ML-Aided Computational Screening of 2D Materials for Photocatalytic Water Splitting

Yatong Wang,^{†,‡,§} Murat Cihan Sorkun,^{†,§} Geert Brocks,^{‡,¶} and Süleyman Er^{*,†}

†DIFFER – Dutch Institute for Fundamental Energy Research, De Zaale 20, Eindhoven 5612 AJ, the Netherlands.

[‡]Materials Simulation and Modeling, Department of Applied Physics, Eindhoven University of Technology, Eindhoven 5600 MB, the Netherlands.

 \P Computational Chemical Physics, Faculty of Science and Technology and MESA+

Institute for Nanotechnology, University of Twente, P.O. Box 217, 7500 AE Enschede, the

Nether lands.

§ These authors contributed equally to this work.

E-mail: s.er@differ.nl

Abstract

The exploration of two-dimensional (2D) materials with exceptional physical and chemical properties is essential for the advancement of solar water splitting technologies. However, the discovery of 2D materials is currently heavily reliant on fragmented studies with limited opportunities for fine-tuning the chemical composition and electronic features of compounds. Starting from the V2DB digital library as a resource of 2D materials, we set up and execute a funnel approach that incorporates multiple screening steps to uncover potential candidates for photocatalytic water splitting. The initial screening step is based upon machine learning (ML) predicted properties, and subsequent steps involve first-principles modeling of increasing complexity, going from DFT to hybrid-DFT to GW calculations. Ensuring that at each stage more complex calculations are only applied to the most promising candidates, our study introduces an effective screening methodology that may serve as a model for accelerating 2D materials discovery within a large chemical space. Our screening process yields a selection of 11 promising 2D photocatalysts.

TOC Graphic



In future energy systems, hydrogen (H₂) is expected to have a prominent position as an energy carrier. As such, it is imperative that we produce hydrogen in an efficient and environmentally sustainable manner. In recent years, solar water splitting as a means of producing H₂ has gained significant interest; however, the efficacy of the process is heavily constrained by the performance of the photocatalysts.^{1,2} Due to the maximal specific surface area, abundant active sites, short carrier migration distance, and appreciable sunlight absorption, 2D materials are regarded as promising candidates for serving as photocatalysts in the conversion of solar energy into valuable H₂ fuel.^{3,4} Motivated by these advantages, the pursuit of new 2D photocatalysts has garnered increasing attention in recent years.^{5–8} There is still a pressing need to find new 2D photocatalysts, given the stringent criteria they should meet, such as possessing a moderate band gap to maximize solar energy absorption and appropriate band edge positions to align with the water redox potentials.^{9,10}

The field of material exploration has evolved into a new era characterized by big data and powerful ML algorithms, in relation to the remarkable growth in computing power and available computational data.^{11,12} In recent years, several computational material databases for 2D materials have been developed,^{13,14} which allow the use of ML techniques to uncover intricate correlations within the material space and accurately predict material properties.^{15,16} With the rapid prediction power of ML models, large chemical space of 2D materials can be screened based on the predictions of key properties and more accurate computational methods such as density functional theory (DFT) can be employed to expeditiously verify these predictions and create new training data for the models to extend the screening boundaries.^{17–21}

In this study, we screen the Virtual 2D Materials Database (V2DB),²² which contains a total of 316,505 potentially stable 2D materials with ML-predicted properties, to identify potential 2D photocatalysts for solar water splitting. To achieve this objective, we design a funnel workflow that leverages the existing data within V2DB, evaluates the uncertainty of ML-predicted properties through a newly developed meta-model, and performs multi-step first-principles calculations for promising candidates. Figure 1 illustrates the three principal stages of our workflow, which are color-coded in blue, orange, and yellow, and each of these stages comprises multiple steps. During the initial stage, we employed a combination of filtering criteria based on ML-predicted electronic properties and manual screening procedures to down-select promising photocatalysts. Next, we conduct PBE-level DFT computations, which encompass structural optimizations, band gap calculations, phonon dispersion calculations, and *ab initio* molecular dynamics (AIMD) simulations to identify stable 2D semiconductor photocatalysts. Finally, we utilize high-level calculation methods, including HSE06 and G_0W_0 , to obtain accurate electronic properties and assess the photocatalytic water splitting potential of the top 2D semiconductor candidates. As a result of our screening approach, we identify 11 2D material candidates as promising photocatalysts candidates for deeper investigation but also presents an effective ML-assisted physics-based computational screening approach that could inspire future searches for functional 2D materials.

The ML-predicted properties within V2DB, such as electronic band gap and band edges, are derived from training on DFT data at the PBE-level. Compared to standard DFT methods, the GW method including many-body effects can give a more accurate description of electronic properties, band gaps and band edge positions in particular, which are key properties for assessing the applicability of materials for solar water splitting. By applying a regression study between PBE and G_0W_0 electronic properties of 2D materials,²² the associated predicted G_0W_0 band gaps and band edge positions were also collected in V2DB, which are used in the current work to screen potential semiconductor photocatalysts. The lightcolored parts of violin maps, shown in Figure 2 (a) and Supporting Information (SI) Figure S1, illustrate the distribution of V2DB materials ML-predicted G_0W_0 band gap ($E_g^{ML(GW)}$) values for different prototypes.

For overall photocatalytic water splitting, the band gap of photocatalysts should exceed the water splitting free energy of 1.23 eV, and at the same time be smaller than 3 eV to



Figure 1: The workflow for scrutinizing V2DB as a possible source of 2D semiconductor photocatalysts for solar water splitting. The blue part contains the first screening layer (based on ML-predicted G_0W_0 properties, from *Candidates-316,505* to *Candidates-13,338*) and the second screening layer (further manual selection, to *Candidates-81*). The orange part refers to computational screening, including the third layer (based on DFT-PBE level calculations, to *Candidates-25*) and the fourth layer (dynamical stability check, to *Candidates-11*). The yellow part represents the high-level (HSE06 and G_0W_0) electronic band structure calculations of the best 11 2D semiconductor candidates for solar water splitting.

efficiently harvest sunlight. Therefore, $1.23 \leq E_{\rm g}^{\rm ML(GW)} \leq 3$ eV is set as the first selection criterion in the current study, which is shown in Figure 2 (a) and Figure S1 as dashed red lines. In addition to band gap, photocatalysts for solar water splitting should also have appropriate band edges to straddle the water redox potentials. Specifically, the valence band maximum (VBM) should be lower than the oxidation potential of O₂/H₂O (-5.67 eV vs vacuum at pH=0), while the conduction band minimum (CBM) should be higher than the reduction potential of H⁺/H₂ (-4.44 eV vs vacuum at pH=0), and these constitute our second and third criteria, respectively. The distribution of ML-predicted G₀W₀ band edges of all V2DB materials is shown in Figure 2 (b) with grey dots, as well as in Figures S2 and S3 split up according to the different 2D material prototypes. The black lines in these figures show the applied cutoffs for the band edges, while the diagonal lines denote the cutoffs for the



Figure 2: Screening based on ML-predicted properties. (a) The violin maps showing the distributions of ML-predicted band gap values for 2D materials from the seven prototypes identified in V2DB. For each prototype, the light-color shaded parts refer to all available candidates in V2DB, while the dark-color shaded parts refer to candidates with small values of ML-predicted band gap uncertainty ($E_{\rm g}^{\rm unc(GW)} < 0.326 \text{ eV}$). The red-dashed lines denote the ML-predicted band gap selection criterion, $1.23 \leq E_{\rm g}^{\rm ML(GW)} \leq 3$ eV, applied in the current study. (b) The distribution of ML-predicted band edges in V2DB. The horizontal and vertical axes represent the VBM $(E_{\text{VBM}}^{\text{ML}(\text{GW})})$ and CBM $(E_{\text{CBM}}^{\text{ML}(\text{GW})})$ values, both relative to the vacuum level, respectively. The horizontal black line at -4.44 eV and the vertical black line at -5.67 eV represent the reduction potential of H^+/H_2 and the oxidation potential of O_2/H_2O at pH=0, respectively. Additionally, two diagonal white lines indicate band gaps of 1.23 and 3 eV. The grey dots represent all 2D materials from V2DB (*Candidates-316,505*). The light-colored dots represent candidates from seven selected prototypes, whereas the darkcolored dots in the triangle represent structures after the first-stage screening (*Candidates*-81). (c) The rose map and structural view of the seven prototypes. The light-colored regions of the rose map depict all structures belonging to the selected prototypes in V2DB, whereas the dark-colored areas and numbers correspond to *Candidates-81*.

band gap screening. Clearly, 2D materials located inside the triangle marked by these lines are ideal photocatalysts, which meet both the band gap and the band edge requirements.

It's important to note that the ML-predicted materials within V2DB are not evenly distributed across its chemical space. Consequently, predictions for different material types exhibit varying levels of uncertainty.^{23,24} While the V2DB study outlines error metrics for the ML models in general, it does not provide uncertainties for property predictions on an individual material basis. Therefore, in the present work, we further developed a metamodel to predict the uncertainty associated with previously predicted G₀W₀ band gap values, allowing us to evaluate the reliability of each individual prediction. The meta-model is based on a consensus of three different algorithms, namely Artificial Neural Network (ANN), Random Forest (RF), and eXtreme Gradient Boosting (XGB). Further technical information is given in the Methods section and Table S1. Using the trained meta-model, we predict the G_0W_0 band gap uncertainty $(E_g^{unc(GW)})$ for all 2D materials in V2DB. To determine the materials with accurately predicted $E_{\rm g}^{\rm ML(GW)}$ values, $E_{\rm g}^{\rm unc(GW)} < 0.326$ eV is used as the fourth filter criterion here. The darker colored parts, shown in Figure 2 (a) and Figure S1, denote the ML-predicted G_0W_0 band gap distribution of V2DB materials that satisfy the fourth criterion. After these four criteria are executed, the total number of 2D materials is reduced from 316, 505 to 13, 338 (*Candidates-13,338*).

In the second layer of the funnel, we impose two screening steps, including prototype selection and structure filtering, to effectively narrow down the candidate space. Anticipating difficulties in experimental synthesis of 2D materials with a complex structure, here, the maximum number of atoms in a unit-cell is limited to three, in order to focus on relatively simple structures. Consequently, all prototypes with up to three atoms are preserved, which are BN, BiTeI, CdI_2 , GeS_2 , GeSe, MoS_2 , and MoSSe. The structures of these selected prototypes are shown in Figure 2 (c). Additionally, the scatter plots for the ML-predicted G_0W_0 band edge values of the V2DB materials from these prototypes are shown in Figure 2 (b) using light colored dots. Considering that all atoms in the structures of these seven

prototypes have more than one covalent bond with neighboring atoms, the existence of H atoms, with only one electron available for bonding, is not suited for the formation of stable 2D-bonded networks. Therefore, all structures that contain H atoms are also excluded in the current study. Subsequently, we remove any duplicate structures in the candidate list and the materials that are present in the ML training dataset. Additionally, 2D materials that have already been reported in literature are also excluded, as we aim here to find novel 2D photocatalysts. Table S2 shows examples of 2D structures that have been excluded during the structure filtering step.

As a result of the filtering with the aforementioned criteria, the first stage screening ended up with 81 potential 2D photocatalysts for solar water splitting (*Candidates-81*), which successfully proceeded to the computational screening stage. The assembled number of candidates in *Candidates-81* from each prototype are shown in a rose map of Figure 2 (c). The ML-predicted G_0W_0 band edge distribution of these 81 candidates is illustrated in Figure 2 (b) in the form of dark colored dots. Besides the colored dots, there are a large number of grey dots located within the triangle shown in Figure 2 (b), which are as well candidate 2D materials for photocatalytic water splitting and awaiting to be explored in future studies. They all satisfy the band gap and band edge criteria applied in here but belong to the fifteen other prototypes in V2DB with more complex structure which are not considered in the current study.

The physics-based computational screening stage starts with structure relaxation of the materials in *Candidates-81*. Since in V2DB virtual 2D materials are generated without any geometrical optimization, it is essential to perform full structural optimization to find the most energy-favorable state of each candidate. The structure optimizations are carried out in multiple steps with increasing precision from low to high until the imposed convergence standards are reached for the total energy and forces exerting on each atom. The details of computational settings are provided in the Methods section. After the automated structural optimization, we manually check each structure to ensure they remain in reasonable 2D



Figure 3: Computational screening and the final selection of 2D materials for water photocatalysis. (a) The top and side views of the TiFCl monolayer structure. (b) The phonon dispersion spectra of TiFCl. (c) The changes in the total free energy of TiFCl during an 8 ps AIMD simulation at 300 K, while the inset shows the top and side views of the equilibrium structure. (d) The element-projected fat band structure of TiFCl with respect to vacuum level, as calculated using the HSE06 functional. The color bar indicates the donation proportion of Ti, and F plus Cl elements. Grey lines depict G_0W_0 band structure relative to the vacuum level according to the HSE06 band gap center energy, E_{BGC}^{HSE} (indicated by the horizontal black dashed line). (e) The band alignments of *Candidates-11* relative to the vacuum level. The colored rectangular columns represent the HSE06 band edges, with different colors indicating the corresponding prototype. The G_0W_0 band edge positions are shown by the outer horizontal grey lines. The red dashed lines represent the reduction potential of H⁺/H₂ and the oxidation potential of O_2/H_2O at pH=0.

geometries. The majority of the candidates keep a structure similar to their corresponding prototypes, although their lattice parameters and atomic positions undergo changes during structural optimization due to elemental differences in chemical compositions. For example, TiFCl monolayer preserves an H-phase sandwich configuration consistent with its prototype of MoS_2 as shown in Figure 3 (a). After determining the most energy-favorable geometry of the materials in *Candidates-81*, the electronic band gaps at PBE level, both with and without spin-orbit-coupling (SOC) effects, are calculated to identify semiconductors. Due to the well-known underestimation of GGA-PBE functional, we don't put a lower bound on the band gap, but keep all candidates with a non-zero band gap. As a result, 25 semiconductors (*Candidates-25*) are identified after the third screening layer (Figure 1). The corresponding PBE-based electronic density of states (DOS), as calculated with the spin-polarization and SOC, are shown in Figures S4 and S5, respectively. In addition, the numerical values for the calculated band gaps are included in Table S3.

It is necessary to assess the dynamical stability of candidate 2D materials in *Candidates*-25. We start with calculating the phonon dispersion curves in the full Brillouin zone (BZ) of *Candidates-25*, where a dynamically stable structure should have no imaginary frequencies. As an example, Figure 3 (b) shows the phonon dispersions of the TiFCl monolayer, where all dispersion curves of the six optical and three acoustic modes have positive frequencies only, demonstrating that TiFCl monolayer in its optimized structure is dynamically stable. However, one needs to stress that in first-principles phonon calculations of 2D materials, the flexural phonon branch, pertaining to out-of-plane motions, is sensitive to the computational parameters, and the existence of small imaginary values near the Γ point is a common issue.^{25,26} The inaccuracy in properly describing such long wave length out-of-plane motions is not relevant for the quantities discussed in this paper, however. Furthermore, 2D structures with small imaginary frequencies near Γ point can be stabilized by extrinsic effects, such as their interactions with substrates.²⁷ Therefore, here we only exclude structures that have large imaginary frequencies in the middle of the BZ, which high likely point to the structure

being unstable.²⁸ The phonon calculation results of the *Candidates-25* are collected in Figure S6, in which the phonon stable 2D structures are labeled with red frames. As can be observed in this figure, stable structures either have no imaginary frequencies, or small imaginary frequencies near $\Gamma (\leq 0.2 \text{ THz}, \text{ such as VCIO or SnSSe})$, whereas unstable structures have imaginary frequencies over a significant part of the Brillouin zone ($\geq 1 \text{ THz}$, such as GaTeI or ZnTe).

Subsequently, we further estimate the thermal stability of these structures by short AIMD simulations at T = 300 K. For example, Figure 3 (c) shows the total energy during AIMD simulation of the TiFCl monolayer, demonstrating that it merely oscillates around a fixed value. Moreover, there are no bond breaking or structural distortions in the final configuration. These results imply that the TiFCl monolayer should be relatively stable. Figure S7 includes a summary of all AIMD simulation results, revealing that two structures exhibiting noticeable distortions and one structure transforming completely from H- to T-phase.

As a final test to assess structural stability, we calculate the elastic constants for the remaining 11 structures, and the results are compiled in Table S4. By applying the criteria for stability on the elastic constants, corresponding to their structural symmetry, 29,30 the mechanical stability of these structures is confirmed. As a result, 11 stable 2D semiconductors are screened out as photocatalyst candidates (*Candidates-11*). The corresponding structure configuration and lattice information of *Candidates-11* are presented individually in Figures S8-S18 (a, b).

In order to further investigate the electronic band structures of the selected potential 2D photocatalysts, the hybrid density functional HSE06 with SOC correction is employed. The calculated HSE06 band gap $(E_{\rm g}^{\rm HSE})$ results of the *Candidates-11* materials are shown in Table 1. To visualize the contributions from each chemical element to the electronic states, we also calculate the projected band structures of the *Candidates-11*, which are shown in Figures S8-S18 (c). One might expect that semiconductor where the VBM and CBM have different characters, *i.e.*, have contributions from different atom layers are more susceptible

Table 1: The V2DB-ID, material formula, prototype information, band character, total magnetic moment (M_{total} ; μ_B per formula unit), calculated spin-polarized PBE band gap ($E_g^{PBE-spin}$), PBE band gap (E_g^{PBE}), HSE06 band gap (E_g^{HSE}), G_0W_0 band gap (E_g^{GW}), and G_0W_0 band edges with respect to the vacuum level of the best 2D material candidates, *Candidates-11*, for photocatalytic water splitting that have been identified in the current study. All the band gap and band edge values are shown in units of eV.

V2DB-ID	material	prototype	character	${\rm M}_{\rm total}$	$E_{\rm g}^{\rm PBE-spin}$	$E_{\rm g}^{\rm PBE}$	$E_{\rm g}^{\rm HSE}$	$E_{\rm g}^{\rm GW}$	$E_{\rm VBM}^{\rm GW}$	$E_{\rm CBM}^{\rm GW}$
129	VFCl	BiTeI	Indirect	3.0	1.00	0.97	4.04	4.24	-6.12	-1.88
145	VClO	BiTeI	Direct	2.0	0.34	0.33	2.44	3.37	-7.16	-3.79
246	MnClBr	BiTeI	Indirect	5.0	1.67	1.62	4.25	5.44	-7.93	-2.49
286	MnIBr	BiTeI	Indirect	5.0	1.24	1.06	3.13	3.95	-6.72	-2.77
1375	MnFF	CdI_2	Direct	5.0	2.37	2.36	4.85	6.92	-8.76	-1.84
1392	MnClI	CdI_2	Indirect	5.0	1.18	0.97	3.03	4.15	-6.57	-2.43
1628	SnSSe	CdI_2	Indirect	0.0	0.99	0.84	1.45	2.71	-7.33	-4.61
1684	MnClI	GeS_2	Indirect	5.0	1.41	1.33	3.38	5.12	-7.40	-2.28
1686	MnBrI	GeS_2	Indirect	5.0	0.74	0.64	3.59	4.95	-7.43	-2.48
1781	TiFCl	MoS_2	Direct	0.0	1.15	1.12	1.48	2.12	-5.82	-3.70
3126	WTeO	MoSSe	Indirect	0.0	0.70	0.53	0.75	1.41	-6.52	-5.11

to spatial separation of the photo-excited carriers, which is an important step in initiating the photocatalytic reactions.^{31,32} For instance, as shown in Figure S14 (c), the CBM of SnSSe has dominant contribution from the Sn atoms, while S and Se atoms contribute to most of the VBM. This indicates that the transition of electrons from VBM to CBM is accompanied by a spatial movement of the charges from the anions to the cations. Moreover, the character of band gap might affect different aspects of the photocatalytic efficiency as well. Direct band gap materials usually show a good solar energy absorption, while indirect band gap materials can have good performance in charge-carrier separation.^{33,34}

The many-body G_0W_0 approach³⁵ with SOC correction based on the PBE wave-function is applied to calculate the G_0W_0 band gap (E_g^{GW}) of *Candidates-11*, and the corresponding results are listed in Table 1. Compared to the HSE06 band gap values, G_0W_0 generally increase the band gaps of the candidate materials. It should be noted, however, that the G_0W_0 band gaps calculated here are for free-standing 2D layers. These calculated values

are likely to form an upper bound, as in an experimental situation the 2D layers will be adsorbed on the substrate, whose dielectric screening will reduce the band gap. In all cases PBE was used to generate the starting point for the G_0W_0 calculations, with the exception of VClO. In that particular case the PBE functional gives a qualitatively wrong description of the character of the bands around the band gap. This error is corrected by the PBE+U functional, see Figure S19, and we use this functional to generate the starting point for the G_0W_0 calculation on VClO. The G_0W_0 band gaps of the complete set of Candidates-11 materials exceed the water splitting free energy of 1.23 eV. In addition, the calculated values for the three 2D semiconductors, including SnSSe, TiFCl and WTeO, are smaller than 3 eV, which suggests that these compounds can efficiently utilize the visible light spectrum. Some candidates, such as MnFF and MnClBr, have relatively large band gap values at G₀W₀ level, which indicates that they are not perfectly suitable for photocatalytic reactions. However, it may still be possible that the photocatalytic performance of these large band gap materials can be improved, for instance through the application of mechanical strain or an electric field,³ or by embedding in heterostructures.³⁶ It may also be possible to optimize such a material for one of the two half reactions of water splitting (either oxygen evolution, or hydrogen evolution), for instance by adjusting the pH values of the solution appropriately.³³

To estimate the absolute values for the G_0W_0 band edges, we employ here the band gap center (BGC) model,^{3,37} which has proven to be useful in previous studies.^{31,38,39} As an example, Figure 3 (d) illustrates the G_0W_0 band structures of the TiFCl monolayer relative to the vacuum level obtained based on the HSE06 BGC energy. Table 1 and Figure 3 (e) give an overview of the calculated G_0W_0 band edge energy values for the *Candidates-11*. In Figure 3 (e), one can assess the water splitting performance of each candidate by comparing the VBM and CBM positions (grey lines) with the redox potentials of oxygen (O_2/H_2O) and hydrogen evolution (H^+/H_2) reactions at pH=0 (red dashed lines), respectively. Accordingly, the majority of *Candidates-11* have G_0W_0 band edges that straddle the redox potentials of water, which in principle make them promising candidates for overall photocatalytic water splitting. Nevertheless, SnSSe and WTeO still only meet the criterium for VBM position, which indicates that they will be suitable only for the oxygen evolution reaction. Interestingly, compared to HSE06, G_0W_0 moves the VBM of TiFCl to below the oxygen evolution potential, implying that this material has inherent capability for overall solar water splitting.

In summary, we employ an ML-aided physics-based computational funnel approach for the discovery of potential 2D solar water splitting photocatalysts from the V2DB database. The initial screening stage utilized ML-predicted electronic properties, including band gap and band edge positions, as well as a newly developed ML model to estimate the uncertainties in ML-predicted band gap values. We further narrowed down the list of candidates by selecting structures from prototypes with no more than three atoms per unit-cell, and excluding materials that have already been reported in the literature. The second stage of the funnel involves DFT calculations, beginning with structural optimization and band gap calculations at the PBE level to identify suitable semiconductors for photocatalytic water splitting. The stability of the remaining candidates is then checked by phonon dispersion and AIMD calculations. Following this two-stage screening process, we identify 11 dynamically stable 2D semiconductors as potential photocatalysts for solar water splitting.

We further study these candidates by performing high-level hybrid-DFT (HSE06) and G_0W_0 calculations. As a result, all 11 candidates are found to have G_0W_0 gap values that exceed the water splitting free energy of 1.23 eV, while three 2D semiconductors SnSSe, TiFCl and WTeO have G_0W_0 gaps that are smaller than 3 eV, which implies a potential effective use of visible sunlight for water photocatalysis. Furthermore, based on the HSE06 BGC energy, the G_0W_0 band edge positions are estimated for the best candidates. Nine out of 11 2D materials are found to have band edges that straddle the water redox potentials, whereas the remaining two materials are found to be suitable only for the oxygen evolution reaction. We conclude that several newly identified 2D photocatalysts in this work hold promise for future validation. Thus, this research could potentially inspire further exploration of novel 2D materials by effectively utilizing both physics-based calculations and ML models.

Methods

Uncertainty Labeling Of The Band Gap Predictions

To label our dataset and train the models, we collected the ML-predicted PBE band gap $(E_{\rm g}^{\rm ML(PBE)})$ values from V2DB and the DFT calculated PBE band gap $(E_{\rm g}^{\rm PBE})$ values from C2DB,¹⁴ a computational 2D materials database, which is the source of ML training data for V2DB. Next, based on the absolute difference between these two values, we quantified the prediction uncertainty of the ML-predicted PBE band gaps $(E_{\rm g}^{\rm unc(PBE)})$. We also transformed the band gap uncertainty from PBE to GW $(E_{\rm g}^{\rm unc(GW)})$ using Equation (1) below.

$$E_{\rm g}^{\rm unc(GW)} = 1.63 \times E_{\rm g}^{\rm unc(PBE)} = 1.63 \times |E_{\rm g}^{\rm ML(PBE)} - E_{\rm g}^{\rm PBE}|$$
 (1)

where 1.63 is the linear regression fitting slope between PBE and GW band gap from V2DB work.²² The set screening criterion for band gap uncertainty at GW level $E_{\rm g}^{\rm unc(GW)} < 0.326$ eV (corresponding to $E_{\rm g}^{\rm unc(GW)} < 0.200$ eV). We used ($E_{\rm g}^{\rm unc(GW)}$) values as our targets when training our meta-models for uncertainty prediction.

Machine Learning Algorithms

We employed a feature vector and three different ML algorithms with the optimized parameters as explained below. Our consensus ML model consists of a combination of these individual ML models, with the final predictions being obtained by calculating the arithmetic average of the predictions made by each of these models.

Feature Vector Configuration

To represent the materials in the latent space we constructed a feature vector by amalgamating three different vectors using the same methodology as described in the V2DB study.²² The first one is the prototype vector, which is a one-hot vector containing the prototype information of the materials. The second vector is the chemical composition vector, which comprises details about the chemical elements present in the materials, along with the composition of these elements. Lastly, the electronegativity vector contains float data values of electronegativity, calculated using the geometric mean of the electronegativity of positively and negatively charged atoms in the compound, and has a length of 2.

Artificial Neural Network (ANN) Model Configuration

We used scikit-learn⁴⁰ version 0.22.2.post1 to train the ANN models and performed parameter optimization using grid search with the configurations given below. To optimize the parameters, we employed 10-fold cross-validation. The parameters selected for the final ANN model that demonstrated the best performance are shown below in bold, while the remaining parameters were set to their default values.

- Activation function: (relu, tanh)
- Alpha: (0.025, 0.05, **0.1**, 0.2)
- Max iteration: (**100**, 200, 400)
- Hidden Layers: [(100),(200),(400),(800),(100,10),(200,10),(400,10),(800,10),(100,20),
 (200,20),(400,20),(800,20)]

Random Forest (RF) Model Configuration

We used scikit-learn⁴⁰ library version 0.22.2.post1 to train the RF models. We used the default values for the parameters of the given version except the number of estimators was set to 1000.

Extreme Gradient Boosting (XGB) Model Configuration

We used xgboost⁴¹ library version 1.0.2 to train the XGB models. Using grid search, we optimized the parameters of the XGB models with the configurations given below. To optimize the given parameters, we used 10-fold cross-validation. The selected parameters for the final XGB model that exhibited the best performance are indicated in bold below, whereas the default values were used for the remaining parameters.

- Max depth: (4, 6, 8, **10**)
- Learning rate: (0.2, **0.3**, 0.4)
- Min child weight: (1, 2)

- gamma: (**0**, 0.1, 0.2, 0.3, 0.4)
- Colsample bytree: (0.3, 0.5, **0.7**)

Consensus Model Configuration

The consensus model comprises ANN, RF, and XGB models. Predictions are generated by averaging the predictions from these three models. Each model is trained independently, with parameters optimized as described earlier. Cross-validation results for both individual models and the consensus model are provided in the SI.

DFT Calculations

All DFT calculations were performed with the Vienna *ab initio* Simulation Package (VASP),^{42,43} and the results were handled by the VASPKIT package.²⁹ We employed the frozen-core projector augmented wave (PAW) method and Perdew-Burke-Ernzerhof (PBE) functional within the generalized gradient approximation (GGA).^{44,45} To avoid interactions with the mirror image, a sufficiently large vacuum space of 15 Å was added perpendicular to the surface in the *c* direction for each 2D structure. The electronic wavefunction was expanded using a plane-wave basis with a kinetic energy cutoff of 500 eV. All the geometric structures were visualized using the VESTA package.⁴⁶

In the structure optimization calculations, the convergence criteria were set to be 10^{-5} eV in energy between the consecutive relaxation steps and 0.01 eV/Å in force remaining on atoms. The 2D BZ was sampled using a $12 \times 12 \times 1$ Γ -centered k-point mesh. For the density of state (DOS) calculations, the effects of magnetism and spin-orbital coupling (SOC) corrections were taken into account, and denser k-point grids of $21 \times 21 \times 1$ were used.

The phonon dispersion curves were calculated using density functional perturbation theory (DFPT) as implemented in the PHONOPY code.⁴⁷ Supercells of $5\times5\times1$ unit cells were employed for this purpose. The AIMD simulations were performed employing the canonical ensemble (*NVT*) at 300 K for 8 ps, at a time step of 1 fs, and using supercells of $4\times4\times1$. The temperature was controlled using the Nosé-Hoover method. Additionally, for magnetic materials, the phonon and AIMD calculations involved the spin-polarization effects. The stress-strain method implemented in VASPKIT is employed to investigate the elastic constants.

HSE06 Calculations

To achieve accurate band structure calculations, the Heyd–Scuseria–Ernzerhof (HSE06) hybrid density functional was employed, which incorporates 25% exact Hartree–Fock (HF) exchange.⁴⁸ In these calculations, the effects of SOC were included, while the band edge positions relative to the vacuum level were accurately computed using the electrostatic potential alignments.

G_0W_0 Calculations

The one-shot G_0W_0 calculations were performed, for which the quasi-particle energies were obtained using DFT-PBE wave functions.³⁵ For these calculations, $12 \times 12 \times 1 \ k$ -grids were applied, which is consistent with several recent high-throughput studies.^{20,49,50} The energy convergence and energy cutoff for the response function were set to 10^{-8} between consecutive electronic steps and 100 eV, respectively. To ensure accurate results, 160 empty bands were included and the correction effect of SOC was taken into account in each computation. The maximally localized Wannier functions were fitted to the quasi-particle band structures using the WANNIER90 package.⁵¹

According to the BGC model, 3,37 the band edges at the G_0W_0 level are calculated as follows

$$E_{\rm VBM}^{\rm GW} = E_{\rm BGC}^{\rm HSE} - \frac{1}{2} E_{\rm g}^{\rm GW}$$
⁽²⁾

$$E_{\rm CBM}^{\rm GW} = E_{\rm BGC}^{\rm HSE} + \frac{1}{2} E_{\rm g}^{\rm GW}$$
(3)

where E_{BGC}^{HSE} is the band gap center energy calculated using the HSE06 hybrid functional, and E_{g}^{GW} is the band gap calculated using the G₀W₀ approximation.

Data availability

The authors declare that the data supporting the findings of this study are available within the paper and its Supporting Information files.

Associated content

The Supporting Information (SI) is available for this paper. The ML-predicted properties and physics-based calculation results are collected in SI. Violin maps of ML-predicted G_0W_0 band gaps (Figure S1); cross-validation results of the uncertainty prediction model (Table S1); scatter plot of ML-predicted G_0W_0 band edges distribution of V2DB 2D structures (Figure S2, S3); examples of 2D structures excluded in the structure filtering step (Table S2). The calculated PBE DOS results with spin-polarization (Figure S4) and SOC effects (Figure S5), and their summary (Table S3). The phonon (Figure S6), AIMD (Figure S7) and mechanical constants (Table S4) calculation results. The results of final selection of 11 candidates are individually shown in Figure S7-S18, including structural views, calculated HSE06 and G_0W_0 band structures. The band structures of VCIO as calculated with (a) PBE; (b) PBE+U, (c) G_0W_0 ; (d) G_0W_0+U methods (Figure S19).

Author contributions statement

Y. W. performed all the physics-based calculations. M. C. S. developed the ML models for band gap uncertainty prediction. S. E. devised the project, and supervised it together with G. B. All authors discussed the results, reviewed and edited the manuscript.

Competing interests

The authors declare no competing interests.

Acknowledgement

We thank Dr. Engin Torun for fruitful discussion. DIFFER is part of the institutes organisation of NWO. We acknowledge funding from the initiative "Computational Sciences for Energy Research" of Shell and NWO (Grant No 15CSTT05). Y. W. acknowledges financial support from China Scholarship Council (Grant No 202006930008). This work was sponsored by NWO Exact and Natural Sciences for the use of supercomputer facilities.

References

- Chen, X.; Shen, S.; Guo, L.; Mao, S. S. Semiconductor-based photocatalytic hydrogen generation. *Chem. Rev.* 2010, *110*, 6503–6570.
- (2) Qu, Y.; Duan, X. Progress, challenge and perspective of heterogeneous photocatalysts. *Chem. Soc. Rev.* 2013, 42, 2568–2580.
- (3) 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.
- (4) Ganguly, P.; Harb, M.; Cao, Z.; Cavallo, L.; Breen, A.; Dervin, S.; Dionysiou, D. D.;
 Pillai, S. C. 2D nanomaterials for photocatalytic hydrogen production. ACS Energy Lett. 2019, 4, 1687–1709.
- (5) Wang, X.; Maeda, K.; Thomas, A.; Takanabe, K.; Xin, G.; Carlsson, J. M.; Domen, K.; Antonietti, M. A metal-free polymeric photocatalyst for hydrogen production from water under visible light. *Nat. Mater.* **2009**, *8*, 76–80.
- (6) Xiang, Q.; Yu, J.; Jaroniec, M. Synergetic effect of MoS₂ and graphene as cocatalysts for enhanced photocatalytic H₂ production activity of TiO₂ nanoparticles. J. Am. Chem. Soc. 2012, 134, 6575–6578.

- (7) Su, T.; Shao, Q.; Qin, Z.; Guo, Z.; Wu, Z. Role of interfaces in two-dimensional photocatalyst for water splitting. ACS Catal. 2018, 8, 2253–2276.
- (8) Wang, Y.; Brocks, G.; Er, S. Data-Driven Discovery of Intrinsic Direct-Gap 2D Materials as Potential Photocatalysts for Efficient Water Splitting. ACS Catal. 2024, 14, 1336–1350.
- (9) Li, Y.; Li, Y.-L.; Sa, B.; Ahuja, R. Review of two-dimensional materials for photocatalytic water splitting from a theoretical perspective. *Catal. Sci. Technol.* 2017, 7, 545–559.
- (10) Fu, C.-F.; Wu, X.; Yang, J. Material design for photocatalytic water splitting from a theoretical perspective. Adv. Mater. 2018, 30, 1802106.
- (11) Agrawal, A.; Choudhary, A. Perspective: Materials informatics and big data: Realization of the "fourth paradigm" of science in materials science. APL Mater. 2016, 4, 053208.
- (12) Chen, A.; Zhang, X.; Zhou, Z. Machine learning: Accelerating materials development for energy storage and conversion. *InfoMat* 2020, 2, 553–576.
- (13) Zhou, J.; Shen, L.; Costa, M. D.; Persson, K. A.; Ong, S. P.; Huck, P.; Lu, Y.; Ma, X.; Chen, Y.; Tang, H., et al. 2DMatPedia, an open computational database of two-dimensional materials from top-down and bottom-up approaches. *Sci. Data* 2019, 6, 1–10.
- (14) 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., et al. The Computational 2D Materials Database: high-throughput modeling and discovery of atomically thin crystals. 2D Mater. 2018, 5, 042002.

- (15) Moosavi, S. M.; Jablonka, K. M.; Smit, B. The role of machine learning in the understanding and design of materials. J. Am. Chem. Soc. 2020, 142, 20273–20287.
- (16) Butler, K. T.; Davies, D. W.; Cartwright, H.; Isayev, O.; Walsh, A. Machine learning for molecular and materials science. *Nature* **2018**, *559*, 547–555.
- (17) Tran, K.; Ulissi, Z. W. Active learning across intermetallics to guide discovery of electrocatalysts for CO₂ reduction and H₂ evolution. *Nat. Catal.* **2018**, *1*, 696–703.
- (18) Masood, H.; Toe, C. Y.; Teoh, W. Y.; Sethu, V.; Amal, R. Machine learning for accelerated discovery of solar photocatalysts. ACS Catal. 2019, 9, 11774–11787.
- (19) Jin, H.; Zhang, H.; Li, J.; Wang, T.; Wan, L.; Guo, H.; Wei, Y. Discovery of novel twodimensional photovoltaic materials accelerated by machine learning. J. Phys. Chem. Lett. 2020, 11, 3075–3081.
- (20) Kumar, R.; Singh, A. K. Chemical hardness-driven interpretable machine learning approach for rapid search of photocatalysts. *npj Comput. Mater.* **2021**, *7*, 1–13.
- (21) Yin, H.; Sun, Z.; Wang, Z.; Tang, D.; Pang, C. H.; Yu, X.; Barnard, A. S.; Zhao, H.; Yin, Z. The data-intensive scientific revolution occurring where two-dimensional materials meet machine learning. *Cell Reports Phys. Sci.* **2021**, *2*, 100482.
- (22) Sorkun, M. C.; Astruc, S.; Koelman, J. V. A.; Er, S. An artificial intelligence-aided virtual screening recipe for two-dimensional materials discovery. *npj Comput. Mater.* 2020, 6, 1–10.
- (23) Varivoda, D.; Dong, R.; Omee, S. S.; Hu, J. Materials property prediction with uncertainty quantification: A benchmark study. *Appl. Phys. Rev.* 2023, 10, 021409.
- (24) Tran, K.; Neiswanger, W.; Yoon, J.; Zhang, Q.; Xing, E.; Ulissi, Z. W. Methods for comparing uncertainty quantifications for material property predictions. *Mach. Learn. Sci. Technol.* **2020**, *1*, 025006.

- (25) Cahangirov, S.; Topsakal, M.; Aktürk, E.; Şahin, H.; Ciraci, S. Two- and One-Dimensional Honeycomb Structures of Silicon and Germanium. *Phys. Rev. Lett.* 2009, 102, 236804.
- (26) Wang, J.; Yip, S.; Phillpot, S. R.; Wolf, D. Crystal instabilities at finite strain. *Phys. Rev. Lett.* 1993, 71, 4182–4185.
- (27) Zólyomi, V.; Drummond, N.; Fal'Ko, V. Electrons and phonons in single layers of hexagonal indium chalcogenides from ab initio calculations. *Phys. Rev. B* 2014, *89*, 205416.
- (28) Mounet, N.; Gibertini, M.; Schwaller, P.; Campi, D.; Merkys, A.; Marrazzo, A.; Sohier, T.; Castelli, I. E.; Cepellotti, A.; Pizzi, G., et al. Two-dimensional materials from high-throughput computational exfoliation of experimentally known compounds. *Nat. Nanotechnol.* **2018**, *13*, 246–252.
- (29) Wang, V.; Xu, N.; Liu, J.-C.; Tang, G.; Geng, W.-T. VASPKIT: A user-friendly interface facilitating high-throughput computing and analysis using VASP code. *Comput. Phys. Commun.* **2021**, 108033.
- (30) Maździarz, M. Comment on 'The Computational 2D Materials Database: highthroughput modeling and discovery of atomically thin crystals'. 2D Mater. 2019, 6, 048001.
- (31) Zhao, P.; Liang, Y.; Ma, Y.; Huang, B.; Dai, Y. Janus Chromium Dichalcogenide Monolayers with Low Carrier Recombination for Photocatalytic Overall Water-Splitting under Infrared Light. J. Phys. Chem. C 2019, 123, 4186–4192.
- (32) Ju, L.; Bie, M.; Tang, X.; Shang, J.; Kou, L. Janus WSSe monolayer: an excellent photocatalyst for overall water splitting. ACS Appl. Mater. Interfaces 2020, 12, 29335– 29343.

- (33) Jin, H.; Zhang, H.; Li, J.; Wang, T.; Wan, L.; Guo, H.; Wei, Y. Data-driven systematic search of promising photocatalysts for water splitting under visible light. J. Phys. Chem. Lett. 2019, 10, 5211–5218.
- (34) Peng, R.; Ma, Y.; Huang, B.; Dai, Y. Two-dimensional Janus PtSSe for photocatalytic water splitting under the visible or infrared light. J. Mater. Chem. A 2019, 7, 603–610.
- (35) van Schilfgaarde, M.; Kotani, T.; Faleev, S. Quasiparticle self-consistent GW theory. Phys. Rev. Lett. 2006, 96, 226402.
- (36) Zhang, X.; Zhang, Z.; Wu, D.; Zhang, X.; Zhao, X.; Zhou, Z. Computational screening of 2D materials and rational design of heterojunctions for water splitting photocatalysts. *Small Methods* **2018**, *2*, 1700359.
- (37) Toroker, M. C.; Kanan, D. K.; Alidoust, N.; Isseroff, L. Y.; Liao, P.; Carter, E. A. First principles scheme to evaluate band edge positions in potential transition metal oxide photocatalysts and photoelectrodes. *Phys. Chem. Chem. Phys.* **2011**, *13*, 16644–16654.
- (38) Liang, Y.; Huang, S.; Soklaski, R.; Yang, L. Quasiparticle band-edge energy and band offsets of monolayer of molybdenum and tungsten chalcogenides. *Appl. Phys. Lett.* 2013, 103, 042106.
- (39) Zhuang, H. L.; Hennig, R. G. Computational search for single-layer transition-metal dichalcogenide photocatalysts. J. Phys. Chem. C 2013, 117, 20440–20445.
- (40) Pedregosa, F. et al. Scikit-learn: Machine Learning in Python. J. Mach. Learn. Res. 2011, 12, 2825–2830.
- (41) Chen, T.; Guestrin, C. XGBoost: A Scalable Tree Boosting System. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2016; pp 785–794.

- (42) Becke, A. D.; Edgecombe, K. E. A Simple Measure of Electron Localization in Atomic and Molecular Systems. J. Chem. Phys. 1990, 92, 5397–5403.
- (43) Kresse, G.; Furthmüller, J. Efficiency of ab-initio total energy calculations for metals and semiconductors using a plane-wave basis set. *Comput. Mater. Sci.* 1996, 6, 15–50.
- (44) Blöchl, P. E. Projector augmented-wave method. Phys. Rev. B 1994, 50, 17953.
- (45) Perdew, J. P.; Chevary, J. A.; Vosko, S. H.; Jackson, K. A.; Pederson, M. R.; Singh, D. J.; Fiolhais, C. Atoms, Molecules, Solids, and Surfaces: Applications of the Generalized Gradient Approximation for Exchange and Correlation. *Phys. Rev. B* 1992, 46, 6671–6687.
- (46) Momma, K.; Izumi, F. VESTA 3 for three-dimensional visualization of crystal, volumetric and morphology data. J. Appl. Crystallogr. 2011, 44, 1272–1276.
- (47) Togo, A.; Tanaka, I. First principles phonon calculations in materials science. Scr. Mater. 2015, 108, 1–5.
- (48) Perdew, J. P.; Ruzsinszky, A.; Csonka, G. I.; Vydrov, O. A.; Scuseria, G. E.; Constantin, L. A.; Zhou, X.; Burke, K. Restoring the density-gradient expansion for exchange in solids and surfaces. *Phys. Rev. Lett.* **2008**, *100*, 136406.
- (49) Zhao, P.; Ma, Y.; Lv, X.; Li, M.; Huang, B.; Dai, Y. Two-dimensional III2-VI3 materials: promising photocatalysts for overall water splitting under infrared light spectrum. *Nano Energy* 2018, 51, 533–538.
- (50) Yang, X.; Singh, D.; Xu, Z.; Wang, Z.; Ahuja, R. An emerging Janus MoSeTe material for potential applications in optoelectronic devices. J. Mater. Chem. C 2019, 7, 12312– 12320.
- (51) Pizzi, G. et al. Wannier90 as a community code: new features and applications. J. Physics: Condens. Matter 2020, 32, 165902.

ML-Aided Computational Screening of 2D Materials for Photocatalytic Water Splitting

Yatong Wang,^{1,2,‡} Murat Cihan Sorkun,^{1,‡} Geert Brocks,^{2,3} and Süleyman Er^{1,*}

¹DIFFER – Dutch Institute for Fundamental Energy Research, De Zaale 20, Eindhoven 5612 AJ, the Netherlands.

²Materials Simulation and Modeling, Department of Applied Physics, Eindhoven University of Technology, Eindhoven 5600 MB, the Netherlands.

³Computational Chemical Physics, Faculty of Science and Technology and MESA+ Institute for Nanotechnology, University of Twente, P.O. Box 217, 7500 AE Enschede, the Netherlands.

[‡]These authors contributed equally to this work.

*E-mail: s.er@differ.nl

Model	MAE	RMSE
ANN	0.114	0.199
RF	0.104	0.199
XGB	0.106	0.200
Consensus	0.103	0.190

Table S1. Cross-validation results of the uncertainty prediction model. Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE) 10-fold cross-validation results (eV) of the individual models and the consensus model.



Figure S1. Violin maps of ML-predicted G_0W_0 band gap ($E_g^{ML(GW)}$) distribution of the candidates from the V2DB prototypes which have not been included in the current study. For each prototype, the light color shaded part refers to all the available candidates in V2DB, while the dark color shaded part refers to the candidates that have ML-predicted G_0W_0 band gap uncertainty ($E_g^{unc(GW)}$) values smaller than 0.326 eV. The red horizontal dashed lines denote the limits for the band gap selection criteria that have been applied in the current study.



Figure S2. The distribution of ML-predicted G_0W_0 band edges of all 2D structures from the seven V2DB prototypes that have been considered in the current study. The color bars show the $E_g^{\text{unc}(GW)}$ values, while the grey dots within the scatter plots denote the compounds with $E_g^{\text{unc}(GW)} > 0.326 \text{ eV}$.



Figure S3. The distribution of ML-predicted G_0W_0 band edges of all 2D structures from the V2DB prototypes that have not been considered in the current study. The color bars show the $E_g^{unc(GW)}$ values, while the grey dots within the scatter plots denote the compounds with $E_g^{unc(GW)} > 0.326$ eV.

V2DB-ID	Material	Prototype	$E_{\rm g}^{\rm ML(GW)}$	E _{VBM} ^{ML(GW)}	E _{CBM} ^{ML(GW)}	$E_{g}^{unc(GW)}$
1402	MnII ^[b]	Cdl ₂	2.02	-5.85	-3.83	0.212
1627	SnSS ^[b]	Cdl ₂	2.95	-7.28	-4.33	0.212
1687	MnII ^[b]	GeS ₂	2.50	-6.00	-3.50	0.196
1940	HfSCI ^[a]	MoS ₂	2.27	-5.97	-3.70	0.130
3075	HfSCI ^[a]	MoSSe	2.41	-5.97	-3.56	0.163
3120	WSS ^{[b][c]}	MoSSe	2.59	-6.27	-3.68	0.277
3121	WSSe ^{[b][c]}	MoSSe	2.80	-6.01	-3.21	0.277
3228	GeTe ^{[b][c]}	GeSe	2.95	-5.78	-2.82	0.147

^[a]Repetitive structures in V2DB.

^[b]Exist in training dataset.

^[c]Reported in literature.

Table S2. Examples of 2D structures, which have been excluded in the manual filtering step, and their corresponding electronic properties as obtained from the V2DB database.

Discussion: Here, for example, HfSCI monolayer with V2DB-IDs of 1940 and 3075 have different 2D prototype structures of MoS_2 and MoSSe, respectively. However, since their actual 2D structures will share the same configurations in *H*-phase, we treat them as duplicates and keep one of them for further optimization and analysis in the current study.

As we aim to find novel 2D photocatalysts here, the structures that have been included in the V2DB training dataset (marked with [b]) and the structures that have been reported in the literature (marked with [c]) are also excluded.



Figure S4. The calculated spin-polarized DFT (PBE) electronic density of states (DOS) for the *Candidates-25*. The origin of the energy scale was set at the Fermi level and shown with horizontal dotted lines.



Figure S5. The calculated DFT (PBE) electronic density of states (DOS) also including the spin-orbitalcoupling (SOC) relativistic effects for the *Candidates-25*. The origin of the energy scale was set at the Fermi level and shown with horizontal dotted lines.

	Material	Prototype	Magnetic	PBE_spin	(eV)	PBE_soc (eV)	
V200-10	Material	Trototype	moment (μ _B)	E ^{total} g	$E_{\rm g}^{\rm up}$	E ^{dn} g	Eg
129	VFCI	BiTel	3.0	1.00	1.00	5.66	0.97
145	VCIO	BiTel	2.0	0.34	0.34	3.37	0.33
166	VBrF	BiTel	3.0	0.78	0.78	4.62	0.77
246	MnClBr	BiTel	5.0	1.67	3.40	3.05	1.62
286	MnlBr	BiTel	5.0	1.24	2.86	2.12	1.06
657	GaTel	BiTel	0.0	0.42	0.42	0.42	0.17
659	GalSe	BiTel	0.0	0.88	0.88	0.88	0.54
1062	SbOI	BiTel	0.0	1.62	1.62	1.62	1.07
1375	MnFF	Cdl ₂	5.0	2.37	2.37	6.23	2.36
1379	MnFBr	Cdl ₂	5.0	1.76	2.17	3.00	1.68
1381	MnFl	Cdl ₂	5.0	1.07	1.35	1.90	0.83
1392	MnCll	Cdl ₂	5.0	1.18	2.72	2.03	0.97
1628	SnSSe	Cdl ₂	0.0	0.99	0.99	0.99	0.84
1684	MnCll	GeS ₂	5.0	1.41	2.93	1.83	1.33
1686	MnBrl	GeS ₂	5.0	0.74	2.68	1.90	0.64
3187	MnSe	GeSe	5.0	0.07	0.94	2.20	0.02
3221	ZnTe	GeSe	0.0	1.22	1.22	1.22	1.16
1784	TiFCI	MoS ₂	0.0	1.15	1.15	1.15	1.12
1986	AlSel	MoSSe	0.0	0.87	0.87	0.87	0.50
2027	ScFSe	MoSSe	0.0	1.05	1.05	1.05	0.98
2076	TiNBr	MoSSe	0.0	0.67	0.67	0.67	0.59
2250	MnFBr	MoSSe	5.0	1.34	1.74	2.10	1.22
2301	MnIF	MoSSe	5.0	0.48	0.75	1.19	0.24
2303	MnICI	MoSSe	5.0	0.82	2.11	1.51	0.58
3126	WTeO	MoSSe	0.0	0.70	0.70	0.70	0.53

Table S3. A summary of the calculated DFT (PBE) results for the Candidates-25.



Figure S6. The calculated phonon dispersion spectra for the *Candidates-25*. The phonon-stable 2D semiconductors are labeled with a red frame.



Figure S7. The energy profiles for the phonon-stable 2D structures from AIMD simulations at 300 K. The thermal-dynamically stable 2D semiconductors, *Candidates-11*, are labeled with a red frame.

V2DB-ID	Material	Prototype	Space group	C11 (N m ⁻¹)	C12 (N m ⁻¹)	C22 (N m ⁻¹)	C66 (N m ⁻¹)	Mechanical stability criteria ^[1,2]	Stable?
129	VFCI	BiTel	P3m1	47.27	12.81	(=C11)	17.23 (=(C11-C12)/2)	C11>0 & C11> C12	Yes
145	VCIO	BiTel	P3m1	113.29	38.36	(=C11)	37.47 (=(C11-C12)/2)	C11>0 & C11> C12	Yes
246	MnClBr	BiTel	P3m1	35.58	13.31	(=C11)	11.13 (=(C11-C12)/2)	C11>0 & C11> C12	Yes
286	MnlBr	BiTel	P3m1	31.39	10.49	(=C11)	10.45 (=(C11-C12)/2)	C11>0 & C11> C12	Yes
1375	MnFF	Cdl ₂	P-3m1	46.90	15.23	(=C11)	15.83 (=(C11-C12)/2)	C11>0 & C11> C12	Yes
1392	MnCll	Cdl ₂	P3m1	33.55	11.95	(=C11)	10.80 (=(C11-C12)/2)	C11>0 & C11> C12	Yes
1628	SnSSe	Cdl ₂	P3m1	64.22	15.92	(=C11)	24.15 (=(C11-C12)/2)	C11>0 & C11> C12	Yes
1684	MnCll	GeS ₂	Pmm2	24.31	0.24	14.83	0.89	C11>0 & C66>0 & C11*C22>C12*C12	Yes
1686	MnBrl	GeS ₂	Pmm2	20.00	0.53	15.30	0.89	C11>0 & C66>0 & C11*C22>C12*C12	Yes
1781	TiFCI	MoS ₂	P3m1	100.15	19.81	(=C11)	40.17 (=(C11-C12)/2)	C11>0 & C11> C12	Yes
3126	WTeO	MoSSe	P3m1	158.80	25.30	(=C11)	61.75 (=(C11-C12)/2)	C11>0 & C11> C12	Yes

Table S4. A summary of the calculated elastic constants for the *Candidates-11*. The mechanical stability criteria of each structure are based on its crystal system^[1,2].





Figure S8. (a, b) The top and side views of VFCI monolayer; (c) HSE06-calculated fat band structure of each element within the 2D material; (d) G_0W_0 -calculated band structure. In (c) and (d), the origin of the energy scale was set at the Fermi level and shown with horizontal dashed lines.

VCIO



Figure S9. (a, b) The top and side views of VCIO monolayer; (c) HSE06-calculated fat band structure of each element within the 2D material; (d) G_0W_0 -calculated band structure. In (c) and (d), the origin of the energy scale was set at the Fermi level and shown with horizontal dashed lines.

MnCIBr



Figure S10. (a, b) The top and side views of MnClBr monolayer; (c) HSE06-calculated fat band structure of each element within the 2D material; (d) G_0W_0 -calculated band structure. In (c) and (d), the origin of the energy scale was set at the Fermi level and shown with horizontal dashed lines.

MnIBr



Figure S11. (a, b) The top and side views of MnIBr monolayer; (c) HSE06-calculated fat band structure of each element within the 2D material; (d) G_0W_0 -calculated band structure. In (c) and (d), the origin of the energy scale was set at the Fermi level and shown with horizontal dashed lines.





Figure S12. (a, b) The top and side views of MnFF monolayer; (c) HSE06-calculated fat band structure of each element within the 2D material; (d) G_0W_0 -calculated band structure. In (c) and (d), the origin of the energy scale was set at the Fermi level and shown with horizontal dashed lines.





Figure S13. (a, b) The top and side views of MnCII monolayer; (c) HSE06-calculated fat band structure of each element within the 2D material; (d) G_0W_0 -calculated band structure. In (c) and (d), the origin of the energy scale was set at the Fermi level and shown with horizontal dashed lines.

SnSSe



Figure S14. (a, b) The top and side views of SnSSe monolayer; (c) HSE06-calculated fat band structure of each element within the 2D material; (d) G_0W_0 -calculated band structure. In (c) and (d), the origin of the energy scale was set at the Fermi level and shown with horizontal dashed lines.





Figure S15. (a, b) The top and side views of MnCII monolayer; (c) HSE06-calculated fat band structure of each element within the 2D material; (d) G_0W_0 -calculated band structure. In (c) and (d), the origin of the energy scale was set at the Fermi level and shown with horizontal dashed lines.

MnBrl



Figure S16. (a, b) The top and side views of MnBrI monolayer; (c) HSE06-calculated fat band structure of each element within the 2D material; (d) G_0W_0 -calculated band structure. In (c) and (d), the origin of the energy scale was set at the Fermi level and shown with horizontal dashed lines.





Figure S17. (a, b) The top and side views of TiFCI monolayer; (c) HSE06-calculated fat band structure of each element within the 2D material; (d) G_0W_0 -calculated band structure. In (c) and (d), the origin of the energy scale was set at the Fermi level and shown with horizontal dashed lines.

WTeO



Figure S18. (a, b) The top and side views of WTeO monolayer; (c) HSE06-calculated fat band structure of each element within the 2D material; (d) G_0W_0 -calculated band structure. In (c) and (d), the origin of the energy scale was set at the Fermi level and shown with horizontal dashed lines.



Figure S19. The band structures of VCIO as calculated with (a) PBE; (b) PBE+U, (c) G_0W_0 ; (d) G_0W_0+U methods. For all calculations the spin-orbit-coupling (SOC) effects have been included. The G_0W_0 calculation is carried out using the converged PBE wavefunction.

Discussion: For a comparison of Hubbard correction effects on the VCIO compound, we performed PBE and G_0W_0 band structure calculations both with and without +U parameter, where for V the U = 3 value was employed.

Although, the PBE and G_0W_0 band structures without the +U effect, as shown in (a) and (c), are consistent, they are qualitatively different from the HSE06 band structure shown in Figure S9 (c).

Applying the Hubbard correction in PBE and G_0W_0 calculations for VCIO has changed the band structure configurations as shown in (b) and (d), respectively. Importantly, the band configuration of the top two valence bands has become consistent with HSE06 band structure. Therefore, we conclude that for the VCIO monolayer, inclusion of +U effect is necessary when calculating its G_0W_0 band structure based on the PBE wave-function.

References

[1] Wang, V.; Xu, N.; Liu, J.-C.; Tang, G.; Geng, W.-T. VASPKIT: A user-friendly interface facilitating high-throughput computing and analysis using VASP code. *Comput. Phys. Commun.* **2021**, 108033.

[2] Maździarz, M. Comment on 'The Computational 2D Materials Database: high-throughput modeling and discovery of atomically thin crystals'. 2D Mater. **2019**, 6, 048001.