Machine-learning accelerated identification of exfoliable two-dimensional materials

IOP Publishing Mach. Learn.: Sci. Technol. 3 (2022) 045014 https://doi.org/10.1088/2632-2153/ac9bca

MACHINE LEARNING
Science and Technology

OPEN ACCESS

RECEIVED
18 July 2022

REVISED
7 September 2022

ACCEPTED FOR PUBLICATION
18 October 2022

PUBLISHED
8 November 2022

PAPER

Machine-learning accelerated identification of exfoliable two-dimensional materials

Mohammad Tohidi Vahdat[1,2], **Kumar Varoon Agrawal**[2] and **Giovanni Pizzi**[1,3,*]

[1] Theory and Simulation of Materials (THEOS) and National Centre for Computational Design and Discovery of Novel Materials (MARVEL), EPFL, Lausanne, Switzerland
[2] Laboratory of Advanced Separations (LAS), École Polytechnique Fédérale de Lausanne (EPFL), Sion, Switzerland
[3] Laboratory for Materials Simulations (LMS), Paul Scherrer Institut (PSI), CH-5232 Villigen PSI, Switzerland
[*] Author to whom any correspondence should be addressed.

E-mail: giovanni.pizzi@epfl.ch

Keywords: two-dimensional materials, exfoliation, crystal structure, binding energy, online tool

Supplementary material for this article is available online

Abstract

Two-dimensional (2D) materials have been a central focus of recent research because they host a variety of properties, making them attractive both for fundamental science and for applications. It is thus crucial to be able to identify accurately and efficiently if bulk three-dimensional (3D) materials are formed by layers held together by a weak binding energy that, thus, can be potentially exfoliated into 2D materials. In this work, we develop a machine-learning (ML) approach that, combined with a fast preliminary geometrical screening, is able to efficiently identify potentially exfoliable materials. Starting from a combination of descriptors for crystal structures, we work out a subset of them that are crucial for accurate predictions. Our final ML model, based on a random forest classifier, has a very high recall of 98%. Using a SHapely Additive exPlanations analysis, we also provide an intuitive explanation of the five most important variables of the model. Finally, we compare the performance of our best ML model with a deep neural network architecture using the same descriptors. To make our algorithms and models easily accessible, we publish an online tool on the Materials Cloud portal that only requires a bulk 3D crystal structure as input. Our tool thus provides a practical yet straightforward approach to assess whether any 3D compound can be exfoliated into 2D layers.

1. Introduction

Two-dimensional (2D) materials have been extensively explored for various applications across several disciplines [1–7]. In part, the scientific excitement about 2D materials originates from their ultimate thinness, making them exceptionally promising for applications, e.g. in electronics [8–10], in catalysis [11–13], or in molecular separations [14–16]. Moreover, the physical properties of monolayers are often distinct from those of their parent three-dimensional (3D) materials, since charge and heat transports are restricted to a plane [17]. Therefore, this makes them attractive not only for applications, but also for fundamental science [18–21]. Furthermore, the possibility of realizing stacks of 2D layers into van-der-Waals (vdW) heterostructures opens the way to a combinatorially large number of novel properties to be explored [22–26].

Decades ago, it was demonstrated that layered vdW materials, such as metal dichalcogenides, can be mechanically and chemically exfoliated into a few or even single layers [27]. Since the experimental demonstration of the exfoliation of graphene [28], the interest in these materials and in the experimental approaches to exfoliate them has sparked [29–32]. However, experimentally synthesizing these materials is time-consuming, only a fraction of them have ever been synthesized, resulting in a limited diversity of the 2D materials that are investigated experimentally. First-principle simulations, and in particular density-functional theory (DFT) calculations, have been thus recently employed to determine the binding-energy strength between layers and consequently assess whether a structure can be exfoliated, resulting in the

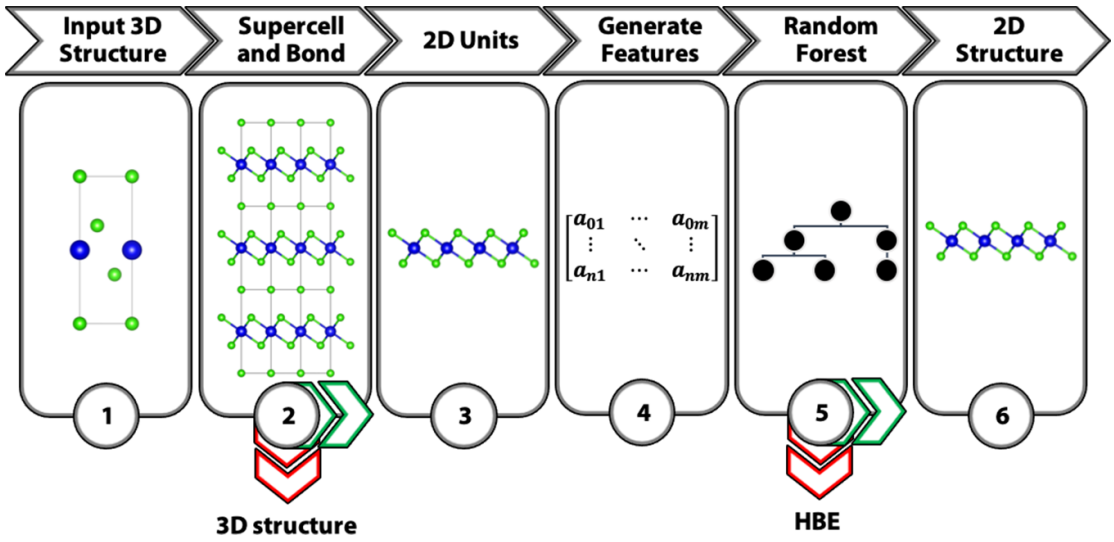


Figure 1. Schematic illustrating the steps of our algorithm to efficiently determine the exfoliability of 3D materials. From left to right: 1 the procedure initiates from the unit cell of a bulk crystal structure. 2 A $3 \times 3 \times 3$ supercell of the bulk primitive cell is constructed, and chemical bonds are identified by comparing the bond distances with the atomic vdW radii. 3 Low-dimensional subunits are determined based the rank of the lattice vectors embedded into chemically bonded units, and a 2D primitive cell is built (*LowDimFinder* geometrical pre-filtering step). 4 The crystal structures passing the previous steps are represented with numeric features, used as input of a random forest classifier (step 5) trained and tested on a set of over 2000 binding energies calculated with DFT. Our ML model predicts if the material has a high binding energy (HBE) and should be discarded, thus filtering out potential 2D structures for further analysis (step 6).

creation of databases of potential 2D materials [33–39]. However, such calculations are quite expensive and require deep understanding of the simulation methods to obtain accurate results, as well as the capability of using a workflow infrastructure to manage the resulting large number of calculations [40–43]. Therefore, first-principles simulations might be a barrier when needing a fast, yet accurate, prediction of the exfoliability of candidate materials.

Recently, different approaches based on machine learning have been employed to optimize the synthesis process of 2D materials [44–50]. Here, we develop a combined geometrical and machine-learning (ML) model that can provide accurate and fast predictions on whether a bulk 3D material is formed by layers held together by a weak binding energy and, thus, is potentially exfoliable into 2D layers. Our multi-step procedure, illustrated in figure 1, starts by pre-screening layered structures based on geometrical criteria requiring only the atomic positions of the atoms in the structure. The resulting filtered structures are featurized, and finally a ML model based on a random forest classifier is applied to assess whether the material can be exfoliated or, instead, has a high binding energy (HBE).

We train and then evaluate the performance of the model using a set of 2392 structures from [37], obtained from the preliminary geometrical step applied on a large initial set of 108 423 unique 3D structures originating from the Inorganic Crystal Structure Database (ICSD) [51] and Crystallography Open Database (COD) [52]. To capture the underlying structural patterns that correlate with the exfoliability of a material, we first reduce the number of descriptors needed by the model, and then use a SHapely Additive exPlanations (SHAP) analysis to explain the most relevant features. We also compare our best-performing ML model with a deep neural network. Finally, in order to make our methodology and code immediately available to all researchers without the need of any software installation, we implement and publish an online web tool on the Materials Cloud web platform [53], available at http://ml-layer-finder.materialscloud.io and working directly in the web browser. The tool provides the opportunity, even for non-experts of simulations, to just upload a crystal structure in a variety of file formats and rapidly evaluate if it can be exfoliated into 2D layers or not.

2. Methods

2.1. Description of the crystal structures and the method used to identify layered compounds

In the following, we will indicate exfoliable 2D materials as a positive hit of a model, while a HBE material as a negative hit. Therefore, false positives indicate HBE systems that are instead predicted to be exfoliable (2D), while false negatives indicate exfoliable systems that the model suggests to discard as HBE.

Our screening procedure (steps 1 and 2 of figure 1) starts from the comprehensive initial set of 108 423 unique bulk 3D crystals extracted from ICSD and COD, used in the study by Mounet *et al* [37, 54]. In [37], the preliminary geometric step (that we will refer to as the *LowDimFinder*) identifies chemically-bonded subunits by comparing interatomic distances d_{AB} between pairs of atoms A and B in the structure with the corresponding atomic vdW radii r_A and r_B (radii are taken from [55]). We consider two atoms as bonded if $d_{AB} \leqslant r_A + r_B - \Delta$, where Δ is a tunable parameter (see below). The dimensionality of each connected manifold is then obtained from the rank of the set of 3D lattice vectors embedded into the manifold (see [37] for more details). One main advantage of this approach is identifying subunits of any dimensionality (i.e. 2D layers, 1D chains, and 0D clusters). Moreover, in the case of 2D layers, no specific orientation is assumed for the 2D plane with respect to the Cartesian axes.

For the purposes of preliminary filtering, we wish to have a very low rate of false negatives (i.e. a very high recall rate) so as not to miss important materials, while at the same time we can accept a moderate amount of false positives, because these can be further filtered later (first by the ML model discussed here, and possibly later by accurate DFT simulations following the application of our ML model). Therefore, also considering that this first filtering is only aware of distances between atoms and not of their chemistry (except for the atomic vdW radii), we consider five values for the tunable parameter, namely $\Delta = 1.1, 1.2, 1.3, 1.4, 1.5$ Å. Any crystal structure recognized as 2D for at least one of the five values of Δ is then handed to the ML model in the following steps.

2.2. Choice of descriptors for 3D compounds

A crucial preliminary step before implementing and training a ML model is to represent crystal structures in terms of numerical features representing the material chemistry and geometry.

Different strategies have been developed in the literature [56–59]. In this work, all the descriptors are chosen from those available and implemented in the matminer package [60]. Choosing the right descriptors requires care and testing, and is dependent on the details of the descriptors and the properties of interest. As such, there is not a single unique correct way to represent crystal structures. We start from a set of potentially relevant features and, as a first step, we identify which of these are most relevant (see supplementary information section 1 for details). Our final set of features are based on those proposed by Ward *et al* [61], where space is partitioned into Wigner–Seitz cells using Voronoi tessellation. These cells (including all points in space closer to a central atom than to any other atom) are then used to assign neighbors and then construct local descriptors of the environment that are only weakly sensitive to small changes that might occur during a geometry relaxation. Derived from the Voronoi tessellation, several attributes are proposed in [61]. In the following we will restrict to three different sets of features that we selected for our final model: chemical ordering, maximum packing efficiency, and local environment attributes. The first one, chemical ordering, describes how much the placement of species in a structure deviates from random, using Warren–Cowley-like ordering parameters [62]. The second one, maximum packing efficiency, measures the largest sphere that fits inside each Voronoi cell. The last, local environment attributes, are based on comparing elemental properties (such as the electronegativity and 21 other elemental properties) of the central atom to its neighbors, weighted by the surface area of the faces of the Wigner–Seitz cells. More technical details can be found in [61].

2.3. Model selection and training

To achieve a ML model that can classify potentially exfoliable materials, we train the model on binding energies for the 2392 structures prefiltered by the *LowDimFinder*, that were computed using DFT in [37]. We first center the features at zero and rescale them using their mean and standard deviation, respectively. In order to determine the quality of the model on an unseen data set, train-test splitting is performed randomly, and the size of the train set is 70%. All the statistics reported later are computed by averaging over ten different random seeds used for train-test splitting.

We tested various learning approaches, including deep learning architectures. Although some of the architectures perform well (see supplementary information section 2), we decided to employ and discuss further in the following a random forest classifier ML model [63], as it is relatively simple and, at the same time, it demonstrates an excellent performance for our goal. The random forest classifier is trained with a number of trees of 600 and a maximum depth of trees of 17.

3. Discussion and results

Mounet *et al* [37] performed binding energy calculations on 2662 prospective layered structures and identified those that are held together by weak interactions and are ready for exfoliation. We consider 2392 structures out of 2662 (two different vdW DFT functionals are used in [37]; here, we restrict only to those for

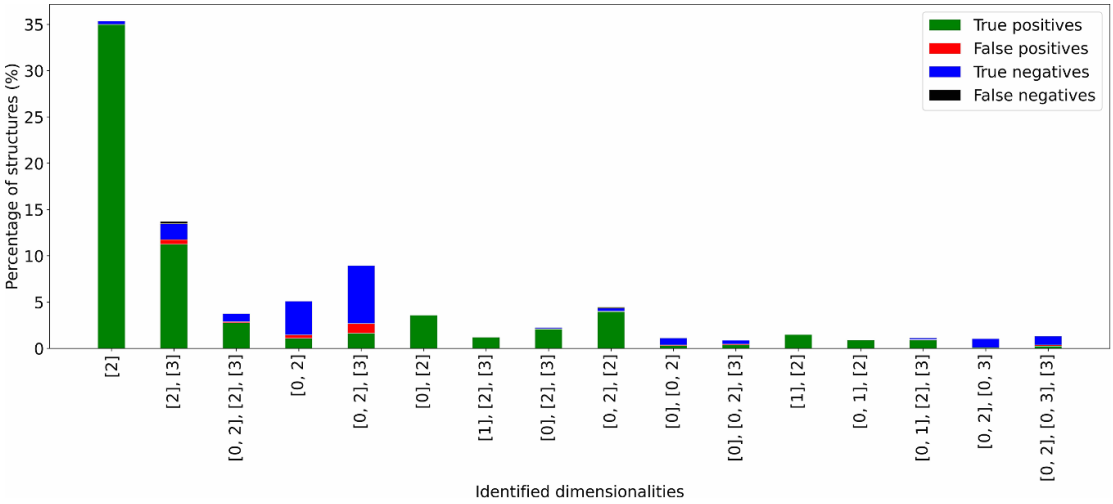


Figure 2. Evaluation of the performance of our ML-based model with respect to accurate results on the binding energies calculated using DFT. The bar labels combine the dimensionalities obtained for all values of Δ used to determine atom bonds in the preliminary geometrical *LowDimFinder* screening (see main text). For example, if the obtained dimensions of subunits for the five values of Δ are [0,2], [0,2], [0,2], [2] and [2], the resulting label reported here is [0,2],[2]. Structures are color-coded based on the results of the ML classifier. Green: True positives (i.e. structures that are correctly recognized as exfoliable by the ML model). Red: False positives (i.e. compounds incorrectly identified as exfoliable 2D). Blue: True negatives (i.e. structures correctly determined as HBE structures). Black: False negatives (i.e. structures incorrectly classified as HBE structures). The value on the *y* axis is the percentage of structures with respect to the total number of structures with at least one 2D subcomponent in the *LowDimFinder* analysis.

Table 1. Performance of our ML model (based on a random forest classifier) on the test data set. Our model captures 513 out of 525 structures correctly (true positives) resulting in a very high recall rate of 98% (i.e. the model has very few false negatives). Additionally, it reduces the number of false positives by $\approx$ 50% with respect to the initial *LowDimFinder* geometrical screening.

Label	Precision	Recall	F1-score	Total number of structures
HBE	85%	49%	62%	139
2D	88%	98%	93%	525

which binding energies are calculated using the vdW-DF2-C09 functional [64–66], so as to have a consistent calculation method of binding energies for all structures in the dataset). In particular, our portfolio contains 1813 structures that are determined to be exfoliable into 2D layers by DFT simulations including vdW interactions (vdW-DFT), defined as those whose binding energy is lower than 130 meV Å^{-2}, while 579 structures have HBE (although the first *LowDimFinder* preliminary filtering had not discarded them).

Figure 2 illustrates the combined performance of the preliminary *LowDimFinder* geometrical filtering followed by our ML model, with respect to the accurate vdW-DFT calculations, resolved with respect to the set of dimensionalities of the sub-units identified by *LowDimFinder* (the performance of *LowDimFinder* alone can be better inspected in supplementary figure S1). In particular, in figure 2 we see that systems that are tagged by *LowDimFinder* as only having 2D subcomponents for any value of Δ (labeled as `[2]`) are those with highest probability of displaying low binding energy, and thus promising for exfoliation. Moreover, this set also has a very low rate of false positives. On the other hand, structures marked also displaying a zero-dimensional subunit for at least one value of Δ (and in particular `[0,2]` and `[0,2],[3]`) include a relatively large fraction of systems that are confirmed to have HBE by vdW-DFT. Nevertheless, these results demonstrate that, while the dimensionalities found by *LowDimFinder* alone can provide an indication of the probability of finding false positives, they are not sufficient (alone) to provide a reliable method to reduce their number without also increasing significantly the rate of false negatives (we remind that our design strategy is to have the smallest amount possible of false negatives, but avoiding to miss important 2D candidates).

Therefore, we train our ML model (based on the random forest classifier) to improve the classification of structures in 2D and HBE, still at a negligible fraction of the cost of a full DFT simulation. As visible in figure 2, the ML model is able to correctly label most HBE systems that had passed the *LowDimFinder* prefilter (true negative systems, in blue). Detailed results on the performance of our ML model is reported in table 1. In the test dataset, we have 664 structures, including 139 HBE structures and 525 exfoliable structures. Our model managed to correctly predict 513 structures as true positives (i.e. a recall of 98%, very close to our

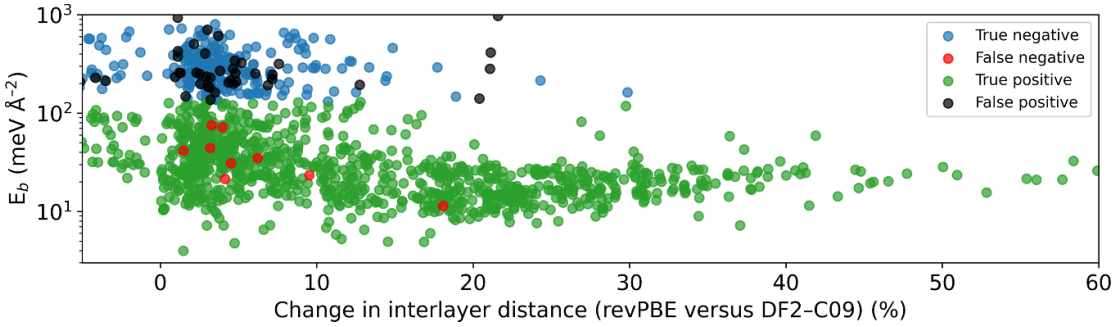


Figure 3. Binding energy versus change in the interlayer distance (for the 3D structures relaxed using the vdW-aware DF2-C09 functional with respect to the relaxation with the revPBE functional, as defined in figure 2(c) of [37]). Using our trained ML model, materials are classified as exfoliable or with high binding energy and reported in different colors. Green and blue circles show structures that are classified correctly both with vdW-DFT and our ML model as exfoliable and high binding energy structures (true positives and true negatives, respectively). Red circles indicate exfoliable structures according to vdW-DFT, but tagged as HBE by our ML model (false negatives), while black circles represent HBE structures as determined by vdW-DFT, while the ML model tags them as exfoliable (false positives).

ideal goal of 100%), while at the same time reducing the number of false positives of the *LowDimFinder* by about 50%. Therefore, our model reaches our goal of significantly filtering the number of candidates that can go to an ultimate screening step using DFT, thus reducing the associated computational cost.

To better understand the distribution of false positives and false negatives of our ML model, we first show in figure 3 the binding energies for all compounds included in our training set, as a function of the change in the interlayer distance with two DFT functionals (including or not vdW interactions, respectively, as an approximate proxy for the strength of vdW interactions in the material). This figure mirrors figure 2(c) of [37], while adding information on the performance of our ML model. The figure visualizes graphically the model performance, and shows that false positives and negatives do not have a specific distribution with respect to the two variables of the scatter plot.

We then use the SHAP technique [67] to gain a deeper understanding of what the model learned in the training process. SHAP can disclose how the features used by the model affect the predictions. Figure 4 shows the five most important variables of our model: maximum packing efficiency, maximum number of unfilled *s* states, chemical ordering (first neighbor), mean of periodic table row, and chemical ordering (third neighbor). The importance here is defined as the mean absolute SHAP value of all the points in the dataset. The colors in the figure denote the value of the input variable, where red indicates high and blue indicates low values. Consequently, red on the right-hand (left-hand) side of the plot implies a positive (negative) correlation with exfoliation.

Figure 4 shows that maximum packing efficiency is the most important feature and has a negative correlation with exfoliation, consistent with what one can expect as this feature indicates the largest sphere that fits inside each Voronoi cell, and structures with high packing efficiency will be most probably not layered (packing ratio was already used as a filtering criterion in one of the early DFT studies to predict 2D materials [33]). The second most important feature is the maximum number of unfilled *s* states. Indeed, as the number of unfilled *s* states increases, the material will tend to form stronger bonds and will be consequently less prone to break them and be exfoliated, justifying its negative correlation. Another important variable is chemical ordering (both for first and third neighbors). This variable is associated with the arrangement of the crystal structure, where structures with ordered arrangements will have a larger value (closer to 1), while random arrangements will have a value closer to zero. Also in this case, the negative correlation visible in figure 4 can be expected, because structures very ordered configurations are typically strongly bonded 3D materials that are not composed by exfoliable layers. Finally, another important feature is the mean of the periodic table row, indicating a tendency of lighter elements (i.e. small value for the periodic table row) to form exfoliable materials.

Finally, since our aim is to reduce as much as possible the number of false negatives, we report explicitly the 12 exfoliable structures that are considered as HBE by our ML model. These are schematically represented in figure 5 (the complete list is shown in supplementary information table S2). We highlight in particular that 6 structures out of 12 contain halogen elements, and 3 contain sulfur. This result can be used as a guide for further improvements of the model, e.g. to include further descriptors that can help in correctly identifying structures with low binding energy (we note, however, that other structures including the same chemical elements exist, where the model works correctly: therefore their sole presence cannot be used as a simple criterion to identify false negatives).

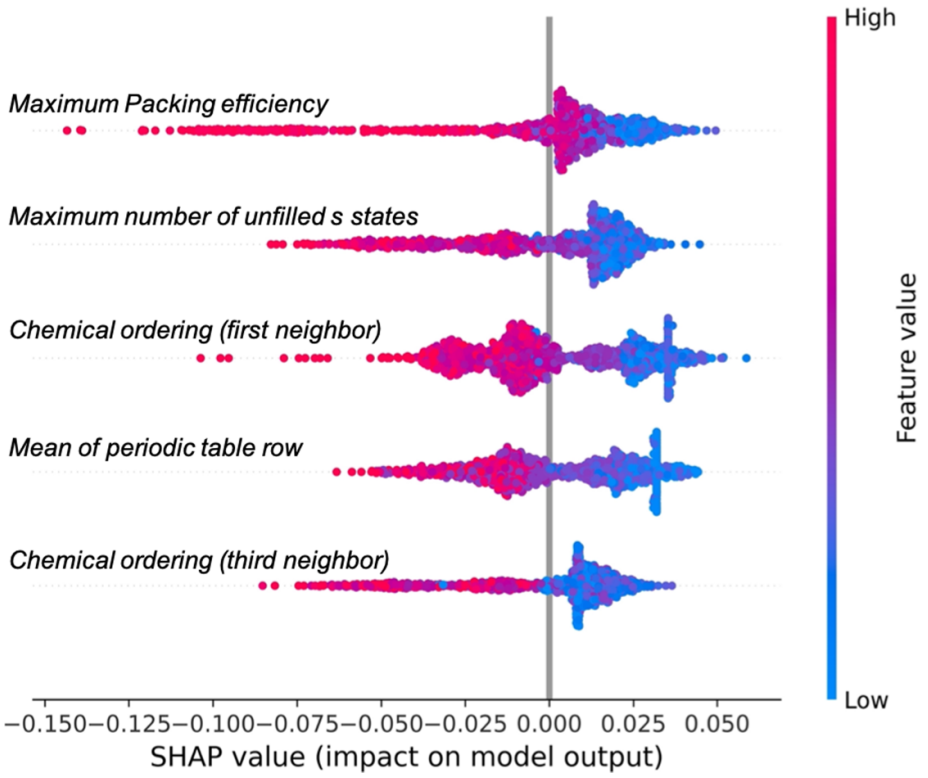


Figure 4. SHAP plot summarizing the five most important variables for our model, evaluated on all structures in the dataset. The color corresponds to the value of each input variable and reveals if there is a positive or negative correlation with the possibility of exfoliation (e.g. red circles on the left-hand side of the plot indicate a negative correlation). Additionally, the violin plots show the distribution of the importance, i.e. the spread of the SHAP values (along the *x* axis) and how many samples have similar SHAP values (the thickness of the violin).

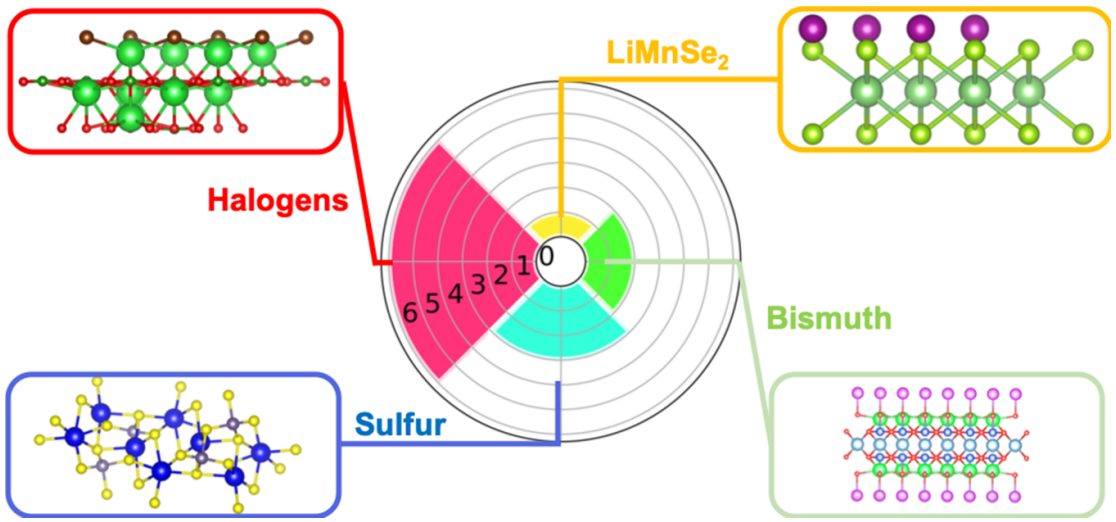
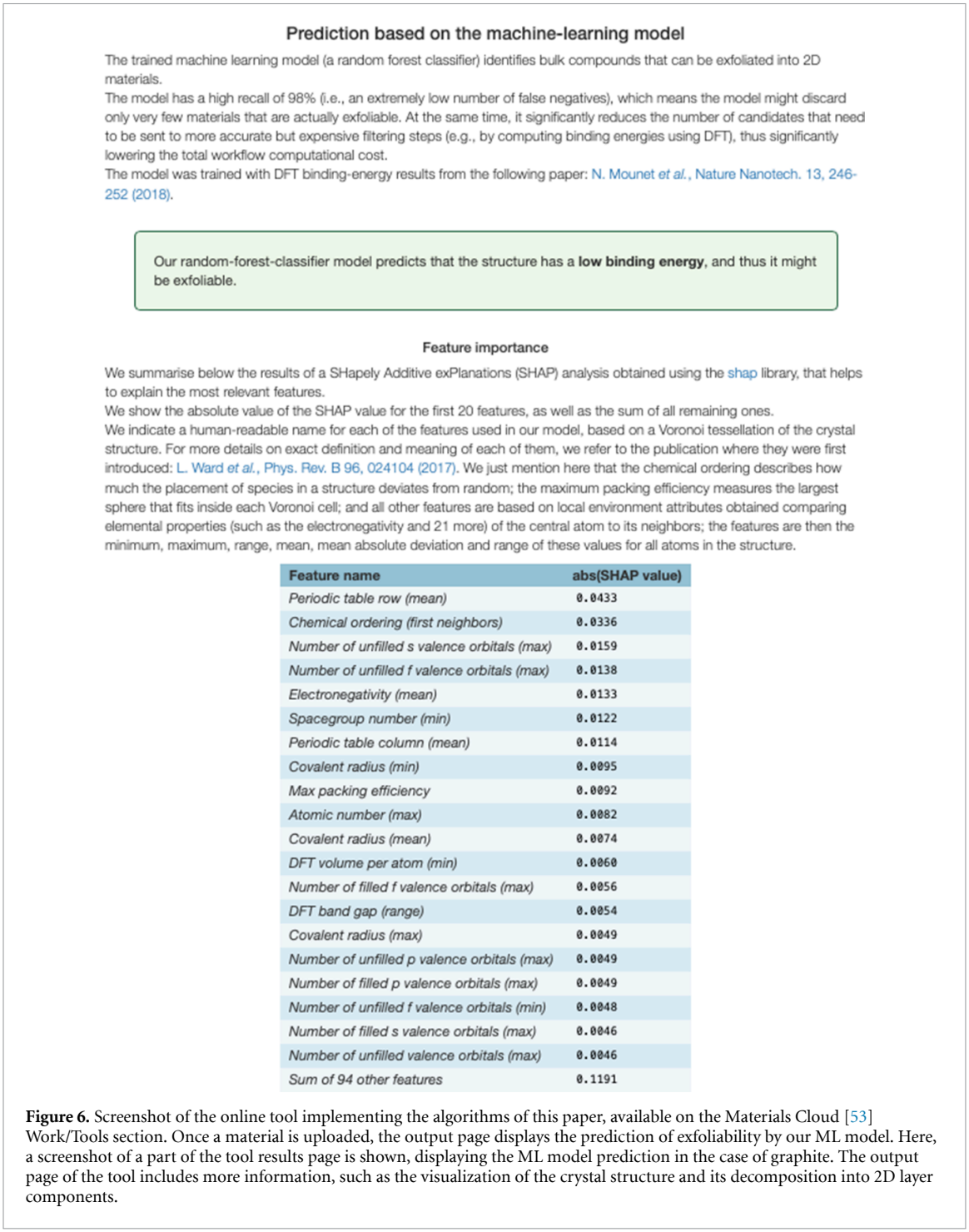


Figure 5. Summary of the 12 false negatives of our ML model (out of the 525 exfoliable structures in our test set). In addition to $LiMnSe_2$, six structures include halogens, three include sulfur and two include bismuth.

4. Online tool

In order to give easy access to our model and its predictions without the need of preliminary expertise, we also implement an online web tool that we publish on the Materials Cloud web platform [53] at the URL http://ml-layer-finder.materialscloud.io. The tool works directly in the browser and does not require any installation.

In the first selection page, the user can upload a bulk crystal structure. Several common formats are supported thanks to the parsers implemented in the Atomic Simulation Environment (ASE) [68] and



Prediction based on the machine-learning model

The trained machine learning model (a random forest classifier) identifies bulk compounds that can be exfoliated into 2D materials.
The model has a high recall of 98% (i.e., an extremely low number of false negatives), which means the model might discard only very few materials that are actually exfoliable. At the same time, it significantly reduces the number of candidates that need to be sent to more accurate but expensive filtering steps (e.g., by computing binding energies using DFT), thus significantly lowering the total workflow computational cost.
The model was trained with DFT binding-energy results from the following paper: N. Mounet *et al.*, Nature Nanotech. 13, 246-252 (2018).

Our random-forest-classifier model predicts that the structure has a **low binding energy**, and thus it might be exfoliable.

Feature importance

We summarise below the results of a SHapely Additive exPlanations (SHAP) analysis obtained using the shap library, that helps to explain the most relevant features.
We show the absolute value of the SHAP value for the first 20 features, as well as the sum of all remaining ones.
We indicate a human-readable name for each of the features used in our model, based on a Voronoi tessellation of the crystal structure. For more details on exact definition and meaning of each of them, we refer to the publication where they were first introduced: L. Ward *et al.*, Phys. Rev. B 96, 024104 (2017). We just mention here that the chemical ordering describes how much the placement of species in a structure deviates from random; the maximum packing efficiency measures the largest sphere that fits inside each Voronoi cell; and all other features are based on local environment attributes obtained comparing elemental properties (such as the electronegativity and 21 more) of the central atom to its neighbors; the features are then the minimum, maximum, range, mean, mean absolute deviation and range of these values for all atoms in the structure.

Feature name	abs(SHAP value)
Periodic table row (mean)	0.0433
Chemical ordering (first neighbors)	0.0336
Number of unfilled s valence orbitals (max)	0.0159
Number of unfilled f valence orbitals (max)	0.0138
Electronegativity (mean)	0.0133
Spacegroup number (min)	0.0122
Periodic table column (mean)	0.0114
Covalent radius (min)	0.0095
Max packing efficiency	0.0092
Atomic number (max)	0.0082
Covalent radius (mean)	0.0074
DFT volume per atom (min)	0.0060
Number of filled f valence orbitals (max)	0.0056
DFT band gap (range)	0.0054
Covalent radius (max)	0.0049
Number of unfilled p valence orbitals (max)	0.0049
Number of filled p valence orbitals (max)	0.0049
Number of unfilled f valence orbitals (min)	0.0048
Number of filled s valence orbitals (max)	0.0046
Number of unfilled valence orbitals (max)	0.0046
Sum of 94 other features	0.1191

Figure 6. Screenshot of the online tool implementing the algorithms of this paper, available on the Materials Cloud [53] Work/Tools section. Once a material is uploaded, the output page displays the prediction of exfoliability by our ML model. Here, a screenshot of a part of the tool results page is shown, displaying the ML model prediction in the case of graphite. The output page of the tool includes more information, such as the visualization of the crystal structure and its decomposition into 2D layer components.

pymatgen [69] libraries. Once the bulk structure is uploaded (or selected from already available examples), the tool performs the needed computations to first determine if the material is layered, and then to predict if it is exfoliable. In particular, the tool (following the workflow of figure 1) first computes the bonds and then identifies the dimensionality of each connected sub-manifold using the *LowDimFinder*. Any subunit that is a 2D layer is displayed on the output web page. The tool then generates feature vectors for these structures, and then it uses the pre-trained ML model to predict if the crystal structure can be exfoliated or has a HBE. The result is provided on output, together with a table summarizing the SHAP value of the most important features for the uploaded structure (see figure 6).

5. Conclusion

We presented an efficient yet accurate method combining a preliminary geometrical screening followed by a random-forest-classifier ML model to identify bulk compounds that can be exfoliated into 2D materials.

The procedure that we developed provides an effective and straightforward prefilter to assess whether any bulk structure has the potential to be exfoliated: thanks to its high recall of 98%, our method has an extremely low rate of false negatives, i.e. of exfoliable materials that would be discarded by the model and thus be missed in a search for interesting candidates. Yet, it helps in significantly reducing the number of candidates that need to be sent to more accurate but expensive filtering steps (e.g. by computing binding energies using DFT), thus significantly lowering the total workflow computational cost. By means of a SHAP analysis, we also provide an explanation of the most important features used by our model to determine exfoliability. These include the maximum packing efficiency, the maximum number of unfilled *s* sates, the chemical ordering and the mean of the periodic table row. Finally, in order to enable even non-expert fellows to run our model and predict exfoliability of new materials, we implemented and deployed it as an online web tool published on the Materials Cloud web platform, only requiring to provide the parent crystal structure as input.

Data availability statement

Supplementary information is available for this paper. Data that were used and are needed to reproduce this study are deposited on the Materials Cloud Archive [53] and are available at https://doi.org/10.24435/materialscloud:m4-7f.

Acknowledgments

This research was supported by the NCCR MARVEL, a National Centre of Competence in Research, funded by the Swiss National Science Foundation (Grant No. 205602). The authors gratefully acknowledge Nicola Marzari and Seyed Mohamad Moosavi for fruitful discussions, and Nicolas Mounet and Philippe Schwaller for the initial implementation of the *LowDimFinder* code used in [37].

Code availability

The code for parsing, featurization of structures and data analysis is available free of charge on Github at https://github.com/Mohammad-vahdat/Machine_learning_accelerated_identification_of_exfoliable_two_dimensional_materials.

ORCID iDs

Mohammad Tohidi Vahdat https://orcid.org/0000-0003-1113-5894
Giovanni Pizzi https://orcid.org/0000-0002-3583-4377

References

[1] Xia F, Wang H, Xiao D, Dubey M and Ramasubramaniam A 2014 *Nat. Photon.* **8** 899–907
[2] Fiori G, Bonaccorso F, Iannaccone G, Palacios T, Neumaier D, Seabaugh A, Banerjee S K and Colombo L 2014 *Nat. Nanotechnol.* **9** 768–79
[3] Saito Y, Nojima T and Iwasa Y 2016 *Nat. Rev. Mater.* **2** 1–18
[4] Manzeli S, Ovchinnikov D, Pasquier D, Yazyev O V and Kis A 2017 *Nat. Rev. Mater.* **2** 17033
[5] Mannix A J, Kiraly B, Hersam M C and Guisinger N P 2017 *Nat. Rev. Chem.* **1** 1–15
[6] Vahdat M T, Campi D, Colonna N, Villalobos L F, Marzari N and Agrawal K V 2020 *J. Mater. Chem.* A **8** 17747–55
[7] Vahdat M T, Campi D, Colonna N, Marzari N and Agrawal K V 2021 *J. Phys. Chem.* C **125** 18896–904
[8] Radisavljevic B, Radenovic A, Brivio J, Giacometti V and Kis A 2011 *Nat. Nanotechnol.* **6** 147–50
[9] Chhowalla M, Jena D and Zhang H 2016 *Nat. Rev. Mater.* **1** 1–15
[10] Pizzi G, Gibertini M, Dib E, Marzari N, Iannaccone G and Fiori G 2016 *Nat. Commun.* **7** 1–9
[11] Wang X, Han X, Zhang J, Wu X, Liu Y and Cui Y 2016 *J. Am. Chem. Soc.* **138** 12332–5
[12] Deng D, Novoselov K S, Fu Q, Zheng N, Tian Z and Bao X 2016 *Nat. Nanotechnol.* **11** 218–30
[13] Zhu Y, Peng L, Fang Z, Yan C, Zhang X and Yu G 2018 *Adv. Mater.* **30** 1–19
[14] Varoon K *et al* 2012 *Science* **334** 72–76
[15] Li H, Song Z, Zhang X, Huang Y, Li S, Mao Y, Ploehn H J, Bao Y and Yu M 2013 *Science* **342** 95–98
[16] Peng Y, Li Y, Ban Y, Jin H, Jiao W, Liu X and Yang W 2014 *Science* **346** 1356–9
[17] Butler S Z 2013 *ACS Nano* **7** 2898–926
[18] Moore J E 2010 *Nature* **464** 194–8
[19] Kou L, Wu S-C, Felser C, Frauenheim T, Chen C and Yan B 2014 *ACS Nano* **8** 10448–54
[20] Walsh L A and Hinkle C L 2017 *Appl. Mater. Today* **9** 504–15
[21] Marrazzo A, Gibertini M, Campi D, Mounet N and Marzari N 2018 *Phys. Rev. Lett.* **120** 117701
[22] Nayak P K *et al* 2017 *ACS Nano* **11** 4041–50
[23] Tran K *et al* 2019 *Nature* **567** 71–75

[24] Alexeev E M 2019 *Nature* **567** 81–86
[25] Pizzi G, Milana S, Ferrari A C, Marzari N and Gibertini M 2021 *ACS Nano* **15** 12509–34
[26] Marrazzo A and Gibertini M 2022 *npj 2D Mater. Appl.* **6** 30
[27] Frindt R F 1966 *J. Appl. Phys.* **37** 1928–9
[28] Novoselov K S, Geim A K, Morozov S V, Jiang D, Zhang Y, Dubonos S V, Grigorieva I V and Firsov A A 2004 *Science* **306** 666–9
[29] Kaniyoor A, Baby T T and Ramaprabhu S 2010 *J. Mater. Chem.* **20** 8467–9
[30] Coleman J N 2013 *Acc. Chem. Res.* **46** 14–22
[31] Yi M and Shen Z 2015 *J. Mater. Chem.* A **3** 11700–15
[32] Magda G Z, Petø J, Dobrik G, Hwang C, Biró L P and Tapasztó L 2015 *Sci. Rep.* **5** 3–7
[33] Lebègue S, Björkman T, Klintenberg M, Nieminen R M and Eriksson O 2013 *Phys. Rev.* X **3** 1–7
[34] Ashton M, Paul J, Sinnott S B and Hennig R G 2017 *Phys. Rev. Lett.* **118** 1–6
[35] Cheon G, Duerloo K A N, Sendek A D, Porter C, Chen Y and Reed E J 2017 *Nano Lett.* **17** 1915–23
[36] Choudhary K, Kalish I, Beams R and Tavazza F 2017 *Sci. Rep.* **7** 5179
[37] Mounet N *et al* 2018 *Nat. Nanotechnol.* **13** 246–52
[38] Haastrup S 2018 *2D Mater.* **5** 1–36
[39] Zhou J *et al* 2019 *Sci. Data* **6** 86
[40] Adhianto L, Banerjee S, Fagan M, Krentel M, Marin G, Mellor-Crummey J and Tallent N R 2010 *Concurr. Comput.* **22** 685–701
[41] Ong S P, Richards W D, Jain A, Hautier G, Kocher M, Cholia S, Gunter D, Chevrier V L, Persson K A and Ceder G 2013 *Comput. Mater. Sci.* **68** 314–9
[42] Larsen A H *et al* 2017 *J. Phys.: Condens. Matter.* **29** 273002
[43] Huber S P *et al* 2020 *Sci. Data* **7** 1–18
[44] Lin X *et al* 2018 *Nano Res.* **11** 6316–24
[45] Yang H, Zhang Z, Zhang J and Zeng X C 2018 *Nanoscale* **10** 19092–9
[46] Frey N C, Wang J, Vega Bellido G I, Anasori B, Gogotsi Y and Shenoy V B 2019 *ACS Nano* **13** 3031–41
[47] Han B 2020 *Adv. Mater.* **32** 2000953
[48] Tang B, Lu Y, Zhou J, Chouhan T, Wang H, Golani P, Xu M, Xu Q, Guan C and Liu Z 2020 *Mater. Today* **41** 72–80
[49] Jin H, Zhang H, Li J, Wang T, Wan L, Guo H and Wei Y 2020 *J. Phys. Chem. Lett.* **11** 3075–81
[50] Haraguchi Y, Igarashi Y, Imai H and Oaki Y 2022 *Digit. Discovery* **1** 26–34
[51] Inorganic Crystal Structure Database (ICSD) (available at: http://www.fiz-karlsruhe.com/icsd.html)
[52] Gražulis S, Daškevič A, Merkys A, Chateigner D, Lutterotti L, Quirós M, Serebryanaya N R, Moeck P, Downs R T and Le Bail A 2012 *Nucleic Acids Res.* **40** 420–7
[53] Talirz L *et al* 2020 *Sci. Data* **7** 299
[54] Mounet N *et al* 2020 Materials Cloud Archive 2020.158 (available at: https://archive.materialscloud.org/record/2020.158)
[55] Alvarez S 2013 *Dalton Trans.* **42** 8617–36
[56] Ward L, Agrawal A, Choudhary A and Wolverton C 2016 *npj Comput. Mater.* **2** 1–7
[57] Meredig B, Agrawal A, Kirklin S, Saal J E, Doak J W, Thompson A, Zhang K, Choudhary A and Wolverton C 2014 *Phys. Rev.* B **89** 1–7
[58] Bartók A P, Kondor R and Csányi G 2013 *Phys. Rev.* B **87** 1–16
[59] De S, Bartók A P, Csányi G and Ceriotti M 2016 *Phys. Chem. Chem. Phys.* **18** 13754–69
[60] Ward L *et al* 2018 *Comput. Mater. Sci.* **152** 60–69
[61] Ward L, Liu R, Krishna A, Hegde V I, Agrawal A, Choudhary A and Wolverton C 2017 *Phys. Rev.* B **96** 1–12
[62] Cowley J M 1950 *Phys. Rev.* **77** 669–75
[63] Breiman L 2001 *Mach. Learn.* **45** 1573–0565
[64] Lee K, Murray E D, Kong L, Lundqvist B I and Langreth D C 2010 *Phys. Rev.* B **82** 1–4
[65] Cooper V R 2010 *Phys. Rev.* B **81** 1–4
[66] Hamada I and Otani M 2010 *Phys. Rev.* B **82** 1–4
[67] Lundberg S M and Lee S I 2017 A unified approach to interpreting model predictions *Advances in Neural Information Processing Systems 30*, ed I Guyon, U V Luxburg, S Bengio, H Wallach, R Fergus, S Vishwanathan and R Garnett (Red Hook, NY: Curran Associates, Inc.) pp 4765–74
[68] Hjorth Larsen A *et al* 2017 *J. Phys.: Condens. Matter.* **29** 273002
[69] Ong S P, Richards W D, Jain A, Hautier G, Kocher M, Cholia S, Gunter D, Chevrier V L, Persson K A and Ceder G 2013 *Comput. Mater. Sci.* **68** 314–9