in DT and RF models were determined by the ratios of training samples and their classifications in each leaf. Probability estimates in NNs were determined by evaluating the softmax activation function after training with the cross-entropy loss function.

Unlike in the SISSO framework, the optimization procedures of SVM, RF, and NN are stochastic. Randomness is introduced into the SVM fitting by the internal cross-validation step that enables probability estimates with pairwise coupling [61,67]. The many decision trees comprising a random forest are each fitted with a randomized subset of data by sampling with replacement [63]. The weights and biases of neural networks are optimized using a stochastic gradient-based algorithm implemented in Adam, [68]. To better understand the predictive power of SVMs, DTs, RFs, and NNs, we retrain each model 25 times with identical parameters, except for the random seed and record all predictions (available in the supplemental information).

5. Results and discussion

Coupled with the cross-validation technique, the SISSO framework generated millions of equations and returned the six equations with the smallest overlap between domains, as summarized in Table 1. In several rounds, equations with unphysical units such as $\text{\AA}^{5/7}$ were filtered out in post-processing. All equations are reported in the supplementary material. Across all rounds in both the original and oversampling cases, the generated SISSO equations have high accuracy in predicting tilting behavior in their respective testing samples. The overall 85% accuracy with cross-validation indicates that the SISSO framework can generalize to unseen data. Notably, the equation $x = r_a^3 - r_a r_b (r_a + r_b)$ appears in four cases (5.4, 5.5, 6.5, 6.6) with a consistently high testing accuracy across three folds – half of the available data. Electronegativity appears in only one round each (5.2 and 6.2), while dielectric polarizability does not appear at all, suggesting that they are relatively unimportant for uniquely describing octahedral tilting.

Partially correct predictions, included in both Table 1 and Fig. 4, are defined as the instances where two categories are assigned nearly equal probability by a model and the correct classification is included. In this work, we take the criteria for comparable probabilities to be a maximum difference of 10%. A prediction is given full credit where the probability for the correct class is more than 10% greater than the

Table 1

Summary of SISSO equations as identified by cross-validation and performance on the corresponding 10-sample testing sets. Results with imbalanced training data and with the oversampling scheme are shown in the left and right columns, respectively. Abbreviations: U = Units, C = number of correct predictions, P = number of partially correct predictions.

		Imbalanced Training Data				Oversampling Scheme					
		Expression	U	C	P			Expression	U	C	P
Round 1	(5.1)	$\sqrt{r_a r_b} - 2r_a + r_b$	$\text{\AA}$	7	1	(6.1)		$\sqrt{r_a r_b} - 2r_a + r_b$	$\text{\AA}$	7	1
Round 2	(5.2)	$\frac{v_b \chi_a^3 r_b}{4\chi_a}$	$\text{\AA}$	9	0	(6.2)		$\frac{v_b \chi_a^3 r_b}{9\chi_a}$	$\text{\AA}$	9	0
Round 3	(5.3)	$\frac{(r_a - r_b)^3(r_a + r_b)}{r_b}$	$\text{\AA}^3$	8	0	(6.3)		$\frac{(r_a - r_b)^3(r_a + r_b)}{r_b}$	$\text{\AA}^3$	8	0
Round 4	(5.4)	$r_a^3 - r_a r_b(r_a + r_b)$	$\text{\AA}^3$	10	0	(6.4)		$\frac{(r_a + r_b)r_a}{r_b} - \frac{r_b \chi_a}{\chi_a + \chi_b}$	$\text{\AA}$	8	0
Round 5	(5.5)	$r_a^3 - r_a r_b(r_a + r_b)$	$\text{\AA}^3$	9	1	(6.5)		$r_a^3 - r_a r_b(r_a + r_b)$	$\text{\AA}^3$	9	1
Round 6	(5.6)	$\frac{r_a^3}{r_b^2} + r_a + r_b$	$\text{\AA}$	8	0	(6.6)		$r_a^3 - r_a r_b(r_a + r_b)$	$\text{\AA}^3$	10	0

remaining probabilities. A “correct” prediction with less than a 10% difference in probability between either of the other categories is given no credit due to the lack of confidence. Conceptually, partially correct predictions are expected to occur when classifying materials that are very close to a decision boundary. We then define a composite accuracy as % composite accuracy = $(C + 0.5P)/N$, where C is the number of correct predictions, P is the number of partially correct predictions, and N is the number of samples.

Although rounds 4 and 6 yield different equations with the oversampling scheme compared to the original imbalanced case, oversampling offers no improvement in the overall CV accuracy. With the same 10 testing materials from fold 4, equation 6.4 performs worse than equation 5.4. On the other hand, equation 6.6 performs better than equation 5.6 in predicting tilting in fold 6. To distinguish whether these differences arise from differences in the generation of equations or the ranking, we analyzed the intermediate SO results for each round. These intermediate results each contained 1000 equations, ranked by the size of the overlapping region between convex hulls as well as the number of training samples within the overlapping regions. We found that, in all rounds, all equations in Table 1 appeared among the 1000 equations with different rankings. As evident in rounds 4 and 6, it appears that the ranking criterion used by SISSO is sometimes insufficient for finding equations with the best generalizability or extrapolative capacity. Moreover, rather than one highly generalizable equation, SISSO identified 5 different equations with dissimilar functional forms when provided with six different partitions of available data. This indicates that SISSO is sensitive to the input data, and therefore a key implication is that the simple holdout method with one training and one testing set is inadequate when using SISSO with a small dataset.

The SISSO results of Round 6 are plotted in Fig. 3, including both equation 5.6 to the left and equation 6.6 to the right with a multiplicative coefficient of -1 for visualization purposes. The SISSO evaluations of the training data are displayed at the top, with dashed lines at k nearest-neighbor decision boundaries for qualitative reference. Testing predictions are displayed at the bottom with background color gradients corresponding to the continuous probability estimates calculated by the k nearest-neighbor method ($k = 10$). The visualization of 5.6 and 6.6 on the testing data shows a distinct difference in predictions on materials with antiphase tilting, which span the smallest region in both models. Despite having a smaller antiphase domain, both relative to the other classes and in absolute value, equation 6.6 clearly separates LaNiO_3 , LaAlO_3 , and LaCoO_3 from the other two domains, while equation 5.6 cannot. In addition, equation 5.6 exhibits peculiar multimodal probability functions with the k nearest-neighbor method due to the poor separation between testing materials near the decision

boundaries. Across both training and testing sets, equation 5.6 correctly predicts 49/60 samples with three partially correct predictions close to the marked decision boundaries: BaTiO_3 , SrTiO_3 , and LaCoO_3 . On the other hand, equation 6.6 correctly predicts 49/60 samples with only one partially correct prediction: $\text{Sr}(\text{Co}_{1/3}\text{Ta}_{2/3})\text{O}_3$. Although the accuracy of equation 6.6 across the entire dataset is not appreciably higher than equation 5.6 or other equations in Table 1, its repeated identification by the SISSO framework across rounds 4, 5, and 6 indicate that it is the most generalizable and transferable equation out of the millions generated using this set of sixty samples.

The classification accuracies for the imbalanced data and oversampling schemes are summarized in Table 2 for SISSO and the seven alternative methods. For stochastic algorithms, we report the respective performance for the two runs, out of 25, with the best and worst composite accuracies. The SISSO equations collectively achieve a composite accuracy of 86.7% across the six cross-validation folds. In both imbalanced and oversampling cases, the contribution of partially correct predictions to the accuracy is small (half of 3.3%). The composite accuracies of the SVM trials are the lowest among all methods, while the corresponding partially correct accuracies are the highest. This observation may be an artifact of the instability in the pairwise coupling probability estimates, which were calibrated using relatively few training samples in each round. The nonlinear SVM is slightly more accurate than the linear SVM, likely due to the increased complexity afforded by the nonlinear kernel transformation. The DT models are higher in accuracy than their RF counterparts in all cases, with a modest 6.6% difference in the oversampling case. The DT with oversampling has the highest accuracy among the non-SISSO methods, with a composite accuracy of 83.3%. The lack of partially correct predictions in the DT case is unsurprising because the multiple discrete decision boundaries in DTs are optimized to maximize the separation of classes among leaves, which inherently penalizes the partially correct cases. On the other hand, disagreement between DTs in the RF ensemble yields up to 15.0% partially correct predictions. The lack of cluster performance of the ensemble method compared to the individual classifier may be due in part to the sufficiency of the DT for the problem at hand. Moreover, the constituent trees in the RF are trained with randomized subsets of the available data, which, as discussed regarding the SISSO equations, is scarce and sensitive to over or underrepresentation of training data. The NN methods are comparable to the DT and RF in performance, with the ReLU NN achieving a maximum composite accuracy of 81.7%. The higher performance of the ReLU NN and tanh NN, compared to the linear NN, is most likely due to increased nonlinearity afforded by the ReLU and tanh activation functions.

The oversampling scheme slightly improves accuracy in the

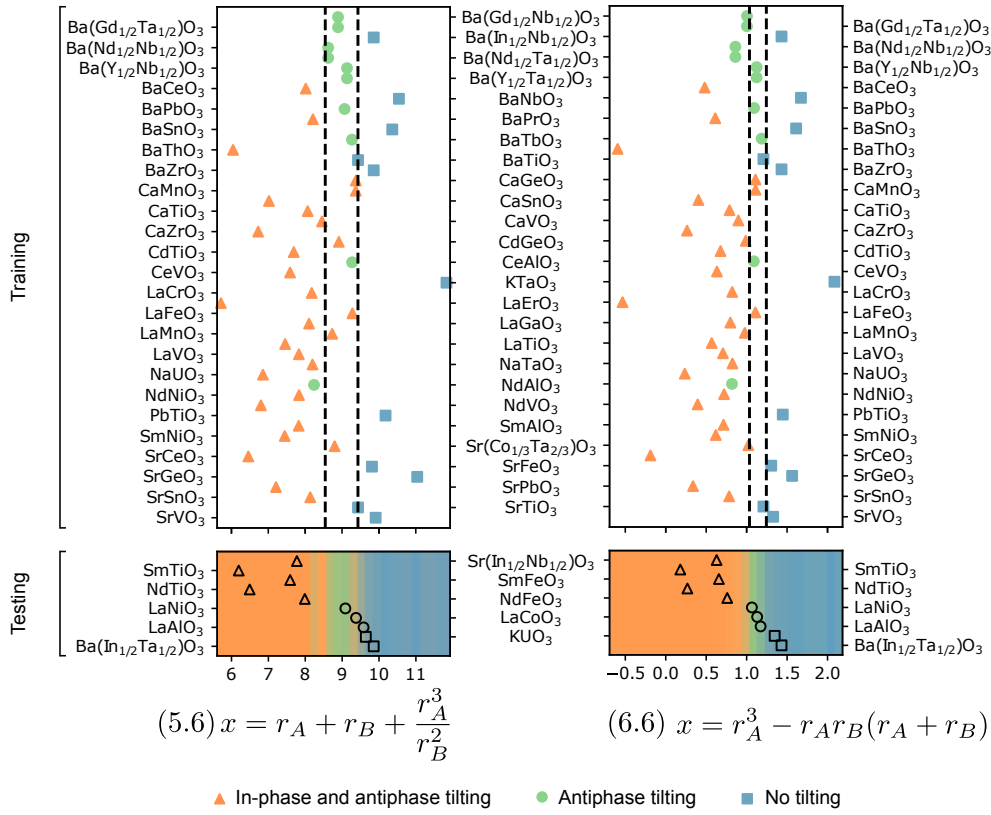


Fig. 3. Comparison of quantitative separation and accuracy between equations 5.6 and 6.6. The latter equation is mirrored across $x = 0$ for visualization purposes. Testing predictions are shown with black symbols with background colors corresponding to the k nearest-neighbor probability estimates along the x -coordinate.

nonlinear SVM and DT methods but otherwise tends to decrease composite accuracy, the number of correct predictions, and the number of partially correct predictions. The percent discrepancy between best and worst runs in the SVM trials are among the smallest (4.20% in the oversampling nonlinear SVM case to 7.50% in the linear SVM case). The same metric in the RF case ranges from 15.90% with oversampling to 18.40% without oversampling. The linear NN trial with oversampling exhibits the greatest discrepancy (25.0%) between the lowest and highest observed composite accuracies. The variation in accuracy attributed to the stochastic optimization is smaller in the SVM implementation than in the RF and NN implementations.

Predictions for select classification models are displayed in Fig. 4 for six cross-validation rounds. The corresponding true experimental values are shown in the first row for reference, while the predictions using the tolerance factor of equation (2) and SPuDS are shown in the last two rows for comparison. With consistently incorrect predictions across all methods, NdAlO_3 and CaMnO_3 are clear outliers when evaluated as unseen testing data. This suggests irregularity in their input features or in some combination of features, which might warrant further investigation and experimental confirmation. The Goldschmidt tolerance factor, coupled with the decision boundaries determined by Reaney et al., achieves a composite accuracy of 73.3% against the 60 materials. Notably, the 60 predictions made by the Goldschmidt tolerance factor include all three classes in similar proportions as opposed to the overrepresentation of in-phase and antiphase tilting. On the other hand, SPuDS predicts only 9 materials correctly alongside two partially correct predictions. SPuDS overpredicts the antiphase tilting class and in-phase tilting class while underpredicting the in-phase and antiphase

tilting and no tilting classes. One explanation for this low accuracy may be the breakdown of the assumption of rigid octahedra that SPuDS relies on to determine many degrees of freedom.

The analytical equations identified by SISO (5.1–6.6) can accurately classify tilting in perovskites with over 80% accuracy using only 2–4 physical parameters. In comparison, the Goldschmidt tolerance factor uses four physical parameters, and SPuDS uses two physical parameters and one empirical parameter, and both yield lower accuracies. The SVM, DT, RF, and NN models each have greater mathematical complexity than the SISO equations but are also slightly lower in accuracy. In particular, the equation $x = r_a^3 - r_a r_b (r_a + r_b)$ (5.4, 5.5, 6.5, 6.6) exhibits 81.7% accuracy (49/60) across the 60 samples and was repeatedly selected by SISO across multiple rounds of cross-validation.

Much like the original Goldschmidt tolerance factor from 1926, this SISO equation relates a physical phenomenon to the atomic radii and, indirectly, the theoretical bond lengths. The Goldschmidt tolerance factor uses the ratio of the A-X bond to the B-X bond to determine physical stability, while equation 5.4 relies on the radius of atom A, the radius of atom B, and the length of the A-B bond to predict octahedral tilting. When the radius of B is smaller than the radius of A, the value of the expression is large, and the model predicts no tilting. As the radius of B approaches the radius of A, the expression goes to zero, and the model predicts both in-phase and antiphase tilting. For intermediate values, where the radius of B is slightly smaller than the radius of A, antiphase tilting is observed. Using only two physical parameters, this equation is physically interpretable as a geometric relationship between the radii of A and B with units of volume.

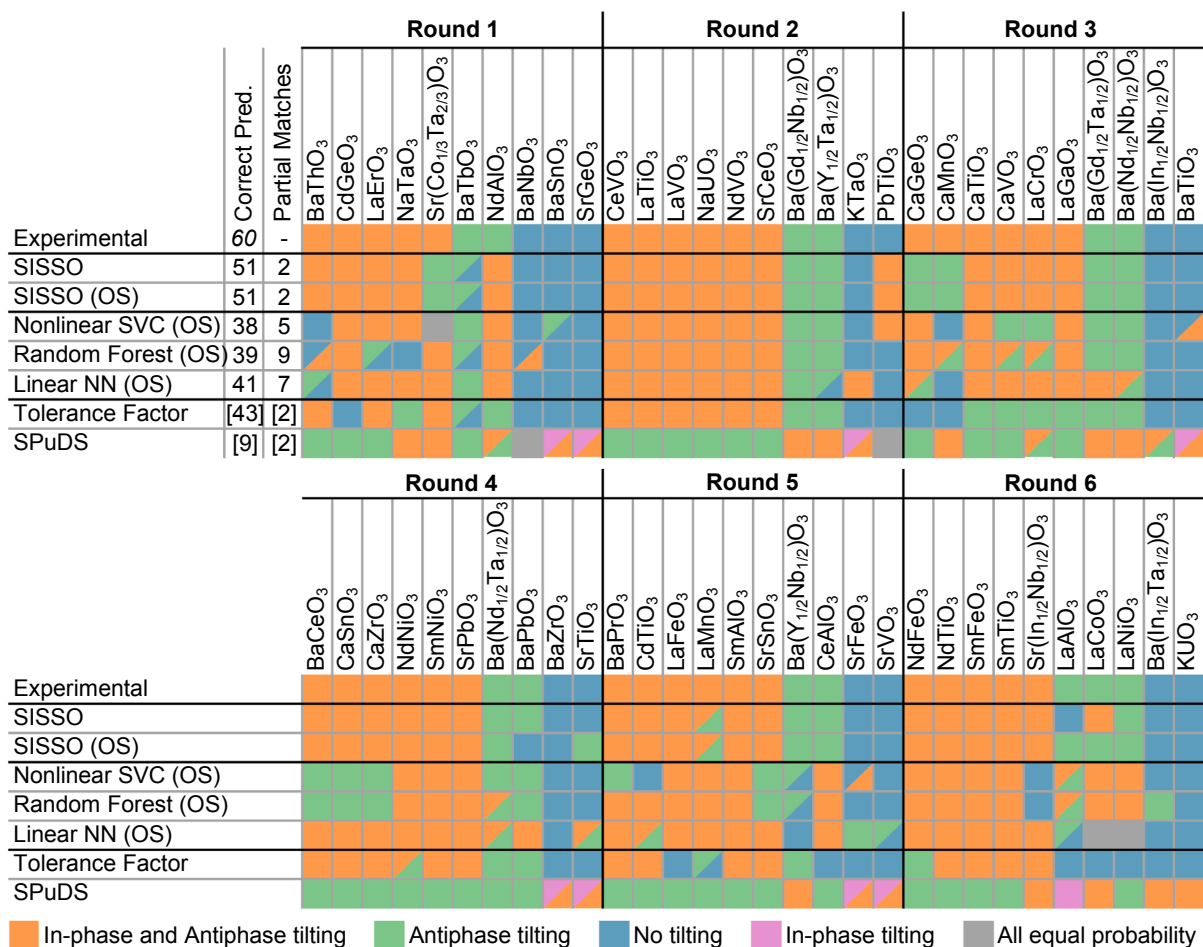


Fig. 4. Testing predictions from cross-validation for select models. Experimental classifications (reference data) comprise the first row. Each additional row corresponds to the testing predictions from a different method, segmented into results from the six rounds of cross-validation. The Goldschmidt tolerance factor and SPuDS predictions on the same materials, without cross-validation, are also shown for comparison. Partially correct predictions are displayed as doubly shaded cells with arbitrary ordering.

Table 2

Comparison between alternative machine learning models and their optimized hyperparameters. Descriptions of hyperparameters can be found in the supplemental material.

Method	Imbalanced Training Data			Oversampling Scheme		
	Composite Accuracy	Correct	Partially Correct	Composite Accuracy	Correct	Partially Correct
SISSO	86.7%	85.0%	3.3%	86.7%	85.0%	3.3%
Linear SVM	67.5%	51.7%	31.7%	65.0%	58.3%	13.3%
	60.0%	48.3%	23.3%	58.3%	50.0%	16.7%
	68.3%	58.3%	20.0%	70.0%	65.0%	10.0%
NonlinearSVM	61.7%	55.0%	13.3%	65.8%	61.7%	8.3%
	80.0%	80.0%	0.0%	83.3%	83.3%	0.0%
	79.2%	71.7%	15.0%	76.7%	73.3%	6.7%
DT	60.8%	56.7%	8.3%	60.8%	53.3%	15.0%
	77.5%	71.7%	11.7%	77.5%	75.0%	5.0%
	60.8%	50.0%	21.7%	52.5%	41.7%	21.7%
ReLU NN	81.7%	81.7%	0.0%	80.0%	80.0%	0.0%
	68.3%	68.3%	0.0%	65.0%	65.0%	0.0%
	80.0%	80.0%	0.0%	79.2%	78.3%	1.7%
tanh NN	70.0%	70.0%	0.0%	71.7%	71.7%	0.0%

The octahedral tilting dataset, SISSO inputs, SISSO outputs, and Python post-processing tools are available on Github at <https://github.com/henniggroup/symbolic-regression-utilities>.

6. Conclusions

Using symbolic regression within the SISSO framework, we identified a new descriptor for predicting the tilting of oxide perovskites, which may be used to screen perovskites for electronic and magnetic applications. Fit with a relatively small dataset of 60 octahedral tilting classifications from literature, our SISSO-generated model exceeds those produced with other modern machine learning methods in cross-validated testing accuracy with far less mathematical complexity. Compared to the existing Goldschmidt tolerance factor and SPuDS methods, the model achieves a higher classification accuracy on the available data with comparable complexity and physical interpretability. Finally, the straightforward functional form of our machine-learned classifier reinforces our understanding of octahedral tilting in perovskites by suggesting a primary basis in physical geometry.

CRedit authorship contribution statement

Stephen R. Xie: Conceptualization, Methodology, Software, Writing - original draft. **Parker Kotlarz:** Data curation, Writing - original draft. **Richard G. Hennig:** Conceptualization, Methodology, Supervision, Writing - review & editing. **Juan C. Nino:** Conceptualization, Supervision, Writing - review & editing.

Declaration of Competing Interest

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

Acknowledgements

The work by PK and JCN is partially supported by the National Science Foundation under grant ECCS-1709641. The work by SRX and RGH is supported by the National Science Foundation under grant DMR-1609306. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

References

- [1] National Science and Technology Council (U.S.), Materials Genome Initiative for Global Competitiveness, Executive Office of the President, National Science and Technology Council, 2011.
- [2] S. Curtarolo, et al., AFLOWLIB.ORG: A distributed materials properties repository from high-throughput ab initio calculations, *Comput. Mater. Sci.* 58 (2012) 227–235.
- [3] A. Jain, et al., Commentary: the Materials Project: A materials genome approach to accelerating materials innovation, *APL Mater.* 1 (1) (2013) 011002.
- [4] J.E. Saal, S. Kirklin, M. Aykol, B. Meredig, C.M. Wolverton, Materials design and discovery with high-throughput density functional theory: the open quantum materials database (OQMD), *JOM* 65 (11) (2013) 1501–1509.
- [5] A.P. Bartók, M.C. Payne, R. Kondor, G. Csányi, Gaussian approximation potentials: the accuracy of quantum mechanics, without the electrons, *Phys. Rev. Lett.* 104 (13) (2010) 136403.
- [6] S. Honrao, B.E. Anthonio, R. Ramanathan, J.J. Gabriel, R.G. Hennig, Machine learning of ab-initio energy landscapes for crystal structure predictions, *Comput. Mater. Sci.* 158 (2019) 414–419.
- [7] L.M. Ghiringhelli, J. Vybiral, S.V. Levchenko, C. Draxl, M. Scheffler, Big data of materials science: critical role of the descriptor, *Phys. Rev. Lett.* 114 (10) (2015) 105503.
- [8] G. Hautier, C.C. Fischer, A. Jain, T. Mueller, G. Ceder, Finding Nature's missing ternary oxide compounds using machine learning and density Functional Theory, *Chem. Mater.* 22 (12) (2010) 3762–3767.
- [9] S. Curtarolo, D. Morgan, K. Persson, J. Rodgers, G. Ceder, Predicting crystal structures with data mining of quantum calculations, *Phys. Rev. Lett.* 91 (13) (2003) 135503.
- [10] L. Yang, G. Ceder, Data-mined similarity function between material compositions, *Phys. Rev. B* 88 (22) (2013) 224107.
- [11] B. Meredig, C. Wolverton, Dissolving the periodic table in cubic zirconia: data mining to discover chemical trends, *Chem. Mater.* 26 (6) (2014) 1985–1991.
- [12] M. Rupp, A. Tkatchenko, K.-R. Müller, O.A. von Lilienfeld, Fast and accurate modeling of molecular atomization energies with machine learning, *Phys. Rev. Lett.* 108 (5) (Jan. 2012) 058301.
- [13] G. Pilania, C. Wang, X. Jiang, S. Rajasekaran, R. Ramprasad, Accelerating materials property predictions using machine learning, *Sci. Rep.* 3 (2013) 2810.
- [14] G. Pilania, A. Mannodi-Kanakkithodi, B.P. Uberuaga, R. Ramprasad, J.E. Gubernatis, T. Lookman, Machine learning bandgaps of double perovskites, *Sci. Rep.* 6 (2016) 19375.
- [15] S.R. Xie, G.R. Stewart, J.J. Hamlin, P.J. Hirschfeld, R.G. Hennig, Functional Form of the Superconducting Critical Temperature from Machine Learning, *Phys. Rev. B* 100 (17) (2019) 174513.
- [16] R. Ouyang, E. Ahmetcik, C. Carbogno, M. Scheffler, L.M. Ghiringhelli, Simultaneous Learning of Several Materials Properties from Incomplete Databases with Multi-Task SISSO, *J. Phys. Mater.* 2 (2) (2019) 024002.
- [17] R. Ouyang, S. Curtarolo, E. Ahmetcik, M. Scheffler, L.M. Ghiringhelli, SISSO: a compressed-sensing method for identifying the best low-dimensional descriptor in an immensity of offered candidates, *Phys. Rev. Mater.* 2 (8) (2018) 083802.
- [18] M. Andersen, S.V. Levchenko, M. Scheffler, K. Reuter, Beyond Scaling Relations for the Description of Catalytic Materials, *ACS Catal.* 9 (4) (2019) 2752–2759.
- [19] C.J. Bartel, et al., Physical descriptor for the Gibbs energy of inorganic crystalline solids and temperature-dependent materials chemistry, *Nat. Commun.* 9 (1) (2018) 4168.
- [20] Acosta, Carlos Mera, et al. "Analysis of topological transitions in two-dimensional materials by compressed sensing." *ArXiv:1805.10950 Cond-Mat*, May 2018.
- [21] C. J. Bartel et al., "New tolerance factor to predict the stability of perovskite oxides and halides," *Sci. Adv.*, vol. 5, no. 2, p. eaav0693, Feb. 2019.
- [22] V.M. Goldschmidt, Die Gesetze der Krystallochemie, *Naturwissenschaften* 14 (21) (May 1926) 477–485.
- [23] R.E. Cohen, Origin of ferroelectricity in perovskite oxides, *Nature* 358 (6382) (1992) 136–138.
- [24] J.G. Bednorz, K.A. Müller, Possible high T_c superconductivity in the Ba–La–Cu–O system, *Z. Für Phys. B Condens. Matter* 64 (2) (Jun. 1986) 189–193.
- [25] J.C. Loudon, N.D. Mathur, P.A. Midgley, Charge-ordered ferromagnetic phase in La_{0.5}Ca_{0.5}MnO₃, *Nature* 420 (6917) (2002) 797.
- [26] J.-H. Park, E. Vescovo, H.-J. Kim, C. Kwon, R. Ramesh, T. Venkatesan, Direct evidence for a half-metallic ferromagnet, *Nature* 392 (6678) (1998) 794.
- [27] J. Hwang, R.R. Rao, L. Giordano, Y. Katayama, Y. Yu, Y. Shao-Horn, Perovskites in catalysis and electrocatalysis, *Science* 358 (6364) (2017) 751–756.
- [28] D.Y. Suárez, I.M. Reaney, W.E. Lee, Relation between tolerance factor and T_c in Aurivillius compounds, *J. Mater. Res.* 16 (11) (2001) 3139–3149.
- [29] I.M. Reaney, E.L. Colla, N. Setter, Dielectric and structural characteristics of Ba- and Sr-based complex perovskites as a function of tolerance factor, *Jpn. J. Appl. Phys.* 33 (7R) (1994) 3984.
- [30] W. Travis, E.N.K. Glover, H. Bronstein, D.O. Scanlon, R.G. Palgrave, On the application of the tolerance factor to inorganic and hybrid halide perovskites: a revised system, *Chem. Sci.* 7 (7) (2016) 4548–4556.
- [31] A.M. Glazer, Simple ways of determining perovskite structures, *Acta Crystallogr. Sect. A* 31 (6) (1975) 756–762.
- [32] C.J. Howard, H.T. Stokes, Group-theoretical analysis of octahedral tilting in perovskites, *Acta Crystallogr. B* 54 (6) (1998) 782–789.
- [33] A.M. Glazer, The classification of tilted octahedra in perovskites, *Acta Crystallogr. B* 28 (11) (1972) 3384–3392.
- [34] E.L. Colla, I.M. Reaney, N. Setter, DiP256: The temperature coefficient of the relative permittivity of complex perovskites and its relation to structural transformations, *Ferroelectrics* 133 (1) (1992) 217–222.
- [35] E.L. Colla, I.M. Reaney, N. Setter, Effect of structural changes in complex perovskites on the temperature coefficient of the relative permittivity, *J. Appl. Phys.* 74 (5) (1993) 3414–3425.
- [36] E. Bousquet, et al., Improper ferroelectricity in perovskite oxide artificial superlattices, *Nature* 452 (7188) (2008) 732–736.
- [37] E. Dagotto, Nanoscale Phase Separation and Colossal Magnetoresistance: The Physics of Manganites and Related Compounds, Springer-Verlag, Berlin Heidelberg, 2003.
- [38] P.M. Woodward, Octahedral Tilting in Perovskites. II. Structure Stabilizing Forces, *Acta Crystallogr. B* 53 (1) (1997) 44–66.
- [39] I.D. Brown, Bond valences—a simple structural model for inorganic chemistry, *Chem. Soc. Rev.* 7 (3) (1978) 359–376.
- [40] A. Salinas-Sanchez, J.L. Garcia-Muñoz, J. Rodriguez-Carvajal, R. Saez-Puche, J.L. Martinez, Structural characterization of R₂BaCuO₅ (R = Y, Lu, Yb, Tm, Er, Ho, Dy, Gd, Eu and Sm) oxides by X-ray and neutron diffraction, *J. Solid State Chem.* 100 (2) (1992) 201–211.
- [41] M.W. Lufaso, P.M. Woodward, Prediction of the crystal structures of perovskites using the software program SPuDS, *Acta Crystallogr. B* 57 (6) (2001) 725–738.
- [42] G.H. Rao, K. Bäcker, I.D. Brown, Bond-valence analysis on the structural effects in magnetoresistive manganese perovskites, *J. Phys. Condens. Matter* 10 (48) (1998) L757–L763.
- [43] N. Kallel, M. Hazzaz, N. Ihzaz, Crystal structure, magnetic and electrical properties of half-doped chromium manganite La_{0.5}Sr_{0.5}Mn_{0.5}Cr_{0.5}O₃, *J. Supercond. Nov. Magn.* 32 (8) (2019) 2623–2631.
- [44] J.B. Philipp, et al., Structural and doping effects in the half-metallic double perovskite A₂CrW₆ (A = Sr, Ba, and Ca), *Phys. Rev. B* 68 (14) (2003) 144431.

- [45] M.W. Lufaso, P.M. Woodward, J. Goldberger, Crystal structures of disordered $A_2Mn_3 + M_5 + O_6$ ($A = Sr, Ca$; $M = Sb, Nb, Ru$) perovskites, *J. Solid State Chem.* 177 (4) (2004) 1651–1659.
- [46] H.J. Xiang, M. Guennou, J. Íñiguez, J. Kreisel, L. Bellaiche, Rules and mechanisms governing octahedral tilts in perovskites under pressure, *Phys. Rev. B* 96 (5) (2017) 054102.
- [47] P. Villars, K. Cenzual, NaSbO₃ Crystal Structure: Datasheet from “PAULING FILE Multinaries Edition – 2012” in SpringerMaterials (https://materials.springer.com/isp/crystallographic/docs/sd_1407944). Springer-Verlag Berlin Heidelberg & Material Phases Data System (MPDS), Switzerland & National Institute for Materials Science (NIMS), Japan.
- [48] H.J. Lee, H.M. Park, Y.W. Song, Y.K. Cho, S. Nahm, J.-D. Byun, Microstructure and dielectric properties of barium strontium magnesium niobate, *J. Am. Ceram. Soc.* 84 (9) (2001) 2105–2110.
- [49] W.T. Fu, D.J.W. IJdo, The structure of CeAlO₃ by Rietveld refinement of X-ray powder diffraction data, *J. Solid State Chem.* 177 (9) (2004) 2973–2976.
- [50] G.L. Murphy, et al., Structure and phase transition in BaThO₃: A combined neutron and synchrotron X-ray diffraction study, *J. Alloys Compd.* 727 (Dec. 2017) 1044–1049.
- [51] M. Ahtee, C.N.W. Darlington, Structures of NaTaO₃ by neutron powder diffraction, *Acta Crystallogr. B* 36 (5) (1980) 1007–1014.
- [52] S. Zhang, Y. Hu, L. Chen, G. Ju, Z. Wang, J. Lin, Luminescent properties of a green emitting persistent phosphor CdGeO₃:Tb³⁺, *Opt. Mater.* 47 (2015) 203–210.
- [53] E. Ruiz-Trejo, M.S. Islam, J.A. Kilner, Atomistic simulation of defects and ion migration in LaYO₃, *Solid State Ion.* 123 (1) (1999) 121–129.
- [54] A.M. Glazer, S.A. Mabud, Powder profile refinement of lead zirconate titanate at several temperatures. II. Pure PbTiO₃, *Acta Crystallogr. B* 34 (4) (1978) 1065–1070.
- [55] E.A. Zhurova, Y. Ivanov, V. Zavodnik, V. Tsirelson, Electron density and atomic displacements in KTaO₃, *Acta Crystallogr. B* 56 (4) (2000) 594–600.
- [56] R.H. Mitchell, M.D. Welch, A.R. Chakhmouradian, Nomenclature of the perovskite supergroup: A hierarchical system of classification based on crystal structure and composition, *Mineral. Mag.* 81 (3) (2017) 411–461.
- [57] R.D. Shannon, Revised effective ionic radii and systematic studies of interatomic distances in halides and chalcogenides, *Acta Crystallogr. A* 32 (5) (1976) 751–767.
- [58] S. Lu, Q. Zhou, Y. Ouyang, Y. Guo, Q. Li, J. Wang, Accelerated discovery of stable lead-free hybrid organic-inorganic perovskites via machine learning, *Nat. Commun.* 9 (1) (2018) 1–8.
- [59] F. Pedregosa et al., “Scikit-learn: Machine Learning in Python,” *Mach. Learn. PYTHON*, pp. 6.
- [60] C. Cortes, V. Vapnik, Support-vector networks, *Mach. Learn.* 20 (3) (1995) 273–297.
- [61] C.-C. Chang, C.-J. Lin, LIBSVM: A library for support vector machines, *ACM Trans. Intell. Syst. Technol.* 2 (3) (2011) 1–27.
- [62] J. R. Quinlan, C4.5: Programs for Machine Learning, 1 edition. San Mateo, Calif: Morgan Kaufmann, 1992.
- [63] I. Barandiaran, The random subspace method for constructing decision forests, *IEEE Trans Pattern Anal Mach Intell* 20 (8) (1998) 1–22.
- [64] L. Breiman, Random Forests, *Mach. Learn.* 45 (1) (2001) 5–32.
- [65] D.E. Rumelhart, G.E. Hinton, R.J. Williams, Learning representations by back-propagating errors, *Cogn. Model.* 5 (3) (1988) 1.
- [66] R. H. R. Hahnloser, R. Sarpeshkar, M. A. Mahowald, R. J. Douglas, and H. S. Seung, “Digital selection and analogue amplification coexist in a cortex-inspired silicon circuit,” vol. 405, 2000, pp. 5.
- [67] T.-F. Wu, C.-J. Lin, R. C. Weng, Probability Estimates for Multi-class Classification by Pairwise Coupling, pp. 31.
- [68] D. P. Kingma, J. Ba, “Adam: A Method for Stochastic Optimization,” *ArXiv:1412.6980 Cs*, 2014.