ACS APPLIED MATERIALS & INTERFACES

www.acsami.org Research Article

Predicting Van der Waals Heterostructures by a Combined Machine Learning and Density Functional Theory Approach

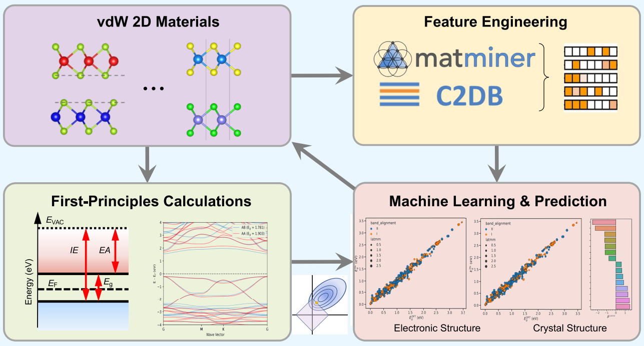
Daniel Willhelm, Nathan Wilson, Raymundo Arroyave, Xiaoning Qian, Tahir Cagin,* Ruth Pachter,* and Xiaofeng Qian*

Cite This: *ACS Appl. Mater. Interfaces* 2022, 14, 25907−25919

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Van der Waals (vdW) heterostructures are constructed by different two-dimensional (2D) monolayers vertically stacked and weakly coupled by van der Waals interactions. VdW heterostructures often possess rich physical and chemical properties that are unique to their constituent monolayers. As many 2D materials have been recently identified, the combinatorial configuration space of vdW-stacked heterostructures grows exceedingly large, making it difficult to explore through traditional experimental or computational approaches in a trial-and-error manner. Here, we present a computational framework that combines first-principles electronic structure calculations, 2D material database, and supervised machine learning methods to construct efficient data-driven models capable of predicting electronic and structural properties of vdW heterostructures from their constituent monolayer properties. We apply this approach to predict the band gap, band edges, interlayer distance, and interlayer binding energy of vdW heterostructures. Our data-driven model will open avenues for efficient screening and discovery of low-dimensional vdW heterostructures and moiré superlattices with desired electronic and optical properties for targeted device applications.



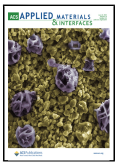
KEYWORDS: *van der Waals heterostructures, 2D materials, density functional theory calculations, machine learning, data-driven models*

INTRODUCTION

The seminal discovery of graphene has motivated the search for two-dimensional (2D) materials beyond graphene.[1] Several libraries of 2D compounds[2−8] have been established with a diverse set of compositions and crystal structures, offering exciting opportunities for theoretical and experimental exploration of their physical and chemical properties[9] as well as potential applications in areas such as flexible electronics,[10] photonics,[11] and topological quantum materials.[12] The unique weak van der Waals (vdW) interaction in layered 2D materials allows one to arbitrarily stack different 2D monolayers, forming van der Waals heterostructures. These vdW heterostructures may possess emergent physical properties capable of significantly expanding the applicaitons of low-dimensional systems.[13]

Unlike traditional heterostructures, vdW heterostructures are not limited by lattice mismatch. Their properties may be strongly influenced by the interlayer twist angle and stacking order, which can provide a large material design space. Tuning these variables allows for the realization of atomically sharp interfaces[14] and the precise control of material properties. Such fine tunability of electronic band structure has been observed in twisted bilayer graphene leading to intriguing superconductivity and correlated physics at the so-called "magic angle".[15,16] Stacking configuration has also been shown to have significant effects on the vdW heterostructure and offers a unique avenue for tuning material properties.[17] The atomic registry of the individual layers may depend on their crystal symmetry and relative in-plane shift, e.g., AA and AB stacking configurations. Theoretical studies have shown that these different stacking configurations not only affect interlayer binding energies and interlayer distances, but also significantly impact electronic structures and other properties.[18] Thus, combining and stacking different monolayers into vdW heterostructures may allow the fabrication of artificial low-dimensional materials with tailored electronic band gap and band alignment for 2D electronic and optoelectronic applications. For example, vdW heterostructures with Type I band alignment could offer potential 2D materials for engineering dielectric environment to confine and tune the intralayer exciton,[19] while vdW heterostructures with Type II

Received: March 10, 2022
Accepted: May 13, 2022
Published: May 27, 2022



ACS Publications

https://doi.org/10.1021/acsami.2c04403

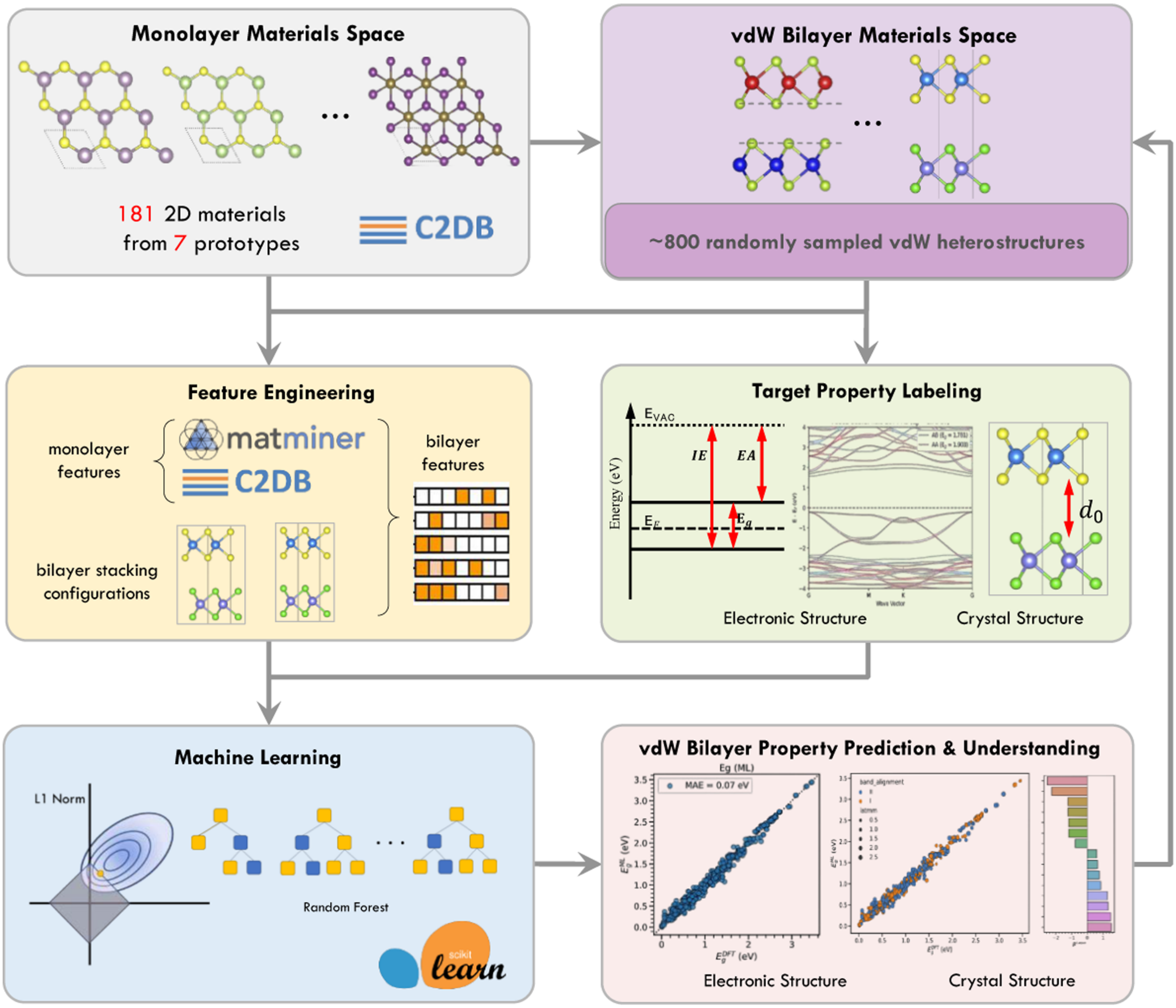


Figure 1. General schematic of computational and machine learning workflow.

band alignment can be useful for enabling the interlayer exciton with long lifetime.[20]

Given the growing number of 2D monolayers, the vdW heterostructure space grows combinatorially. Fabrication and characterization of vdW heterostructures via trial-and-error become an increasingly difficult task. Even high-throughput first-principles approaches become extremely expensive. Recently, machine learning methods have been applied to circumvent this problem. Tawfik et al.[21] constructed 267 bilayers out of 53 different monolayers and trained machine learning models to predict the interlayer distance and band gap by using the property-labeled materials-fragments (PLMF) descriptors.[22] Their best-performing models achieve a testing R^2 score of 0.9 and a mean-squared error (MSE) of 0.012 Å^{2}for predicting the interlayer distance using 267 training data. For the band gap model, they achieved a testing R^2 score of 0.9 and MSE of 0.04 eV^2 using only 53 training data. Choudhary et al.[23] leveraged a 2D material database in the JARVIS-DFT library to predict the band edges and work function of monolayers by using classical force-field inspired descriptors (CFID). Their model predicts the monolayer valence band maximum (VBM) and conduction band minimum (CBM) with a mean-absolute-error of 0.67 eV and 0.62 eV, respectively. Dong et al.[24] recently proposed a low-cost method of obtaining the electronic band structure of bilayers via band-folding. They applied their model to 703 vdW bilayer heterostructures based on 1T and 2H prototype structures and performed a comprehensive analysis of their electronic structures. Thus, it is highly desirable to establish accurate machine learning models that can predict a rich set of material properties for a large number of vdW heterostructures with a variety of stacking configurations and prototype structures from the existing 2D material database.

Here, we present a data-driven framework combining first-principles electronic structure calculations, existing 2D material database, and supervised machine learning, which allows for efficient and accurate prediction of electronic and structural properties of vdW heterostructures. Our framework can predict the physical properties of heterostructures such as ionization energy (IE), electron affinity (EA), and electronic band gap (E_g), interlayer distance (d_0), and interlayer binding energy (E_b) by only using the properties of the constituent

https://doi.org/10.1021/acsami.2c04403
ACS Appl. Mater. Interfaces 2022, 14, 25907−25919

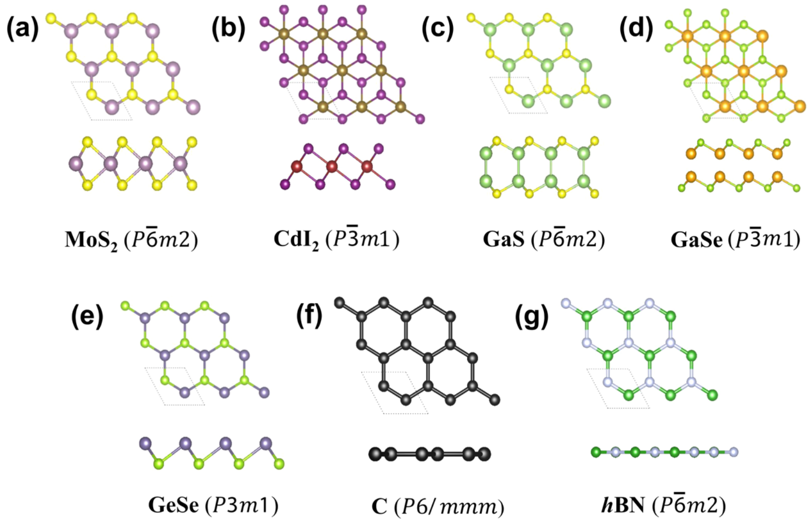


Figure 2. Crystal structures and symmetries of the monolayer prototypes used to create vdW heterostructures. (a) 1H-MoS_2, (b) 1T-CdI_2, (c) hexagonal GaSe, (d) hexagonal GaSe, (e) γ-GeSe, (f) graphene, and (g) *h*BN.

monolayers. The approach reduces the computational costs of first-principles calculations and enables an efficient way to explore larger numbers of theoretical configurations of vdW heterostructures while avoiding brute-force exploration of large combinatorial space. Our model goes beyond Anderson's rule for electronic band gap prediction.[25]

■ METHODS

General Workflow. At a high level, a diagram of our workflow is shown in Figure 1. First, material data and ground-state crystal structures are mined from 2D material databases. The ground-state monolayer structures are stacked vertically to create a large material design space of unique bilayer heterostructures, which include both AA and AB stacking configurations. We then sample the vdW heterostructure design space. The bilayer unit cell and atom positions are subsequently optimized, and the corresponding electronic band structures are calculated using high-throughput density functional theory (DFT) calculations.[26] From our DFT calculations, we determine five target material properties: electronic band gap (E_g), electron affinity (EA), ionization energy (IE), interlayer distance (d_0), and interlayer binding energy (E_b). Concurrently, the monolayer material data is featurized and aggregated to construct a set of unique bilayer machine learning descriptors for the entire material design space. Using nearly 800 vdW bilayer heterostructures and their corresponding DFT target properties, we train supervised machine learning models for the five target properties above. The data-driven models are then used to make predictions on the remaining, unlabeled bilayer material space.

2D Material Data Mining for Establishing vdW Heterostructure Design Space. The 2D monolayer crystal structures used in our framework were obtained from the Computational 2D Materials Database (C2DB),[5] an extensive, open-source database that contains crystal structures and material properties from high-throughput DFT calculations and many-body perturbation theory calculations within GW approximation.[27] The C2DB database contains over 3000 different 2D monolayers found in 50 different crystal structure prototypes based on the symmetry and stoichiometry of common 2D materials, such as hexagonal boron nitride (*h*BN) and transition metal dichalcogenides (TMDs). For this work, we select monolayers with unary and binary compounds from seven 2D prototypes based on the crystal structures of MoS_2, CdI_2, GaS, GaSe, GeSe, hBN, and graphene, as shown in Figure 2 and detailed in Table 1. We then filter this subset to include only those 2D monolayers that satisfy five criteria: (1) possessing hexagonal lattice symmetry, (2) having a nonzero band gap, (3) excluding monolayers with Cr, Fe, V, Co, Mn, or Ni elements, (4) excluding monolayers with antiferromagnetic or ferromagnetic stable phase, and (5) excluding monolayers with low thermodynamic or kinetic stability as defined by the C2DB database. While the magnetic heterostructures are of great interest, the proper electronic structure calculations require extensive computational resources to carefully check different magnetic configurations, Hubbard-*U* corrections, and magnetic anisotropies such as exchange and single-ion anisotropy for each heterostructure. For this reason, it would be too challenging and highly demanding for the present work, thus Criteria (3) and (4) were applied. For the same consideration of high computational cost, Criteria (1) was also included. After filtering, we obtained 181 unique monolayers from which vdW bilayer heterostructures are built. A complete list of the 181 monolayers used in the work is listed in Table S1. The monolayer crystals obtained from the C2DB possess a wide range of lattice constants and optoelectronic properties. Except for the graphene prototype (C), all of the monolayers are binary compounds with a generalized chemical formula of XY, XY_2, and X_2Y_2. In many cases, the X species is a metal atom. The monolayers include both direct and indirect band gap semiconductors with band gaps ranging from 0 to 5.94 eV and in-plane lattice constants ranging from 2.46 to 5.23 Å.

Table 1. Prototype Names, Chemical Formula, Space Group, and Number of the Selected 2D Monolayers

2D prototypes	general formula	space group	no. of 2D monolayers
MoS_2	XY_2	$P\bar{6}m2$	57
CdI_2	XY_2	$P\bar{3}m1$	53
GaS	X_2Y_2	$P\bar{6}m2$	28
GaSe	X_2Y_2	$P\bar{3}m1$	19
GeSe	XY	$P3m1$	17
C	X_2	$P6/mmm$	2
hBN	XY	$P\bar{6}m2$	5

After performing data mining of the 181 monolayers, we use their crystallographic information to construct a large set of vdW heterostructures from the 16 290 possible combinations of monolayers. In this project, we construct commensurate heterostructures, in which the crystal structures of the constituent monolayers are scaled to a common in-plane lattice constant. To minimize the strain effects due to large lattice mismatch, as well as mitigate the computational costs of DFT, we focus on those heterobilayers that form commensurate stacking with minimal lattice scaling. Therefore, we accept only pairs of monolayers with a lattice mismatch of 3% or less, where the lattice mismatch is defined as $|a_1 - a_2|/a_1$, where a_1 and a_2 represent the lattice constants of the two individual monolayers with

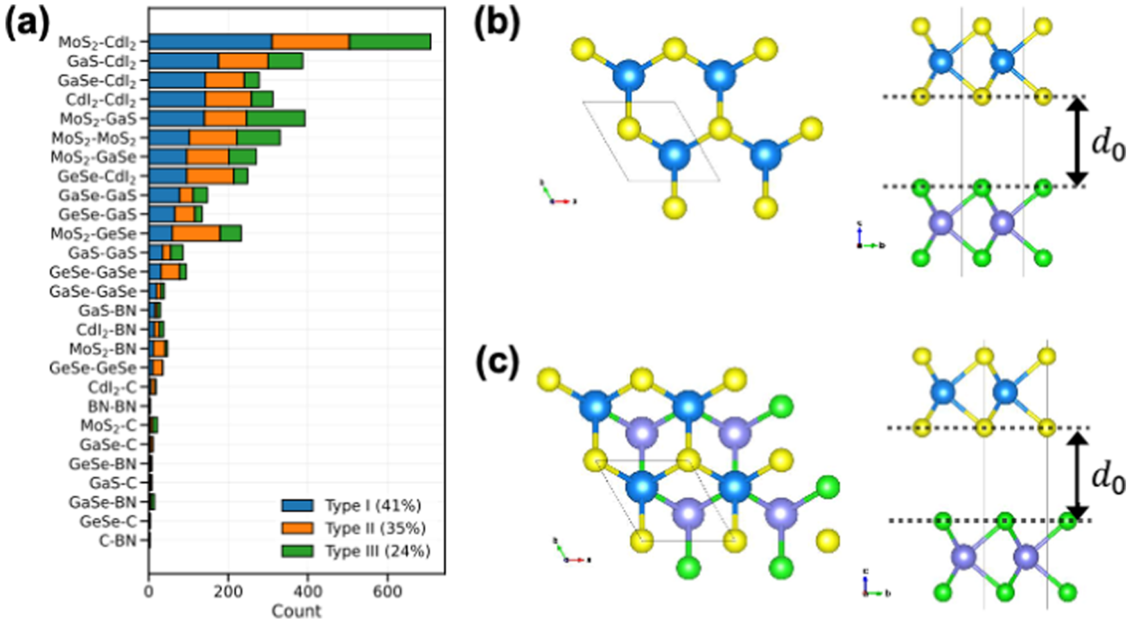


Figure 3. (a) Distribution of bilayer vdW heterostructures with Type I, II, and III band alignment. An illustration of top and side views of an example vdW heterostructure bilayer in the (b) AA stacking configuration and (c) AB stacking configuration. Interlayer distance d_0 is calculated as the vertical distance between the lowest atom of the top layer and the highest atom of the bottom layer.

a_1 being the smaller lattice constant. The heterostructure in-plane lattice constant is determined by the average in-plane lattice value of the two monolayers, and each is scaled accordingly. The scaled monolayers are stacked vertically with a starting interlayer distance of 3.40 Å. To reduce the image interaction from the periodic boundary condition used in DFT, a large vacuum is added along the out-of-plane direction (i.e., *c*-axis). We ensure that all of the heterostructures have the same out-of-plane lattice length (c = 38 Å), which guarantees at least 18 Å of vacuum for each. Since the in-plane lattice constants of each constituent monolayer are identical after scaling, stacking configurations are well defined for the heterostructures. Therefore, it is possible to create different stacking configurations through simple translations and rotations of the top monolayer with respect to the bottom layer following their point group symmetry. In this work, we consider the AA and AB stacking configurations only. Ultimately, our bilayer material space consists of 1950 AA and 1950 AB structures with minimal lattice mismatch.

Each bilayer can be classified by the combination of their monolayer prototypes, yielding 27 different heterobilayer prototypes, as shown in Figure 3a. The C−C bilayer class was excluded during the filtering process. As many monolayers are mostly of the MoS_2 and CdI_2 prototypes, bilayers from these prototypes dominate the proposed design space. Furthermore, we classify the types of band offset (i.e., Type I, Type II, or Type III) for all bilayers according to Anderson's rule based on the monolayer band edges (see Figure 3a). Our bilayer material space consists of 41% Type I, 35% Type II, and 24% Type III heterostructures. Figure 3b,c shows the vdW heterobilayer structure of the MoS_2−MoS_2 prototype in the AA and AB stacking configuration, respectively. According to Anderson's rule, the vacuum levels of two semiconductors on either side of the heterostructure should be aligned at the same energy level. We thus align the vacuum level of the two monolayers to be at 0 eV and establish the position of E_{CBM} and E_{VBM} for both monolayers with respect to the vacuum level. The band gap of vdW heterostructures was then estimated by the difference between the lower CBM energy and the higher VBM energy of the two monolayers. This approach works in the limit of large vacuum separation between the two semiconductors. However, given the relatively vdW weak interaction in vdW heterostructure, it may work qualitatively but not quantitatively since the charge transfer or chemical bonding effects are often minimal in vdW heterostructures. Nevertheless, this Anderson's rule-based approach is a crude estimation, compared to direct first-principles DFT calculations where all of the atomic coordinates, interlayer distance, and Fermi level are fully equilibrated across the vdW heterostructure with all interlayer interactions included such as potential charge transfer.

Labeling Bilayer vdW Heterostructures using High-Throughput DFT Calculations. We sampled over 800 AA and AB stacked bilayers (∼20% of the bilayer material space) with the sampling weights proportional to the size of the structural prototypes of bilayer vdW heterostructures (see Figure 3a), ensuring that the sampled data is representative of the total material design space and that there are, at least, a few bilayers from each prototype in the training data. This subset of bilayers was then cast into the DFT for structural optimization and electronic band structure calculation through our high-throughput workflow. The DFT calculations[26] were carried out using the Vienna ab initio simulation package (VASP).[28] Structural optimization and electronic structure calculations were performed using a plane-wave cutoff energy of 500 eV, the Perdew−Burke−Ernzerhof (PBE) exchange-correlation functional within the generalized gradient approximation (GGA),[29] and a Monkhorst−Pack k-point grid of 11 × 11 × 1. To account for the weak vdW interlayer interaction, we employed the optB86b-vdW functional,[30] which has been shown to give good results for vdW heterostructures.[31] For the structural relaxation, the atomic force convergence threshold was set to 0.02 eV/Å and the electronic energy convergence threshold was set to 1×10^{-6} eV. The electronic band structure was calculated along the high-symmetry k path of Γ−K−M−Γ. To mitigate image interactions due to the periodic boundary conditions used in the plane-wave DFT calculations, we applied a dipole correction to the total energy along the out-of-plane z-direction.[32,33]

Five target material properties (E_g, EA, IE, d_0, and E_b) are extracted from the DFT calculations. The interlayer distance is the vertical distance between the lowest atom of the top layer and the top atom of the bottom layer (Figure 3b), which ensures that the distance is consistent for all structure prototypes. E_b is defined by

$$E_{\mathrm{b}} = \frac{(E^{\mathrm{ML1}} + E^{\mathrm{ML2}} - E^{\mathrm{Het}})}{A} \tag{1}$$

where E^{Het} is the total energy of the heterostructure bilayer, E^{ML1} and E^{ML2} are the total energy of the constituent monolayers, and A is the surface area. It is important to note that the above definition of E_{b}

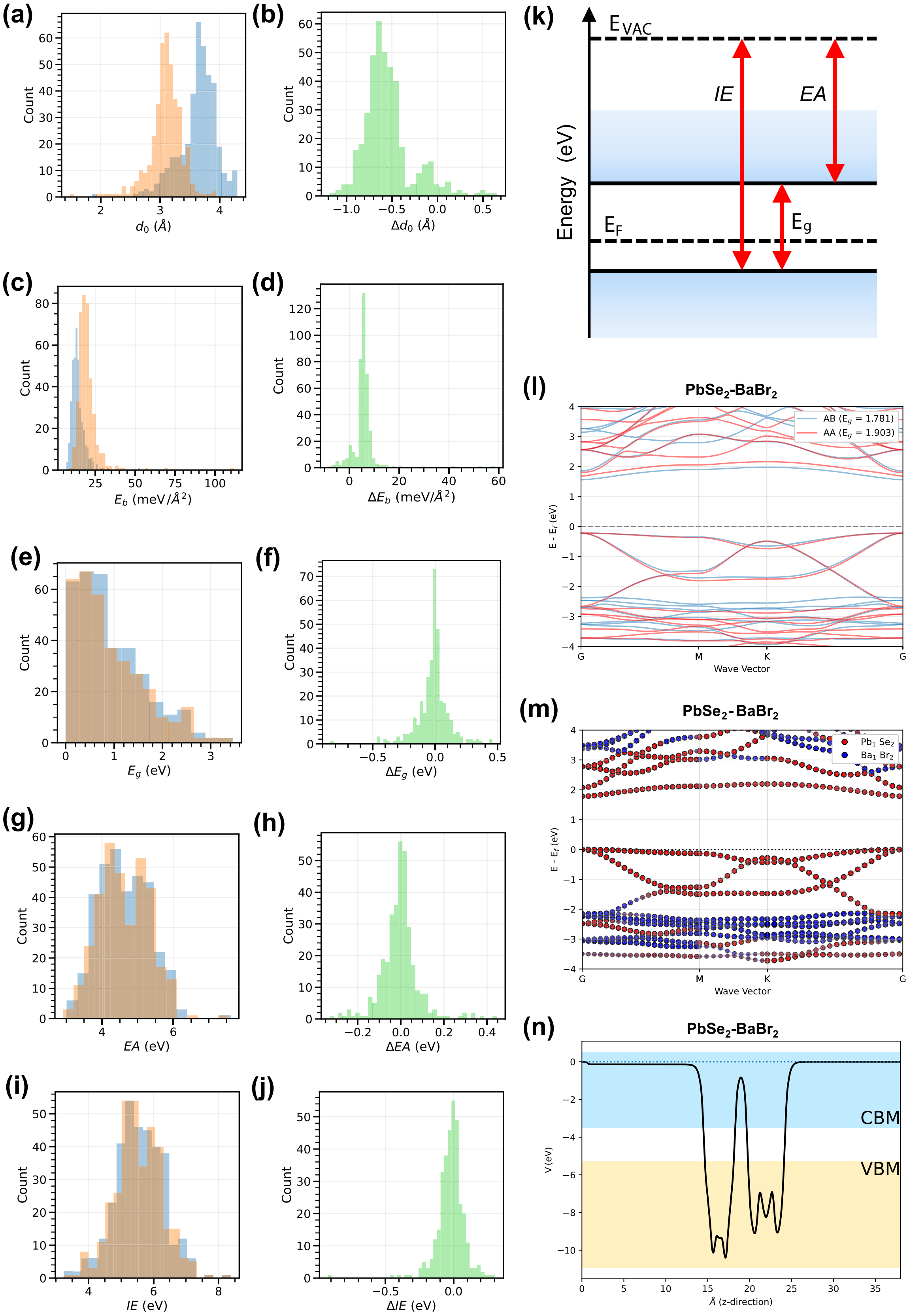


Figure 4. Distribution of the (a, b) interlayer distance d_0 and its difference between AA and AB ($\Delta d = d_0^{AB} - d_0^{AA}$), (c, d) interlayer binding energy E_b and its difference between AA and AB ($\Delta E_b = E_b^{AB} - E_b^{AA}$), (e, f) band gap E_g and its difference between AA and AB ($\Delta E_g = E_g^{AB} - E_g^{AA}$), (g, h) ionization energy IE and its difference between AA and AB ($\Delta \mathrm{IE} = \mathrm{IE}^{AB} - \mathrm{IE}^{AA}$), and (i, j) electron affinity EA and its difference between AA and AB ($\Delta \mathrm{EA} = \mathrm{EA}^{AB} - \mathrm{EA}^{AA}$). (k) Definitions of E_g, EA, and IE with respect to the vacuum level. From the high-throughput DFT, we collect the (l) AA vs AB band structure plot, (m) project band structure plot with respect to the constituent monolayers, and (n) local potential plot for each vdW heterostructure.

results in a positive value for those bilayers with favorable attractive interaction across the interface (i.e., the ground-state total energy of the bilayer is lower than that of the monolayers). The band gap is taken as the difference between VBM, E_{VBM}, and CBM, E_{CBM}. The IE and EA of the bilayer vdW heterostructure are determined by

$$\mathrm{IE} = E_{\mathrm{vac}} - E_{\mathrm{VBM}} \tag{2}$$

$$\mathrm{EA} = E_{\mathrm{vac}} - E_{\mathrm{CBM}} \tag{3}$$

where E_{vac} is the vacuum level. In some bilayers, the formation of a dipole results in two vacuum energy levels. For these cases, we average the vacuum energy level when determining the EA and IE. Additionally, the E_g, EA, and IE are related by

$$E_g = \mathrm{IE} - \mathrm{EA} \tag{4}$$

We exploit this relationship later in our machine learning models to predict E_g indirectly by using the two separate predictions for IE and EA.

Descriptors and Machine Learning of Bilayer vdW Heterostructures. Material descriptors are an essential part of any machine learning model and must capture the underlying material information. For our purposes, we construct a series of bilayer descriptors based on the descriptors of the constituent monolayers. Monolayer descriptors are constructed using material properties from the C2DB database and Matminer Python package.[34] From the Matminer library, we featurize the monolayer density of state (DOS) using the *DOSFeaturizer* and *Hybridization* featurizers, as well as the atomic composition using the *ElementalProperty* featurizer. The *Hybridization* featurizer quantifies the s, p, d, and f orbital character of the band edges. The *ElementalProperty* featurizer calculates composition-based features using various fundamental elemental properties such as the melting point, covalent radius, and electronegativity. The *DOSFeaturizer* featurizes the valence and conduction band edges, returning the major atomic orbital character contributor (e.g., "s"), the corresponding fraction or "score" that the major orbital contributes to the band edge, the major atomic species contributor (e.g., "Mo"), and the amount of hybridization determined by an entropy score (i.e., $x \ln x$). We further engineer this descriptor by encoding the orbital characters (s, p, d, or f) using one-hot encoding, a commonly used categorical data encoder that creates a binary "dummy variable" for each categorical feature.[35] The atomic species are encoded by using several atomic properties (i.e., Mendeleev number, atomic mass, electronegativity, atomic radius, etc.).

To compose a descriptor vector for a bilayer, the unique feature vectors of the monolayers are combined using an aggregation function. An aggregation function is a simple calculation that combines multiple data points into one, such as the average, median, mode, range, etc. Several aggregation methods were tested, as well as combinations of different aggregations functions, but this approach yielded an excessively high-dimensional feature space consisting of many highly correlated features. Therefore, in this work, the monolayer features are aggregated by taking the average only. This approach ensures a unique bilayer descriptor vector that is invariant with respect to stacking order (e.g., $\mathrm{MoS_2{-}MoSe_2}$ and $\mathrm{MoSe_2{-}MoS_2}$) while still encoding the physical and electronic characteristics of its constituent monolayers.

In addition, we include the CBM and VBM of the constituent monolayers (min_cbm, max_cbm, min_vbm, max_vbm). We also take the difference between the minimum CBM and the maximum VBM (min_cbm-max_vbm). These descriptors attempt to encode Anderson's rule as the monolayer band edges are aligned with respect to the vacuum energy. We construct a few more descriptors to distinguish the bilayer stacking configurations and band edge alignment. These include the stacktype (encoded as AB = 0 and AA = 1) and one-hot encoding of the vdW heterostructure band alignment (Type I, II, and III) from Anderson's rule. After building our descriptors, we filter out those descriptors with zero/low variance and other descriptors with extreme skewness (or asymmetry). After filtering, we obtained 91 bilayer descriptors, which are used to train our machine learning models. A complete list of our descriptors can be found in Table S2, including a discussion on the work function feature.

From the labeled training data, we remove any bilayers with interlayer distance $d_0 < 2.5$ Å and interlayer binding energy $E_b > 40$ meV/Å^2, as these bilayers are generally unstable or form chemical bonds at the interface, which leaves us with 776 vdW bilayers. We use all of the 776 vdW bilayers to train the models for an interlayer distance d_0 and an interlayer binding energy E_b. For training the models for electronic properties (i.e., E_g, EA, and IE), we remove metallic and Type III bilayers, as the zero band gap vdW heterostructures may complicate the machine learning models, leaving 595 vdW bilayers for training E_g, EA, and IE. The training data is finally scaled to have zero mean and unit standard deviation.

Before applying various supervised learning models, feature or descriptor selection is first performed using the least absolute shrinkage and selection operation (LASSO). The LASSO estimator is given by

$$\hat{\beta}^{\mathrm{Lasso}} = \arg\min_{\beta}\left\{\frac{1}{2}\sum_{i=1}^{N}\left(y_i - \beta_0 - \sum_{j=1}^{p} x_{ij}\beta_j\right)^2 + \lambda\sum_{j=1}^{p}|\beta_j|\right\} \tag{5}$$

where the first term is a cost function based on least-square linear fitting and the second term is the l_1 penalty term. The l_1 penalty function can force certain coefficients corresponding to less predictive feature variables to be exactly zero, yielding a sparse model consisting of a subset of the original feature variables. Hyperparameter λ controls the extent of regularization, whereas when $\lambda = 0$, the resulting model is reduced to the ordinary least-squares (OLS) regression. The λ value is determined through a five-fold cross-validation in our experiments.

The LASSO is a regularization technique commonly adopted for descriptor selection in machine learning, which can improve both prediction accuracy and interpretability by descriptor subset selection through model coefficient regularization. To select a suitable subset of descriptors, we first apply LASSO linear regression on the complete list of descriptors and fit the model. The LASSO selected descriptors are then fed into other machine learning models. We use the *scikit-learn* Python library[36] to build our machine learning pipelines. A comprehensive list of the models is used in this work including AdaBoost, Elastic Net (EN), Gradient Boosted Trees (GBT), K-Neighbor Regression (KNN), Kernel Ridge Regression (KRR), LASSO, Random Forest (RF), Ridge Regression (Ridge), Support Vector Regression (SVR), Stacked Ensemble Method (SEM). More details can be found in Table S3. We include both linear, nonlinear, and ensemble machine learning models to predict our target properties.

To estimate each model's performance on unseen data (i.e., the generalization error), we employ a repeated *k*-fold cross-validation. In a repeated *k*-fold cross-validation, a conventional train/validation/test split is repeated multiple times. For smaller data sets, the repeated *k*-fold validation can better estimate the generalization error compared to other cross-validation techniques like single *k*-fold or leave-one-out.[37−39] For this work, we apply a five-fold cross-validation ($k = 5$) over 20 repetitions (for total 100 fits) and take the average test score as our estimate of the generalization error estimate. To measure the model's accuracy, we use the mean-absolute error (MAE) as our primary metric, but we also report the mean-squared error (MSE), root-mean-squared error (RMSE), and the coefficient of determination (R^2).

RESULTS AND DISCUSSION

Hight-Throughput DFT Results. The DFT-calculated equilibrium interlayer distance (d_0) for those bilayers ranges between 1.48 and 4.39 Å, with an average of 3.36 Å. These results agree with other DFT studies of vdW heterostructures.[40,41] The d_0 is strongly influenced by the stacking configuration. The average d_0 for the AA and AB stacking is 3.64 and 3.09 Å, respectively, as displayed in Figure 4a. The difference in d_0 between the AA and AB stacked bilayers (see

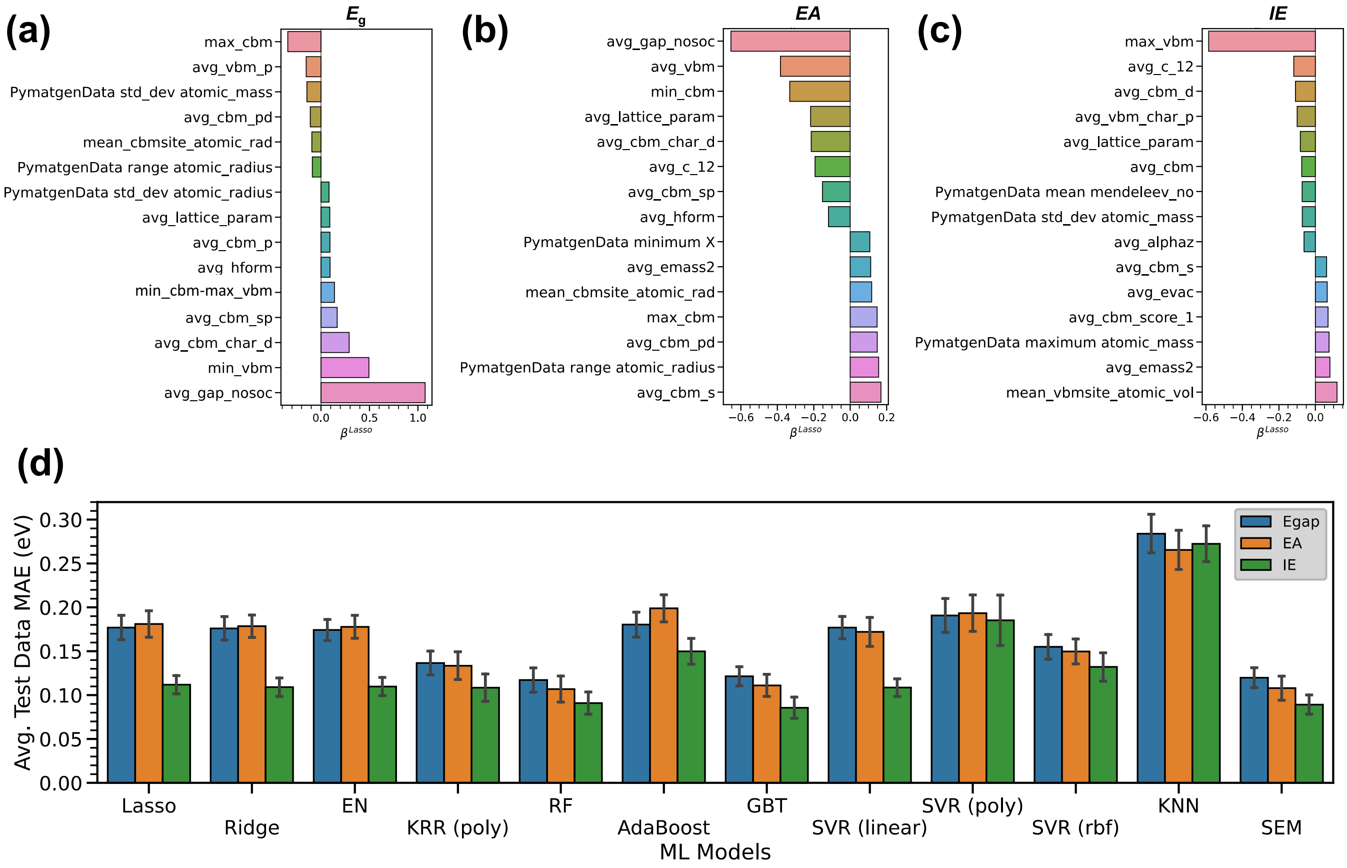


Figure 5. (a−c) Top 15 descriptors for the E_g, EA, and IE and their LASSO coefficients. (d) Average repeated k-fold cross-validation MAE for different machine learning algorithms. The black vertical line represents the standard deviation of the testing MAE.

Figure 4b) can be as large as 1.1 Å in favor of the AB stacking configuration. The preference for the AB stacking is likely due to steric effects from the atoms facing each other across the vdW gap. As shown in Figure 4c,d, E_b is also influenced by the stacking configuration. All of the calculated E_b for our set of bilayers are positive (see eq 1), suggesting that all display attractive forces between the constituent monolayers as expected. The AB stacking configuration is the lower-energy configuration for the majority of vdW heterostructures (∼96%) displaying larger interlayer binding energy, consistent with the above shorter interlayer distance d_0. The average binding energy for the AA and AB stacking is 14.0 and 19.8 meV/Å^2, respectively. The difference between the AA and AB stacking is ∼6 meV/Å^2 but can be as high as 10−20 meV/Å^2. The prevalence of a low-energy AB stacking configuration in contrast to AA stacking is also likely due to steric effects. Generally, we found that the E_b will increase monotonically as the d_0 decreases. Compounds that displayed $d_0 < 2.5$ Å generally did not form stable bilayers and instead are likely to form chemical bonds at the interface. Correspondingly, these compounds displayed $E_b > 40$ meV/Å^2. The large difference in the equilibrium interlayer distance in the AA and AB stacking configurations is likely to have other important effects on the bilayer geometry. For example, the energy difference can drive the reconstruction of the moiré superlattice.[42,43] Our results indicate that the moiré superlattice made of these vdW heterostructures will undergo atomic reconstruction with spatially varying out-of-plane bending and in-plane tension and compression, which will ultimately impact the interactions between the quasiparticles in the moiré superlattice.

The electronic properties (i.e., E_g, EA, and IE) are collected from the DFT calculations and the band edges are shifted relative to the vacuum energy (see Figure 4k). We find that the electronic structure of the bilayers is not as strongly influenced by the bilayer stacking configuration as their interlayer distance and interlayer binding energy, as indicated by the strongly overlapped distribution in Figure 4e−j. We find that the band gap energies range from ∼0 eV to nearly 4 eV, allowing for photon absorption in the infrared to the visible regions. The change in band gaps between the AA and AB stacked bilayer $|\Delta E_g|$ can be as large as 82 meV, suggesting that stacking configuration does affect band gap, which is particularly important for the formation of Moiré exciton in twist bilayer and multilayer vdW heterostructures.[44] For all of the bilayers used in high-throughput DFT calculations, we collect the relevant target properties as well as plot the electronic band structure and local potential that are provided in the Supporting Information. Among the 595 vdW bilayers, we identify 85 vdW heterostructures with direct band gap. A complete list of 85 direct band gap vdW heterostructures can be found in Table S4.

Predictions of Electronic Structure Properties Using the Machine Learning Models. After the DFT calculations and filtering steps, 595 bilayers are used to train the machine learning models. First, these structures are passed with the full set of features to the LASSO estimator. The LASSO estimator reduced the number of descriptors from 91 to 59 descriptors for the E_g pipeline. For the EA and IE models, the descriptor space is reduced to 55 for both targets. A complete list of the E_g, EA, and IE LASSO selected descriptor variables and their

Pearson's correlation can be found in Figures S4−S9. Since the descriptor space is normalized, the magnitudes of the LASSO coefficients can be used to assess the "importance" of the descriptors relative to the other descriptors. The top 15 descriptors for the E_g, EA, and IE and their LASSO coefficients are displayed in Figure 5a−c, respectively. We see that VBM and CBM of the monolayers are two of the best descriptors for the electronic structure properties. The maximum VBM of the two monolayers is a strong predictor of the bilayer IE. Likewise, the minimum CBM of the two monolayers is a strong predictor of the bilayer EA. This arises from the fact that when the band edges of the monolayers are aligned, we can use the difference between the highest (maximum) VBM and lowest (minimum) CBM of the two constituent monolayers to obtain a crude estimate of the bilayer band gap energy, i.e., Anderson's rule. Furthermore, the average band gap of the two monolayers is also a strong predictor of bilayer band gap energy. We find that the stacking configuration descriptor ("stacktype") is a weaker predictor of the band gap, because the band gap differences between the AA and AB stacking configurations are relatively smaller compared to the band gap and band edge difference in the different 2D monolayers.

The LASSO descriptors are used to train a series of additional linear and nonlinear models. To compare each model prediction performance, we evaluate the models using a repeated *k*-fold cross-validation. The average test MAE for the E_g, EA, and IE models is displayed in Figure 5d, while Table 2 provides the average testing and training MAE. With the exception of the KRR model with the rbf kernel, all of the tested machine learning models yield predictions with MAEs as low as 0.09 eV across all target properties (E_g, EA, and IE). Additionally, we see that the standard deviation of the test error is small (only within a few tenths of eV), suggesting that many of the models are stable (i.e., robust against small changes in the data due to the resampling). Since we can achieve multiple models with low prediction error, model selection is a question of the strengths and weaknesses of the individual model. For example, the tree-based methods like RF and GBT are capable of fitting nonlinear data and are generally robust against high-dimensional data sets and the curse of dimensionality. However, these models tend to capture the noise along with the underlying relationship (i.e., overfitting) and are unable to extrapolate beyond the training data. On the other hand, linear models are unable to map nonlinear relationships, but they are less complex and easy to interpret. In addition, overfitting can be mitigated in linear models by employing regularization methods to minimize model variance at a cost of bias. The KRR model with the rbf kernel is discarded due to its high error for the EA and IE. The RF and GBT models yield the lowest testing error. However, these models have a greater discrepancy between the training and testing error, suggesting overfitting (high bias, low variance). In contrast, the regularized linear models (LASSO, Ridge, and Elastic Net) show higher errors on the test data compared to the more complex models but display less overfitting when comparing the training and test errors.

Table 2. Average Testing and Training MAE from Repeated *k*-Fold Cross-Validation for the Band Gap (E_g), Electron Affinity (EA), and Ionization Energy (IE) Models[a]

machine learning algorithm	E_g MAE (eV)	EA MAE (eV)	IE MAE (eV)
Random Forest Regressor	0.117 (0.044)	0.107 (0.040)	0.090 (0.034)
Gradient Boosting Trees	0.122 (0.065)	0.113 (0.064)	0.085 (0.046)
Stacked Ensemble Regressor	0.122 (0.075)	0.112 (0.069)	0.090 (0.064)
Kernel Ridge (polynomial kernel)	0.138 (0.082)	0.133 (0.081)	0.109 (0.061)
SVR (rbf kernel)	0.154 (0.097)	0.151 (0.098)	0.133 (0.083)
Elastic Net	0.175 (0.155)	0.178 (0.156)	0.109 (0.097)
Ridge Regression	0.176 (0.154)	0.178 (0.155)	0.109 (0.096)
LASSO	0.176 (0.155)	0.179 (0.157)	0.111 (0.098)
SVR (linear kernel)	0.176 (0.153)	0.170 (0.147)	0.108 (0.097)
AdaBoost	0.180 (0.162)	0.199 (0.179)	0.149 (0.134)
SVR (polynomial kernel)	0.191 (0.125)	0.198 (0.129)	0.186 (0.120)
Kernel Ridge (rbf kernel)	0.212 (0.161)	n/a	n/a
K-Neighbors Regressor	0.277 (0.212)	0.267 (0.202)	0.271 (0.207)

[a] The training MAEs are shown inside the parenthesis.

To balance the strengths and weaknesses of different models, we proceed to employ a stacked ensemble method (SEM) rather than deploying a single model for predictions on the unlabeled vdW bilayer heterostructures. In machine learning, an SEM is an approach where multiple base machine learning models are combined (or "stacked") into a single, higher-level meta-learner. The meta-learner learns the best way to combine the predictions of the base models, which then makes the final prediction. SEM has been shown to yield better machine learning performance when compared to individual models.[45] Here, we use the GBT, Ridge Regression, linear SVR, and a polynomial kernel ridge regression as the base models. These models were chosen as the base models to promote diversity to include a variety of model types (i.e., linear, tree-based, kernel-methods) and they also gave low prediction error when applied individually. The base model predictions were passed to an ordinary least-squares regression meta-estimator. As shown in Table 2, the testing (training) MAE for the SEM model is 0.122 (0.075), 0.112 (0.069), and 0.09 (0.064) eV for the E_g, EA, and IE properties, respectively. The MAE, MSE, RSME, and R^2 testing and training metrics for all E_g, EA, and IE models can be found in Tables S5−S7, respectively.

Owing to the diversity of the monolayer prototypes and compositions, the vdW bilayer material space is equally diverse, consisting of many different bilayer families (i.e., MoS_2−CdI_2, GaSe−CdI_2). However, some monolayer prototypes, such as the MoS_2 and CdI_2 prototypes, have more compounds than others. This means that the vdW heterostructure material space is composed of many bilayers from these prototypes. This introduces an imbalance with respect to bilayer prototypes (see Figure 3a). In machine learning models, it is important to assess if this bias in the data will adversely affect the model performance. To evaluate if our SEM machine learning model is biased toward the larger prototype classes, we compared the residuals for the top seven prototype classes and the combined residuals for the remaining 20 classes, as shown in Figure 6a. We find that the residuals are comparable across the prototype classes, where the middle 50% of residuals

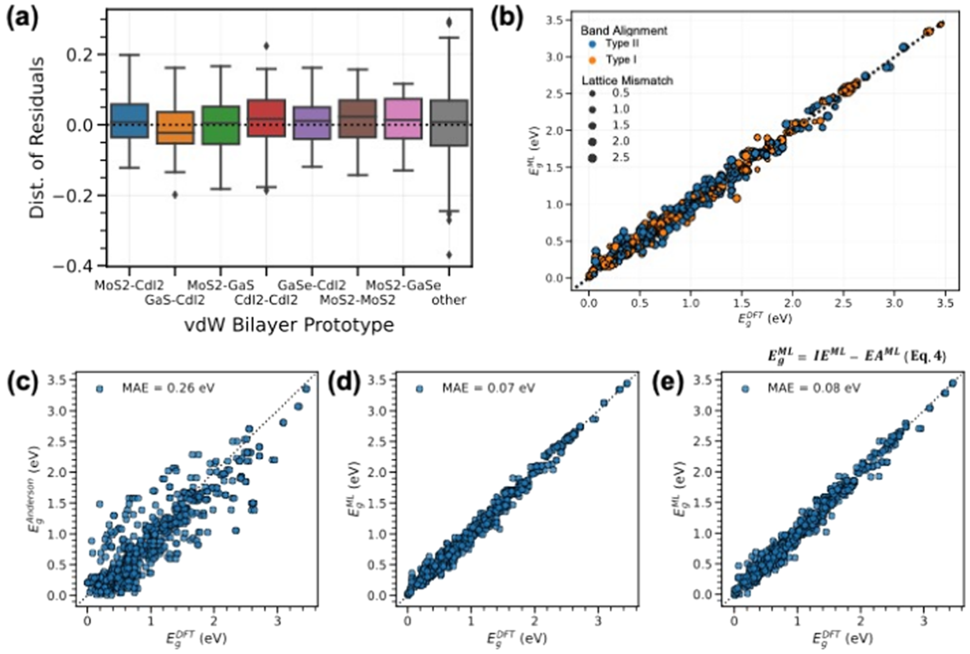


Figure 6. (a) Distribution of band gap residuals for different bilayer prototypes. (b) DFT- and ML-predicted band gap where the orange and blue colors represent Type I and Type II band alignments, respectively. The size of the datapoint corresponds to the lattice mismatch percent. The prediction of E_g using (c) Anderson's rule, (d) machine learning model of E_g, and (e) machine learning model of EA and IE with E_g defined by E_g = EA − IE.

are within roughly ±0.1 eV. Comparing the residual distributions of the top seven and the remaining 20 combined, with the exception of a few outliers, we find that the residuals of the smaller classes are not significantly greater than the top seven. We further evaluate the model with respect to the band alignment class. From Figure 6b, we see that the model can accurately predict both the Type I and Type II bilayers found in the training data. Furthermore, we see no relationship between the model error and the extent of lattice mismatch. Considering these results along with our previous repeated *k*-fold cross-validation, we are confident that the SEM model is accurate and able to generalize successfully on the unseen vdW heterostructure space.

Figure 6c−e shows the band gap estimations/predictions using Anderson's rule, the machine learning model of E_g, and the machine learning model of EA and IE (with E_g defined by E_g = EA − IE), respectively. Anderson's rule estimates the band gap by the difference between the minimum CBM and the maximum VBM of the two monolayers. While this approach does give qualitative agreement with the trend of the DFT-calculated results, it does not account for important interlayer interactions, such as charge transfer and dipole formation, which may affect the band structure of vdW heterostructures. Using Anderson's rule, we can make crude estimates of the band gap energy. However, using machine learning models, we can directly predict the E_g of bilayer vdW heterostructures with the direct band gap model or indirectly with the IE and EA models. Our machine learning models show a significant decrease in the MAE compared to the application of Anderson's rule. Furthermore, our model contains many different monolayers from a variety of bilayer prototypes. This suggests that our machine learning model is robust against the monolayer composition and structure (i.e., monolayer prototypes).

Anderson's rule[25,46] assumes no interlayer interactions or charge transfer between two materials at the junction. Therefore, it may work in the case of weak coupling such as vdW interaction in the vdW heterostructures.[41,47,48] However, additional effects such as charge transfer, charge redistribution, and interfacial dipoles may need to be taken into account. For example, charge transfer has been experimentally observed in MoS_2−WS_2[49] and graphene−WS_2 vdW heterostructures.[50] Our machine learning models show that while Anderson's rule works well for qualitative prediction of band alignment, it does show a larger MAE from Anderson's rule compared to our machine learning model, which not only considers the band edges of the individual monolayers as Anderson's rule does but also includes other factors that are embedded in our material descriptors beyond the band edges of monolayers.

Predictions of Interlayer Distance and Interlayer Binding Energy. After the DFT calculations and filtering steps, 776 bilayers are used to train the machine learning models for the interlayer distance d_0 and interlayer binding energy E_b, the LASSO estimator reduces the descriptor space from 91 descriptors to 64 and 50, respectively. Similarly, the relative descriptor importance can be gauged from the β^{Lasso}. The top 15 descriptors for the d_0 and E_b and their corresponding β^{Lasso} values are shown in Figure 7a,b. A complete list of the d_0 and E_b LASSO selected descriptors and Pearson's correlation can be found in Figures S10−S13. Unlike the electronic property models, the interlayer distance d_0 strongly depends on the stacking configuration. This is successfully learned by the LASSO estimator where the "stacktype" descriptor is one of the top descriptors for both the interlayer distance and interlayer binding energy models. The stacking configuration is encoded as 0 for AB and 1 for AA. Therefore, for the interlayer distance model, the *positive* model parameter associated with the "stacktype" descriptor suggests that interlayer distance d_0 *increases* with the AA type,

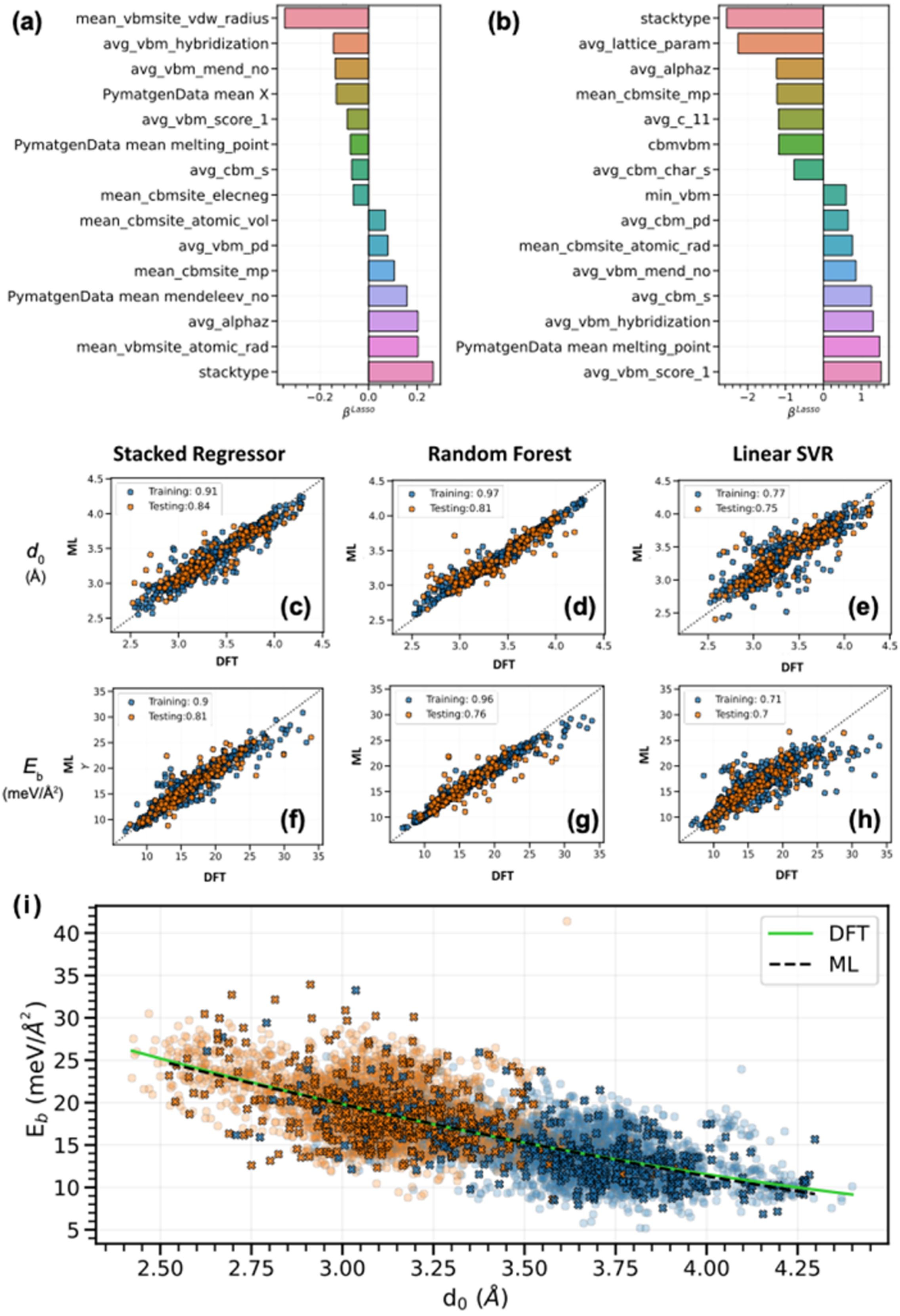


Figure 7. Top 15 descriptors for the (a) d_0 and (b) E_b models via the LASSO estimator descriptor selection. Comparison of the DFT and ML predictions using the Stacked Regressor, RF, and linear SVR models for the (c−e) d_0 and (f−h) E_b target properties. (i) Binding energy (E_b) vs the interlayer distance (d_0). The "x" data points correspond to the training data, while "o" represents the predictions on the unlabeled bilayers. Orange and blue colors represent AB and AA stacked bilayer vdW heterostructures, respectively.

while for the interlayer binding energy model, the *negative* model parameter associated with the "stacktype" descriptor indicates that interlayer binding energy E_b *decreases* with the AA type. The average vdW radius is correlated with a decrease in d_0. We also see that the higher the average hybridization of the monolayer VBM (avg_vbm_hybridization), the lower d_0, suggesting that an increase in hybridization will reduce the vdW distance between the two layers, which is consistent with our DFT calculations and physical intuition.

We again evaluate several machine learning models by repeated k-fold cross-validation. We apply five-fold cross-validation repeated over 20 iterations. Like the previous case, we use GBT, SVR (linear), KRR (polynomial), and Ridge as base models for the Stacked Ensemble Method (SEM). The average training and testing errors are given in Table 3. The MAE for the d_0 predictions ranges from 0.119 Å for the SEM model to 0.294 Å for the KNN model. The MAE for the E_b ranges from 1.39 to 2.953 meV/Å^2. For all metrics, refer to Tables S8 and S9 for the d_0 and E_b models, respectively. Figure 7c−h shows the DFT vs ML predictions with the corresponding training and testing R^2 score for the Stacked Ensemble Method, Random Forest, and Linear SVR models for predictions on the unlabeled bilayers. We observe that the residuals are greater in the d_0 and E_b models, compared to the electronic property models (E_g, EA, and IE). The machine learning models can predict the interlayer distance within a few tenths of Å. The E_b is predicted to be within a few meV/Å^2 of the DFT-calculated energy, especially at larger E_b. It is important to note that the models for the d_0 and E_b are more sensitive to the class imbalance of the vdW heterostructure classes. Figure S14 shows that much of the prediction variance is from those bilayers coming from the smaller bilayer classes. We also see that the machine learning models tend to slightly underfit the E_b compared with the DFT-calculated energies. Like the electronic predictions, the RF and GBT ensemble methods display a lower testing MAE but greater variance when compared to the linear models (i.e., bias-variance trade-off). The SEM model displays the lowest testing MAE when predicting both d_0 and E_b.

Table 3. Average Testing (Training) MAE from Repeated k-Fold Cross-Validation for Interlayer Distance d_0 and Interlayer Binding Energy E_b Models

machine learning algorithm	d_0 MAE (Å)	E_b MAE (meV/Å^2)
Stacked Ensemble Method (SEM)	0.119 (0.078)	1.387 (0.902)
Random Forest	0.119 (0.044)	1.514 (0.564)
Gradient Boosted Trees	0.122 (0.067)	1.496 (0.817)
Kernel Ridge (poly)	0.130 (0.078)	1.438 (0.826)
SVR (linear)	0.136 (0.122)	1.707 (1.534)
Elastic Net	0.139 (0.124)	1.784 (1.631)
Ridge	0.140 (0.125)	1.786 (1.638)
LASSO	0.141 (0.125)	1.807 (1.633)
Ordinary Least Squares	0.142 (0.125)	1.799 (1.633)
SVR (rbf)	0.152 (0.099)	1.974 (1.639)
AdaBoost	0.156 (0.133)	2.278 (1.980)
SVR (poly)	0.193 (0.121)	2.458 (1.971)
K-Neighbors Regressor	0.294 (0.237)	2.953 (2.392)

We deploy the models on the unlabeled bilayers and obtained the predictions of the d_0 and E_b. Figure 7i displays the binding energy vs the interlayer distance. There is a clear separation between the AA (blue) and AB (orange) interlayer distances near 3.40 Å. We attempt to construct a crude mathematical relationship between d_0 and E_b by fitting a 2nd-order polynomial using the DFT-calculated results and the machine learning predictions. The fitting yields $E_b^{\mathrm{DFT}}(d_0) = 1.12d_0^2 - 16.35d_0 + 58.73$ and $E_b^{\mathrm{ML}}(d_0) = 1.65d_0^2 - 19.86d_0 + 64.51$, for the DFT and machine learning-predicted data, respectively. The similarity between the DFT and ML fits suggests that the machine learning approach is sufficient to make confident predictions on the unlabeled vdW heterostructure design space. Conventionally, the equilibrium interlayer distance is found by a multistep calculation where the interlayer distance is manually altered, obtaining the corresponding system energy at each step until a local minimum is found. These machine learning models offer the opportunity to reduce the number of calculations required to determine an equilibrium interlayer distance by knowing beforehand the equilibrium interlayer distance. The clear separation of the AA and AB stacking configurations in the equilibrium interlayer distance by the machine learning model predictions supports the earlier observation in the high-throughput DFT calculations. That is, the majority of moiré superlattices made of twisted vdW heterostructures are expected to undergo atomic reconstruction, which will further impact the underlying electronic structure.

■ CONCLUSIONS

In summary, we presented a computational framework to establish the relationship between 2D monolayers and their vdW heterostructures by combining first-principles DFT calculations, machine learning approach, and existing 2D material database. We constructed a large material library of vdW bilayer heterostructures, which were encoded into a series of descriptors from their constituent monolayers. To predict the atomic and electronic structure of vdW heterostructures, our models were built on top of our large and diverse dataset of nearly 4,000 unique bilayer structures from seven different hexagonal monolayer prototypes. The developed machine learning models were found to predict the bilayer E_g, IE, EA, d_0, and E_b with low error, providing a valuable tool for screening the vast vdW heterostructure material space with a significantly reduced computational cost for a wide range of optoelectronic applications. Our results also suggest that the majority of the vdW heterostructures have a large difference in the equilibrium interlayer distance for AA and AB stacking configurations. Consequently, their corresponding twisted moiré heterostructures are expected to undergo atomic reconstruction with spatially varying out-of-plane bending and in-plane tension and compression, which will ultimately impact the interactions between the quasiparticles in the moiré superlattice.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acsami.2c04403.

> List of monolayers, machine learning descriptors, machine learning model descriptions, direct band gap bilayers, LASSO feature selection tornado charts, prediction error metrics, and analysis of bilayer class vs model predictions (PDF)

AUTHOR INFORMATION

Corresponding Authors

Tahir Cagin − *Department of Material Science and Engineering, Texas A&M University, College Station, Texas 77843, United States*; Email: tcagin@tamu.edu

Ruth Pachter − *Air Force Research Laboratory, Materials and Manufacturing Directorate, Wright-Patterson Air Force Base, Ohio 45433, United States*; orcid.org/0000-0003-3790-4153; Email: ruth.pachter@us.af.mil

Xiaofeng Qian − *Department of Material Science and Engineering and Department of Electrical and Computer Engineering, Texas A&M University, College Station, Texas 77843, United States*; orcid.org/0000-0003-1627-288X; Email: feng@tamu.edu

Authors

Daniel Willhelm − *Department of Material Science and Engineering, Texas A&M University, College Station, Texas 77843, United States*; orcid.org/0000-0002-9365-0744

Nathan Wilson − *Department of Material Science and Engineering, Texas A&M University, College Station, Texas 77843, United States*

Raymundo Arroyave − *Department of Material Science and Engineering, Texas A&M University, College Station, Texas 77843, United States*

Xiaoning Qian − *Department of Electrical and Computer Engineering, Texas A&M University, College Station, Texas 77843, United States*

Complete contact information is available at:
https://pubs.acs.org/10.1021/acsami.2c04403

Author Contributions

Xiaofeng Qian, R.P., and T.C. supervised the project. D.W. performed the theoretical DFT calculations and trained and tested machine learning models. All authors conducted theoretical analysis and analyzed the results. D.W., R.P., T.C., and Xiaofeng Qian wrote the manuscript with the help of others.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

The authors gratefully acknowledge the support of the US Air Force Research Laboratory (AFRL) Minority Leaders Program under subcontract 165852-19F5830-19-02-C1, the US National Science Foundation (NSF) under Award Numbers OAC-1835690 and DMR-2103842, and the US NSF NRT-DESE: Data-Enabled Discovery and Design of Energy Materials (D3EM) under Grant Number 1545403.

REFERENCES

(1) Novoselov, K. S.; Geim, A. K.; Morozov, S. V.; Jiang, D.; Zhang, Y.; Dubonos, S. V.; Grigorieva, I. V.; Firsov, A. A. Electric Field in Atomically Thin Carbon Films. *Science* **2004**, *306*, 666−669.

(2) Rasmussen, F. A.; Thygesen, K. S. Computational 2D Materials Database: Electronic Structure of Transition-Metal Dichalcogenides and Oxides. *J. Phys. Chem. C* **2015**, *119*, 13169−13183.

(3) 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.

(4) Ashton, M.; Paul, J.; Sinnott, S. B.; Hennig, R. G. Topology-Scaling Identification of Layered Solids and Stable Exfoliated 2D Materials. *Phys. Rev. Lett.* **2017**, *118*, No. 106101.

(5) Haastrup, S.; Strange, M.; Pandey, M.; Deilmann, T.; Schmidt, P. S.; Hinsche, N. F.; Gjerding, M. N.; Torelli, D.; Larsen, P. M.; Riis-Jensen, A. C.; Gath, J.; Jacobsen, K. W.; Mortensen, J. J.; Olsen, T.; Thygesen, K. S. The Computational 2D Materials Database: High-Throughput Modeling and Discovery of Atomically Thin Crystals. *2D Mater.* **2018**, *5*, No. 042002.

(6) 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.

(7) Choudhary, K.; Garrity, K. F.; Reid, A. C. E.; DeCost, B.; Biacchi, A. J.; Hight Walker, A. R.; Trautt, Z.; Hattrick-Simpers, J.; Kusne, A. G.; Centrone, A.; Davydov, A.; Jiang, J.; Pachter, R.; Cheon, G.; Reed, E.; Agrawal, A.; Qian, X.; Sharma, V.; Zhuang, H.; Kalinin, S. V.; Sumpter, B. G.; Pilania, G.; Acar, P.; Mandal, S.; Haule, K.; Vanderbilt, D.; Rabe, K.; Tavazza, F. The Joint Automated Repository for Various Integrated Simulations (JARVIS) for Data-Driven Materials Design. *npj Comput. Mater.* **2020**, *6*, No. 173.

(8) Singh, A. K.; Mathew, K.; Zhuang, H. L.; Hennig, R. G. Computational Screening of 2D Materials for Photocatalysis. *J. Phys. Chem. Lett.* **2015**, *6*, 1087−1098.

(9) Li, W.; Qian, X.; Li, J. Phase Transitions in 2D Materials. *Nat. Rev. Mater.* **2021**, *6*, 829−846.

(10) Akinwande, D.; Petrone, N.; Hone, J. Two-Dimensional Flexible Nanoelectronics. *Nat. Commun.* **2014**, *5*, No. 5678.

(11) Sun, Z.; Martinez, A.; Wang, F. Optical Modulators with 2D Layered Materials. *Nat. Photonics* **2016**, *10*, 227−238.

(12) Qian, X.; Liu, J.; Fu, L.; Li, J. Quantum Spin Hall Effect in Two-Dimensional Transition Metal Dichalcogenides. *Science* **2014**, *346*, 1344−1347.

(13) Novoselov, K. S.; Mishchenko, A.; Carvalho, A.; Castro Neto, A. H. 2D Materials and Van der Waals Heterostructures. *Science* **2016**, *353*, 461.

(14) Haigh, S. J.; Gholinia, A.; Jalil, R.; Romani, S.; Britnell, L.; Elias, D. C.; Novoselov, K. S.; Ponomarenko, L. A.; Geim, A. K.; Gorbachev, R. Cross-Sectional Imaging of Individual Layers and Buried Interfaces of Graphene-Based Heterostructures and Superlattices. *Nat. Mater.* **2012**, *11*, 764−767.

(15) Bistritzer, R.; MacDonald, A. H. Moiré Bands in Twisted Double-Layer Graphene. *Proc. Natl. Acad. Sci. U.S.A.* **2011**, *108*, 12233.

(16) Cao, Y.; Fatemi, V.; Demir, A.; Fang, S.; Tomarken, S. L.; Luo, J. Y.; Sanchez-Yamagishi, J. D.; Watanabe, K.; Taniguchi, T.; Kaxiras, E.; et al. Correlated Insulator Behaviour at Half-Filling in Magic-Angle Graphene Superlattices. *Nature* **2018**, *556*, 80−84.

(17) Frisenda, R.; Navarro-Moratalla, E.; Gant, P.; Pérez De Lara, D.; Jarillo-Herrero, P.; Gorbachev, R. V.; Castellanos-Gomez, A. Recent Progress in the Assembly of Nanodevices and Van Der Waals Heterostructures by Deterministic Placement of 2D Materials. *Chem. Soc. Rev.* **2018**, *47*, 53−68.

(18) Hsu, W.-T.; Lin, B.-H.; Lu, L.-S.; Lee, M.-H.; Chu, M.-W.; Li, L.-J.; Yao, W.; Chang, W.-H.; Shih, C.-K. Tailoring Excitonic States of Van Der Waals Bilayers through Stacking Configuration, Band Alignment, and Valley Spin. *Sci. Adv.* **2019**, *5*, No. eaax7407.

(19) Kumar, R.; Verzhbitskiy, I.; Giustiniano, F.; Sidiropoulos, T. P. H.; Oulton, R. F.; Eda, G. Interlayer Screening Effects in WS_2/WSe_2 Van Der Waals Hetero-Bilayer. *2D Mater.* **2018**, *5*, No. 041003.

(20) Rivera, P.; Yu, H.; Seyler, K. L.; Wilson, N. P.; Yao, W.; Xu, X. Interlayer Valley Excitons in Heterobilayers of Transition Metal Dichalcogenides. *Nat. Nanotechnol.* **2018**, *13*, 1004−1015.

(21) Tawfik, S. A.; Isayev, O.; Stampfl, C.; Shapter, J.; Winkler, D. A.; Ford, M. J. Efficient Prediction of Structural and Electronic Properties of Hybrid 2D Materials Using Complementary DFT and Machine Learning Approaches. *Adv. Theory Simul.* **2019**, *2*, No. 1800128.

(22) Isayev, O.; Oses, C.; Toher, C.; Gossett, E.; Curtarolo, S.; Tropsha, A. Universal Fragment Descriptors for Predicting Properties of Inorganic Crystals. *Nat. Commun.* **2017**, *8*, No. 15679.

(23) Choudhary, K.; Garrity, K. F.; Pilania, G.; Tavazza, F. Efficient Computational Design of 2D Van der Waals Heterostructures: Band-Alignment, Lattice-Mismatch, Web-app Generation and Machine-learning, 2020. arXiv:2004.03025. https://arxiv.org/abs/2004.03025.

(24) Dong, R.; Jacob, A.; Bourdais, S.; Sanvito, S. High-Throughput Bandstructure Simulations of Van Der Waals Hetero-Bilayers Formed by 1T and 2H Monolayers. *npj 2D Mater. Appl.* **2021**, *5*, No. 26.

(25) Anderson, R. L. Germanium-Gallium Arsenide Heterojunctions. *IBM J. Res. Dev.* **1960**, *4*, 283.

(26) Kohn, W.; Sham, L. J. Self-Consistent Equations Including Exchange and Correlation Effects. *Phys. Rev.* **1965**, *140*, A1133−A1138.

(27) Hybertsen, M. S.; Louie, S. G. First-Principles Theory of Quasiparticles: Calculation of Band-Gaps in Semiconductors and Insulators. *Phys. Rev. Lett.* **1985**, *55*, 1418−1421.

(28) Kresse, G.; Furthmuller, J. Efficient Iterative Schemes for Ab Initio Total-Energy Calculations Using a Plane-Wave Basis Set. *Phys. Rev. B* **1996**, *54*, 11169−11186.

(29) Perdew, J. P.; Burke, K.; Ernzerhof, M. Generalized Gradient Approximation Made Simple. *Phys. Rev. Lett.* **1996**, *77*, 3865−3868.

(30) Klimeš, J.; Bowler, D. R.; Michaelides, A. Chemical Accuracy for the Van der Waals Density Functional. *J. Phys.: Condens. Matter* **2010**, *22*, No. 022201.

(31) Tawfik, S. A.; Gould, T.; Stampfl, C.; Ford, M. J. Evaluation of Van der Waals Density Functionals for Layered Materials. *Phys. Rev. Mater.* **2018**, *2*, No. 034005.

(32) Makov, G.; Payne, M. C. Periodic Boundary Conditions in Ab Initio Calculations. *Phys. Rev. B* **1995**, *51*, 4014−4022.

(33) Neugebauer, J.; Scheffler, M. Adsorbate-Substrate and Adsorbate-Adsorbate Interactions of Na and K Adlayers on Al(111). *Phys. Rev. B* **1992**, *46*, 16067−16080.

(34) Ward, L.; Dunn, A.; Faghaninia, A.; Zimmermann, N. E. R.; Bajaj, S.; Wang, Q.; Montoya, J.; Chen, J.; Bystrom, K.; Dylla, M.; Chard, K.; Asta, M.; Persson, K. A.; Snyder, G. J.; Foster, I.; Jain, A. Matminer: An Open Source Toolkit for Materials Data Mining. *Comput. Mater. Sci.* **2018**, *152*, 60−69.

(35) James, G.; Witten, D.; Hastie, T.; Tibshirani, R. *An Introduction to Statistical Learning with Applications in R*, Springer, 2013; Vol. *112*.

(36) Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V. Scikit-Learn: Machine Learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825−2830.

(37) Braga-Neto, U. M.; Dougherty, E. R. Is Cross-Validation Valid for Small-Sample Microarray Classification? *Bioinformatics* **2004**, *20*, 374−380.

(38) Kim, J.-H. Estimating Classification Error Rate: Repeated Cross-Validation, Repeated Hold-out and Bootstrap. *Comput. Stat. Data Anal.* **2009**, *53*, 3735−3745.

(39) Krstajic, D.; Buturovic, L. J.; Leahy, D. E.; Thomas, S. Cross-Validation Pitfalls When Selecting and Assessing Regression and Classification Models. *J. Cheminf.* **2014**, *6*, 10.

(40) Chen, J.; He, X.; Sa, B.; Zhou, J.; Xu, C.; Wen, C.; Sun, Z. III−VI Van der Waals Heterostructures for Sustainable Energy Related Applications. *Nanoscale* **2019**, *11*, 6431−6444.

(41) Haldar, S.; Vovusha, H.; Yadav, M. K.; Eriksson, O.; Sanyal, B. A Systematic Study of Structural, Electronic and Optical Properties of Atomic Scale Defects in 2D Transition Metal Dichalcogenides MX_2 (M = Mo,W; X = S, Se, Te). *Phys. Rev. B* **2015**, *92*, No. 235408.

(42) Quan, J.; Linhart, L.; Lin, M.-L.; Lee, D.; Zhu, J.; Wang, C.-Y.; Hsu, W.-T.; Choi, J.; Embley, J.; Young, C.; Taniguchi, T.; Watanabe, K.; Shih, C.-K.; Lai, K.; MacDonald, A. H.; Tan, P.-H.; Libisch, F.; Li, X. Phonon Renormalization in Reconstructed MoS_2 Moiré Superlattices. *Nat. Mater.* **2021**, *20*, 1100−1105.

(43) Yoo, H.; Engelke, R.; Carr, S.; Fang, S.; Zhang, K.; Cazeaux, P.; Sung, S. H.; Hovden, R.; Tsen, A. W.; Taniguchi, T.; Watanabe, K.; Yi, G.-C.; Kim, M.; Luskin, M.; Tadmor, E. B.; Kaxiras, E.; Kim, P. Atomic and Electronic Reconstruction at the Van der Waals Interface in Twisted Bilayer Graphene. *Nat. Mater.* **2019**, *18*, 448−453.

(44) Förg, M.; Baimuratov, A. S.; Kruchinin, S. Y.; Vovk, I. A.; Scherzer, J.; Förste, J.; Funk, V.; Watanabe, K.; Taniguchi, T.; Högele, A. Moiré Excitons in $MoSe_2$-WSe_2 Heterobilayers and Heterotrilayers. *Nat. Commun.* **2021**, *12*, No. 1656.

(45) Singh, S. K.; Bejagam, K. K.; An, Y.; Deshmukh, S. A. Machine-Learning Based Stacked Ensemble Model for Accurate Analysis of Molecular Dynamics Simulations. *J. Phys. Chem. A* **2019**, *123*, 5190−5198.

(46) Anderson, R. L. Experiments on Ge-GaAs Heterojunctions. *Solid-State Electron.* **1962**, *5*, 341−351.

(47) Koda, D. S.; Bechstedt, F.; Marques, M.; Teles, L. K. Trends on Band Alignments: Validity of Anderson's Rule in SnS_2- and $SnSe_2$-Based Van der Waals Heterostructures. *Phys. Rev. B* **2018**, *97*, No. 165402.

(48) Xu, K.; Xu, Y.; Zhang, H.; Peng, B.; Shao, H.; Ni, G.; Li, J.; Yao, M.; Lu, H.; Zhu, H.; Soukoulis, C. M. The Role of Anderson's Rule in Determining Electronic, Optical and Transport Properties of Transition Metal Dichalcogenide Heterostructures. *Phys. Chem. Chem. Phys.* **2018**, *20*, 30351−30364.

(49) Ceballos, F.; Bellus, M. Z.; Chiu, H.-Y.; Zhao, H. Probing Charge Transfer Excitons in a $MoSe_2$−WS_2 Van Der Waals Heterostructure. *Nanoscale* **2015**, *7*, 17523−17528.

(50) He, J.; Kumar, N.; Bellus, M. Z.; Chiu, H.-Y.; He, D.; Wang, Y.; Zhao, H. Electron Transfer and Coupling in Graphene−Tungsten Disulfide Van der Waals Heterostructures. *Nat. Commun.* **2014**, *5*, No. 5622.

■ NOTE ADDED IN PROOF

The data sets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

https://doi.org/10.1021/acsami.2c04403
ACS Appl. Mater. Interfaces 2022, 14, 25907−25919