ACS NANO
Cite This: ACS Nano 2019, 13, 3031−3041
www.acsnano.org

ARTICLE

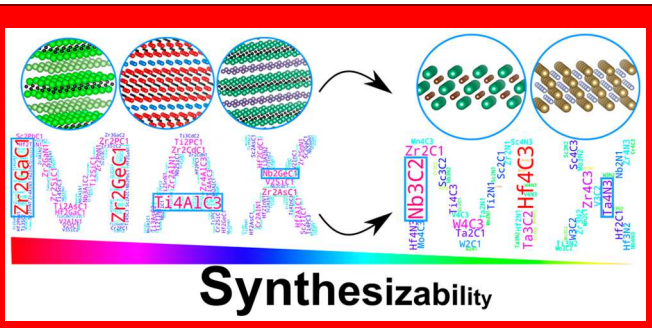
Prediction of Synthesis of 2D Metal Carbides and Nitrides (MXenes) and Their Precursors with Positive and Unlabeled Machine Learning

Nathan C. Frey,[†] Jin Wang,[†] Gabriel Iván Vega Bellido,[†,‡] Babak Anasori,[§] Yury Gogotsi,[§] and Vivek B. Shenoy*[,†]

[†]Department of Materials Science and Engineering, University of Pennsylvania, Philadelphia, Pennsylvania 19104, United States
[‡]Department of Chemical Engineering, University of Puerto Rico at Mayagüez, Mayagüez 00681, Puerto Rico
[§]Department of Materials Science and Engineering and A.J. Drexel Nanomaterials Institute, Drexel University, Philadelphia, Pennsylvania 19104, United States

Supporting Information

Growing interest in the potential applications of two-dimensional (2D) materials has fueled advancement in the identification of 2D systems with exotic properties. Increasingly, the bottleneck in this field is the synthesis of these materials. Although theoretical calculations have predicted a myriad of promising 2D materials, only a few dozen have been experimentally realized since the initial discovery of graphene. Here, we adapt the state-of-the-art positive and unlabeled (PU) machine learning framework to predict which theoretically proposed 2D materials have the highest likelihood of being successfully synthesized. Using elemental information and data from high-throughput density functional theory calculations, we apply the PU learning method to the MXene family of 2D transition metal carbides, carbonitrides, and nitrides, and their layered precursor MAX phases, and identify 18 MXene compounds that are highly promising candidates for synthesis. By considering both the MXenes and their precursors, we further propose 20 synthesizable MAX phases that can be chemically exfoliated to produce MXenes.



machine learning, semisupervised learning, 2D materials, materials synthesis, MXene, DFT

The explosion of progress in high-throughput computational screening of materials has led to accelerated identification of many promising candidates for energy storage,[1] electrocatalysis,[2] photovoltaic absorbers,[3] and a staggering variety of other applications.[4] There is a growing availability of experimental and theoretical data on 3D crystals in popular databases,[5,6] and recent efforts have dramatically increased the collection of proposed 2D materials and their predicted properties.[7−10] While high-throughput density functional theory (DFT) studies are a powerful method for targeted materials design and characterization without requiring costly and time-consuming synthesis,[4,11−13] an imposing challenge remains: using computation to accelerate the synthesis of materials. Machine learning (ML) has emerged as a promising way forward in this respect, as shown in studies on predicting thermodynamic stability of arbitrary compositions[14] and reaction successes in inorganic−organic hybrid material synthesis,[15] identifying trends in synthesis conditions for metal oxides,[16] and searching for high-temperature ferroelectric perovskites.[17] However, ML has not yet been exploited to provide insights into the synthesis of 2D materials.

To make actionable predictions about the synthesis of possible 2D materials, we first identify a family of 2D materials that is well-suited for analysis *via* ML, namely, a family comprising a large chemical search space, with examples of successful synthesis. The 2D transition metal carbides, carbonitrides, and nitrides (MXenes)[18,19] with the general formula $M_{n+1}X_nT_x$ (n = 1−3) and their parent MAX phases (layers of MXenes interleaved with A-element atoms)[20,21] are an ideal choice to satisfy these constraints. The large variety of chemical compositions, number of layers, pure (single

Received: October 19, 2018
Accepted: March 4, 2019
Published: March 4, 2019

DOI: 10.1021/acsnano.8b08014
ACS Nano 2019, 13, 3031−3041

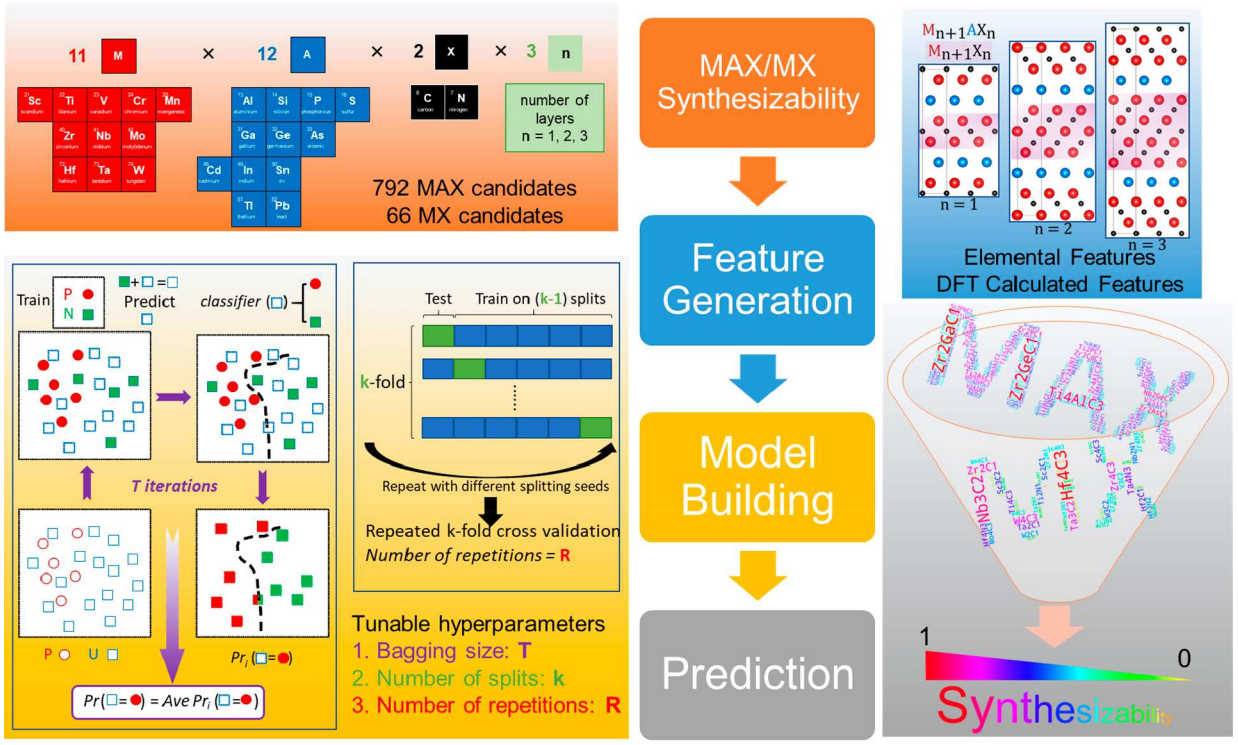


Figure 1. Schematic of the chemical search space and computational workflow including PU learning. The MAX/MXene search space (upper left) is determined by all possible combinations of M, A, and X atoms and n = 1, 2, or 3. Structural models of bulk MAX phase (upper right) and the corresponding MXene highlighted in pink. Schematic of the PU learning process (lower left) and model parameters. Word clouds (lower right) where large (small) font size and red (orange) color correspond to high (low) predicted synthesizability of the named compound.

transition element, M, atom), solid solution, and ordered phases (more than one M element) define a vast materials search space. More than 20 of the many theoretically predicted MXenes have been synthesized,[19] providing successful examples to consider. Perhaps most importantly, exciting experimental results in electrochemical applications[19,22] and predictions of attractive electronic,[23] optical,[24,25] and magnetic[26−29] properties for theoretically proposed MXenes invigorate interest in expanding the family of synthesized MXenes.

In this study, we make use of state-of-the-art positive and unlabeled (PU) ML[30−33] to quantify the degree of "synthesizability" of theoretically predicted MAX and MXene compounds. We consider the 66 single M atom MXenes, but each of these is paired with 12 potential MAX precursors, yielding a total of nearly 800 potential pathways to achieve synthesis of MXenes. Of the compounds that have not yet been successfully synthesized, we predict 111 MAX phases and 18 MXenes with a high probability of synthesis success. We also identify 20 MAX/MXene pairs with high combined synthesizability. Elemental information as well as structural, thermodynamic, and electronic structure data from DFT calculations are considered as possible inputs to our models. With these simple inputs as potential features to characterize candidate materials, we first applied physical and chemical intuition to identify the most relevant features for predicting synthesis success. Then statistical learning in the form of principal component analysis (PCA) was performed to isolate the most important model inputs, *e.g.*, lattice parameters, formation energies, and atomic Bader charges. The key features were then used to train clustering and classification PU learning models. The models were *k*-fold cross-validated with their performance measured by their true positive rates (TPR) tested against the 63 (10) experimentally synthesized MAX (single M MXene) compounds (Table S1). It is important to note that this study relies on small positive sample sizes, so we have included a variety of external methods to validate the model predictions, including phase stability tests and comparisons to experiment. Our proposed computational framework can be applied to potentially mechanically exfoliable 2D materials[7] to provide quantitative insights beyond DFT-calculated exfoliation energies, or more generally to the liquid exfoliation of any layered materials.[34] We expect the predictions and approach presented in this paper will guide future experimental efforts and help to bridge the most crucial remaining gap between theoretical and experimental materials science: materials synthesis.

RESULTS

First, we will describe the workflow used here and the details of the PU learning process as applied to the materials synthesis problem. The computational framework and the PU learning algorithm are represented schematically in Figure 1. PU learning is particularly well suited to the problem of providing insight into 2D materials synthesis, as it is a so-called "semisupervised" method relying only on positive data (experimentally synthesized materials), while all yet-to-be synthesized materials are "unlabeled". We start by specifying a chemical search space defined by 11 transition metal M atoms, 12 A group elements, carbon or nitrogen X atoms, and n = 1, 2, or 3 layers of X with n + 1 layers of M. This yields 792 potential single M MAX phase candidates and, after removing

DOI: 10.1021/acsnano.8b08014
ACS Nano 2019, 13, 3031−3041

the A group elements, 66 single M MXene candidates. While there are examples of successfully synthesized solid solutions and ordered double transition metal structures, here we restrict the search to single M systems to reduce computational complexity. The addition of solid solutions and ordered double M element systems would add only 9 positive samples,[35] while vastly increasing the number of unlabeled samples, causing an imbalanced data set and unstable PU learning models that overfit to the limited positive samples. To characterize each material candidate, we consider an exhaustive set of structural, thermodynamic, electronic structure, and elemental data. Elemental features such as atomic masses, electronegativities, chemical potentials, and atom-in-a-box energies are easily obtained for constituent atoms. DFT calculations were used to relax structures and obtain quantities such as interlayer distances, bond lengths, formation energies, and atomic Bader charges. In this way, over 80 features are generated to describe each material system. The relative importance of these features and the feature engineering process that was used to select the key features for model building are discussed in detail below.

The framework of PU learning has been adopted and modified for use in areas using real world (not simulated) data, such as drug discovery,[36−38] text classification,[39] and time series data classification,[40] with the common thread being a data imbalance given a smaller number of positive samples and a relative abundance of unlabeled samples. In protein interaction networks, the positive sample size is as low as 25,[41] and in gene identification studies the positive sample sizes are as small as 11 samples[41] or 20 samples.[42] These successes motivate our approach to apply PU learning to the materials synthesis problem, where positively labeled data are scarce, negative samples are not available, and high-throughput computational screening provides a rich variety of unlabeled data. The approach to PU learning applied here is based on imposing penalties for misclassifying data during the learning process. This approach comes in a variety of flavors including class-weighted learning,[30] bagging support vector machine (SVM),[31] and unbiased learning with risk estimators to optimize the strategy for weighting unlabeled samples.[33] In this study, we are particularly constrained by the dearth of positive samples available, so we implement a variation of transductive bagging SVM, which has demonstrated advantages over other PU learning schemes when the size of positive samples is much smaller than that of unlabeled ones.[31] We modify the bagging scheme to use a decision tree as the base classifier, rather than the standard SVM classifier, because decision trees implicitly perform feature selection and generate feature importance metrics during model training, giving explicit physical and chemical insights. To show the robustness of the PU learning approach for the materials synthesis problem, we also implemented an alternative algorithm for PU learning and show that similar model performance was achieved (with both models outperforming *k*-means clustering), meaning that the results are not dependent on a specific algorithm or implementation.

With this scheme in place, we built and trained the ML model. The procedure is illustrated in Figure 1. The sample space is represented abstractly by randomly distributed positive data (red circles, synthesizable) and unlabeled data (blue squares). In each iteration of the PU learning process, some of the unlabeled samples are randomly labeled negative (green squares, not synthesizable). A decision tree base classifier is constructed based on these data to classify the remaining unlabeled samples as either positive or negative. This procedure is repeated T times, where T is the number of iterations (bagging size), and a new random selection of data is labeled negative in each iteration. We define the "synthesizability score" of a given unlabeled compound as the average of the predictive score from all the decision tree classifiers that do not contain that sample. A score greater than 0.5 corresponds to a positive prediction, while a compound scoring less than 0.5 is labeled negative. To improve the stability and accuracy of the model, a repeated *k*-fold cross-validation is performed with different splits of samples in each repetition. The model training, composed of T iterations, is repeated R times with the data split into k number of folds. We define the TPR, that is, the percentage of correctly classified positive samples, as the validation metric

$$\mathrm{TPR} = \frac{1}{R}\sum_{r=1}^{R}\frac{1}{K}\sum_{k=1}^{K}\mathrm{TPR}_{k,r} \tag{1}$$

where K is the number of splits, R is the number of repetitions, and the TPR is first generated from the prediction of each out-of-fold validation sample and then averaged across the *k*-folds. The final TPR is then averaged across all the repetitions. Further details of the general PU learning method are presented in the Methods section.

Feature Engineering. As mentioned above, one advantage of implementing a decision tree as the base classifier is obtaining quantitative insight into the importance of features in the PU learning model. Additionally, we are not restricted in the number of features the model can consider as long as overfitting is avoided *via* the *k*-fold cross-validation detailed above. We begin the discussion of feature engineering by briefly noting some of the salient details of MXene synthesis, as they relate to the choice of relevant properties to compute or collect for model input. The M−A bond in layered MAX precursor phases is metallic, which excludes the possibility of making MXenes by mechanical shearing of their parent phases.[19] Instead, MXene synthesis exploits the chemical activity of the M−A bonds compared to the stronger, more inert M−X bonds. Selective etching of the A-elements can be achieved using acidic solutions or using an electrochemical approach in basic solutions.[42] However, with the exception of Ti_3SiC_2,[43] only Al-containing MAX phases have been successfully etched to synthesize MXenes.[44] Experiments have also shown that MAX phases with larger n and heavier M atoms tend to require longer etching times and stronger solutions, which has been related to the larger number of M valence electrons.[44]

With this synthesis procedure and these empirical findings in mind, we identify a few classes of features that may be particularly relevant to this synthesis problem. Structurally, the interlayer (out-of-plane) distances between M atoms, M and X, M and A atoms, *etc.*, as well as nearest-neighbor bond lengths, quantify the strength of the relevant bonds. Thermodynamic data including the total, formation, cohesive, and per-atom energies all give at least a simple picture of a compound's stability. For the reasons mentioned above, the mass and n descriptors are heavily weighted in the MAX model to represent the relative difficulty (ease) of synthesizing MAX phases with more (fewer) layers. And finally, the per-atom Bader charges provide electronic structure information related to charge transfer and the character of bonds. The feature

DOI: 10.1021/acsnano.8b08014
ACS Nano 2019, 13, 3031−3041

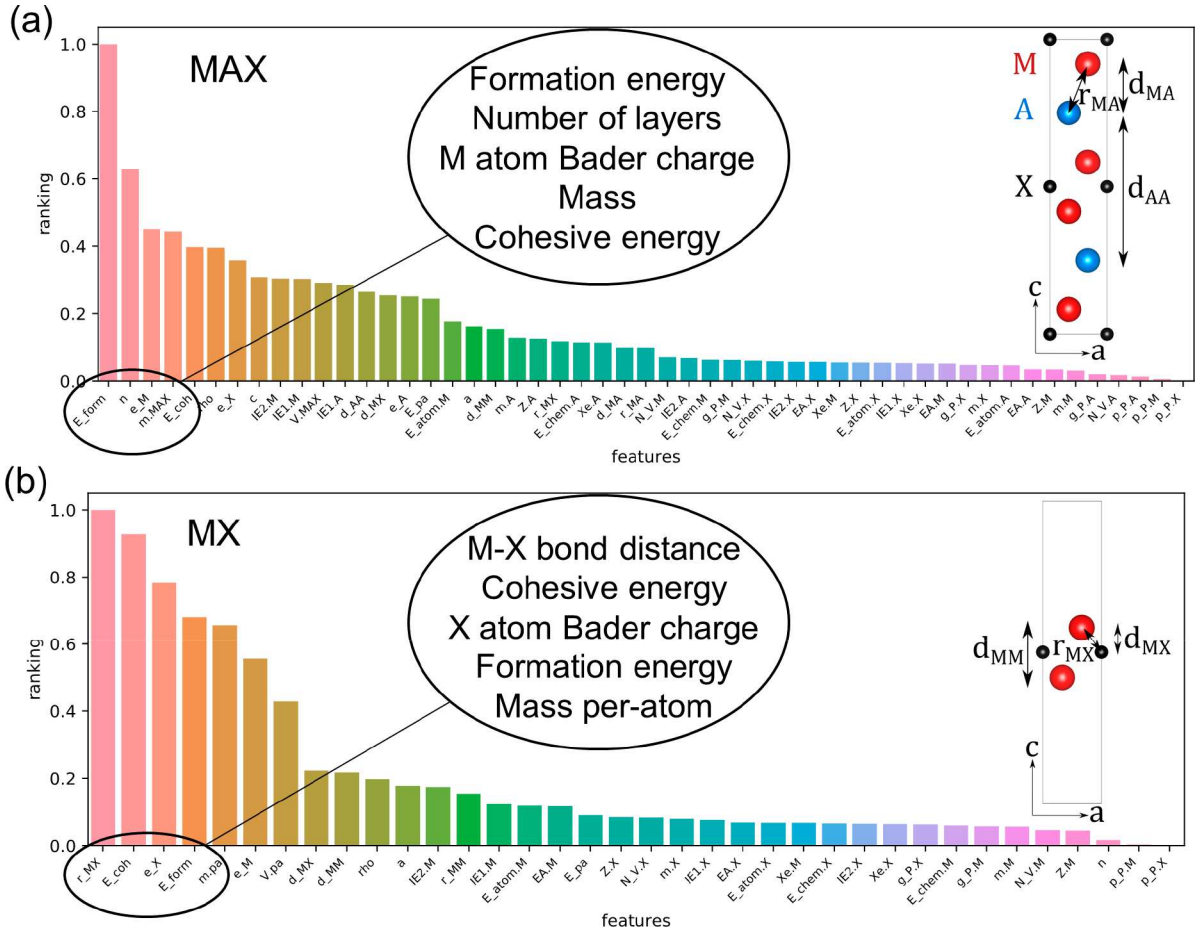


Figure 2. Feature importance ranking for (a) MAX and (b) MXene models. Layer distances and bond lengths are labeled in the unit cell diagrams. See Table 1 for definitions of all feature labels.

importance ranking of the top 50 features for the best-performing MAX and MXene models is depicted in Figure 2. Definitions of these features are given in Table 1. Feature importance reveals how much weight the model assigns to a particular descriptor when making predictions, thereby giving direct chemical insight into which features are most relevant for predicting synthesizability.

For the MAX phases, we find that the top five most important features are the formation energy, number of layers, M atom Bader charge, system mass, and cohesive energy. The X atom Bader charge, M atom ionization potentials, and *c* lattice constant are also in the top 10 most important features. The high rankings of these features reveal that the ML model is relying heavily on the thermodynamic stability and data related to the M−X bond to make its classifications. This is in line with our chemical intuition and shows that we can safely neglect features other than the 50 shown in Figure 2a in the classification. Moreover, these quantitative data bolster our chemical intuition when choosing a more limited feature set for use in alternative PU schemes, in which base classifiers other than a decision tree are used. As another visual representation of feature importance, we plot the feature correlation matrix for the MAX model in Figure S1a.

The feature importance plot for the MXene model (Figure 2b) is consistent with previous empirical observations based on successful MXene syntheses. Among the most important features we find the M−X bond length, cohesive and formation energies, per-atom mass, and the Bader charges for the M and

Table 1. Detailed Description of Computed and Elemental Features Used in the PU Learning Model

DFT-calculated features		elemental features	
n	number of layers	Z	atomic number
a	in-plane lattice constant	N_V	number of valence electrons
c	out-of-plane lattice constant	g_P	group number
d_MX	layer distances	p_P	period number
d_MM		Xe	electronegativity
d_MA		EA	electron affinity
d_AA		IE1	1st ionization potential
r_MX	bond lengths	IE2	2nd ionization potential
r_MA		E_chem	chemical potential
m	mass	E_atom	atom-in-box potential
V	volume		
rho	density		
Energy	total energy		
E_pa	energy per atom		
E_form	formation energy		
E_coh	cohesive energy		
e_M	Bader charges on M, A, X atoms, respectively		
e_A			
e_X			

DOI: 10.1021/acsnano.8b08014
ACS Nano 2019, 13, 3031−3041

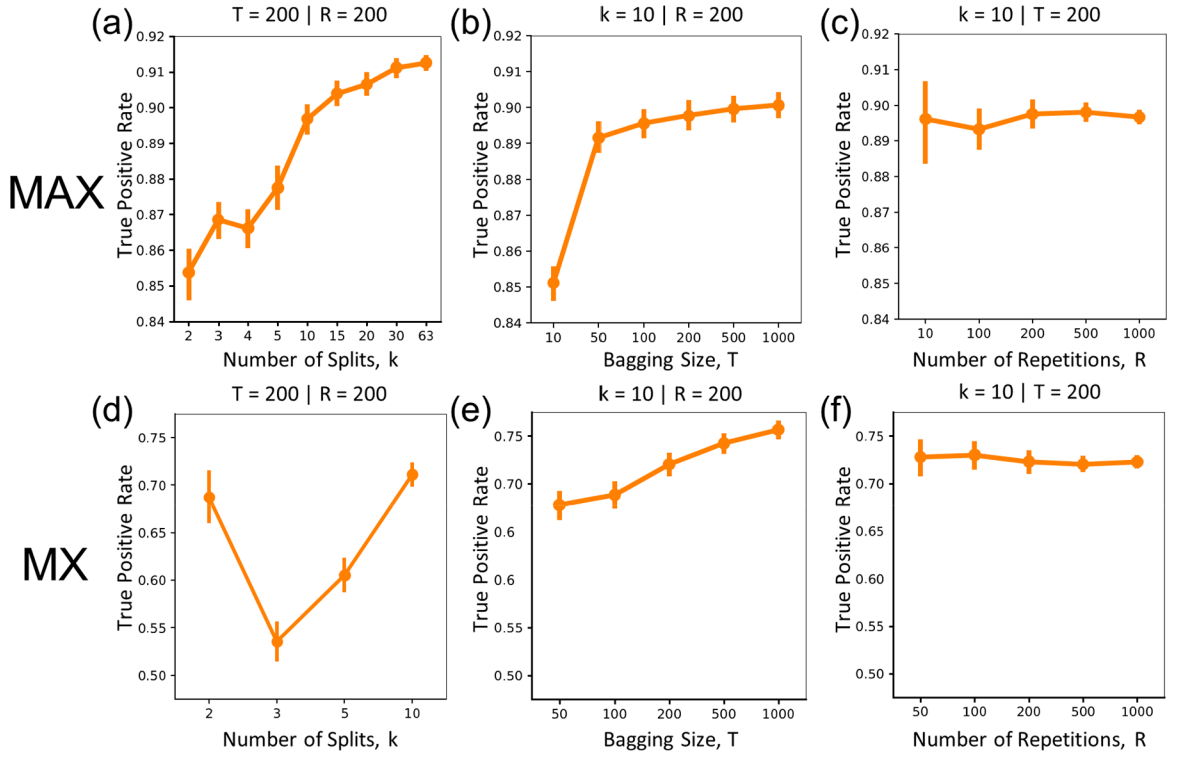


Figure 3. Model evaluation for per-atom normalized features. Model performance measured by true positive rate is tested as a function of number of splits (k), bagging size (T), and number of repetitions (R).

X atoms. These features encode information about the strength of the M−X bond and the overall thermodynamic stability. Interestingly, the simple metric of number of M atom valence electrons is not important, although previous experiments have suggested this feature is strongly tied to synthesis success.[44] Instead, the atomic Bader charges are among the most important features, as these calculated quantities give a more realistic estimate of the distribution of charge in the system. The MXene feature correlation matrix is shown in Figure S1b.

Model Building. With reasonable features generated, next we turn to the actual construction of the PU learning model. This process involves three interrelated steps: (1) testing possible feature sets, (2) tuning hyperparameters, and (3) evaluating model performance. To construct a robust model for synthesizability predictions, extensive and careful testing over all three steps must be performed. As discussed above, using the decision tree as a base classifier allows for an unrestricted feature space dimensionality, but there may be redundancies in descriptors or numerical differences based on weighting/normalization schemes, such that it is still important to construct multiple feature sets and test their performance. Of the many tested feature sets, the best performance was achieved after normalizing descriptors by the number of atoms in the system. This simple reweighting can dramatically increase the model performance. The model performance as measured by the TPR is uniformly better after normalization, over all tested hyperparameter values. The per-atom normalization reduces bias based on system size, helping to isolate the real chemical and structural differences underlying successful synthesis, thereby improving the TPR. For comparison, model training performance for a feature set without normalization is shown in Figure S2.

The hyperparameters k, T, and R were tested for convergence for both MAX and MXene PU models by varying one parameter and fixing the other two at a large value. This testing is summarized in Figure 3, where performance is shown in orange. We limit the discussion of model performance to the results for the per-atom normalized feature set, as it outperforms other feature sets in all respects. The number of splits, k, can range from 2 to the number of positively labeled samples. For the MAX model, a TPR of nearly 0.90 is achieved for $k = 10$ (Figure 3a), with marginal increases of ∼1% beyond that point. Cross-validation poses a unique challenge for the MXene data set, where there are only 10 positive samples. The optimal performance was reached using a "leave-one-out cross-validation" scheme (Figure 3d), where k equals the number of positive samples, and within each k-fold the model is trained on all but one sample, which is then used as the test set. The bagging size, T, determines the number of base models generated for the ensemble averaging, and a sharp increase in TPR was observed between $T = 10$ and $T = 50$ for the MAX model (Figure 3b). For $T > 50$, the TPR increase is again ∼1%. The MXene model TPR showed a stronger dependence on T (Figure 3e). An increase in the TPR of over 5% was seen between $T = 50$ and $T = 1000$. Finally, model performance with respect to number of repetitions, R, was tested. While increasing k and T incurs only a small cost to the runtime of model training, R controls the number of iterations of the entire model training procedure, and thus it is desirable to find a lower bound for convergence with respect to R. Fortunately, the TPR is stable over a large range of R (up to $R = 1000$) for both MAX and MXene models.

As a further check against bias in the model, we repeated this entire training and evaluation procedure on the MAX and MXene data sets for two other ML algorithms: k-means clustering and robust ensemble SVM (RESVM).[32] k-means clustering is a conceptually simple method in which samples are partitioned into one of k clusters based on the distance

DOI: 10.1021/acsnano.8b08014
ACS Nano 2019, 13, 3031−3041

between the vector of features that describes the sample and the mean of each cluster. In our case, we know *a priori* that there are only two possible clusters: predicted positive and predicted negative. The *k*-means clustering attempts to sort all the samples into one of these clusters based on its feature vector's proximity to the means of the clusters. The transductive bagging PU learning approach assigns a numeric score between 0 and 1 to each sample that can be visualized in 3D as a function of three features to graphically depict "clusters" of similar samples (Figure S3a), but a similar plot using *k*-means clustering generates a much more easily interpreted representation. Plotting the clusters against formation energy, M−A layer distance, and M−X bond length, *k*-means clustering reveals a clear division in this feature space between predicted positive and predicted negative samples (Figure S3b). Most of the true positives clearly sit in the predicted positive cluster, while most of the false negatives overlap with the predicted negative cluster, revealing that the model has misclassified these samples due to their distance in feature space from the predicted positive cluster. RESVM is yet another ML scheme that can be applied to PU learning problems (details of the algorithm are provided in the Methods section). This approach has not been previously applied to materials science problems, but we have adapted the algorithm to our synthesis problem and used it to generate synthesizability scores (Figure S3c) in the same manner as in PU learning with a decision tree base classifier. Unlike the decision-tree-based model, using SVM as the base classifier imposes limits on the feature space dimensionality, so we must consider around 10 features, rather than the 50 used in the decision tree method. However, similar model performance was achieved with the two PU learning methods (and both outperformed simple *k*-means clustering), so we conclude that the PU learning approach is a robust method for tackling the synthesis problem, and the results are not algorithm-specific.

After exhaustive hyperparameter testing over multiple feature sets, optimal model performances of TPR > 90% and TPR > 75% were found for $(k, T, R) = (10, 500, 200)$ and $(k, T, R) = (10, 1000, 200)$ for MAX and MXene models, respectively. Perhaps unsurprisingly, while the model performance for the MAX phases is quite good, the performance for MXenes is less so. To provide clarity on the predictive power of these models, we first analyze the model predictions and provide a breakdown by chemical composition, number of layers, and the most relevant model descriptors. We then combine the insights from this analysis with a holistic picture of MAX/MXene precursor/product pairs to highlight the most promising synthesis pathways.

Model Predictions. From the 792 MAX phases, accounting for the 63 true positives, there are 729 unlabeled samples. The PU learning model predicts that 111 of these unlabeled samples are positive, *i.e.*, promising candidates for synthesis. The 111 predicted positive compounds are listed in Table S2. This result provides significant guidance in future MAX synthesis, reducing the possible space of unlabeled samples to 17% of its original size, while still providing a large number and chemical variety of proposed candidates. Ten of the 11 possible M species (all except W) are represented, as well as all 12 possible A species. Twenty-nine of the predicted positives are nitrides, and there are $n = 1$, 2, and 3 phases present. Surprisingly, four Mn-based compounds are predicted positive (*e.g.*, Mn_2AlC and Mn_2AlN), although there are no examples of successfully synthesized Mn-based MAX phases. These compounds could be promising precursors for MXenes with exotic properties for which no parent phases currently exist.

To further evaluate our model's predictive power, we considered the thermodynamic and elastic stability of the predicted positive MAX phases and their stability compared to their competing binary phases; 87% of the predicted positives were classified "viable crystals", satisfying both the Cauchy−Born elastic stability criteria and having negative heats of formation compared to the MAX phase constituent elements in their most stable unary forms.[45] Using the MaterialsWeb database,[46] which reports phase stability of M_2AX compounds, we then applied a further test by looking at the phase stability of the predicted positives against their competing phases. Of the predicted positives contained in the database, 89% are stable or metastable (energy difference per atom between the compound and their decompositional products less than 100 meV/atom),[47,48] and 11% are unstable. The phase stable and metastable compounds are listed in Table S3. It should also be noted that reported phonon calculations on MAX phases found no negative frequencies, indicating dynamic stability.[49]

The MXene model yielded 18 predicted positive compounds out of the 56 unlabeled samples. These systems are listed in Table 2. The predicted positive compounds offer a wide chemical variety while providing important guidance to future synthesis efforts, reducing the space of 56 unlabeled candidates to nearly a third of that size. We again consider the stability of the predicted positives as a check on the model predictions. Fourteen of the 18 compounds have formation energies below 200 meV/atom and so are considered stable under the proposed threshold for 2D material stability.[50] The four unstable systems (W_4C_3, Ta_2C, W_3C_2, and Mo_4C_3) may be stabilized by surface functionalization. Sixteen of the 18 have medium or high dynamical stability according to phonon calculations reported in the Computational 2D Materials Database.[8]

Table 2. Predicted Positive MXene Compounds with Synthesizability Score > 0.5

MXene predicted positives		
Hf_4C_3	Ta_4N_3	Sc_3C_2
Nb_3C_2	Ta_2C	Ti_2N
Zr_2C	Hf_4N_3	Sc_2C
Ta_3C_2	Ti_4C_3	W_3C_2
W_4C_3	Hf_2C	Nb_2N
Zr_4C_3	Sc_4C_3	Mo_4C_3

Most interestingly, the model predicts certain Ta, Hf, and Nb nitrides, as well as two Sc-based MXenes (Sc_3C_2 and Sc_2C) and two W-based MXenes (W_4C_3 and W_3C_2), as positive, although no such nitrides or pure phase Sc- or W-based MXenes have been reported. The predicted positive W-based MXenes, for which no existing (or even predicted) MAX phases can act as precursors, may indicate that non-MAX phase precursors, such as the recently reported W-based nanolaminated ternary phase, $(W, Ti)_4C_{4-x}$,[51] are better starting points to successfully synthesize these compounds. Mo_2C, Zr_3C_2, and Hf_3C_2 MXenes have previously been synthesized by such methods.[52−54] The PU learning approach employed here outputs an easily interpreted "synthesis probability" between 0 and 1, where 0 corresponds to a negative (not synthesizable) compound and 1 is a positive

DOI: 10.1021/acsnano.8b08014
ACS Nano 2019, 13, 3031−3041

(synthesized) compound. This allows for a variety of analyses based on comparison to experimental data and the distributions of synthesis probability and features for predicted positive and negative compounds.

The scaled errors of machine learning model predictions decrease as the number of samples increases, so the sample size is an important factor affecting a model's predictive accuracy. Typical examples of machine learning in materials science use between 100 and 10^4 training samples.[55] Additionally, traditional *k*-fold cross-validation is overly optimistic when evaluating materials discovery machine learning models, because materials data sets are rarely uniformly distributed.[56] With these points in mind, it is particularly important and necessary in this study to compare the model predictions to experiments and other external measures of validation, because of the small size of positive samples.

Next, we test our model predictions against experimental observations by comparing the "predicted negative" MXenes (Table S4) to compounds for which we have unsuccessfully attempted synthesis. Many MXenes have not been successfully synthesized mostly due to the lack of available synthesized MAX phase precursors. For example, the MAX phase precursors for Hf_3N_2, Cr_3C_2, Cr_3N_2, and Mo_3N_2 have not been reported in the literature. Even if the MAX phase exists, the attempted wet-chemical etching in hydrofluoric acid (HF)-containing aqueous solutions is not always successful. Although the Cr_2AlC MAX phase was synthesized long before the discovery of MXenes,[57] attempts to make Cr_2C from Cr_2AlC by HF etching have been unsuccessful. Based on our knowledge to date, no selective etching of Al is observed for Cr_2AlC immersion in different concentrations of HF, and the MAX phase powder is dissolved after a few hours.[58] This observation suggests that Cr_2C synthesis, if possible, is not likely to be successful *via* the HF route. Ti_4N_3 was similarly difficult to synthesize (dissolution in HF), and we were able to synthesize it *via* molten salt etching of Ti_4AlN_3.[59] We have tried the molten salt route for Cr_2C synthesis, and it was not successful (not published).

These experimental findings suggest that the MXenes in Table S4 are difficult to synthesize and should have correspondingly low synthesizability scores. Indeed, Cr_2C is predicted negative with a synthesizability score of 0.25. More broadly, the low synthesizability scores of many nitride MXenes agree with our experimental observations of the difficulty in synthesizing such compounds, either because of the lack of MAX phase precursor or dissolution of the MAX powder instead of selective etching in the current MXene synthesis routes. This agreement with experimental data provides added validation of the model predictions.

The violin plots in Figure 4 summarize the predicted synthesis probability distributions as a function of atomic species. Some trends across both MAX and MXene systems are readily apparent. Distributions for carbides (green) are always at least equal or skew to higher values than those for nitrides (blue). This is expected because of all the successfully synthesized MAX phases, only 9 (less than 15%) are nitrides. The situation is even worse for MXenes, of which only one nitride, Ti_4N_3, has been synthesized from a MAX phase precursor.[59] While this fact certainly biases the model predictions toward carbides, it also underscores the importance of identifying nitride compounds with high synthesis probabilities. In the violin plots, the median is represented by the thicker dashed line, and the interquartile is contained by

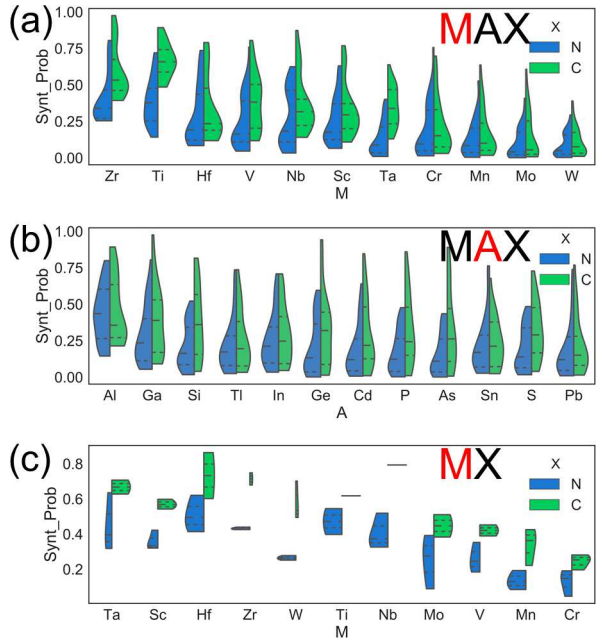


Figure 4. Violin plots of synthesis probabilities as a function of constituent atomic species (X = N shown in blue, X = C shown in green). (a) MAX synthesis probabilities as a function of M atom. (b) MAX synthesis probabilities as a function of A atom. (c) MXene synthesis probabilities as a function of M atom.

the thinner dashed lines. For many species, the interquartile covers a range of low synthesis probability, and a much smaller distribution extends to higher synthesis probabilities. Again, this agrees with intuition from experiment; we expect that most of the not-yet synthesized compounds are difficult to synthesize, and we are most interested in the relatively fewer unidentified systems with a high synthesis probability.

Figure 4a shows the distributions for MAX phases as a function of M atom. There is a clear trend of decreasing synthesizability with increasing group number, where group 4 Ti-, Zr-, and Hf-based MAX phases have high synthesizability, and group 6 Mo- and W-based systems have relatively low synthesis probability. For W in particular, the interquartile is completely below 0.2. This agrees with experimental results, in which the majority of synthesized MAX compounds have M atoms from group 4, and no W-based systems have been synthesized. The A atom dependence (Figure 4b) is less apparent. The distributions for carbides are in general spread out over the whole range of scores. This result is interesting in the sense that, although only Al-based MAX phases have been successfully etched into MXenes,[19] there are a large number of predicted positive compounds containing a different A-group element that may be promising MXene precursors. The previously clear trend for MAX phases disappears entirely for the MXenes. In Figure 4c, there are a surprising number of predicted positive samples from all the early transition metal elements except for V, Mn, and Cr. However, this stark contrast between the 3d block and other transition metal compounds may be an artifact due to the fact that many of the strongly correlated 3d block transition metal MXenes are predicted to be magnetic,[26−28] and only nonmagnetic ground states were considered in this work. The Ti and Nb carbide "violin plots" are in fact simply lines at high synthesis probability, because only one of the layered MXene systems of each of these families has not yet been synthesized.

DOI: 10.1021/acsnano.8b08014
ACS Nano 2019, 13, 3031−3041

Box plots of feature distributions in the MAX model (Figure S4) for predicted negative (score <0.5) and predicted positive (score >0.5) samples further elucidate the relationship between specific descriptors and synthesizability. In general, feature distributions for predicted positive samples are much more tightly distributed than those for predicted negatives. While the total energies calculated from DFT increase linearly with increasing n, the formation energies within one standard deviation of the median of all predicted positives are between 0 and −1 eV, and the cohesive energies are between −5.5 and −7.5 eV, regardless of n. In contrast, formation and cohesive energies for predicted negatives span much larger ranges in energy. Similarly, stark differences between negative and positive samples exist with respect to interlayer distances and bond lengths. Irrespective of n, predicted positives have M−X interlayer separations, M−A interlayer separations, and M−X bond lengths closely distributed around 1.2, 4.5, and 2.2 Å, respectively. With respect to the atomic Bader charges, the predicted positives have median values between 2 and 3 e^- smaller for M atoms and slightly smaller median values for A and X atoms. This finding agrees well with experimental observations, where compounds with heavier M elements and more valence electrons are more difficult to synthesize and require stronger etching conditions.

The combined synthesizability scores for MAX and MXenes offer a strategy for overcoming the scarcity of data in the pure M MXene family of 2D materials. Of the 729 unlabeled MAX phases, 111 were predicted positive by our PU learning model. Taking these as potential precursors for the synthesis of MXenes, we generate 111 (MAX, MXene) pairs to identify the most promising synthesis pathways. The static exfoliation energy has been used to characterize the possibility of chemically etching a bulk MAX phase into 2D MXenes.[49] We compute the etching energy as

$$E_{etch} = \frac{E_{max} - 2E_{MX} - 2\mu_A}{4S} \quad (2)$$

where E_{max} is the total energy of the MAX phase, E_{MX} is the total energy of a MXene unit cell, μ_A is the chemical potential of the MAX phase A atom from its most stable crystal phase, $S = \frac{\sqrt{3}a^2}{2}$ is the surface area, and a is the MAX phase in-plane (a) lattice constant. It is important to note that the surfaces of synthesized MXenes are functionalized by O, F, and OH groups, which affect their formation and etching energies.[50] The etching energy is used as a metric to further discriminate between (MAX, MXene) pairs, especially considering pairs with different precursor phases but the same MXene. These three factors, MAX synthesizability, MXene synthesizability, and etching energy, were used as input to a k-means clustering model to group all possible pairs of predicted positive MAX phases and their MXene partners into positive and negative (MAX, MXene) pairs. The model also considers the differences in bond lengths, interlayer distances, and Bader charges between each MAX parent and its corresponding MXene. In this way, the k-means clustering groups together (MAX, MXene) pairs with similarly favorable features (high combined synthesizability, low etching energy, *etc.*) into the "positive" cluster and similarly unfavorable features (low combined synthesizability, high etching energy, *etc.*) into the "negative" cluster.

The results of k-means clustering are presented in Figure 5, where predicted positive pairs are shown in green and predicted negative pairs are in orange. The etching energy is plotted along the z-axis as an aide for visualizing the clusters. The top 20 pairs with highest combined synthesizability scores are highlighted in dark green. The k-means clustering was repeated with the synthesizability scores from RESVM (Figure S5). Table 3 contains the top 20 pairs with their respective MAX and MXene scores, as well as E_{etch}. Each of these pairs represents a not-yet-synthesized MAX phase and corresponding MXene that can both be synthesized with high probability. This provides further guidance to experiment, reducing the

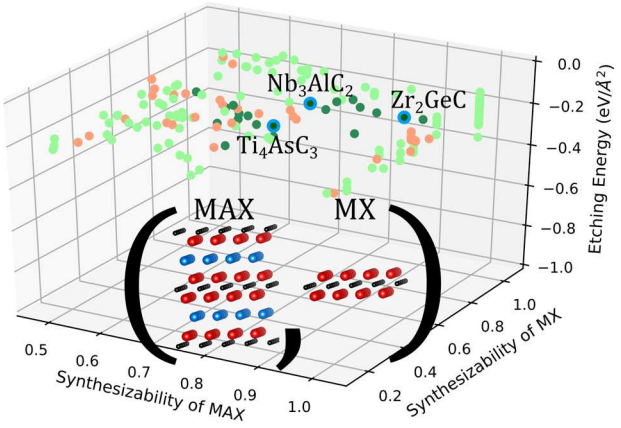


Figure 5. k-means clustering of (MAX, MXene) pair synthesizability as a function of the required chemical etching energy, and individual MAX and MXene synthesizability. Predicted positive samples are shown in light green, top 20 predicted positive samples are in dark green, and predicted negative samples are shown in orange. Zr_2GeC, Ti_4AsC_3, and Nb_3AlC_2 data points (circled in blue) are shown as examples for the top predicted positive samples.

Table 3. Most Preferable MAX Phase Precursors and the Synthesizability of Their Corresponding MXenes, as Well as the Etching Energy (eV/Å²) of the (MAX, MXene) Pair[a]

MAX	synthesizability of MAX	synthesizability of MX	etching energy
Zr_2GaC	0.975	0.748	−0.143
Zr_2GeC*	0.942	0.748	−0.152
Zr_2AsC	0.890	0.748	−0.150
Zr_2PC	0.860	0.748	−0.138
Zr_2CdC	0.846	0.748	−0.108
Zr_2SiC	0.816	0.748	−0.131
Nb_3AlC_2*	0.751	0.793	−0.187
Zr_4AlC_3	0.819	0.680	−0.135
Ti_4AlC_3	0.893	0.618	−0.185
Ti_4AsC_3*	0.742	0.618	−0.225
Ti_4SnC_3	0.678	0.618	−0.189
Ti_4PC_3	0.676	0.618	−0.214
Sc_2PbC	0.768	0.543	−0.219
Ti_4InC_3	0.652	0.618	−0.171
Sc_2TlC	0.737	0.543	−0.178
Ti_4SC_3	0.636	0.618	−0.252
Ti_2SnN	0.720	0.545	−0.146
Zr_4SiC_3	0.574	0.680	−0.158
Ti_2TlN	0.693	0.545	−0.108
Sc_2GaC	0.677	0.543	−0.305

[a]Three representative samples highlighted in Figure 5 are denoted here with an asterisk.

DOI: 10.1021/acsnano.8b08014
ACS Nano 2019, 13, 3031−3041

space from 111 predicted positive MAX phases to the 20 specific phases that can also be chemically etched into MXenes with high probability. There are some notable omissions in this list, which contains only one-third of the predicted positive MXenes. This is because there are no predicted positive MAX precursors for compounds such as W_3C_2. The model predictions indicating these MXenes as promising synthesis candidates means that non-MAX phase precursors are preferable to yield these compounds.

CONCLUSIONS

In this work, PU machine learning was adapted and applied to the problem of bulk and 2D materials synthesis. We chose the family of 2D MXenes and their parent MAX phases as an ideal chemical search space because of the large dimensionality, chemical variety, and numerous examples of successful synthesis. Elemental data and descriptors from DFT calculations were used to characterize each material, and extensive testing was performed to optimize the model hyperparameters and feature set to achieve robust, validated predictions. Two PU learning algorithms were employed, yielding similar results, with the bulk of the discussion devoted to our own implementation of transductive bagging with a decision tree base classifier. This is the most transparent algorithm because it offers a detailed breakdown of feature importance and an easily interpreted synthesizability score between 0 and 1 for each unlabeled sample. Analyzing the trained model revealed that features related to thermodynamic stability, bond strength, and charge distribution were most important in generating model predictions. The resulting classifications showed trends that agree well with experimental findings, in which earlier group transition metal compounds with smaller charge densities on the M atoms are easier to synthesize. The PU learning model predicted 111 MAX phases and 18 MXenes as synthesizable, including some systems such as Hf_4N_3, Sc_3C_2, and W_4C_3, although no Hf nitride or Sc- or W-based single M MXenes yet exist. Considering both MAX phase synthesizability and the score of the corresponding MXene, we also identified the top 20 most promising MAX systems that can be synthesized with high probability and etched to form previously unavailable MXenes. This work provides a computational workflow based on high-throughput DFT and PU learning to make actionable predictions about the synthesis of bulk and 2D materials and further bridge the gap between theory and experimental realization. More specifically, we have applied this framework to the family of 2D MXenes and provided insight and guidance on which materials systems are most likely to be synthesized. By accelerating materials discovery, design, and now synthesis, the availability of 2D materials with exotic properties can be rapidly expanded and exploited for use in next-generation technologies.

METHODS

The PU learning algorithm implemented in this paper is a variant of the well-established transductive bagging SVM proposed by Mordelet *et al.*[31] Bagging is also commonly used to create ensembles of decision tree classifiers,[60,61] so it is a natural extension to adapt the transductive bagging scheme to decision trees. The technical details of the implementation of this model are as follows: Denoting $\boldsymbol{P}$ as the positive sample set, $\boldsymbol{U}$ as the unlabeled sample set, K as the number of positive samples, and T as the number of bootstraps, *i.e.*, the bagging size. The model iteratively generates T decision tree classifiers as base models. In each iteration, a random subsample of $\boldsymbol{U}$, denoted as $\boldsymbol{U}_s$, is generated and treated as the negative sample set in this iteration. A binary decision tree classifier is trained using $\boldsymbol{P}$ and $\boldsymbol{U}_s$, and this classifier is used to predict the score for the remaining unlabeled samples, $\boldsymbol{U} \setminus \boldsymbol{U}_s$. In practice, the size of $\boldsymbol{U}_s$ is chosen to be equal to K, which is the same as that of $\boldsymbol{P}$, to ensure a balanced training. After T iterations, the score of any unlabeled sample x is obtained by averaging the predictive score from the decision tree classifiers trained on subsamples that do not contain x. If the average predictive score of the unlabeled sample x is greater than 0.5, then the label of x is predicted to be positive; otherwise a negative label is assigned to x. We have taken measures to minimize the impact of the small positive sample sizes by using ensemble methods, performing k-fold cross-validation and extensive testing of model performance with respect to the bagging size and number of training repetitions. Moreover, we validated our approach through comparisons to experimental evidence and tests of thermodynamic, elastic, and phase stability.

The RESVM method developed by Claesen *et al.*[32] was implemented using class weighted SVM (CWSVM) base models. The optimization problem for training CWSVMs in PU learning uses manually tuned misclassification penalties C_P and C_U for positive and unlabeled samples, respectively, with $C_P > C_U$. The base models were trained using collections of random samples of size n_P and n_U from $\boldsymbol{P}$ and $\boldsymbol{U}$, respectively. The hyperparameter w_P then determines C_P according to

$$C_P = C_U w_P \frac{n_U}{n_P} \tag{3}$$

An additional hyperparameter, γ, was used in the radial basis functions for the base models. Hyperparameters were tuned to maximize model performance (measured by TPR) by conducting a grid search for multiple feature sets starting from the optimal values for n_U and C_U suggested by Claesen *et al.*[32] RESVM results were evaluated using k-fold cross-validation with $k = 10$ folds and $T = 100$ iterations to determine the TPR. The RESVM workflow begins with training samples being fed to the base models, where each model assigns decision values to the remaining samples. The decision values are then aggregated to produce the final predictions in an ensemble scheme that reduces the bias each individual model has for the data on which it was trained. Majority voting was used for the aggregation, whereby the fraction of positive votes for an unlabeled sample x is given by

$$v(x) = \frac{n_{\text{models}} + \sum_{i=1}^{n_{\text{models}}} \text{sgn}(\psi_i(x))}{2n_{\text{models}}} \tag{4}$$

where n_{models} is the total number of CWSVM base models and ψ_i is the decision function of the ith model. Samples are predicted positive if $v(x) > 0.5$ and predicted negative if $v(x) \leq 0.5$. A schematic of the RESVM workflow is given in Figure S6.

DFT calculations were performed within the Vienna ab-Initio Simulation Package (VASP)[62] using the Perdew–Burke–Ernzerhof (PBE)[63] exchange–correlation functional and projector-augmented wave (PAW) pseudopotentials.[64] Structural relaxations were performed by straining around experimental lattice constant values and fitting energies to an equation of state, while final static calculations were done with a 520 eV plane-wave energy cutoff, an $8 \times 8 \times 1$ k-point mesh, and a total energy convergence criterion of 10^{-6} eV.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acsnano.8b08014.

Additional figures and tables (PDF)

DOI: 10.1021/acsnano.8b08014
ACS Nano 2019, 13, 3031–3041

AUTHOR INFORMATION

Corresponding Author

*E-mail: vshenoy@seas.upenn.edu.

ORCID

Nathan C. Frey: 0000-0001-5291-6131
Babak Anasori: 0000-0002-1955-253X
Yury Gogotsi: 0000-0001-9423-4032

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This work is supported primarily by contract W911NF-16-1-0447 from the Army Research Office (V.B.S.) and also by grants EFMA-542879 and CMMI-1727717 from the U.S. National Science Foundation. N.C.F. was supported by the Department of Defense (DoD) through the National Defense Science & Engineering Graduate Fellowship (NDSEG) Program. G.I.V.B. was supported by grant MRSEC DMR-1720530 from the U.S. National Science Foundation and by the National Institute of General Medical Sciences of the National Institutes of Health under Award Number T34GM008419. Experimental work at Drexel University was supported by the U.S. National Science Foundation under grant number DMR-1740795.

REFERENCES

(1) Liu, M.; Rong, Z.; Malik, R.; Canepa, P.; Jain, A.; Ceder, G.; Persson, K. A.; Liu, M. Spinel Compounds as Multivalent Battery Cathodes: A Systematic Evaluation Based on *Ab Initio* Calculations. *Energy Environ. Sci.* **2015**, *8*, 964−974.

(2) Greeley, J.; Jaramillo, T. F.; Bonde, J.; Chorkendorff, I.; Nørskov, J. K. Computational High-Throughput Screening of Electrocatalytic Materials for Hydrogen Evolution. *Nat. Mater.* **2006**, *5*, 909−913.

(3) Yu, L.; Zunger, A. Identification of Potential Photovoltaic Absorbers Based on First-Principles Spectroscopic Screening of Materials. *Phys. Rev. Lett.* **2012**, *108*, 68701.

(4) Curtarolo, S.; Hart, G. L. W.; Nardelli, M. B.; Mingo, N.; Sanvito, S.; Levy, O. The High-Throughput Highway to Computational Materials Design. *Nat. Mater.* **2013**, *12*, 191.

(5) Gražulis, S.; Daškevič, A.; Merkys, A.; Chateigner, D.; Lutterotti, L.; Quirós, M.; Serebryanaya, N. R.; Moeck, P.; Downs, R. T.; Le Bail, A. Crystallography Open Database (COD): An Open-Access Collection of Crystal Structures and Platform for World-Wide Collaboration. *Nucleic Acids Res.* **2012**, *40*, D420−D427.

(6) Jain, A.; Ong, S. P.; Hautier, G.; Chen, W.; Richards, W. D.; Dacek, S.; Cholia, S.; Gunter, D.; Skinner, D.; Ceder, G.; Persson, K. A. Commentary: The Materials Project: A Materials Genome Approach to Accelerating Materials Innovation. *APL Mater.* **2013**, *1*, 011002.

(7) Mounet, N.; Gibertini, M.; Schwaller, P.; Campi, D.; Merkys, A.; Marrazzo, A.; Sohier, T.; Castelli, I. E.; Cepellotti, A.; Pizzi, G.; Marzari, N. Two-Dimensional Materials from High-Throughput Computational Exfoliation of Experimentally Known Compounds. *Nat. Nanotechnol.* **2018**, *13*, 246−252.

(8) Haastrup, S.; Strange, M.; Pandey, M.; Deilmann, T.; Schmidt, P. S.; Hinsche, N. F.; Gjerding, M. N.; Torelli, D.; Larsen, P. M.; Riis-Jensen, A. C.; Gath, J.; Jacobsen, K. W.; Mortensen, J. J.; Olsen, T.; Thygesen, K. S. The Computational 2D Materials Database: High-Throughput Modeling and Discovery of Atomically Thin Crystals. *2D Mater.* **2018**, *5*, 042002.

(9) Ashton, M.; Paul, J.; Sinnott, S. B.; Hennig, R. G. Topology-Scaling Identification of Layered Solids and Stable Exfoliated 2D Materials. *Phys. Rev. Lett.* **2017**, *118*, 106101.

(10) Saal, J. E.; Kirklin, S.; Aykol, M.; Meredig, B.; Wolverton, C. Materials Design and Discovery with High-Throughput Density Functional Theory: The Open Quantum Materials Database (OQMD). *JOM* **2013**, *65*, 1501−1509.

(11) Franceschetti, A.; Zunger, A. The Inverse Band-Structure Problem of Finding an Atomic Configuration with given Electronic Properties. *Nature* **1999**, *402*, 60−63.

(12) Jóhannesson, G. H.; Bligaard, T.; Ruban, A. V.; Skriver, H. L.; Jacobsen, K. W.; Nørskov, J. K. Combined Electronic Structure and Evolutionary Search Approach to Materials Design. *Phys. Rev. Lett.* **2002**, *88*, 255506.

(13) Jain, A.; Shin, Y.; Persson, K. A. Computational Predictions of Energy Materials Using Density Functional Theory. *Nat. Rev. Mater.* **2016**, *1*, 15004.

(14) Meredig, B.; Agrawal, A.; Kirklin, S.; Saal, J. E.; Doak, J. W.; Thompson, A.; Zhang, K.; Choudhary, A.; Wolverton, C. Combinatorial Screening for New Materials in Unconstrained Composition Space with Machine Learning. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2014**, *89*, 94104.

(15) Raccuglia, P.; Elbert, K. C.; Adler, P. D. F.; Falk, C.; Wenny, M. B.; Mollo, A.; Zeller, M.; Friedler, S. A.; Schrier, J.; Norquist, A. J. Machine-Learning-Assisted Materials Discovery Using Failed Experiments. *Nature* **2016**, *533*, 73−76.

(16) Kim, E.; Huang, K.; Saunders, A.; McCallum, A.; Ceder, G.; Olivetti, E. Materials Synthesis Insights from Scientific Literature via Text Extraction and Machine Learning. *Chem. Mater.* **2017**, *29*, 9436−9444.

(17) Balachandran, P. V.; Kowalski, B.; Sehirlioglu, A.; Lookman, T. Experimental Search for High-Temperature Ferroelectric Perovskites Guided by Two-Step Machine Learning. *Nat. Commun.* **2018**, *9*, 1668.

(18) Naguib, M.; Kurtoglu, M.; Presser, V.; Lu, J.; Niu, J.; Heon, M.; Hultman, L.; Gogotsi, Y.; Barsoum, M. W. Two-Dimensional Nanocrystals Produced by Exfoliation of Ti_3AlC_2. *Adv. Mater.* **2011**, *23*, 4248−4253.

(19) Anasori, B.; Lukatskaya, M. R.; Gogotsi, Y. 2D Metal Carbides and Nitrides (MXenes) for Energy Storage. *Nat. Rev. Mater.* **2017**, *2*, 16098.

(20) Barsoum, M. W. *MAX Phases: Properties of Machinable Ternary Carbides and Nitrides*; Wiley-VCH, 2013.

(21) Barsoum, M. W.; Radovic, M. Elastic and Mechanical Properties of the MAX Phases. *Annu. Rev. Mater. Res.* **2011**, *41*, 195−227.

(22) Xia, Y.; Mathis, T. S.; Zhao, M. Q.; Anasori, B.; Dang, A.; Zhou, Z.; Cho, H.; Gogotsi, Y.; Yang, S. Thickness-Independent Capacitance of Vertically Aligned Liquid-Crystalline MXenes. *Nature* **2018**, *557*, 409−412.

(23) Khazaei, M.; Ranjbar, A.; Arai, M.; Sasaki, T.; Yunoki, S. Electronic Properties and Applications of MXenes: A Theoretical Review. *J. Mater. Chem. C* **2017**, *5*, 2488−2503.

(24) Rajan, A. C.; Mishra, A.; Satsangi, S.; Vaish, R.; Mizuseki, H.; Lee, K.-R.; Singh, A. K. Machine-Learning-Assisted Accurate Band Gap Predictions of Functionalized MXene. *Chem. Mater.* **2018**, *30*, 4031−4038.

(25) Zhang, H.; Yang, G.; Zuo, X.; Tang, H.; Yang, Q.; Li, G. Computational Studies on the Structural, Electronic and Optical Properties of Graphene-like MXenes (M_2CT_2, M = Ti, Zr, Hf; T = O, F, OH) and Their Potential Applications as Visible-Light Driven Photocatalysts. *J. Mater. Chem. A* **2016**, *4*, 12913−12920.

(26) Khazaei, M.; Arai, M.; Sasaki, T.; Chung, C.-Y.; Venkataramanan, N. S.; Estili, M.; Sakka, Y.; Kawazoe, Y. Novel Electronic and Magnetic Properties of Two-Dimensional Transition Metal Carbides and Nitrides. *Adv. Funct. Mater.* **2013**, *23*, 2185−2192.

(27) Kumar, H.; Frey, N. C.; Dong, L.; Anasori, B.; Gogotsi, Y.; Shenoy, V. B. Tunable Magnetism and Transport Properties in Nitride MXenes. *ACS Nano* **2017**, *11*, 7648−7655.

(28) Frey, N. C.; Kumar, H.; Anasori, B.; Gogotsi, Y.; Shenoy, V. B. Tuning Noncollinear Spin Structure and Anisotropy in Ferromagnetic Nitride MXenes. *ACS Nano* **2018**, *12*, 6319−6325.

(29) Frey, N. C.; Bandyopadhyay, A.; Kumar, H.; Anasori, B.; Gogotsi, Y.; Shenoy, V. B. Surface Engineered MXenes: Electric Field

DOI: 10.1021/acsnano.8b08014
ACS Nano 2019, 13, 3031−3041

Control of Magnetism and Enhanced Magnetic Anisotropy. *ACS Nano* **2019**, DOI: 10.1021/acsnano.8b09201
(30) Elkan, C.; Noto, K. Learning Classifiers from Only Positive and Unlabeled Data. In *Proceeding of the 14th ACM SIGKDD international conference on knowledge discovery and data mining - KDD 08*; 2008; p 213.
(31) Mordelet, F.; Vert, J.-P. A Bagging SVM to Learn from Positive and Unlabeled Examples. *Pattern Recognit. Lett.* **2014**, *37*, 201−209.
(32) Claesen, M.; De Smet, F.; Suykens, J. A. K.; De Moor, B. A Robust Ensemble Approach to Learn from Positive and Unlabeled Data Using SVM Base Models. *Neurocomputing* **2015**, *160*, 73−84.
(33) Niu, G.; Plessis, M. C. du.; Sakai, T.; Ma, Y.; Sugiyama, M. Theoretical Comparisons of Positive-Unlabeled Learning against Positive-Negative Learning. In *NIPS'16 Proceedings of the 30th International Conference on Neural Information Processing Systems*; 2016; pp 1207−1215.
(34) Nicolosi, V.; Chhowalla, M.; Kanatzidis, M. G.; Strano, M. S.; Coleman, J. N. Liquid Exfoliation of Layered Materials. *Science* **2013**, *340*, 1226419.
(35) Zhang, X.; Zhang, Z.; Zhou, Z. MXene-Based Materials for Electrochemical Energy Storage. *J. Energy Chem.* **2018**, *27*, 73−85.
(36) Lan, W.; Wang, J.; Li, M.; Liu, J.; Li, Y.; Wu, F.-X.; Pan, Y. Predicting Drug−target Interaction Using Positive-Unlabeled Learning. *Neurocomputing* **2016**, *206*, 50−57.
(37) Peng, L.; Zhu, W.; Liao, B.; Duan, Y.; Chen, M.; Chen, Y.; Yang, J. Screening Drug-Target Interactions with Positive-Unlabeled Learning. *Sci. Rep.* **2017**, *7*, 8087.
(38) Hameed, P. N.; Verspoor, K.; Kusljic, S.; Halgamuge, S. Positive-Unlabeled Learning for Inferring Drug Interactions Based on Heterogeneous Attributes. *BMC Bioinf.* **2017**, *18*, 140.
(39) Denis, F.; Gilleron, R. Text Classification and Co-Training from Positive and Unlabeled Examples. *Proceeding of the ICML 2003 Work. Contin. from Labeled to Unlabeled Data* 2003; pp 80−87.
(40) Nguyen, M. N.; Li, X. L.; Ng, S. K. Ensemble Based Positive Unlabeled Learning for Time Series Classification. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*; Springer: Berlin, Heidelberg, 2012; Vol. *7238* LNCS, pp 243−257.
(41) Kiliç, C.; Tan, M. Positive Unlabeled Learning for Deriving Protein Interaction Networks. *Netw. Model. Anal. Heal. Inf. Bioinform* **2012**, *1*, 87−102.
(42) Yang, S.; Zhang, P.; Wang, F.; Ricciardulli, A. G.; Lohe, M. R.; Blom, P. W. M.; Feng, X. Fluoride-Free Synthesis of Two-Dimensional Titanium Carbide (MXene) Using A Binary Aqueous System. *Angew. Chem., Int. Ed.* **2018**, *57*, 15491−15495.
(43) Alhabeb, M.; Maleski, K.; Mathis, T. S.; Sarycheva, A.; Hatter, C. B.; Uzun, S.; Levitt, A.; Gogotsi, Y. Selective Etching of Silicon from Ti_3SiC_2 (MAX) Produces 2D Titanium Carbide (MXene). *Angew. Chem., Int. Ed.* **2018**, *57*, 5444−5448.
(44) Naguib, M.; Mochalin, V. N.; Barsoum, M. W.; Gogotsi, Y. 25th Anniversary Article: MXenes: A New Family of Two-Dimensional Materials. *Adv. Mater.* **2014**, *26*, 992−1005.
(45) Aryal, S.; Sakidja, R.; Barsoum, M. W.; Ching, W. Y. A Genomic Approach to the Stability, Elastic, and Electronic Properties of the MAX Phases. *Phys. Status Solidi B* **2014**, *251*, 1480−1497.
(46) Ashton, M.; Hennig, R. G.; Broderick, S. R.; Rajan, K.; Sinnott, S. B. Computational Discovery of Stable M_2AX Phases. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2016**, *94*, 54116.
(47) Palmquist, J. P.; Li, S.; Persson, P. O.; Emmerlich, J.; Wilhelmsson, O.; Högberg, H.; Katsnelson, M. I.; Johansson, B.; Ahuja, R.; Eriksson, O.; Hultman, L.; Jansson, U. $M_{n+1}AX_n$ Phases in the Ti-Si-C System Studied by Thin-Film Synthesis and *Ab Initio* Calculations. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2004**, *70*, 165401.
(48) Keast, V. J.; Harris, S.; Smith, D. K. Prediction of the Stability of the $M_{N+1}AX_n$ Phases from First Principles. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2009**, *80*, 214113.
(49) Khazaei, M.; Ranjbar, A.; Esfarjani, K.; Bogdanovski, D.; Dronskowski, R.; Yunoki, S. Insights into Exfoliation Possibility of MAX Phases to MXenes. *Phys. Chem. Chem. Phys.* **2018**, *20*, 8579−8592.
(50) Ashton, M.; Mathew, K.; Hennig, R. G.; Sinnott, S. B. Predicted Surface Composition and Thermodynamic Stability of MXenes in Solution. *J. Phys. Chem. C* **2016**, *120*, 3550−3556.
(51) Anasori, B.; Lu, J.; Rivin, O.; Dahlqvist, M.; Halim, J.; Voigt, C.; Rosen, J.; Hultman, L.; Barsoum, M. W.; Caspi, E. N. A Tungsten-Based Nanolaminated Ternary Carbide: $(W,Ti)_4C_{4-X}$. *Inorg. Chem.* **2019**, *58*, 1100−1106.
(52) Halim, J.; Kota, S.; Lukatskaya, M. R.; Naguib, M.; Zhao, M.-Q.; Moon, E. J.; Pitock, J.; Nanda, J.; May, S. J.; Gogotsi, Y.; Barsoum, M. W. Synthesis and Characterization of 2D Molybdenum Carbide (MXene). *Adv. Funct. Mater.* **2016**, *26*, 3118−3127.
(53) Zhou, J.; Zha, X.; Chen, F. Y.; Ye, Q.; Eklund, P.; Du, S.; Huang, Q. A Two-Dimensional Zirconium Carbide by Selective Etching of Al_3C_3 from Nanolaminated $Zr_3Al_3C_5$. *Angew. Chem.* **2016**, *128*, 5092−5097.
(54) Zhou, J.; Zha, X.; Zhou, X.; Chen, F.; Gao, G.; Wang, S.; Shen, C.; Chen, T.; Zhi, C.; Eklund, P.; Du, S.; Xue, J.; Shi, W.; Chai, Z.; Huang, Q. Synthesis and Electrochemical Properties of Two-Dimensional Hafnium Carbide. *ACS Nano* **2017**, *11*, 3841−3850.
(55) Zhang, Y.; Ling, C. A Strategy to Apply Machine Learning to Small Datasets in Materials Science. *npj Comput. Mater.* **2018**, *4*, 25.
(56) Meredig, B.; Antono, E.; Church, C.; Hutchinson, M.; Ling, J.; Paradiso, S.; Blaiszik, B.; Foster, I.; Gibbons, B.; Hattrick-Simpers, J.; Mehta, A.; Ward, L. Can Machine Learning Identify the next High-Temperature Superconductor? Examining Extrapolation Performance for Materials Discovery. *Mol. Syst. Des. Eng.* **2018**, *3*, 819−825.
(57) Bloom, D. S.; Grant, N. J. The System Chromium-Carbon. *JOM* **1950**, *2*, 41−46.
(58) Abdelmalak, M. N. MXenes: A New Family of Two-Dimensional Materials and Its Application as Electrodes for Li-Ion Batteries; Ph.D. Thesis, Drexel University, 2014.
(59) Urbankowski, P.; Anasori, B.; Makaryan, T.; Er, D.; Kota, S.; Walsh, P. L.; Zhao, M.; Shenoy, V. B.; Barsoum, M. W.; Gogotsi, Y. Synthesis of Two-Dimensional Titanium Nitride Ti_4N_3 (MXene). *Nanoscale* **2016**, *8*, 11385−11391.
(60) Banfield, R. E.; Hall, L. O.; Bowyer, K. W.; Kegelmeyer, W. P. A Comparison of Decision Tree Ensemble. *IEEE Trans. Pattern Anal. Mach. Intell.* **2007**, No. 1, *29*173−180.
(61) Kotsiantis, S. B. Decision Trees: A Recent Overview. *Artif. Intell. Rev.* **2013**, *39*, 261−283.
(62) Kresse, G.; Furthmüller, J. Efficient Iterative Schemes for *Ab Initio* Total-Energy Calculations Using a Plane-Wave Basis Set. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1996**, *54*, 11169−11186.
(63) Perdew, J. P.; Burke, K.; Ernzerhof, M. Generalized Gradient Approximation Made Simple. *Phys. Rev. Lett.* **1996**, *77*, 3865−3868.
(64) Kresse, G.; Joubert, D. From Ultrasoft Pseudopotentials to the Projector Augmented-Wave Method. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1999**, *59*, 1758−1775.

DOI: 10.1021/acsnano.8b08014
ACS Nano 2019, 13, 3031−3041