

Machine-Learning-Driven Expansion of the 1D van der Waals Materials Space

Yanbing Zhu, Evan R. Antoniuk, Dylan Wright, Fariborz Kargar, Nicholas Sasing, Austin D. Sendek, Tina T. Salguero, Ludwig Bartels, Alexander A. Balandin, Evan J. Reed, and Felipe H. da Jornada*



Cite This: *J. Phys. Chem. C* 2023, 127, 21675–21683



Read Online

ACCESS |



Metrics & More

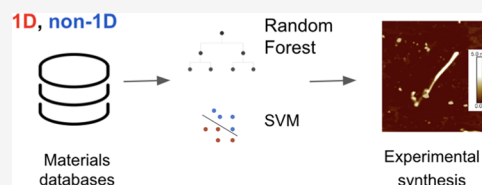


Article Recommendations



Supporting Information

ABSTRACT: One-dimensional (1D) van der Waals (vdW) materials display electronic and magnetic transport properties that make them uniquely suited as interconnect materials and for low-dimensional optoelectronic applications. However, there are only around 700 1D vdW structures in general materials databases, making database curation approaches ineffective for 1D discovery. Here, we utilize machine-learning techniques to discover 1D vdW compositions that have not yet been synthesized. Our techniques go beyond discovery efforts involving elemental substitutions and instead start with a composition space of 4741 binary and 392,342 ternary formulas. We predict up to 3000 binary and 10,000 ternary 1D compounds and further classify them by expected magnetic and electronic properties. Our model identifies MoI_3 , a material we experimentally confirm to exist with wire-like subcomponents and exotic magnetic properties. More broadly, we find several chalcogen-, halogen-, and pnictogen-containing compounds expected to be synthesizable using chemical vapor deposition and chemical vapor transport.



1. INTRODUCTION

There has been an increased use of machine learning to predict novel material compositions. Despite significant advances in materials databases, including the Materials Project,¹ Inorganic Crystal Structure Database (ICSD),² Open Quantum Materials Database (OQMD),³ Crystallography Open Database (COD),⁴ and Cambridge Structural Database (CSD),⁵ there remains a much larger region in the chemical composition space of potentially undiscovered materials, up to 10K for relevant binary and ternary compositions.

Databases have been curated where compositions and structures from previous experimental published reports are aggregated. This space is commonly augmented with simple elemental substitution, and materials discovery seeks to uncover more novel compositions. While one-dimensional (1D) materials offer exciting application potential, predicting additional compositions is difficult due to the restricted set of known 1D materials. Here, we expand the space by machine learning, trained on existing known compositions, and tackle a small-data problem more common in machine learning applied to the physical sciences.⁶

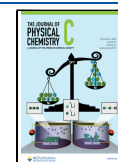
We focus our work on 1D van der Waals materials, a subset of the general family of van der Waals-bonded two-dimensional (2D) planes, 1D wires, and zero-dimensional (0D) clusters. While such structures can be present in freestanding forms and synthesized from bottom-up processes, as in quasi-1D carbon nanotubes or 0D quantum dots, we concentrate on structures where the low-dimensional subcomponents exist in weakly bonded bulk form, e.g., a 2D plane of graphene that forms graphite. We hence focus on systems wherein the bonds

between the subcomponents are primarily van der Waals in characters. Conversely, bonds within each individual layer, wire, or cluster are primarily ionic or covalent. Several classes of 2D materials have been previously studied and characterized, including systems with a planar honeycomb structure such as graphene (such as silicene and borophene), a puckered honeycomb structure (such as phosphorene and monochalcogenides), transition-metal dichalcogenides, materials of the MXene family, etc. High-throughput screening has revealed an even wider scope of compositions, including 1D vdW counterparts that are studied less frequently, in part due to their limited availability.^{7,8} While 2D materials represent around 4% of all materials, 1D materials, in their bulk, nonexfoliated form, represent 1–2% of all experimentally synthesized materials. Since the number of 1D materials computed so far is even more limited than 2D materials (less than 1000), it would be useful to extend such a space and deduce other potentially new 1D compounds that have not yet been computationally predicted or experimentally synthesized. Previously, similar ideas have been successful in expanding the space of 2D vdW materials⁹ by more than 1500 and closing the loop between computation and experimental in other material applications.^{10,11}

Received: June 8, 2023

Revised: September 4, 2023

Published: October 27, 2023



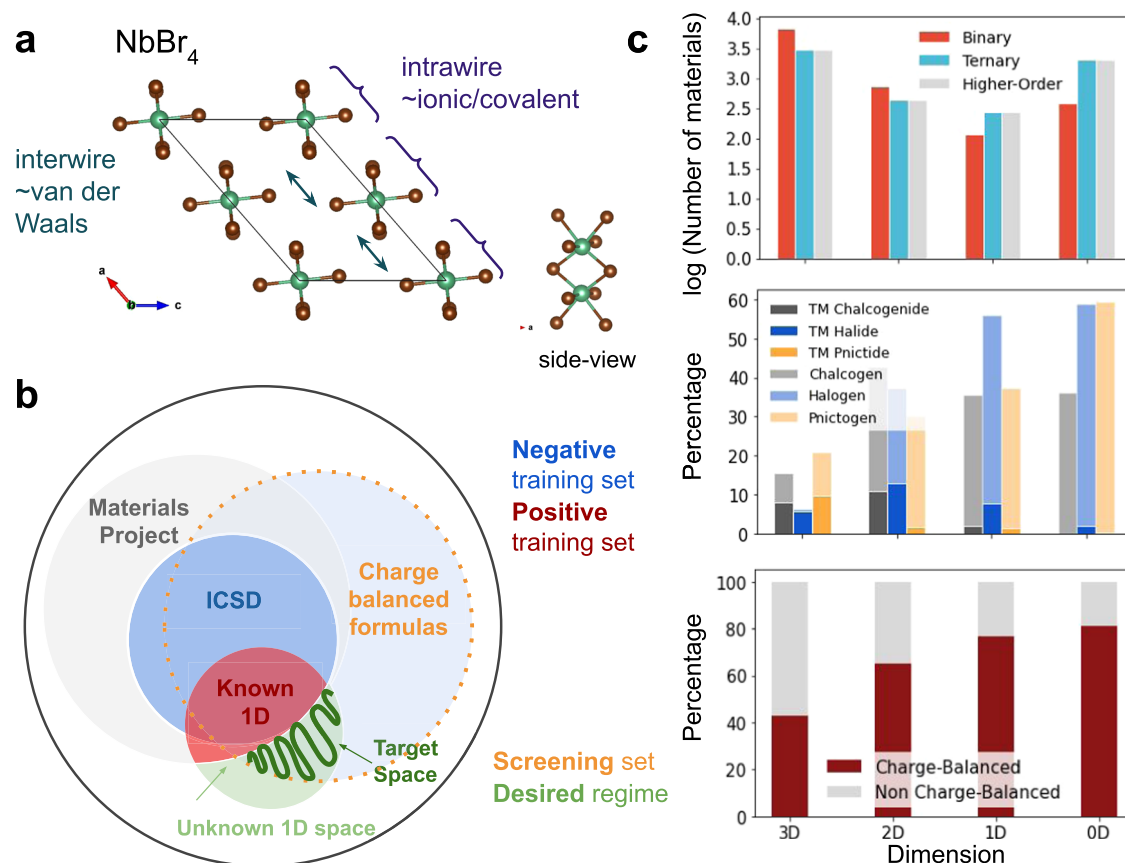


Figure 1. (a) Example of a one-dimensional vdW material, NbBr_4 corresponding to Materials Project ID mp-1080428 and ICSD 239640, and the associated structure is shown. (b) The material space under consideration is presented. We start with materials in the Materials Project database with corresponding ICSD numbers, where the dimensionality of the materials is known. We seek for additional low-dimensional materials in the larger space of randomly generated compositions, targeting those that are likely conductive or magnetic. (c) The percentage of materials by family (containing at least one chalcogen, halogen, and pnictogen and containing a transition metal of the mentioned families) is shown by dimension. Since the counts refer to the presence of a family type and each material composition can contain elements of multiple families, the total percentage exceeds 100. Materials in these classes constitute a majority of low-dimensional materials, which motivates the separate model training process. The number of materials classified by dimension is shown and further broken down by binary, ternary, or higher-order compositions. 1D materials make up a small number of existing materials, between 1 and 2%. Out of this, a large fraction is captured by the binary and ternary compositions. We further observe that a majority of 1D compositions are charge-balanced, which motivates the charge-balanced application subspace.

Our starting point is in a recent work categorizing existing materials based on dimensionality, which revealed a number of bulk structures that support 2D and 1D vdW subcomponents and are potentially exfoliable.^{8,12,13} Such materials are of interest for next-generation computing technologies¹⁴ due to stackability, desirable scaling properties, and emergent phenomena and are relevant for optics and photonics applications.¹⁵ In particular, the 1D magnetic material Fe_3O_4 ¹⁶ has been synthesized, but there exists limited work toward a more diverse range of analogous 1D magnetic structures from such curation efforts.

Such efforts would likely be rewarding given the wealth of correlated electronic states in 1D, ranging from Luttinger-liquid physics to correlated magnetism, Peierls instability, and topology (based on the Zak phase, for instance). Likewise, the possibility of engineering 1D materials with unusual transport is of key importance for interconnection or heterojunction applications¹⁷ and motivates the search for novel 1D materials beyond our restricted set of experimentally synthesized and first-principles computed structures.

We note that our effort is complementary to recent ML-based investigations which aim at expanding 1D material space^{18,19} by training over crystal structures. We take a different approach of

omitting the middle step of predicting structures and instead utilizing density functional theory (DFT) as a postverification process. This is motivated by the fact that the final crystal structure is not a control parameter in common materials synthesis techniques such as chemical vapor deposition (CVD) and chemical vapor transfer (CVT). Instead, experimentalists typically have control of only the input precursor materials, suggesting that the chemical composition (formula) can be enough to predict novel synthesizable 1D materials.

In particular, energetic stability, deduced from ground-state DFT calculations in minimal primitive unit cells, is not sufficient to guarantee the experimental stability of a given 1D structure. This is evident by the fact that 1D materials are prone to displaying charge-density waves, Peierls instability, magnetism, and other broken-symmetry states. Given the larger unit cell necessary to simulate such lower-symmetry structures, it is computationally prohibitive to attempt approaches such as *ab initio* random structure search (AIRSS)²⁰ to study the landscape of 1D systems systematically.

Our approach is not to predict stable crystal structures or potentially higher-fractional composition ratio for a particular class or family of material, which is proven to be difficult for ML tasks but advantageous for DFT calculations.²¹ Instead, we

Table 1. Intrinsic Dimension Distributions and Machine-Learning Results^a

positive (size)	negative (size)	baseline %	SVM accuracy	SVM precision	RF accuracy	RF precision	SVM P/B	RF P/B	SVM AUC	RF AUC
1D (647)	non-1D (40,636)	1.59	0.919	0.131	0.985	0.667	8.239	41.950	0.830	0.576
1D	bulk (19,088)	3.39	0.959	0.438	0.979	0.853	12.920	25.162	0.912	0.722
1D	2D (1378)	46.95	0.783	0.64	0.788	0.75	1.363	1.597	0.829	0.714
2D	bulk	7.5	0.889	0.373	0.955	0.793	4.973	10.573	0.873	0.737
2D	non-2D	3.57	0.877	0.202	0.972	0.837	5.658	23.445	0.872	0.824
1D (561)	non-1D (27,948)	2.0	0.918	0.154	0.981	0.6	7.7	30	0.609	0.552
charge-balanced	charge-balanced									
binary 1D 1–3 (75)	binary non-1D 1–3 (1548)	4.84	0.939	0.429	0.957	1.0	8.864	<i>x</i>	0.743	0.563
binary 1D 1–5 (115)	binary non-1D 1–5 (1953)	5.89	0.889	0.31	0.947	1.0	5.263	<i>x</i>	0.824	0.542
ternary 1D 1–3 (112)	ternary non-1D 1–3 (4355)	2.57	0.96	0.375	0.973	1.0	14.591	<i>x</i>	0.858	0.662

^aIntrinsic class distributions by material dimension and machine-learning results are provided, including the results for the restricted binary and ternary compositions. The precision metric suggests that out of the model-predicted 1D materials, this given fraction will truly be 1D in nature. The precision increases significantly after the charge-balanced binary and ternary restrictions, though the intrinsic fraction of 1D materials is higher. Classification of 1D versus bulk yields a much higher precision than 1D versus all non-1D (including 2D, 0D, or intercalated classifications). The results are shown for the support vector regression (SVR) and the random forest model (RF). P/B indicates the precision over the baseline occurrence percentage of 1D materials, which can measure how much the model improves over random choice. AUC is the area under the curve, the area under the receiver operating characteristic (ROC).

directly seek to identify approximately 1–2% of all compositions that may have the potential to be 1D vdW in nature. We have limited our application space to binary formulas with a coordination number of up to 5 and ternary compounds with a coordination number of up to 3.

To our knowledge, while other works, including on 2D materials,^{22,23} have substantially helped understanding the potential structure space of low-dimensional van der Waals materials, this work can directly predict compositions that can aid experimentalists in the search of novel 1D materials starting from a large space of almost 400,000 possible formulas. In particular, our approach of directly training on the formulas is also agnostic to the emergence of charge-density waves and other broken-symmetry states, which could be incorrectly missed in DFT calculations.

Our method of directly learning stable compositions is based on previous successful efforts involving 2D materials.⁹ While this approach does not directly yield crystal structures, the ML-generated compositions can directly inform the synthesis of stable materials through common CVD and CVT methods. In addition, such an approach yields a probability output between 0 and 1 that determines the robustness of our predictions, which can be tuned to produce conservative guesses to a more exploratory set of materials. We showcase this technique can indeed identify a 1D material, MoI₃, which was subsequently validated in experiments. Our work is organized as follows: detailing the existing space of materials, describing our machine-learning methods, and observing how these output predictions compare with existing families.

II. METHODS

To identify the low-dimensional materials of interest, we consider entries from the Materials Project database and classify them by dimension (3D bulk, 2D, and 1D) using a method we developed,⁸ determined by how connectivity scales with supercell size. We include materials in the training data only if they contain a corresponding Inorganic Crystal Structure Database (ICSD) number, which indicates prior experimental synthesis. The composition of the initial training set by

dimensionality and characteristics relevant to our work (including binary or ternary compositions, coordination number, and charge balancing) is shown Figure 1c. The positive training set is taken as the 1D materials in this training set (Figure 1b).

We considered two machine-learning approaches: random forest and support vector machine (SVM) models. For the random forest approach, we consider a model with 100 trees and a reduced feature set of the top 20 features selected via the resultant feature importance. For the SVM approach, we used a radial basis kernel (RBF) with hyperparameter tuning across a range of input *c*-values and *γ*-values, which determines the misclassification versus the smoothness of the decision hyperplane and the impact of a single training example, respectively. For both models, we assessed the robustness by focusing on the precision metric. We find high overlap between the two different types of models across different subsets of the atomic feature space representation.

The trained models are then applied to the larger search space of all charge-balanced binary and ternary formulas up to certain coordination numbers (5 for binary, 3 for ternary). As a secondary extension, to better match the application search space to the training space, we further tested another training method where portions of the application space are introduced in the negative (non-1D) training set. Given that the intrinsic percentage of 1D materials is so low (1–2%), the cost of potentially introducing a true positive 1D material to the negative training set is outweighed by the likely better match between the training space and the application space.

III. RESULTS AND DISCUSSION

The results of training between different dimensions and restricting compositions to charge-balanced binary and ternary cases using the random forest and SVM models are shown in Table 1. We find that predicted positive examples contain a large overlap with SVM-predicted positive examples, which indicates model-agnostic results. From the performance metrics, we observe that 1D materials are separable from bulk materials and closer in character to 2D vdW compositions. This is in line with

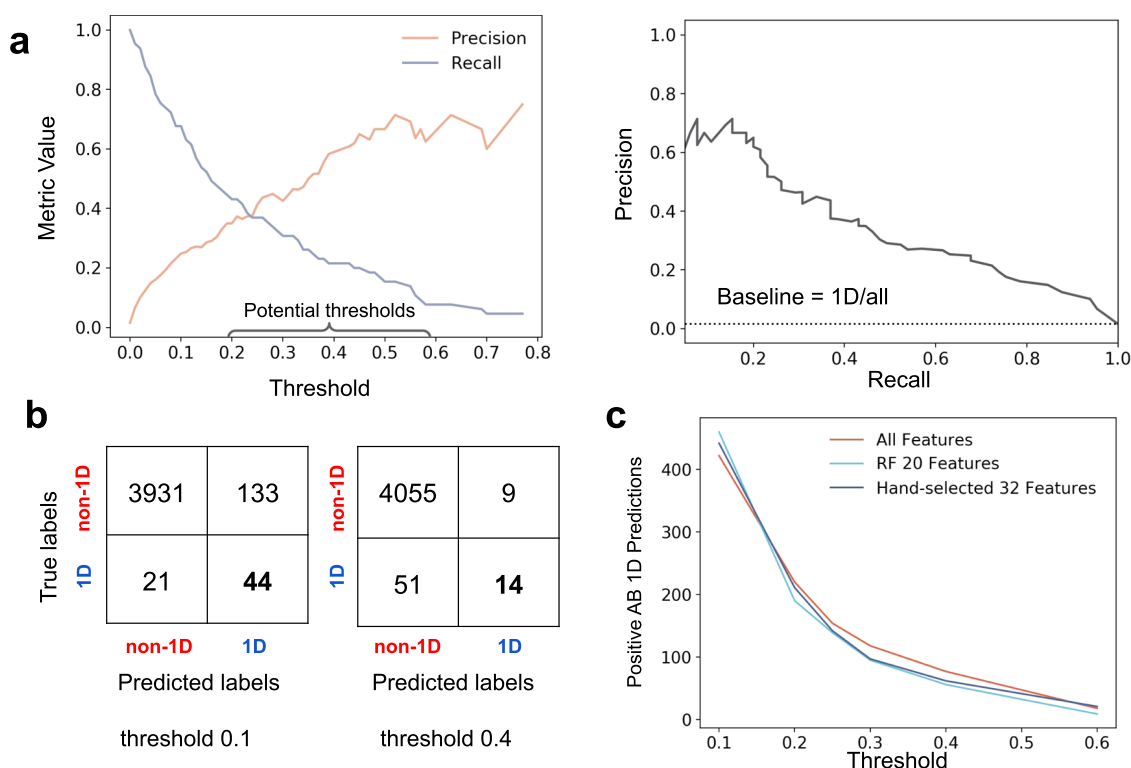


Figure 2. (a) Precision–recall curve is shown for the random forest model with the top 20 features selected using feature importance. The precision and recall are plotted as a function of the threshold for determining whether a material is 1D or not. (b) The confusion matrix is displayed at two threshold values. (c) The number of positive predicted AB binary compositions is given as a function of the threshold value.

materials such as tellurium (Te) capable of showing both 1D and 2D polymorphism.²⁴ The random forest model demonstrates higher precision even when using a reduced feature set, compared to the SVM after hyperparameter tuning.

Following model training, the task of interest is the classification of 1D versus non-1D applied to a larger space of randomly generated formulas taken using the method of ref 9. Within such an approach, all binary and ternary compositions of the form A_xB_y or $A_xB_yC_z$, where x , y , and z are coordination numbers, respectively, are taken using atomic elements up to number 94, excluding the radioactive elements Tc, Po, Ra, and At and the group 18 elements of rare gases. A large number of such random formulas are not expected to yield stable structures since they do not satisfy basic charge balancing. We accordingly filter these structures and only consider random formulas that are charge-balanced, utilizing the Pymatgen python package,²⁵ i.e., there is a combination of the common oxidation states known for each element that sums to zero. Since a high percentage of 1D materials can be captured by charge-balanced binary or ternary compositions, the random subspace is correspondingly restricted. Due to the large presence of materials containing a chalcogen, pnictogen, or halogen element (or one of these elements combined with a transition metal), we train separate models on these specific subclasses. Since 1D materials have larger coordination numbers than 2D materials, this space is expanded using the coordination from (1–5) instead of (1–3), which was previously used in efforts involving 2D materials. These compositions are termed randomly generated since they are not experimentally verified compositions.

Due to the limited number of positive samples, different methods of feature selection are used to reduce the feature set

and limit overfitting. A set of 16 hand-selected features likely to play a role in 1D material formation is identified based on physical intuition. The sum and the product variants of these features, as in the method of ref 26, yields 32 total features. For the random forest model, as feature importance can be assigned, a set based on the top 20 features is implemented for the full data set. A reduced set of the top 10 features is used for the individual chalcogen-, pnictogen-, and halogen-based feature sets due to the smaller data size. We find the relevant overlap between our hand-selected set, and the random forest-selected top features are whether a constituent element is a metal, electronegativity, whether a constituent element is an alkaline metal, atomic volume, Mendelev number, first ionization energy, periodic table column number, and covalent radius, indicating that these characteristics of the composing elements are predictive for the structural 1D nature.

To determine the relevant metric of interest, we note that the quality of a machine-learning model can be defined in terms of precision and recall. Precision is the fraction of predicted 1D materials that are indeed 1D and is the metric of interest in our approach since we target 1D materials for experimental synthesis where false positives are costly. On the other hand, recall is the ratio of 1D materials that are captured out of the total available space. Given the limited number of present experimental efforts on 1D vdW structures, it is unlikely this latter space is saturated, and correspondingly, we place less emphasis on the recall.

For predictions based on SVM models, instead of using a boolean output value indicating 1D or non-1D material based on the location of a feature vector relative to the separating hyperplane, we generate a continuous classification between 0 and 1 using Platt scaling.²⁷ For random forest models, a probability can likewise be determined by the percentage of

individual decision trees (100 used) where the individual tree predicts a composition to be 1D. The relevant threshold of the probability prediction can be chosen not only based on the precision–recall curve for the test set but also on the number of predicted positive 1D materials when the model is used for screening. A higher threshold yields a shorter candidate list, which often contains materials that are related by atomic substitution but likely contain a larger fraction of materials that are indeed 1D. Conversely, a lower threshold allows for more diverse candidates that are not likely to be related to one another by simpler atomic substitutions. The precision and recall metrics as a function of the probability cutoff threshold and the precision–recall curve are shown in Figure 2a. The confusion matrix of test set predictions (true positives, false positives, false negatives, and true negatives) is compared for two different cutoff thresholds in Figure 2b. Figure 2c demonstrates the cutoff threshold determining the number of predicted positive 1D compositions for the random forest model using different feature sets. The predicted 1D binary materials for different thresholds of positive classification are included in the Supporting Information (SI).

Since the targeted application of our approach is the experimental synthesis of novel 1D materials, we further separate the compositions by the material class, determined by commonly available precursors for chemical vapor transport (CVT) and chemical vapor deposition (CVD), examining compounds containing members of the pnictogen (group 15), chalcogen (group 16), and halogen (group 17) classes as well as a transition-metal element combined with a member of these groups. The class, number of materials, and resultant machine-learning metrics are given in Table 2. Predicted 1D materials from the chalcogen-only trained model are SbS_2 , SbS_3 , SbTe_2 , Te_2As , Te_2I_3 , Te_2S , Te_2Se , Te_2Se_3 , Te_3As , Te_3Br , Te_3Br_2 , Te_3Cl , Te_3I , Te_3I_2 , Te_3P , Te_3S , Te_3Se , Te_3Se_2 , TeBr , TeSe , TeSe_2 , and TeSe_3 . Predicted materials from the halogen-trained model include AsBr , AsI , CrI_3 , GeCl_2 , MnI_3 , MoI_3 , PtBr_2 , SbBr , TaI_3 , Te_3Br , Te_3F , Te_3I_2 , and VI_3 . Pnictogen-trained model compounds include AsI , SbI , SbS_2 , and Te_2As . These do not appear in the Materials Project with corresponding ICSD entries.

Table 2. Random Forest Performance by Material Class^a

class	accuracy	precision	recall
pnictogens	0.983	1	0.16
chalcogens	0.977	0.75	0.25
halogens	0.962	0.75	0.25
TM chalcogenides	0.985	1.0	0.5
TM halides	0.865	0.5	0.2

^aPerformance based on class is shown for the random forest model using the top 10 features selected from training on the entire composition set. The top features for the chalcogen, halogen, and pnictogen classes are similar to the features trained on the entire set due to the larger number of materials in these classes, whereas the features for the transition metal (TM) versions of these classes differ.

One issue that arises in our analysis is 1D materials are intrinsically rarer by nature: 1–2% of all materials, judging by the number of materials in the Materials Project that have ICSD tags associated with their experimental synthesis. However, the material space in ICSD does not necessarily reflect the search space. For instance, only 80% of the non-1D material in ICSD is either binary or ternary, but all of the materials we study are

either ternary or binary. This motivates us to further validate the performance of our models by mixing a portion of random formulas that are not 1D into the negative class of the training space (see the Supporting Information for details). We keep the test set unchanged to ensure a consistent performance metric between the original model and the one with the negative class mixed back into the training set. The unbalanced positive–negative class distribution motivates this approach, which can be seen as a variant similar to semisupervised or partially supervised learning. We see a large amount of predicted overlap results from nonmixed models, demonstrating prediction robustness.

To further address imbalance of the positive versus negative classes, we investigated the assignment of different class weights and implemented oversampling with the synthetic minority oversampling technique (SMOTE),²⁸ undersampling the non-1D class with Tomek Links,²⁹ as well as the combined method SMOTE-Tomek Links using the imbalance-learn python library.³⁰ While there are slight metric changes, the overall output-predicted 1D compositions remain comparable. In particular, we find that the original balanced random forest classifier, where weights are assigned inversely proportional to frequency, outperforms the SMOTE and Tomek Links methods with a higher area under the precision–recall curve (AUC-PR). This indicates better performance across all probability thresholds, whereas the SMOTE-based methods can show increased metric performance at particular thresholds. There is a high (>70%) overlap between the predicted output binary compositions across the different sampling methods with additional details in the SI.

Our machine-learning technique can not only predict new 1D materials but also find materials with particular electronic and magnetic properties of interest. One popular method is to directly train students on the limited subset of materials that have the desired properties. This would involve repeating the machine-learning technique applied in this work but on a more restricted positive class. We found this approach not to be robust due to the small positive class size. We instead use another approach, where we train a second model, as a postprocessing step, on the original set of predicted materials as transfer learning has been shown to yield robust results.³¹ Table 3 displays the test metrics for magnetic and conductive predictions using SVM classifiers on whether a particular composition contains a magnetic moment larger than $1 \mu\text{B}$ or a 0 eV band gap, respectively, according to data from the Materials Project database. We furthermore train an SVM regressor for values of the electronic band gap with the results shown in Table 4.

The 1D materials, using a random forest model, separated by the predicted band gap are given in Table 4. Compositions predicted using other models and feature sets are included in SI.

Finally, we explore the potential to predict 1D vdW materials with multiple phases, which is useful for phase-change devices. Examining known 1D materials, there are 45 compositions supporting multiple 1D phases and 131 compositions with another phase that is not 1D. Similarly to the case of determining the electronic and magnetic properties of novel 1D systems, we introduce a secondary random forest model trained on data from the Materials Project to identify whether a particular composition displays two different crystal structures.

We found a number of predicted 1D compositions to have multiple phases: AsSe_2 , Te_2As , Te_3P_2 , and TeSe_2 . This prediction is based on a stringent threshold of 0.55 in our random forest model, which best matches the ratio of 1D structures predicted to host multiple phases, compared to data in

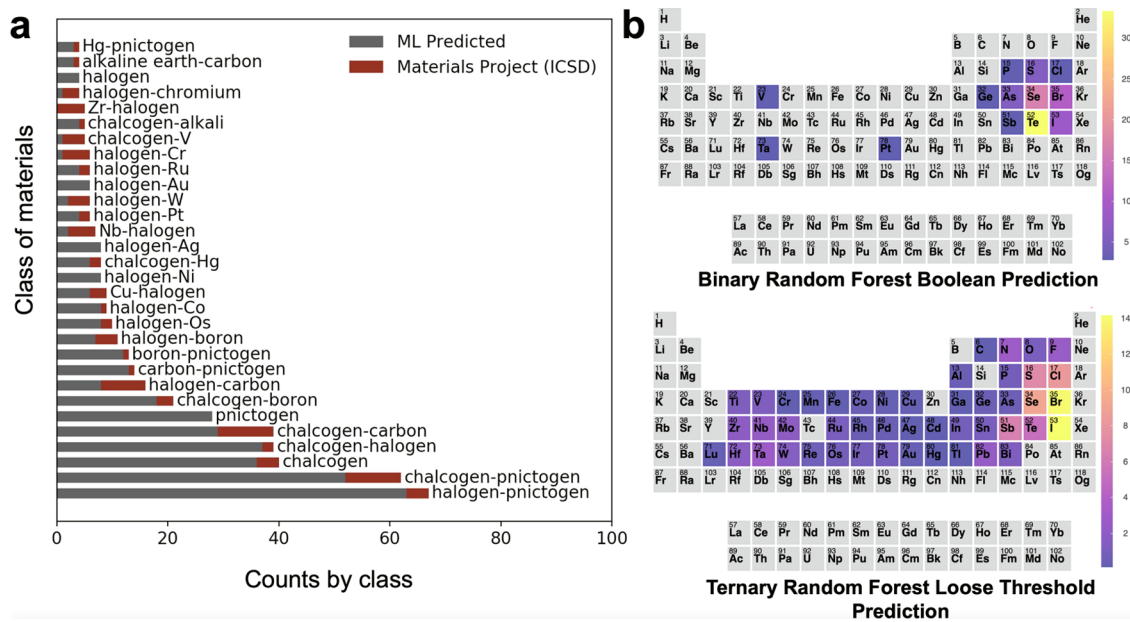


Figure 3. (a) Existing Materials Project 1D binary compositions and the predicted 1D compositions are grouped by class. Each transition metal is distinct, whereas other elements are classified by their periodic table group and named using the representative element with the lowest atomic number. For the 2D binary compositions, a unique class of material with no previous synthesized equivalents, nitrogen-based compounds, appears. (b) The distribution of elemental frequency across the periodic table is shown for binary compositions (top) and then for ternary compositions using a looser threshold (bottom), demonstrating the tunable distribution shift and increased diversity of composition types possible.

Table 3. Magnetic and Conductive Performance^a

dimension	positive	negative	baseline	accuracy	precision	recall
Magnetic						
1D	156	491	31.8	0.788	0.55	0.688
2D	341	1081	31.5	0.812	0.577	0.857
3D	5301	13825	38.3	0.892	0.755	906
Conductive						
1D	98	549	17.9	0.8	0.4	0.6
2D	451	971	4.65	0.826	0.698	0.804
3D	13210	5916		0.883	0.942	0.886

^aSince we aim for magnetic and/or conductive materials, we train on band gaps and magnetic moments calculated by density functional theory, separating the materials based on the dimensionality class. The cutoff for our classification of a material as magnetic is 1 Bohr magneton (μ_B) per unit cell. Metrics are specified for the test set. For magnetic properties, the precision is highest for the bulk (3D) materials, but the difference is not very significant across different dimension classes. For conductivity, the precision is significantly higher for non-1D and non-0D materials, posing a potential challenge for prediction.

the Materials Project database. We note that chalcogens such as Te and Se are known to participate in compounds that display multiple structures, which is captured by our model. We further predict up to 10 000 ternary formulas that host multiple structures.

To demonstrate the novelty of the predicted materials and determine how much overlap is present between the predicted 1D materials compared to the existing classes, Figure 3 presents the number of both existing and predicted materials for common families. Materials are classified into families based on the presence of specific transition metals, and for all of the other elements, they are classified according to their periodic table group. The results show an overlap of common families with the

Table 4. Binary Composition Band Gap Predictions^a

band gap	predicted 1D compositions
<0.5 eV	Pb ₂ S ₃ , RuI ₃ , SbTe ₂ , TaI ₃ , Te ₂ S, Te ₂ Se, Te ₂ Se ₃ , Te ₃ S ₂ , Te ₃ Se ₂ , TeSe, VI ₃
0.5 ≤ x < 1.5 eV	AsSe ₂ , AsSe ₃ , MoI ₃ , Sb ₂ Se, Sb ₃ Se ₂ , SbS ₂ , SbS ₃ , SbSe, SbSe ₂ , SbTe ₃ , Te ₂ As, Te ₂ O, Te ₃ As, Te ₃ Br, Te ₃ Br ₂ , Te ₃ Cl, Te ₃ I, Te ₃ I ₂ , Te ₃ Se, TeBr, TeS, TeSe ₂ , TeSe ₃
1.5 ≤ x < 2.5 eV	Sb ₂ I, Sb ₂ I ₃ , Sb ₃ I, Sb ₃ I ₂ , SbI, SbSe ₃
2.5 ≤ x	Sb ₂ Br, Sb ₂ Br ₃ , Sb ₃ Br, Sb ₃ Br ₂ , SbBr, SbCl

^a1D binary compositions, generated by a random forest model trained on all 1D materials, are classified by the predicted band gap, determined using a separate support vector regression model applied as a poststep. For all of the materials in this list, there exists an existing known material of the same family class. Band gap predictions for model results trained on the separated chalcogen, pnictogen, and halogen classes are included in the SI.

existing 1D materials, indicating that the predictions are heuristically reasonable, while also revealing novel families.

In particular, we note the appearance of a predicted class of nitrogen-based pnictogen compounds. This is a surprising result, as this form of compound was not included in the training data as Materials Project database entries with the corresponding ICSD number. Yet there are previous reports of nitrogen-involved 1D wire-like chains formed under high pressures³² or nitrogen-involved other low-dimensional forms,³³ likely indicating model predictive power. On the other hand, the predictions of pure halogens by the model, which are less likely to exist, might suggest a lower bound for the precision in the model, below which one expects considerably lower confidence in the predicted compounds.

One such highlighted 1D material, MoI₃, was subsequently synthesized using CVT and verified to be one-dimensional, supporting the applicability of our machine-learning predictions.

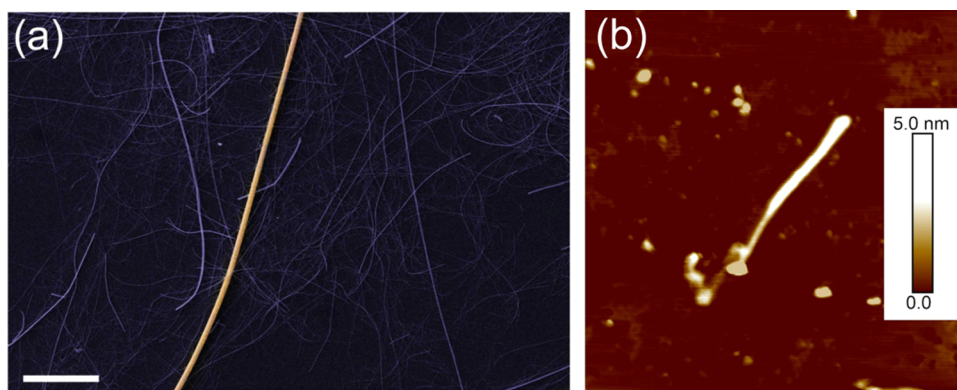


Figure 4. (a) Scanning electron microscopy (SEM) and (b) atomic force microscopy (AFM) characterization of MoI_3 nanowires after liquid-phase exfoliation in butyl alcohol. The scale bar in the SEM image is $20\ \mu\text{m}$. Note the one-dimensional morphology of the threads with high aspect ratios. The AFM image shows the presence of nanowires with a diameter as small as $0.9\ \text{nm}$, which is close to the thickness of bundles with a few atomic chains.

This composition was chosen for experimental verification based on it being predicted by our models even after employing a high threshold value for the Platt probability, suggesting that it is highly likely to exist as 1D in nature. Previous reports suggested a 2D layered structure of this material, with only one recent report of chain-like synthesis,³⁴ alongside structures of cluster chains and connected 2D sheets of Mo–I-based structures, with no mention of “1D”. This reference was found through a literary search on the predicted shortlist but was not yet included in existing databases. While this material is present in the Materials Project as a submitted entry, it does not have an associated ICSD number, indicating that the machine-learning model did not come across this material in training. This is a successful case of the model behavior, given that the material was selected from the predicted composition list and then confirmed to be one-dimensional in nature.

To verify the one-dimensional structure of the synthesized crystals, we conducted scanning electron microscopy (SEM) and atomic force microscopy (AFM) characterization of the CVT-grown crystals exfoliated via liquid-phase exfoliation (LPE) in butyl alcohol, as shown in Figure 4. The SEM image clearly shows the threads of MoI_3 nanowires with high aspect ratios. Among these nanowires, we located some threads with a thickness as small as $0.9\ \text{nm}$, close to the thickness of a few single-atom chain structures. While this material is still undergoing further characterization, it highlights the strength of our machine-learning-based prediction of novel 1D materials and the feasibility of such predictions. Further details about the experimental characterization are described in ref 35

IV. CONCLUSIONS

In conclusion, we reveal predicted candidates for 1D van der Waals materials, focusing on those with magnetic and conductive properties and chemical compositions that are likely accessible via common experimental synthesis techniques. We find that one such predicted compound, MoI_3 , is experimentally demonstrated to be low-dimensional. To address the concern of potential overfitting, we trained on different material composition subspaces, including mixing-in a portion of the screening space as part of the negative space to best recreate the composition distribution. The predictions using these various subspaces and models show a high amount of overlap, indicating the robustness of this approach.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acs.jpcc.3c03882>.

SVM and random forest model parameters including probability threshold selection; feature selection; sampling; and predicted output compositions and density functional theory calculations; namely; exfoliation energies and electronic band-structures for the synthesized composition MoI_3 (PDF)

■ AUTHOR INFORMATION

Corresponding Author

Felipe H. da Jornada – Department of Materials Science and Engineering, Stanford University, Stanford, California 94305, United States; Stanford PULSE Institute, SLAC National Accelerator Laboratory, Menlo Park, California 94025, United States; orcid.org/0000-0001-6712-7151; Email: jornada@stanford.edu

Authors

Yanbing Zhu – Department of Applied Physics, Stanford University, Stanford, California 94305, United States; orcid.org/0000-0002-6850-2926

Evan R. Antoniuk – Materials Science Division, Physical and Life Sciences Directorate, Lawrence Livermore National Laboratory, Livermore, California 94550, United States; orcid.org/0000-0002-4273-1974

Dylan Wright – Department of Material Science and Engineering, University of California Riverside, Riverside, California 92521, United States

Fariborz Kargar – Department of Electrical Engineering, University of California Riverside, Riverside, California 92521, United States; orcid.org/0000-0003-2192-2023

Nicholas Sasing – Department of Chemistry, University of Georgia, Athens, Georgia 30602, United States; orcid.org/0000-0003-2162-0236

Austin D. Sendek – Department of Materials Science and Engineering, Stanford University, Stanford, California 94305, United States; orcid.org/0000-0003-3338-1615

Tina T. Salguero – Department of Chemistry, University of Georgia, Athens, Georgia 30602, United States; orcid.org/0000-0001-9396-3583

Ludwig Bartels – Department of Chemistry, University of California Riverside, Riverside, California 92521, United States; orcid.org/0000-0002-3964-4538

Alexander A. Balandin – Department of Material Science and Engineering, University of California Riverside, Riverside, California 92521, United States; Department of Electrical Engineering, University of California Riverside, Riverside, California 92521, United States; orcid.org/0000-0002-9944-7894

Evan J. Reed – Department of Materials Science and Engineering, Stanford University, Stanford, California 94305, United States; Stanford PULSE Institute, SLAC National Accelerator Laboratory, Menlo Park, California 94025, United States

Complete contact information is available at:
<https://pubs.acs.org/10.1021/acs.jpcc.3c03882>

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This research was supported by the National Science Foundation (NSF) through the Designing Materials to Revolutionize and Engineer our Future (DMREF) awards DMREF-1921958 and 1922312. Electronic structure calculations of selected 1D materials were supported by the NSF award DMR-2103721. Calculations used computational resources from the Texas Advanced Computing Center (TACC) at The University of Texas at Austin, funded by the NSF award 1818253.

REFERENCES

- (1) 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*, No. 011002, DOI: [10.1063/1.4812323](https://doi.org/10.1063/1.4812323).
- (2) Belsky, A.; Hellenbrandt, M.; Karen, V. L.; Luksch, P. New Developments in the Inorganic Crystal Structure Database (ICSD): Accessibility in Support of Materials Research and Design. *Acta Crystallogr., Sect. B: Struct. Sci.* **2002**, *58*, 364–369.
- (3) Kirklin, S.; Saal, J. E.; Meredig, B.; Thompson, A.; Doak, J. W.; Aykol, M.; Rühl, S.; Wolverton, C. The Open Quantum Materials Database (OQMD): Assessing the Accuracy of DFT Formation Energies. *npj Comput. Mater.* **2015**, *1*, No. 15010, DOI: [10.1038/npjcompumats.2015.10](https://doi.org/10.1038/npjcompumats.2015.10).
- (4) Gražulis, S.; Chateigner, D.; Downs, R. T.; Yokochi, A.; Quirós, M.; Lutterotti, L.; Manakova, E.; Butkus, J.; Moeck, P.; Le Bail, A. Crystallography Open Database—an Open-Access Collection of Crystal Structures. *J. Appl. Crystallogr.* **2009**, *42*, 726–729.
- (5) Groom, C. R.; Bruno, I. J.; Lightfoot, M. P.; Ward, S. C. The Cambridge Structural Database. *Acta Crystallogr., Sect. B: Struct. Sci., Cryst. Eng. Mater.* **2016**, *72*, 171–179.
- (6) Brown, K. A.; Brittan, S.; Maccaferri, N.; Jariwala, D.; Celano, U. Machine Learning in Nanoscience: Big Data at Small Scales. *Nano Lett.* **2020**, *20*, 2–10.
- (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) Cheon, G.; Duerloo, K.-A. N.; Sendek, A. D.; Porter, C.; Chen, Y.; Reed, E. J. Data Mining for New Two- and One-Dimensional Weakly Bonded Solids and Lattice-Commensurate Heterostructures. *Nano Lett.* **2017**, *17*, 1915–1923.
- (9) Cheon, G.; Cubuk, E. D.; Antoniuk, E. R.; Blumberg, L.; Goldberger, J. E.; Reed, E. J. Revealing the Spectrum of Unknown Layered Materials with Superhuman Predictive Abilities. *J. Phys. Chem. Lett.* **2018**, *9*, 6967–6972.
- (10) Oliynyk, A. O.; Antono, E.; Sparks, T. D.; Ghadbeigi, L.; Gaultois, M. W.; Meredig, B.; Mar, A. High-Throughput Machine-Learning-Driven Synthesis of Full-Heusler Compounds. *Chem. Mater.* **2016**, *28*, 7324–7331.
- (11) Iwasaki, Y.; Takeuchi, I.; Stanev, V.; Kusne, A. G.; Ishida, M.; Kirihaara, A.; Ihara, K.; Sawada, R.; Terashima, K.; Someya, H.; et al. Machine-Learning Guided Discovery of a New Thermoelectric Material. *Sci. Rep.* **2019**, *9*, No. 2751.
- (12) Zhu, Y.; Rehn, D. A.; Antoniuk, E. R.; Cheon, G.; Freitas, R.; Krishnapriyan, A.; Reed, E. J. Spectrum of Exfoliable 1D van der Waals Molecular Wires and Their Electronic Properties. *ACS Nano* **2021**, *15*, 9851–9859.
- (13) Balandin, A. A.; Kargar, F.; Salguero, T. T.; Lake, R. K. One-Dimensional van der Waals Quantum Materials. *Mater. Today* **2022**, *55*, 74–91.
- (14) Liu, C.; Chen, H.; Wang, S.; Liu, Q.; Jiang, Y.-G.; Zhang, D. W.; Liu, M.; Zhou, P. Two-Dimensional Materials for Next-Generation Computing Technologies. *Nat. Nanotechnol.* **2020**, *15*, 545–557.
- (15) Liu, X.; Guo, Q.; Qiu, J. Emerging Low-Dimensional Materials for Nonlinear Optics and Ultrafast Photonics. *Adv. Mater.* **2017**, *29*, No. 1605886, DOI: [10.1002/adma.201605886](https://doi.org/10.1002/adma.201605886).
- (16) Han, R.; Li, W.; Pan, W.; Zhu, M.; Zhou, D.; Li, F.-S. 1D Magnetic Materials of Fe₃O₄ and Fe with High Performance of Microwave Absorption Fabricated by Electrospinning Method. *Sci. Rep.* **2014**, *4*, No. 7493, DOI: [10.1038/srep07493](https://doi.org/10.1038/srep07493).
- (17) Sangwan, V. K.; Hersam, M. C. Electronic Transport in Two-Dimensional Materials. *Annu. Rev. Phys. Chem.* **2018**, *69*, 299–325.
- (18) Moustafa, H.; Larsen, P. M.; Gjerding, M. N.; Mortensen, J. J.; Thygesen, K. S.; Jacobsen, K. W. Computational Exfoliation of Atomically Thin One-Dimensional Materials with Application to Majorana Bound States. *Phys. Rev. Mater.* **2022**, *6*, No. 064202, DOI: [10.1103/PhysRevMaterials.6.064202](https://doi.org/10.1103/PhysRevMaterials.6.064202).
- (19) Moustafa, H.; Lyngby, P. M.; Mortensen, J. J.; Thygesen, K. S.; Jacobsen, K. W. Hundreds of New, Stable, One-Dimensional Materials from a Generative Machine Learning Model. **2022**, arXiv:2210.08878. arXiv.org e-Print archive. <https://arxiv.org/abs/2210.08878>.
- (20) Pickard, C. J.; Needs, R. Ab initio Random Structure Searching. *J. Phys.: Condens. Matter* **2011**, *23*, No. 053201, DOI: [10.1088/0953-8984/23/5/053201](https://doi.org/10.1088/0953-8984/23/5/053201).
- (21) Bartel, C. J.; Trewartha, A.; Wang, Q.; Dunn, A.; Jain, A.; Ceder, G. A Critical Examination of Compound Stability Predictions from Machine-Learned Formation Energies. *npj Comput. Mater.* **2020**, *6*, No. 97, DOI: [10.1038/s41524-020-00362-y](https://doi.org/10.1038/s41524-020-00362-y).
- (22) Talirz, L.; Kumbhar, S.; Passaro, E.; Yakutovich, A. V.; Granata, V.; Gargiulo, F.; Borelli, M.; Uhrin, M.; Huber, S. P.; Zoupanos, S.; et al. Materials Cloud, a Platform for Open Computational Science. *Sci. Data* **2020**, *7*, No. 299, DOI: [10.1038/s41597-020-00637-5](https://doi.org/10.1038/s41597-020-00637-5).
- (23) Gjerding, M. N.; Taghizadeh, A.; Rasmussen, A.; Ali, S.; Bertoldo, F.; Deilmann, T.; Knøsgaard, N. R.; Kruse, M.; Larsen, A. H.; Manti, S.; et al. Recent Progress of the Computational 2D Materials Database (C2DB). *2D Mater.* **2021**, *8*, No. 044002, DOI: [10.1088/2053-1583/ac1059](https://doi.org/10.1088/2053-1583/ac1059).
- (24) Qiu, G.; Charnas, A.; Niu, C.; Wang, Y.; Wu, W.; Ye, P. D. The Resurrection of Tellurium as an Elemental Two-Dimensional Semiconductor. *npj 2D Mater. Appl.* **2022**, *6*, No. 17, DOI: [10.1038/s41699-022-00293-w](https://doi.org/10.1038/s41699-022-00293-w).
- (25) Ong, S. P.; Richards, W. D.; Jain, A.; Hautier, G.; Kocher, M.; Cholia, S.; Gunter, D.; Chevrier, V. L.; Persson, K. A.; Ceder, G. Python Materials Genomics (Pymatgen): A Robust, Open-Source Python Library for Materials Analysis. *Comput. Mater. Sci.* **2013**, *68*, 314–319.
- (26) Ward, L.; Agrawal, A.; Choudhary, A.; Wolverton, C. A General-Purpose Machine Learning Framework for Predicting Properties of Inorganic Materials. *npj Comput. Mater.* **2016**, *2*, No. 16028, DOI: [10.1038/npjcompumats.2016.28](https://doi.org/10.1038/npjcompumats.2016.28).

- (27) Platt, J. Probabilistic Outputs for Support Vector Machines and Comparisons to Regularized Likelihood Methods. *Adv. Large Margin Classifiers* **1999**, *10*, 61–74.
- (28) Chawla, N. V.; Bowyer, K. W.; Hall, L. O.; Kegelmeyer, W. P. SMOTE: Synthetic Minority Over-sampling Technique. *J. Artif. Intell. Res.* **2002**, *16*, 321–357.
- (29) Tomek, I. Two Modifications of CNN. *IEEE Trans. Syst., Man, Cybernetics* **1976**, SMC-6, 769–772.
- (30) Lemaitre, G.; Nogueira, F.; Aridas, C. K. Imbalanced-learn: A Python Toolbox to Tackle the Curse of Imbalanced Datasets in Machine Learning. *J. Mach. Learn. Res.* **2017**, *18*, 1–5.
- (31) Gupta, V.; Choudhary, K.; Tavazza, F.; Campbell, C.; Liao, W.-k.; Choudhary, A.; Agrawal, A. Cross-Property Deep Transfer Learning Framework for Enhanced Predictive Analytics on Small Materials Data. *Nat. Commun.* **2021**, *12*, No. 6595.
- (32) Bykov, M.; Bykova, E.; Aprilis, G.; Glazyrin, K.; Koemets, E.; Chuvashova, I.; Kuppenko, I.; Mc-Cammon, C.; Mezouar, M.; Prakapenka, V.; et al. Fe-N System at High Pressure Reveals a Compound Featuring Polymeric Nitrogen Chains. *Nat. Commun.* **2018**, *9*, No. 2756, DOI: [10.1038/s41467-018-05143-2](https://doi.org/10.1038/s41467-018-05143-2).
- (33) Laniel, D.; Winkler, B.; Fedotenko, T.; Pakhomova, A.; Chariton, S.; Milman, V.; Prakapenka, V.; Dubrovinsky, L.; Dubrovinskaia, N. High-Pressure Polymeric Nitrogen Allotrope with the Black Phosphorus Structure. *Phys. Rev. Lett.* **2020**, *124*, No. 216001, DOI: [10.1103/PhysRevLett.124.216001](https://doi.org/10.1103/PhysRevLett.124.216001).
- (34) Ströbele, M.; Thalwitzer, R.; Meyer, H.-J. Facile Way of Synthesis for Molybdenum Iodides. *Inorg. Chem.* **2016**, *55*, 12074–12078.
- (35) Kargar, F.; Barani, Z.; Sesing, N. R.; Mai, T. T.; Debnath, T.; Zhang, H.; Liu, Y.; Zhu, Y.; Ghosh, S.; Biacchi, A. J.; et al. Elemental Excitations in MoI₃ One-Dimensional van der Waals Nanowires. *Appl. Phys. Lett.* **2022**, *121*, No. 221901, DOI: [10.1063/5.0129904](https://doi.org/10.1063/5.0129904).