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 Mayaguez, Mayaguez 00681, Puerto Rico
§Department of Materials Science and Engineering and A.J. Drexel Nanomaterials Institute, Drexel University, Philadelphia, Pennsylvania 19104, United States

Supporting Information

ABSTRACT: 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.

KEYWORDS: 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,‡ electrocatalysis,§ photovoltaic absorbers,§ and a staggering variety of other applications. There is a growing availability of experimental and theoretical data on 3D crystals in popular databases,§,10 and recent efforts have dramatically increased the collection of proposed 2D materials and their predicted properties.‡−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,§,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 compositions14 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 Mn+nXnTn (n = 1−3) and their layered precursor 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
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, providing successful examples to consider. Perhaps most importantly, exciting experimental results in electrochemical applications and predictions of attractive electronic, optical, and magnetic 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 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 to provide quantitative insights beyond DFT-calculated exfoliation energies, or more generally to the liquid exfoliation of any layered materials. 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...
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, 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, 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 2541 and in gene identification studies the positive sample sizes are as small as 11 samples41 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.33 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-fold cross-validation), 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

\[
TPR = \frac{1}{R} \sum_{r=1}^{R} \frac{1}{K} \sum_{k=1}^{K} TPR_{k,r}
\]

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.35 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 Ti3SiC2,43 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 elements, et al., 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
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 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 X atoms.

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.

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

<table>
<thead>
<tr>
<th>DFT-calculated features</th>
<th>elemental features</th>
</tr>
</thead>
<tbody>
<tr>
<td>n</td>
<td>Z</td>
</tr>
<tr>
<td>a</td>
<td>N_V</td>
</tr>
<tr>
<td>c</td>
<td>g_P</td>
</tr>
<tr>
<td>d_MX</td>
<td>p_P</td>
</tr>
<tr>
<td>d_MM</td>
<td>Xe</td>
</tr>
<tr>
<td>d_MA</td>
<td>EA</td>
</tr>
<tr>
<td>d_AA</td>
<td>IE1</td>
</tr>
<tr>
<td>r_MX</td>
<td>IE2</td>
</tr>
<tr>
<td>r_MA</td>
<td>E_chem</td>
</tr>
<tr>
<td>m</td>
<td>E_atom</td>
</tr>
<tr>
<td>V</td>
<td></td>
</tr>
<tr>
<td>rho</td>
<td></td>
</tr>
<tr>
<td>Energy</td>
<td></td>
</tr>
<tr>
<td>E_pa</td>
<td></td>
</tr>
<tr>
<td>E_form</td>
<td></td>
</tr>
<tr>
<td>E_coh</td>
<td></td>
</tr>
<tr>
<td>e_M</td>
<td></td>
</tr>
<tr>
<td>e_A</td>
<td></td>
</tr>
<tr>
<td>e_X</td>
<td></td>
</tr>
</tbody>
</table>
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. The MXene feature correlation matrix is shown in Figure S1b.

**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).

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...
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 synthesiz-
ability 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 perform-
ance 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., Mn2AlC and Mn2AlN), 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. Using the MaterialsWeb
database,66 which reports phase stability of M2AX 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.

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
cost function between 0 and 1, where 0 corresponds to a
negative (not synthesizable) compound and 1 is a positive

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

<table>
<thead>
<tr>
<th>MXene predicted positives</th>
<th>TPR &gt; 0.5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hf2C3</td>
<td>Ta2N1</td>
</tr>
<tr>
<td>Nb2C2</td>
<td>Ta2C2</td>
</tr>
<tr>
<td>Zr2C2</td>
<td>Hf2N1</td>
</tr>
<tr>
<td>Ta2C2</td>
<td>Ti2C3</td>
</tr>
<tr>
<td>W2C3</td>
<td>Hf2C</td>
</tr>
<tr>
<td>Zr2C3</td>
<td>Sc2C1</td>
</tr>
</tbody>
</table>

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 (W2C3, Ta2C2, W2C1, and Mo2C1) 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

Most interestingly, the model predicts certain Ta, Hf, and
Nb nitrides, as well as two Sc-based MXenes (Sc2C1 and Sc2C) and
two W-based MXenes (W2C1 and W2C3), 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
collectors, such as the recently reported W-based nanolaminated ternary phase, (W, Ti)2C4−x,51 are better
starting points to successfully synthesize these compounds.
Mo2C3, Zr2C2, and Hf2C2 MXenes have previously been
synthesized by such methods.52–54 The PU learning approach
employed here outputs an easily interpreted “synthesis prob-
ability” between 0 and 1, where 0 corresponds to a
negative (not synthesizable) compound and 1 is a positive
(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. Additionally, traditional $k$-fold cross-validation is overly optimistic when evaluating materials discovery machine learning models, because materials data sets are rarely uniformly distributed. 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$_2$N$_2$, Cr$_3$C$_2$, Cr$_3$N$_2$, and Mo$_3$N$_2$ 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$_2$AlC MAX phase was synthesized long before the discovery of MXenes, attempts to make Cr$_2$C from Cr$_2$AlC by HF etching have been unsuccessful. Based on our knowledge to date, no selective etching of Al is observed for Cr$_2$AlC immersion in different HF concentrations, and the MAX phase powder is dissolved after a few hours. This observation suggests that Cr$_2$C synthesis, if possible, is not likely to be successful via the HF route. Ti$_2$N$_3$ was similarly difficult to synthesize (dissolution in HF), and we were able to synthesize it via molten salt etching of Ti$_x$AlN$_y$. We have tried the molten salt route for Cr$_2$C 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$_2$C 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 additional 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$_2$N$_3$, has been synthesized from a MAX phase precursor. 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 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 unidentifed 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 systems 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, 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, 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.
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 $n_t$, the formation energies within one standard deviation of the median of all predicted positives are between 0 and $\pm 1$ eV, and the cohesive energies are between $\pm 5.5$ and $\pm 7.5$ eV, regardless of $n_t$. 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 $n_t$, 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

$$E_{\text{etch}} = \frac{E_{\text{max}} - 2E_{\text{MX}} - 2\mu_A}{4S}$$

where $E_{\text{max}}$ is the total energy of the MAX phase, $E_{\text{MX}}$ is the total energy of a MXene unit cell, $\mu_A$ 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. 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 $E_{\text{etch}}$. 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](image-url)
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 W₁₇C₂₃. 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 Hf₆N₁₀, Sc₃C₂, and W₁₇C₂₃, 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, 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 $P$ as the positive sample set, $U$ 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 $U$, denoted as $U_0$, is generated and treated as the negative sample set in this iteration. A binary decision tree classifier is trained using $P$ and $U_0$, and this classifier is used to predict the score for the remaining unlabeled samples, $U \setminus U_0$. In practice, the size of $U_0$ is chosen to be equal to $K$, which is the same as that of $P$, 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 $x$.

An additional hyperparameter, $\gamma$, 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 $n_t$ 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

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

where $n_{\text{models}}$ is the total number of CW SVM base models and $\psi_i$ is the decision function of the $i$th 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 × 8 × 1 k-point mesh, and a total energy convergence criterion of $10^{-6}$ eV.

ASSOCIATED CONTENT

3 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)
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.L.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


(29) Frey, N. C.; Bandypadhyay, A.; Kumar, H.; Anasori, B.; Gogotsi, Y.; Shenoy, V. B. Surface Engineered MXenes: Electric Field
Control of Magnetism and Enhanced Magnetic Anisotropy. ACS Nano 2019, DOI: 10.1021/acsnano.8b09201


