

Efficient prediction of structural and electronic properties of hybrid 2D materials using DFT and machine learning

Sherif Abdulkader Tawfik,^{1, 2, a)} Olexandr Isayev,^{3, b)} Catherine Stampfl,⁴ Joe Shapter,^{5, 6} David A. Winkler,^{7, 8, 9, c)} and Michael J. Ford^{1, d)}

¹⁾*School of Mathematical and Physical Sciences, University of Technology Sydney, Ultimo, New South Wales 2007, Australia*

²⁾*Institute for Biomedical Materials and Devices (IBMD), Faculty of Science, University of Technology Sydney, Sydney, NSW, Australia*

³⁾*Laboratory for Molecular Modeling, Division of Chemical Biology and Medicinal Chemistry, UNC Eshelman School of Pharmacy, University of North Carolina, Chapel Hill, North Carolina 27599, USA*

⁴⁾*School of Physics, The University of Sydney, New South Wales, 2006, Australia*

⁵⁾*Australian Institute for Bioengineering and Nanotechnology, University of Queensland, St. Lucia, Queensland, 4072, Australia*

⁶⁾*School of Chemical and Physical Sciences, Flinders University, Bedford Park, South Australia 5042, Australia*

⁷⁾*Manufacturing, Commonwealth Scientific and Industrial Research Organisation, Bag 10, Clayton South MDC, Victoria 3169, Australia*

⁸⁾*Monash Institute of Pharmaceutical Sciences, Monash University, 381 Royal Parade, Parkville, Victoria 3052, Australia*

⁹⁾*Latrobe Institute for Molecular Science, La Trobe University, Kingsbury Drive, Bundoora, Victoria 3086, Australia*

Materials constructed from different van der Waals two-dimensional (2D) heterostructures offer a wide range of benefits, but these systems have been little studied because of their experimental and computational complexity, and because of the very large number of possible combinations of 2D building blocks. The simulation of the interface between two different 2D materials is computationally challenging due to the lattice mismatch problem, which sometimes necessitates the creation of very large simulation cells for performing density-functional theory (DFT) calculations. Here we use a combination of DFT, linear regression and machine learning techniques in order to rapidly determine the interlayer distance between two different 2D heterostructures that are stacked in a bilayer heterostructure, as well as the band gap of the bilayer. Our work provides an excellent proof of concept by quickly and accurately predicting a structural property (the interlayer distance) and an electronic property (the band gap) for a large number of hybrid 2D materials. This work paves the way for rapid computational screening of the vast parameter space of van der Waals heterostructures to identify new hybrid materials with useful and interesting properties.

I. INTRODUCTION

An ongoing challenge in material research is the design of materials with desired physical and chemical properties. To overcome this challenge, the application of machine learning (ML) is gaining momentum as a feasible tool that can significantly accelerate the material discovery process.^{1, 2} There has recently been an increasing interaction between first principles physical theories, chiefly density functional theory (DFT), and various ML frameworks. The DFT computations have been applied to the generation of computational material data for vast material databases, such as AFLOW,³ with ~ 1.7 million structures, and MaterialsProject,⁴ with $\sim 620,000$

structures. This data is then fed into the ML algorithms for the rapid prediction of material properties, such as the dielectric constant,⁵ superconducting critical temperature,⁶ thermal and mechanical properties,⁷ electronic structure,^{3, 8} and hydrogen storage capacity.⁹

In the past few years, there has been a rising interest in the LEGO-like creation of hybrid 2D heterostructures¹⁰ for various applications in photovoltaics and photonics.^{11–15} A recent data mining study reported the existence of 1,825 2D materials that could be exfoliated from known experimental inorganic compounds, and this set can make ~ 1.7 million bilayer structures (with n 2D materials, the number of possible bilayers is $n(n-1)/2 + n = n(n+1)/2$). The number of possible trilayers, tetralayers, etc. increases very rapidly and quickly becomes an intractable problem for DFT to explore. Is there an efficient method to rapidly predict the properties of these structures without having to perform expensive first principles calculations? Here we answer this question by applying various machine learning

^{a)}Electronic mail: sherif.abbas@uts.edu.au

^{b)}Electronic mail: olexandr@olexandrisayev.com

^{c)}Electronic mail: D.Winkler@latrobe.edu.au

^{d)}Electronic mail: mike.ford@uts.edu.au

(ML) models for the prediction of the structural and electronic properties of layered van der Waals (vdW) materials, specifically bilayer materials that are constructed by layering different 2D materials. The success of the ML approach using a small training set could potentially save a significant amount of time and computational cost, and ultimately experimental effort.

In 2D materials, the interlayer van der Waals (vdW) forces are essential to maintain the equilibrium structure. The key structural quantity that indicates the strength of the vdW force in these materials is the interlayer distance, and a related energetic quantity is the layer binding energy. For DFT to be able to accurately predict these two quantities, it should be corrected by incorporating a vdW correlation potential to the DFT correlation potential.^{17,18} To this end, various forms of the vdW correlation have been proposed and applied to 2D materials, many of which displayed impressive accuracy, such as the Tkatchenko-Scheffler (TS) method^{19,20} and the SCAN+rVV10 method.²¹ The DFT databases, such as the aforementioned AFLOW and MaterialsProject databases, include structures that have been calculated without considering the vdW interaction potential. For example, there are a number of calculations published on the online portals of these databases that have wrong relaxed c lattice parameter, such as MoS₂-WS₂ in which $c = 22.37$ Å, which is extremely large compared to the lattice parameter in other hybrid bilayers such as graphene—boron nitride, which is 3.3 Å.²² This is a result of the use of non-dispersive correlation potentials. Therefore, an essential prerequisite is the construction of a data set of hybrid 2D vdW materials using a dispersion-supported DFT method.

Hybrid 2D heterostructures have recently attracted attention as potential vertical p-n junctions, in which the charge carriers move in the direction perpendicular to the plane of the 2D layers. Several vertical p-n junctions have been reported recently.^{11–15} The 2D materials that have been investigated for p and n doping include MoS₂, MoSe₂, WSe₂. With the growing number of semiconducting vdW materials, the number of possible hybrid bilayers that achieve p-n band alignment is increasing, and the prediction of the band gap of these bilayers using ML models would greatly support the search for new atomically thin p-n junction materials for optoelectronics applications.

A critical problem, however, in the application of DFT to hybrid vdW structures is the problem of creating lattice-matching interfaces between noncommensurate 2D materials. The application of DFT for studying hybrid 2D materials necessitates that the supercell describing the interface between two materials (whether parallel or perpendicular to the plane of the 2D materials) must have commensurate supercells of the two materials. One approach that is commonly adopted to this problem is to search for supercells that minimise the strain in each of the incommensurate monolayers. This is not a trivial problem and often necessitates quite large strains

of the order of a few percent in order to keep the size of the supercell reasonable. The use of ML largely bypasses this problem since DFT calculation are only performed on the isolated individual layers and a small subset of bilayers to generate the training and test sets for the ML phase.

Here, we use DFT for the calculation of the hybrid bilayer vdW structures. Using this data, we apply four ML models for the prediction of the interlayer distance and the band gap. We find that the considered ML models yield predictions of the interlayer distance and the band gap with reasonable accuracy.

II. COMPUTATIONAL DETAILS

A. The first principles approach

We have selected our 2D structures from the large collection of 2D materials in the “2D atlas”.²³ We have used VASP²⁴ to calculate the atomic and electronic structures using the generalized gradient approximation based on the PBE parametrization,²⁵ and accounted for the vdW interaction by adding the Tkatchenko-Scheffler vdW correlation potential.²⁰ We applied a k -point space of $8 \times 8 \times 1$ for unit cells, and $3 \times 3 \times 1$ for supercells, and an energy cutoff of 400 eV. The energy minimization tolerance is 10^{-6} eV, and the force tolerance is 10^{-2} eV/Å. For the 267 bilayers in the data set, we calculate the interlayer distance d as the equilibrium distance separating the two layers (that is, the minimum distance between the two layers), and the band gap. Note that the application of VASP for the geometries provided in the “2D atlas” by Miro *et al.*²³ induces a small strain on the individual layers. However, the presence of small planar strain have insignificant effects d and the band gap.²⁶

For our training set, we perform DFT calculations for the bilayers assembled from 53 monolayer structures, as shown in Table I. In selecting these monolayers, we have focussed on structures that satisfy the following two criteria: (1) It possesses trigonal symmetry, and (2) do not suffer from lattice distortions arising from covalent interaction with the adjacent layers. For example, we remove the CdX and ZnX monolayers (X= S, Se, Te) because of the significant layer distortions they exhibit when stacked with other layered materials.

B. The bilayer data set

For two different 2D materials, their unit cells are mostly incommensurate (that is, they have different values for the a lattice parameter). Therefore, we seek to construct the bilayers such that they are approximately commensurate; that is, for each of the two monolayers, we increase the number of monolayer images for each monolayer until the difference in the lattice constant of

TABLE I. The 53 monolayers. Note that the “1T” prefix denotes the 1T polymorph of transition metal dichalcogenides (TMDCs). The TMDCs without this prefix are of the 2H polymorph.

BoronNitride	NbSe ₂	Silicene	TaS ₂
CdO	NbTe ₂	SiliconCarbide	TaSe ₂
GaS	NiS ₂	1T-HfS ₂	TaTe ₂
GaSe	NiSe ₂	1T-HfSe ₂	TiS ₂
Graphene	NiTe ₂	1T-HfTe ₂	TiSe ₂
HfS ₂	PdS ₂	1T-MoS ₂	TiTe ₂
HfSe ₂	PdSe ₂	1T-MoSe ₂	WCl ₂
HfTe ₂	PdTe ₂	1T-MoTe ₂	WS ₂
InS	PtS ₂	1T-NbS ₂	WSe ₂
InSe	PtSe ₂	1T-NbSe ₂	WTe ₂
MoS ₂	PtTe ₂	1T-NbTe ₂	ZnO
MoSe ₂	ReS ₂	1T-ReS ₂	
MoTe ₂	ReSe ₂	1T-WS ₂	
NbS ₂	ReTe ₂	TaCl ₂	

each monolayer supercell is $< 2\%$.²⁶ We wrote a Java program that automates the search of approximately commensurate supercells, where the number of images of either monolayer does not exceed 5. For simplicity, the twist angle adopted here (that is, the angle by which one monolayer lattice rotates with respect to the other) is 0. The number of bilayers we choose from this set is 267. These bilayers will be used for the prediction of the interlayer distance.

For the prediction of the band gap, we have to take into consideration that 33 out of the 53 monolayers are metallic. This makes the band gap zero for a large number of bilayers, which yields a very skewed data set that is problematic for machine learning. Therefore, we exclude all bilayers that are formed from metallic monolayers. Out of the set of 267 bilayers, the number of bilayers formed from these 20 monolayers is 49.

The application of DFT methods with vdW correlation correction for layered materials has been demonstrated to yield accurate results for the interlayer distances and the binding energies.^{17,27} We use the TS method which accurately predicts the values of interlayer distances compared with the available experimental values as well as the benchmark Random Phase Approximation (RPA) method,²⁷ as shown in Figure 1. In this figure, we compare between the values of the c lattice parameter for the 11 2D materials in Ref.²⁷, predicted using the TS vdW correlation, RPA and the experimental values (obtained from Ref.¹⁷). These values are for pristine bulk materials, not hybrid materials, however. The choice of the right method for hybrid multilayered materials is still an open problem, and is currently the subject of an ongoing theoretical investigation.

To perform the DFT calculations for bilayers, one has to find the most optimal stacking configuration for the bilayer. With respect to stacking, there are two groups of

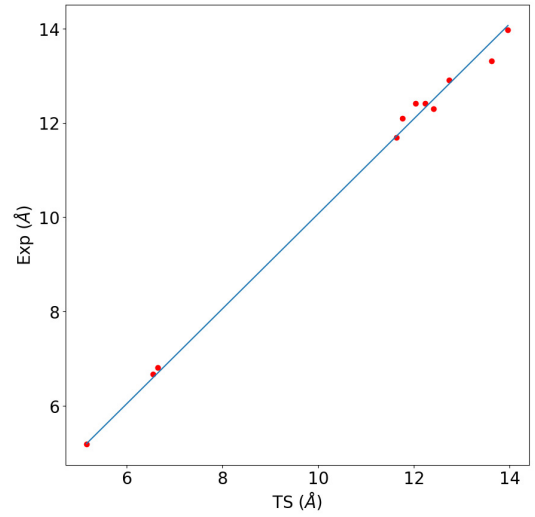


FIG. 1. Comparison between the c lattice parameter for the 11 2D materials in Ref.²⁷, predicted using the TS vdW correlation, and the experimental values.

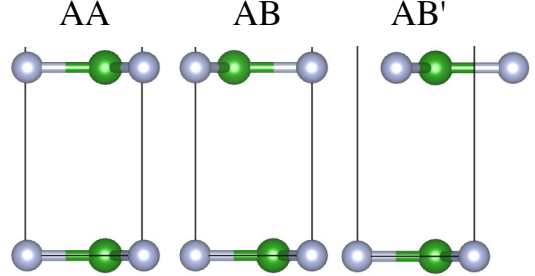


FIG. 2. The three stacking configurations, AA, AB and AB', which were used for obtaining the lowest energy configuration. For boron nitride, we also added the AA', which is similar to AA but in which an N on one layer is faced by a B on the other layer.

bilayers: those with commensurate lattices (and thus the simulation cell is constructed from the unit cells of the two monolayers) and those with incommensurate lattices. For the commensurate lattices, we obtain the equilibrium stacking by performing a geometry relaxation for three stacking configurations: AA, AB and AB'. These configurations are displayed in Figure 2. The structure with lowest energy is then taken as the equilibrium stacking configuration. For the incommensurate unit cells, such as the boron nitride—silicon carbide bilayer, which is formed from 5×5 boron nitride unit cells and 4×4 SiliconCarbide unit cells, sliding one monolayer over the other in such large bilayers does not significantly affect the binding energy, and therefore we do not search for equilibrium stacking configurations in incommensurate bilayers.

We display the number of bilayers in which each monolayer is a component in Figure 3. This figure shows which monolayers are over- or under-represented in the data set.

Here, the InSe monolayer is only present as the bilayer InSe—InSe, and MoTe₂ is the most connected (it exists in 27 bilayers).

A useful method for visualizing high dimensionality data, such as the present data sets, is to apply the t-distributed Stochastic Neighbor Embedding (t-SNE)²⁸ algorithm for generating a 2D plot that clusters the data into groups that are labeled by the values in the output vector. Figure 4 shows a two-dimensional t-SNE projection mapping of chemical space of 267 hybrid 2D materials. t-SNE is a non-linear dimensionality reduction algorithm used for exploring high-dimensional data. It maps multi-dimensional data to two dimensions suitable for human observation. Due to the selection of large number elements as building blocks, we see that materials explore vast chemical space without forming several well-defined clusters. However, we identified three broad regions with specific character: sulphides, tellurides, and CNSi (graphene, carbides, nitrides, and silicenes). On average sulphide materials have a low interlayer distance of 2.1 – 2.7 Å. Tellurides typically have the highest at 3.2–3.8 Å. Two extreme materials PtS₂—PtS₂ and PdTe₂—WTe₂ are marked accordingly.

C. The descriptors vector

The key to implementing a successful machine learning model is the descriptors. Fragment-based descriptors have demonstrated superior performance in enhancing the accuracy of ML models for the case of molecules²⁹ and crystals.³⁰ In this work we adopt the approach of Isayev *et al.*,³⁰ the Property-Labelled Materials Fragments (PLMF), modified for 2D materials and is composed of 1529 descriptors. In the PLMF approach, a crystal structure is represented as a graph, with vertices decorated according to the reference properties of the atoms they represent and nodes are connecting topological neighbors according to the Voronoi tessellation. The adjacency matrix of this graph determines the global topology for a given system, including interatomic bonds and contacts within a crystal. The final descriptor vector to the Machine Learning (ML) model is obtained by partitioning a full graph into smaller subgraphs, that we call fragments by the analogy with fragment-based descriptors in cheminformatics. Every fragment starts from a node (an atom and its properties) and captures a path in the graph through a collection of bonded atoms. See Ref.³⁰ for technical details.

After constructing the monolayer descriptors, the next critical problem is the representation of each bilayer using the monolayer data. That is, is it possible to use the monolayer descriptors to describe the bilayers? Given that the interaction between the monolayers in a bilayer is of a dispersive nature, it does not affect the structure of either of the constituting monolayers. Therefore, it is possible to use the monolayer descriptors to describe the monolayers in the bilayer.

Using the monolayer descriptors, another question is: how to construct the bilayer descriptor vector? An intuitive choice would be to create a descriptors vector which doubles the size of the PLMF vector, and composed of the descriptors for the two monolayers. The problem in this approach is that it is sensitive to the swapping of the bilayers; that is, the descriptor vector for the bilayer A-B, made from monolayers A and B, will be a different vector from that of the bilayer B-A, even though the two bilayers are physically identical.

We approach this problem by the two following approaches:

Bilayer representation 1 (BR1): For each bilayer A-B, we create two data records: one record has A’s descriptors followed by B’s descriptors, and the other with B’s descriptors followed by A’s descriptors. This method of representing multiple representations of the same data item has been previously applied in the context of organic molecules.³¹ Using this representation, for the prediction of d , the data set has a total of $482 = 267 \times 2 - 52$ records (52 records are subtracted, instead of 53, because we have removed the PdS₂—PdS₂ bilayer from the data set due to the appearance of covalent interaction between the S atoms of the adjacent layers), while for the prediction of the band gap it has $78 = 49 \times 2 - 20$ records. Bilayer representation 2 (BR2): Instead of creating a descriptors vector that is double the size of the monolayer descriptor vector, we add the values of the descriptors in both monolayers. This method is intrinsically invariant to changes in the order of the monolayers, and can, in principle, be applied to supercells with more than two layers. Using this representation, for the prediction of d , the data set has a total of 267 records, while for the prediction of the band gap has 49 records. Owing to the small number of records in BR2 for the band gap prediction, we do not perform machine learning on this set.

With 1529 fields, it is crucial to drop the number of descriptors using a dimensionality reduction algorithm. For this purpose we apply the least absolute shrinkage and selection operator (LASSO) algorithm on the dataset.³² Since LASSO is a supervised dimensionality reduction algorithm, it cannot be applied to the set of monolayer descriptors. Instead, we apply LASSO on the bilayer data set in the BR2 against the interlayer distance d . In order to obtain the optimal number of descriptors using LASSO, we vary the value of α in LASSO, then apply the LASSO fitting which yields a number of descriptors, and then use the descriptors to predict using the support vector machine (discussed below). We choose the value of α which yields the lowest R^2 value.

D. The machine learning models

The dataset of bilayers constitutes a highly nonlinear prediction problem. We construct four ML models to learn the functions d and the band gap, which we describe below.

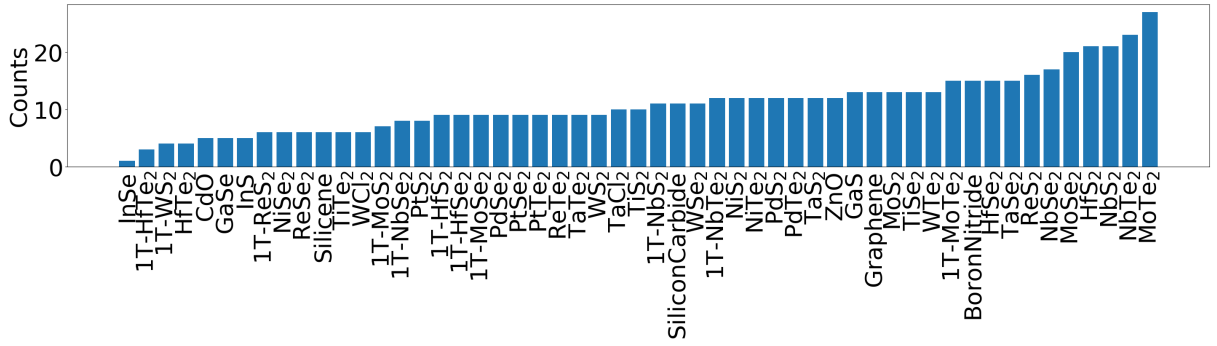


FIG. 3. The number of bilayers in which each monolayer (along the x -axis) is a component.

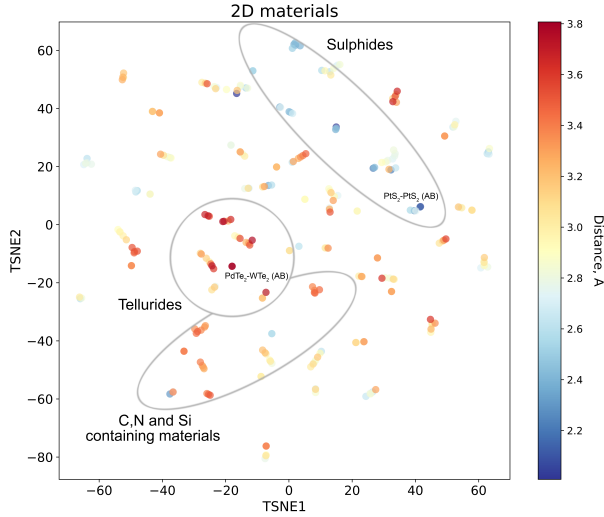


FIG. 4. The t-distributed Stochastic Neighbor Embedding (t-SNE) plot for the data set of 267 structures based on the value of the interlayer distance.

Feedforward Neural Network (NN)³³: NNs simulate the operation of the human brain. They can capture the highly nonlinear associations between input and output variables and can adaptively learn highly complex relationships. In an NN, an input layer receives the numbers in the descriptor vector. This layer is connected to a number of hidden layers, each layer being composed of a number of neurons, and finally the last hidden layer feed to an output layer whose number of neurons is the size of the output vector. Each neuron operates on the data it receives from the previous layer using an activation function which mimics the activation process in biological neurons, and a weight coefficient at the link between each two neurons. Each neuron collects its input from all of the neurons in the previous layer, and so forth up to the output layer. That is, each neuron j receives the following quantity from neurons i : $p_j(t) = \sum_i o_i(t)w_{ij}$, where w_{ij} is the weight connecting neurons i and j , and o_i is the output of neuron i .

These NNs work by using various learning algorithms

that update the values w_{ij} , the simplest being the back-propagation learning algorithm. In this algorithm, w_{ij} are updated according to the formula

$$w_{ij}(t+1) = w_{ij}(t) + \eta \frac{\partial C}{\partial w_{ij}} + \xi(t)$$

where η is the learning rate, C is the loss function, $\xi(t)$ is a stochastic term, and t is the propagation step. For the activation function associated with each neuron, there are several choices available. Of interest to the present work is the logistic sigmoid which is given by $a_{\text{logistic}}(z) = \frac{1}{1+e^{-z}}$, where z is the quantity received by the neuron. We use the Keras³⁴ python platform to implement the NN model. The network we have used for BR1 representation has $35 \times 2 = 70$ input nodes, 5 nodes in a single hidden layer, and one output node. In the BR2 representation, the hidden layer has 35 input nodes. The sigmoid activation function is used in the hidden layer, while the linear activation function is used for the input and output layers. The learning rate is 0.03.

Support Vector Machine (SVM):^{35,36} This is a supervised learning model which was first introduced as a classification model,³⁵ and then modified for regression problems.³⁶ The SVM classifier performs the classification of the data set from selected subsets of samples, called support vectors, in which the characteristic information on class distinction is compressed. In the linear support vector regression problem which we utilize in the present work, the aim is to find the linear function $f(x) = \mathbf{w}x + \mathbf{b}$ that approximates the output vector \mathbf{y} with weights vector \mathbf{w} such that the *primal function* $J(x)$ is minimized, which is given by

$$J(\beta) = \frac{1}{2} \mathbf{w}^T \mathbf{w} + C \sum_{n=1}^N (\eta_n + \text{eta}_n^*) \quad ,$$

given the constraints $|y_n - (\beta x + b)| < \epsilon + \eta_n$ and $|(\beta x + b) - y_n| < \epsilon + \eta_n^*$ for some small ϵ and positive variables η_n and eta_n^* , for all n , where n denotes the data record, and N is the total number of records, η_n and eta_n^* are known as *slack variables* which must be greater

than zero. Both C and ϵ are input parameters to the model.

Relevance Vector Machine (RVM):^{37,38} This is a sparse version of the support vector machine which attempts to amend several of the shortcomings of SVM,³⁸ such as non-probabilistic predictions, low sparsity, and the presence of the two fitting parameters C and ϵ which require cross validation. The RVM was introduced to enhance the sparsity of SVM and introduce a probabilistic weighting of the model weights based on Bayes' rule, assuming a Gaussian distribution of weights.

Random Forest (RF):³⁹ This is an ensemble learning method for classification, regression and other tasks, that operate by constructing a multitude of decision trees at training time and outputting the class that is the mean prediction of the individual trees. The training in RFs is based on *feature aggregating*⁴⁰ method. Given a training set \mathbf{x} with output \mathbf{y} , bagging repeatedly (B times) selects a random sample from \mathbf{x} and \mathbf{y} with replacement of the training set and fits decision trees to these samples. Once the training is finished, the prediction function operates by averaging the predictions from all the individual regression trees. The number of trees and the maximum tree depth are input parameters to the model.

The objective of ML is to build accurate prediction models. The quality of the model is determined by the ability of the model to predict the outcomes for cases which the model never encountered before; that is, how accurately can the model infer the outcomes based on its learning. This can be measured by splitting the data set into two parts: the training set, on which ML algorithm will be performed to build the ML model, and the test set, which will be used to test the quality of the model. The accuracy of the prediction is judged by using one of several *loss functions*. Here we use the following loss functions to assess the accuracy of the training:

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2 \quad (1)$$

$$\text{MARE} = \frac{1}{n} \sum_{i=1}^n \frac{|Y_i - \hat{Y}_i|}{Y_i} \quad (2)$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (Y_i - \hat{Y}_i)^2}{\sum_{i=1}^n (Y_i - \bar{Y})^2} \quad (3)$$

where MSE is the mean square error, MARE is the mean absolute relative error (%), R^2 is the coefficient of determination, Y_i are the original test set outcomes (in our case, the DFT-calculated interlayer distances for the bilayers), \hat{Y}_i are the predicted test set outcomes, and \bar{Y} is the average of the original test set outcomes. The significance of R^2 is that it expresses the proportion of the variance in Y_i that can be predicted from the descriptor vector, and is an important measure of the ML model quality. However, its values are dependent on the size of the data set, and therefore we adopt the three quantities

together, R^2 , MARE and MSE, to gauge the accuracy. For the case of the band gap prediction, we do not use the MARE because some of the values obtained are zero.

Thus, for each of the four models in this work, we train the model on 80% of the data set and use the remaining 20% as the test set. We construct the test set by applying the k-means clustering⁴¹ to the data set (the set of bilayers in BR2), which yields 6 clusters by the Silhouettes analysis.⁴² Then, we randomly choose 20% of each cluster and build our test set. In the NN model, the learning iteration stops when the MSE is below 0.03 (for BR1) and 0.05 (for BR2). Then we compare the accuracy of the models based on the R^2 , the MSE and MARE values.

Each of the four models involved in this work requires the tuning of a number of parameters to ensure optimal performance. We tune the NN parameters manually, and we use the GridSearch algorithm provided by the Python scikit-learn library (GridSearchCV) to tune the parameters of the SVM and RF models. GridSearchCV calculates the best parameter combination by performing a cross-validated grid-search over a parameter grid. The parameters which we optimize for the SVM are: the C and γ , while those of the RF are: number of estimators and maximum depth.

III. RESULTS AND DISCUSSION

A. Prediction of the interlayer distance

Using the LASSO algorithm, we find that the optimal number of descriptors for predicting d is 35 per bilayer. We summarize these descriptors in the Supporting Information.

The accuracy of the prediction of a ML model depends primarily on the quality of the descriptors vector. We find a clear demonstration of this in our comparison between the quality of predictions based on BR1 and BR2. We display the MSE, MARE and R^2 values for the fitting performed on each of the ML models in Tables II and III for both bilayer representations, BR1 and BR2. Comparing the R^2 , MSE and MARE values of the test sets, it can be seen from Tables II and III that the SVM and RVM models exhibit a higher accuracy in BR1 than BR2, while the RF and NN models in BR2 have a higher accuracy than in BR1. Using the BR1 representation (Table II), the RF and SVM models in BR2 are both exhibiting the highest accuracy of the ML models tested in this work, with $R^2(\text{RF}) = 0.82$, $R^2(\text{SVM}) = 0.83$, and $\text{MSE}(\text{RF}) = 0.024$, $\text{MSE}(\text{SVM}) = 0.023$. Such difference in accuracy between BR1 and BR2 can be attributed to the fact that the addition of the descriptor values in BR2 leads to the ambiguity with respect to the identity of the monolayer descriptors added.

Figure 5 displays the correlation between the predicted and the DFT-calculated d values for the bilayers in the test set using the four ML models and using BR1. In order to understand the variation in the errors in the

TABLE II. The R^2 , mean square error (MSE), in Å, and mean absolute relative error (MARE) (%) for each of the four ML models applied to the BR1 bilayer resrepresentation.

BR1	R^2		MSE		MARE %	
	Train	Test	Train	Test	Train	Test
RF	0.96	0.82	0.005	0.024	1.8	3.9
SVM	0.90	0.83	0.012	0.023	1.8	4.1
RVM	0.87	0.79	0.017	0.028	3.4	4.5
NN	0.84	0.82	0.020	0.025	3.5	4.0

TABLE III. The R^2 , mean square error (MSE), in Å, and mean absolute relative error (MARE) (%) for each of the four ML models applied to the BR2 bilayer resrepresentation.

BR2	R^2		MSE		MARE %	
	Train	Test	Train	Test	Train	Test
RF	0.73	0.83	0.035	0.021	4.5	3.7
SVM	0.99	0.67	0.001	0.041	0.6	5.0
RVM	0.84	0.73	0.021	0.034	3.8	4.7
NN	0.88	0.90	0.016	0.012	0.7	2.7

various bilayers, we display in Figure 6, for each monolayer, the average and standard deviation of the MARE values obtained for the bilayers in which the monolayer is a component, comparing the predictions by the four ML models using BR1.

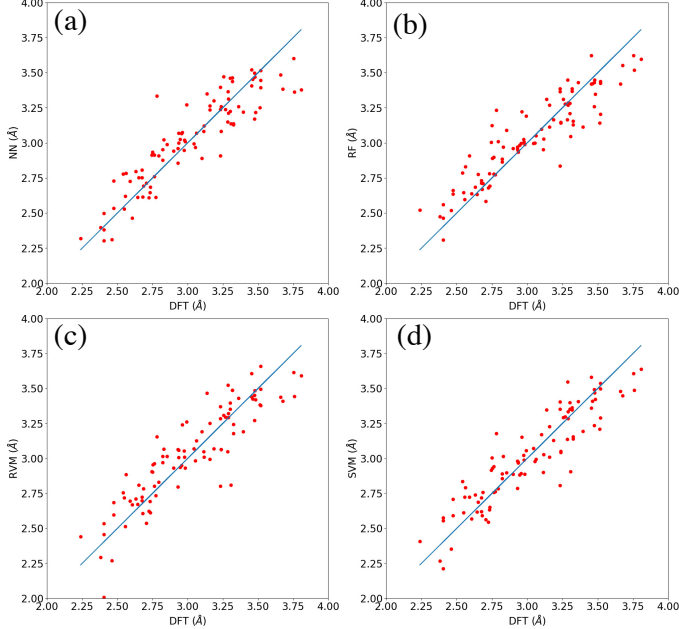


FIG. 5. Comparison between the interlayer distances in the bilayers obtained using DFT, and those using (a) NN, (b) RF, (c) RVM and (d) SVM. All values are in Å.

An interesting feature in Figure 6 is that the presence of some monolayers in bilayers lead to high errors in the prediction of d , irrespective of the ML model used. It

TABLE IV. The summary statistics of the interlayer distances predicted for the 1431 bilayers constructed from the 53 monolayers (values in Å), the MARE for predicting d in bilayers made from identical monolayers (pristine bilayers).

	NN	RF	RVM	SVM
Mean	3.005	3.025	3.035	3.016
Standard deviation	0.280	0.239	0.273	0.260
Minimum	2.095	2.122	1.753	2.018
5% percentile	2.541	2.624	2.578	2.586
50% percentile	3.015	3.023	3.042	3.022
95% percentile	3.461	3.401	3.436	3.424
Maximum	3.676	3.717	3.796	3.740
Pristine bilayer MARE	3.988	2.936	3.750	1.981

can be observed that CdO, TiTe₂, are common to all four ML models in terms of having the largest prediction errors. ReS₂, PdTe₂, are common to RF, SVM and RVM models. In addition, in all four models, the accuracy of predicting the non-metallic monolayers graphene, boron nitride and silicon carbide is in the order, from highest to lowest error: silicon carbide, graphene, boron nitride. GaSe, TaS₂, and boron nitride are common to the four models in terms of having lowest prediction errors.

The values displayed in Figure 3 might be a contributing factor in influencing the accuracy of predicting the bilayer properties (such as in the case of under-represented or over-represented monolayers). However, the trends of errors displayed in Figure 6 show that the accuracy of the models are somewhat independent from the counts in Figure 3.

Using the four ML models developed thus far, we run a prediction of the interlayer distances for all of the possible 1431 bilayers based on BR1, and display the results in Table IV. The NN predicts the minimum mean value of the four models, and the minimum d predicted by the RVM model is the lowest of the four minima (1.753 Å). The SVM gives the most accurate prediction for pristine bilayers (1.981%), followed by RF (2.936%).

Table V shows that, in the set of bilayers with lowest d , all models have bilayers with Pt and Pd atoms. We analyse the resrepresentation of the Pd and Pt atoms in the 5% percentile groups. The NN and SVM models have a lower representation of Pd than the RF and RVM models (that is, the number of Pd atoms in the NN and SVM predictions is 36 and 33 vs. 41 and 39 Pd atoms in the predictions by the RF and RVM, respectively), while all four models have almost the same representation of Pt-based bilayers (24, 25, 25 and 23, in NN, RF, SVM and RVM, respectively). As for the 5 bilayers with the highest d , all models predict that WTe₂—WTe₂ belongs to this group. In the predictions of the RF and RVM models, there are 3 bilayers that have the WTe₂ monolayer, and this is not the case in the other models. However, in the 95% percentile of bilayers, the number of WTe₂ monolayers is largest in the RF model (18 bilayers) while there are 10, 12 and 13 in the NN, SVM and RVM mod-

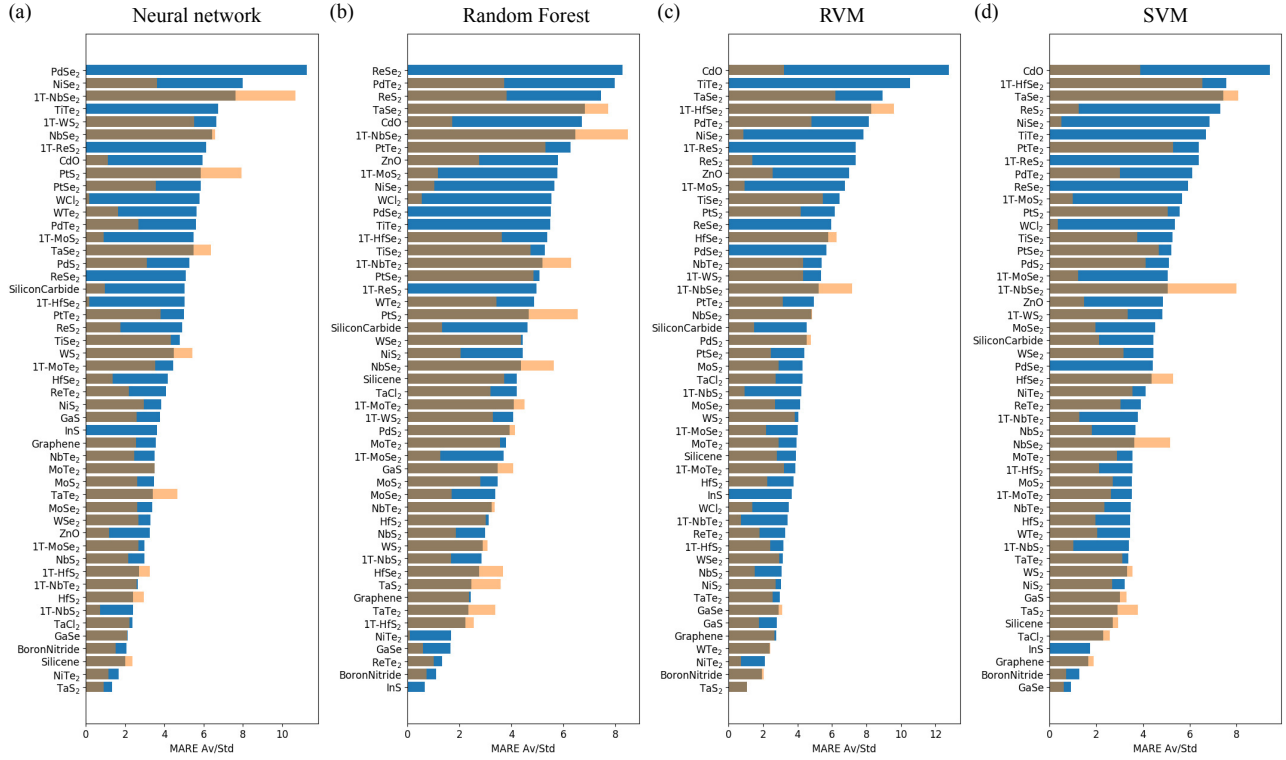


FIG. 6. For each monolayer, the average (blue) and standard deviation (orange) of the MARE for its bilayer data using the (a) NN, (b) RF, (c) SVM and (d) RVM models for prediction.

TABLE V. The bilayers with the smallest 5 and largest 5 predicted d values. All values are in Å.

	NN		RF		RVM		SVM	
smallest 5	PtS ₂ —PtS ₂	2.095	PtS ₂ —PtS ₂	2.122	PdS ₂ —PdS ₂	1.753	PtS ₂ —PtS ₂	2.018
	PdS ₂ —PdS ₂	2.126	PdS ₂ —PtS ₂	2.159	PdS ₂ —NiS ₂	2.008	PdS ₂ —PdS ₂	2.122
	PtS ₂ —PtSe ₂	2.161	PdS ₂ —PdS ₂	2.175	PtS ₂ —PtS ₂	2.063	PdS ₂ —NiS ₂	2.211
	PdS ₂ —PtS ₂	2.217	PdSe ₂ —PtS ₂	2.224	PdS ₂ —NiTe ₂	2.087	PdS ₂ —NbS ₂	2.273
	PdSe ₂ —PtS ₂	2.224	PdSe ₂ —PdS ₂	2.241	PdS ₂ —NbS ₂	2.150	PdS ₂ —T1-NbS ₂	2.278
largest 5	GaSe—WTe ₂	3.627	MoTe ₂ —MoTe ₂	3.626	MoTe ₂ —TaTe ₂	3.658	WTe ₂ —T1-MoTe ₂	3.654
	WTe ₂ —WTe ₂	3.638	MoTe ₂ —PdTe ₂	3.673	PdTe ₂ —TaTe ₂	3.704	ReTe ₂ —WTe ₂	3.655
	PdTe ₂ —BoronNitride	3.648	MoTe ₂ —WTe ₂	3.693	WTe ₂ —WTe ₂	3.760	MoTe ₂ —PdTe ₂	3.680
	PdTe ₂ —Graphene	3.661	ReTe ₂ —WTe ₂	3.704	MoTe ₂ —PdTe ₂	3.768	WTe ₂ —WTe ₂	3.725
	GaSe—WCl ₂	3.676	WTe ₂ —WTe ₂	3.717	MoTe ₂ —WTe ₂	3.796	MoTe ₂ —WTe ₂	3.740

els. This monolayer is also absent in the 5% percentile (except that there is 1 bilayer in the 5% percentile of the RVM model).

As for the chalcogenide atom in TMDCs, the Te atom is the primary vdW repellant, because much more of the Te-based monolayers are in the 95% percentile (52, 54, 48, and 50 bilayers predicted by NN, RF, SVM and RVM, respectively) compared with those in the 5% percentile (8, 13, 11 and 14 bilayers predicted by NN, RF, SVM and RVM, respectively). On the contrary, the presence of the S atom acts as a vdW attractor (it exists in 65, 67, 66 and 68 bilayers predicted by NN, RF, SVM and RVM, respectively in the 5% percentile, compared with

0, 1, 2 and 1 bilayers predicted by NN, RF, SVM and RVM, respectively in the 95% percentile). The Se is almost boarder-line: it exists in 35, 26, 34 and 24 bilayers predicted by NN, RF, SVM and RVM, respectively in the 5% percentile, compared with 12, 14, 16 and 16 bilayers predicted by NN, RF, SVM and RVM, respectively in the 95% percentile.

B. Prediction of the band gap

For the band gap prediction, we applied the BR1 bilayer representation. Using the LASSO algorithm, we

obtained 11 significant descriptors per bilayer (which is much less than the 35 descriptors obtained per bilayer for the d prediction). Those descriptors are listed in the Supporting Information. However, before we discuss the prediction results of the ML models, we discuss the band gap values obtained using DFT.

Some functionals within DFT are known to underestimate the band gap by a considerable amount.⁴³ For example, it predicts a bandgap of ~ 4.5 eV for hexagonal boron nitride, while its experimental band gap is ~ 6 eV.⁴⁴ To overcome this problem, hybrid functionals that mix the DFT exchange with an exact exchange component have been demonstrating impressive agreement with experimental band gaps.⁴⁵ However, the present implementations of hybrid functionals are very costly to run. In the present work, we restrict ourselves to the application DFT for the prediction of the band gap, showing that the success of ML models to predict such quantity using DFT should indicate a similar success in the prediction of band gaps calculated using more accurate functionals, such as hybrid functionals.

The DFT band gap value for the boron nitride bilayer is 3.935 eV for the AB' stacking configuration, which is close to the value of 4.01 eV obtained for same stacking (though using a different vdW method) reported in a recent work.⁴⁶ The equilibrium stacking reported in Ref.⁴⁶ is the AA' stacking, whose band gap is 4.341 eV. However, using the TS vdW method, the lowest energy stacking configuration is AB', and is only 3 meV lower in energy than AA'. To ensure consistency, we adopt the minimum energy stacking of the TS method.

The band gap of each of the bilayers corresponds to a specific band gap alignment of the two constituting monolayers. While the alignments yield a band gap that is smaller than that of the monolayers, as is typically the case in semiconducting interfaces, the following 8 bilayers have a zero band gap: HfS₂—MoTe₂, HfS₂—WTe₂, MoSe₂—TiS₂, MoTe₂—1T-HfS₂, 1T-HfS₂—WTe₂, TiS₂—WSe₂, TiS₂—ZnO, and TiSe₂—WTe₂. These bilayers are interesting because they exhibit a special kind of type III band alignment,⁴⁷ where two semiconducting interfaces form a metallic structure across the vdW vacuum. We will explore these structures in detail in a future contribution. We display the band alignment for two of them, namely HfS₂—MoTe₂ and MoSe₂—TiS₂, in Figure 7(a,b). Here, the conduction band minimum (CBM) of one layer (HfS₂ in Figure 7(a) and TiS₂ in Figure 7(b)) and the valence band maximum (VBM) of the other layer (MoTe₂ in Figure 7(a) and MoSe₂ in Figure 7(b)) are aligned, leading to a zero band gap in an interface between two semiconductors.

Next, we discuss the ML predictions of the band gap, displayed in Table VI and Figure 8. Even with such a small data set, the values of R^2 are quite reasonable for all four ML models, and the MSE values for the test sets are close to those in Tables II and III. The RF and SVM models have the highest prediction accuracy, based on

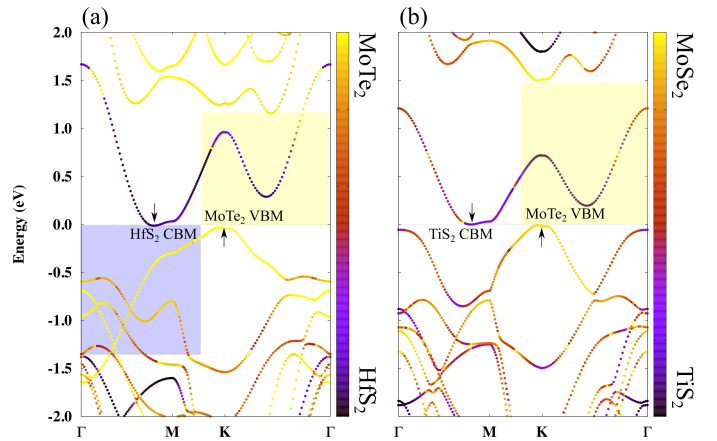


FIG. 7. The band alignment in the (a) HfS₂—MoTe₂ and (b) MoSe₂—TiS₂ bilayers. These bilayers exhibit III alignment in which the CBM of one layer (HfS₂ in (a) and TiS₂ in (b)), the VBM of the other layer (MoTe₂ in (a) and MoSe₂ in (b)) and the Fermi energy are all aligned, leading to a zero band gap in an interface between two semiconductors. Note that the VBM of TiS₂ hybridizes with the MoSe₂ states. The energy is shifted such that $E - E_F = 0$, where E_F is the Fermi energy.

TABLE VI. The R^2 and mean square error (MSE), in eV, for the prediction of the band gap by each of the four ML models applied to the BR1 bilayer representation.

	R^2		MSE	
	Train	Test	Train	Test
RF	0.96	0.90	0.028	0.040
SVM	1.00	0.90	0.000	0.040
RVM	0.98	0.83	0.014	0.070
NN	0.90	0.88	0.063	0.047

the $R^2 = 0.9$ value and the MSE = 0.04 value for each of the two models. However, the fact that $R^2 = 1.0$ for the training set prediction of the SVM alerts us that this model is over-fitted, and might not perform as well for bilayers outside of the present data set.

Figure 8 displays the correlation between the values of the band gap obtained by DFT and those obtained by each of the four ML models. The SVM and RVM models have the highest accuracy in predicting the zero band gaps, as shown in Figure 8(c) and (d) (there are 3 such band gaps in the test set of 14 records), while the RF model is the least accurate in this respect. However, SVM and RVM greatly underestimate the band gap of boron nitride—MoS₂, which is 1.867 eV by 0.6 eV and 0.7 eV, respectively. RF and NN, on the contrary, predict it within an error of 0.3 eV and 0.2 eV, respectively.

We analyze the prediction of the band gap for all of the possible $20 \times 21/2 = 210$ bilayers, based on the BR1 representation, in Table VII. It can be seen that NN, RVM and SVM all predict a negative value for the band gap as the minimum in the set, and this is unphysical. This minimum value is only -0.15 eV, however, and is most

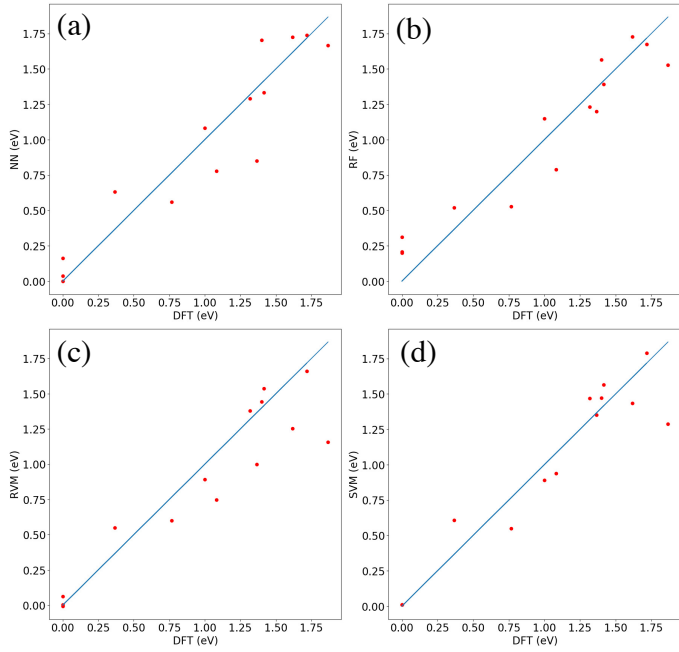


FIG. 8. Comparison between the band gaps distances in the bilayers obtained using DFT, and those using (a) NN, (b) RF, (c) RVM and (d) SVM. All values are in eV.

TABLE VII. The summary statistics of the band gaps predicted for the 210 bilayers constructed from the 20 semiconducting monolayers (values in eV).

	NN	RF	RVM	SVM
Mean	0.556	0.879	0.552	0.881
Standard deviation	0.402	0.378	0.396	0.403
Minimum	-0.150	0.066	-0.150	-0.047
5% percentile	0.088	0.315	0.090	0.270
50% percentile	0.472	0.888	0.468	0.925
95% percentile	1.329	1.491	1.325	1.431
Maximum	3.923	3.337	3.923	3.925

likely because the bilayer is metallic. This is analogous to the test and training sets here the ML models predict bandgaps of 0.1 eV where DFT returns a metallic result. The maximum band gap, which is that of the boron nitride bilayer, is correctly predicted by the NN, RVM and SVM models but not the RF model.

The distribution of the prediction data varies greatly with the ML model, as can be seen in the mean and percentile values. This is unlike the situation in Table IV, where those values were very close across the models. Such volatility can be seen in the values of the standard deviation in Table VII compared to Table IV. This can be attributed to the small size of the data set that was used to train the models for band gap prediction.

IV. CONCLUSION

In the present work we have demonstrated how machine learning approaches can be effectively used to predict structural and electronic properties of van der Waals heterostructures. More specifically, we have used DFT calculations to 53 monolayers and a small subset, 267, of possible bilayers to train machine learning approaches to predict the interlayer spacing and bandgap for bilayers built from all possible 1431 combinations of the monolayers.

The use of property labelled materials fragments³⁰ as descriptors for the monolayers proves to be very effective, yielding prediction R^2 of 0.83 and 0.90 for the interlayer spacing and bandgap respectively.

Hybrid materials built from 2D monolayers are gaining attention as novel materials with tunable properties. The current bottle-neck to exploring the possibilities these materials have is the vast parameter space of possible combinations. While electronic structure calculations are required to answer this question, even the most efficient available methods such as DFT are too time-consuming. The current results show great promise for a combined DFT/machine learning approach to solve this problem.

V. ACKNOWLEDGEMENTS

This work was supported by resources provided by the National Computational Infrastructure (NCI), and Pawsey Supercomputing Centre with funding from the Australian Government and the Government of Western Australia. This work is also supported by the Australian Research Council, ARC DP160101301.

REFERENCES

- ¹Pilania, G.; Wang, C.; Jiang, X.; Rajasekaran, S.; Ramprasad, R., Accelerating materials property predictions using machine learning, Scientific reports, 2013, 3, 2810.
- ²Le, T.C.; Epa, V.C.; Burden, F.R.; Winkler, D.A. Quantitative Structure-Property Relationship Modeling of Diverse Materials Properties, Chem. Rev., 2012, 112, 2889.
- ³Curtarolo, S. et al. AFLOWLIB.ORG: a distributed materials properties repository from high-throughput ab initio calculations. Comput. Mater. Sci., 2012, 58, 227.
- ⁴Jain, A.; Ong, S.P.; Hautier, G.; Chen, W.; Richards, W.D.; Dacek, S.; Cholia, S.; Gunter, D.; Skinner, D.; Ceder, G. The Materials Project: A materials genome approach to accelerating materials innovation, APL Materials, 2013, 1, 011002.
- ⁵Andersen, K.; Latini, S.; Thygesen, K.S. Dielectric Genome of van der Waals Heterostructures, Nano Lett., 2015, 15, 4616.

- ⁶Isayev, O.; Fourches, D.; Muratov, E.N.; Oses, C.; Rasch, K.; Tropsha, A.; Curtarolo, S. Materials Cartography: Representing and Mining Materials Space Using Structural and Electronic Fingerprints, *Chem. Mater.* 2015, 27, 735.
- ⁷Toher, C.; Oses, C.; Plata, J.J.; Hicks, D.; Rose, F.; Levy, O.; de Jong, M.; Asta, M.; Fornari, M.; Nardelli, M.B.; Curtarolo, S. Combining the AFLOWGIBBS and elastic libraries to efficiently and robustly screen thermomechanical properties of solids, *Phys. Rev. Mat.*, 2017, 1, 015401.
- ⁸Ward, L.; Agrawal, A.; Choudhary, A.; Wolverton, C. A general-purpose machine learning framework for predicting properties of inorganic materials, *npj Comput. Mat.*, 2016, 2, 16028.
- ⁹Thornton, A. W.; Simon, C.M.; Kim, J.; Kwon, O.; Deeg, K.S.; Konstas, K.; Pas, S.J.; Hill, M.R.; Winkler, D.A.; Haranczyk, M.; Smit, B. Materials Genome in Action: Identifying the Performance Limits of Physical Hydrogen Storage, *Chem. Mater.* 2017, 29, 2844.
- ¹⁰Geim, A.K.; Grigorieva, I.V. Van der Waals heterostructures, *Nature*, 2013, 499, 419.
- ¹¹Jin, Y.; Keum, D.H.; An, S.J.; Kim, J.; Lee, H.S.; Lee, Y.H. A Van Der Waals homojunction: ideal p-n diode behavior in MoSe₂. *Advanced Materials*, 2015, 27, 5534.
- ¹²Flory, N.; Jain, A.; Bharadwaj, P.; Parzefall, M.; Taniguchi, T.; Watanabe, K.; Novotny, L. A WSe₂/MoSe₂ heterostructure photovoltaic device. *Appl. Phys. Lett.*, 2015, 107, 123106.
- ¹³Ross, J.S.; Rivera, P.; Schaibley, J.; Lee-Wong, E.; Yu, H.; Taniguchi, T.; Watanabe, K.; Yan, J.; Mandrus, D.; Cobden, D.; Yao, W. Interlayer exciton optoelectronics in a 2D heterostructure p-n junction. *Nano Lett.*, 2017, 17, 638.
- ¹⁴Lee, C.H.; Lee, G.H.; Van Der Zande, A.M.; Chen, W.; Li, Y.; Han, M.; Cui, X.; Arefe, G.; Nuckolls, C.; Heinz, T.F.; Guo, J. Atomically thin p-n junctions with van der Waals heterointerfaces. *Nat. Nanotech.*, 2014, 9, 676.
- ¹⁵Choi, M.S.; Qu, D.; Lee, D.; Liu, X.; Watanabe, K.; Taniguchi, T.; Yoo, W.J. Lateral MoS₂ p-n junction formed by chemical doping for use in high-performance optoelectronics. *ACS Nano*, 2014, 8, 9332.
- ¹⁶Mounet, N.; Gibertini, M.; Schwaller, P.; Campi, D.; Merkys, A.; Marrazzo, A.; Sohler, T.; Castelli, I.E.; Cepellotti, A.; Pizzi, G.; Marzari, N. Two-Dimensional Materials From High-Throughput Computational Exfoliation of Experimentally Known Compounds. *Nat. Nano.*, 2018, 13, 246.
- ¹⁷Bjorkman, T.; Gulans, A.; Krashenninnikov, A.V.; Nieminen, R.M. van der Waals Bonding in Layered Compounds from Advanced Density-Functional First-Principles Calculations, *Phys. Rev. Lett.*, 2012, 108, 235502.
- ¹⁸Paul, J.; Singh, A.K.; Dong, Z.; Zhuang, H.; Revard, B.C.; Rijal, B.; Ashton, M.; Linscheid, A.; Blonsky, M. N.; Gluhovic, D.; Guo, J.; Hennig, R.G. Computational methods for 2D materials: discovery, property characterization, and application design, *J. Phys.: Condens. Matt.*, 2017, 29, 473001.
- ¹⁹Gobre, V.V.; Tkatchenko, A. Scaling laws for van der Waals interactions in nanostructured materials. *Nat. Comm.*, 2013, 4, 2341.
- ²⁰Tkatchenko, A.; Scheffler, M. Accurate molecular van der waals interactions from ground-state electron density and free-atom reference data, *Phys. Rev. Lett.*, 2009, 102, 073005.
- ²¹Peng, H.; Yang, Z.-H.; Perdew, J.P.; Sun, J. Versatile van der Waals Density Functional Based on a Meta-Generalized Gradient Approximation. *Phys. Rev. X*, 2016, 6, 041005.
- ²²Leven, I.; Maaravi, T.; Azuri, I.; Kronik, L.; Hod, O. Interlayer Potential for Graphene/h-BN Heterostructures, *J. Chem. Theory Comput.* 2016, 12, 2896.
- ²³Miro, P.; Audiffred, M.; Heine, T. An atlas of two-dimensional materials, *Chem. Soc. Rev.*, 2014, 43, 6537.
- ²⁴Kresse, G.; Furthmuller, J. Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set. *Phys. Rev. B*, 1996, 54, 11169.
- ²⁵Perdew, J.P., Burke, K.; Ernzerhof, M. Generalized gradient approximation made simple, *Phys. Rev. Lett.*, 1996, 77, 3865.
- ²⁶Cai, Y.; Zhang, G.; Zhang, Y. Electronic Properties of Phosphorene/Graphene and Phosphorene/Hexagonal Boron Nitride Heterostructures, *J. Phys. Chem. C*, 2015, 119, 13929.
- ²⁷Tawfik, S.A.; Gould, T.; Stampfl, C.; Ford, M.J. Evaluation of van der Waals density functionals for layered materials, *Phys. Rev. Mater.*, 2018, 2, 034005.
- ²⁸van der Maaten, L.; Hinton, G. Visualizing Data using t-SNE, *J Mach. Learn. Res.*, 2008, 9, 2579.
- ²⁹Gómez-Bombarelli, R.; Wei, J.N.; Duvenaud, D.; Hernández-Lobato, J.M.; Sánchez-Lengeling, B.; Sheberla, D.; Aguilera-Iparraguirre, J.; Hirzel, T.D.; Adams, R.P.; Aspuru-Guzik, A. Automatic chemical design using a data-driven continuous representation of molecules. *ACS Cent. Sci.* 2018, 4, 268.
- ³⁰Isayev, O.; Oses, C.; Toher, C.; Gossett, E.; Curtarolo, S.; Tropsha, A. Universal fragment descriptors for predicting properties of inorganic crystals. *Nat. Comm.*, 2017, 8, 15679.
- ³¹Hansen, K.; Montavon, G.; Biegler, F.; Fazli, S.; Rupp, M.; Scheffler, M.; von Lilienfeld, A.; Tkatchenko, A.; Muller, K. J. *J. Chem. Theory. Comput.* 2013, 9, 3404.
- ³²Tibshirani, R. Regression shrinkage and selection via the Lasso, *J. Roy. Stat. Soc. Ser. B*, 1996, 51, 267.
- ³³Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*, Springer.
- ³⁴Chollet, François et al., Keras, 2015, <https://github.com/keras-team/keras>.
- ³⁵Vapnik, V., *The nature of statistical learning theory*, Springer New York, 1995.
- ³⁶Drucker, H.; Burges, C. J.; Kaufman, L.; Smola, A. J.; Vapnik, V. "Support vector regression machines."

- in Advances in neural information processing systems, 1997, (pp. 155-161).
- ³⁷Tipping, M.E. Sparse Bayesian Learning and the Relevance Vector Machine, *J. Mach. Learn. Res.*, 2001, 1, 211.
- ³⁸Winkler, D.A.; Burden, F.R.; Relevance Vector Machines: Sparse Classification Methods for QSAR, *J. Chem. Inf. Model.* 2015, 55, 1529.
- ³⁹Ho, T.K. Random Decision Forests, *Proceedings of the 3rd International Conference on Document Analysis and Recognition*, Montreal, QC, 1995, 14, 278.
- ⁴⁰Ho, T.K. The Random Subspace Method for Constructing Decision Forests, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1998, 20, 832.
- ⁴¹Faraoun, K. M.; Boukelif, A. Neural networks learning improvement using the K-means clustering algorithm to detect network intrusions, *INFOCOMP*, 2006, 3, 28.
- ⁴²Rousseeuw, P.J. Silhouettes: a Graphical Aid to the Interpretation and Validation of Cluster Analysis, *Computational and Applied Mathematics*, 1987, 20, 53.
- ⁴³Cohen, A.J.; Mori-Sanchez, P.; Yang, W. Challenges for density functional theory. *Chem. Rev.*, 2011, 112, 289.
- ⁴⁴Cassabois, G.; Valvin P.; Gil, B. Hexagonal boron nitride is an indirect bandgap semiconductor. *Nat. Phot.*, 2016, 10, 262.
- ⁴⁵Perdew, J.P, Climbing the ladder of density functional approximations, *MRS Bull.*, 2013, 38, 743.
- ⁴⁶Fujimoto, Y. Saito, S. Band engineering and relative stabilities of hexagonal boron nitride bilayers under biaxial strain, *Phys. Rev. B*, 2016, 94, 245427.
- ⁴⁷Kroemer, H. Nobel Lecture: Quasielectric fields and band offsets: teaching electrons new tricks. *Rev. Mod. Phys.*, 2001, 73, 783.