nature communications

Article

https://doi.org/10.1038/s41467-024-53864-4

High-throughput screening and machine learning classification of van der Waals dielectrics for 2D nanoelectronics

Received: 16 May 2024

Accepted: 24 October 2024

Published online: 04 November 2024

Check for updates

Yuhui Li[1,2,4], Guolin Wan[1,2,4], Yongqian Zhu[1,2], Jingyu Yang[1,2], Yan-Fang Zhang[2], Jinbo Pan[1,2] & Shixuan Du[1,2,3]

Van der Waals (vdW) dielectrics are promising for enhancing the performance of nanoscale field-effect transistors (FETs) based on two-dimensional (2D) semiconductors due to their clean interfaces. Ideal vdW dielectrics for 2D FETs require high dielectric constants and proper band alignment with 2D semiconductors. However, high-quality dielectrics remain scarce. Here, we employed a topology-scale algorithm to screen vdW materials consisting of zero-dimensional (0D), one-dimensional (1D), and 2D motifs from Materials Project database. High-throughput first-principles calculations yielded bandgaps and dielectric properties of 189 0D, 81 1D and 252 2D vdW materials. Among which, 9 highly promising dielectric candidates are suitable for MoS_2-based FETs. Element prevalence analysis indicates that materials containing strongly electronegative anions and heavy cations are more likely to be promising dielectrics. Moreover, we developed a high-accuracy two-step machine learning (ML) classifier for screening dielectrics. Implementing active learning framework, we successfully identified 49 additional promising vdW dielectrics. This work provides a rich candidate list of vdW dielectrics along with a high-accuracy ML screening model, facilitating future development of 2D FETs.

Two-dimensional (2D) semiconductors, characterized by their atomically thin thickness and exceptional gate controllability, have attracted great attention for their potential applications in the next-generation nanoscale field-effect transistors (FETs)[1,2]. 2D transition metal dichalcogenides[3,4], 2D black phosphorus[5], and few-layer InSe[6], possessing high carrier mobility, have been demonstrated to be promising channel materials. Nonetheless, their optimal electrical performance is constrained by the gate dielectrics used. Integrating 2D semiconductors with conventional three-dimensional (3D) dielectrics (e.g., SiO_2, Al_2O_3, and HfO_2) introduces numerous surface dangling bonds that act as carrier scattering centers at the interface, thereby reducing carrier mobility in the channel[7,8]. Van der Waals (vdW) dielectrics, including 0D vdW dielectrics (composed of 0D molecules), 1D vdW dielectrics (composed of 1D atomic chains), and 2D vdW dielectrics (composed of 2D atomic layers) are free of dangling bonds, offering an alternative scheme for integration with 2D semiconductors[8,9]. Encapsulating 2D semiconductors with vdW dielectrics has been observed to result in heightened carrier mobility and improved switching stability compared to those encapsulated with conventional 3D dielectrics[10–12].

The performance of vdW dielectrics in 2D FETs greatly depends on several critical qualities such as dielectric constant, band gap, and band alignment with 2D semiconductors. A large band gap and proper band alignment with 2D semiconductors minimize the gate leakage current, while the high dielectric constant improves gate controllability over the channel. However, the repertoire of suitable vdW dielectrics remains limited to date. Only a handful of vdW dielectrics have been experimentally reported, such as 2D *h*-BN[10,11], 2D MoO_3[13], 2D

[1]Beijing National Laboratory for Condensed Matter Physics, Institute of Physics, Chinese Academy of Sciences, Beijing 100190, China. [2]University of Chinese Academy of Sciences, Chinese Academy of Sciences, Beijing 100190, China. [3]Songshan Lake Materials Laboratory, Dongguan 523808, China. [4]These authors contributed equally: Yuhui Li, Guolin Wan. e-mail: jbpan@iphy.ac.cn; sxdu@iphy.ac.cn

Bi_2SeO_5[14,15], 2D SiP_2[16], 0D Sb_2O_3[12], and 1D vdW organic polymers[17]. Theoretically, a few 0D and 2D vdW dielectrics are predicted[18,19], while 1D inorganic vdW dielectrics are rarely reported. This scarcity, as well as the stringent quality requirements of vdW dielectrics, pose a significant challenge to the development of 2D FETs[8,20]. Consequently, a more abundant list of vdW dielectrics with diverse crystal structures is highly desirable.

In this work, using a topology-scaling algorithm[21], we screened potential vdW materials from the Materials Project (MP) database[22]. Our screening criteria include experimentally synthesized, bandgap exceeding 1.0 eV, atomic number below 50, absence of transition metals and inert elements, and a ratio of total atomic covalent volumes to unit-cell volume larger than 0.14 for 0D compounds. We then obtained bandgaps and dielectric constants (ε) along the vdW direction for 189 0D, 81 1D, and 252 2D vdW materials via high-throughput calculations. Among them, nine highly promising dielectrics are suitable for 2D MoS_2-based FETs, since their band offsets relative to MoS_2 are larger than 1 eV, minimizing the gate tunneling current. Statistical analysis reveals that materials with 1D and 2D motifs, along with strongly electronegative anions and heavy cations, tend to exhibit both large bandgap and ε values. We further developed a machine learning (ML) model using seven relevant feature descriptors to quantitatively screen promising dielectrics. This ML model comprises two sequential screening classifiers for band gap and dielectric constant, both achieving accuracies exceeding 80%. Implementing an active learning framework, we successfully identified an additional 49 promising vdW dielectrics. This work not only provides a rich list of vdW dielectric candidates but also establishes a high-accuracy ML model for efficient screening. These findings hold significant potential for advancing the development and application of 2D FETs.

Results

Database screening and high-throughput calculations

Figure 1 shows the integration of low-dimensional vdW dielectrics and 2D semiconductors. These configurations, featuring clean vdW contact interfaces, preserve the intrinsic electronic properties of 2D semiconductors, such as high carrier mobility. Recent experiments have reported the growth of mixed-dimensional heterostructures[23–25], demonstrating the feasibility of mixed-dimensional FET in practical applications. Furthermore, unlike conventional material integration, which relies on the formation of chemical bonds and requires precise lattice matching, vdW integration circumvents these constraints, significantly broadening the range of possible material combinations for 2D FETs[24]. While our primary focus is on the application of vdW dielectrics in 2D FETs, the vdW materials also show promise for 3D FETs, driven by recent advancements in vdW integration techniques that extend beyond traditional vdW materials[24].

Supplementary Fig. 1 outlines the workflow for screening low-dimensional vdW materials and high-throughput calculations. Starting with 126335 materials from the Materials Project database[22], we focused on potential vdW dielectrics with bandgap larger than 1 eV and ICSD (Inorganic Crystal Structure Database) ID number. Due to the computational complexity, transition metal compounds, along with materials whose atomic number is larger than 50 in unit cells, are excluded. Materials containing inert elements and elemental materials are also omitted due to their weak ionic polarization under an electric field. This initial screening yielded 5753 materials. Subsequently, we used a topology-scaling algorithm[21] to identify their cluster dimensionality. Among the identified low-dimensional materials, only those containing unitary clusters in unit cells are retained as vdW materials[26]. This process yields 452 0D vdW materials, 113 1D vdW materials, and 351 2D vdW materials. We further refined the selection of 0D vdW materials, excluding those with experimental structures measured under extreme conditions (e.g., CO_2, H_2S, and H_2O), as they are unsuitable for device integration. Therefore, we established a criterion, selecting 209 vdW 0D materials out of 452, based on a ratio of total atomic covalent volumes to unit-cell volume exceeding 0.14. Subsequently, we performed the high-throughput calculation, obtaining bandgaps and dielectric values for 189 0D, 81 1D, and 252 2D vdW materials, as shown in Fig. 1d. All calculated data are listed in Supplementary Tables 4 and 5. Dimensionality identification details and calculation methods can be referred to Methods section.

It's worth noting that several experimentally reported vdW dielectrics are also in our high-throughput calculations, such as 2D hBN[10,11], 2D Bi_2SeO_5[14,15], 2D SiP_2[16], and 0D Sb_2O_3[12]. To ensure the accuracy of our high-throughput calculations, we initially compared our computed data with experimental results for these experimentally reported vdW dielectrics, as shown in Supplementary Table 1. The absolute difference of ε between our calculated values and experimental values is within 2. Given that the employing generalized

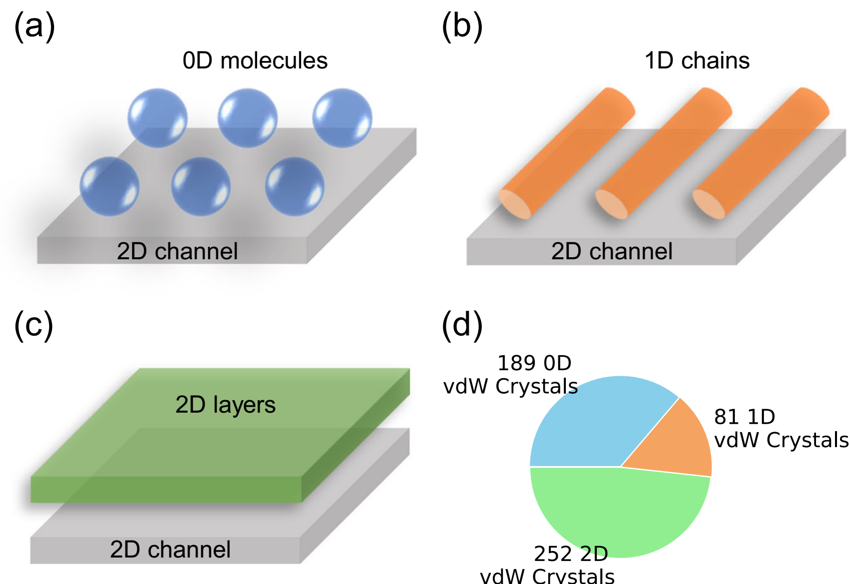

Fig. 1 | Schematic diagrams illustrating the integration of low-dimensional van der Waals (vdW) materials. a Dielectric layer composed of 0D molecules. **b** Dielectric layer composed of 1D chains. **c** Dielectric layer composed of 2D layers. **d** Statistical analysis of cluster dimensions for computationally convergent 522 vdW materials.

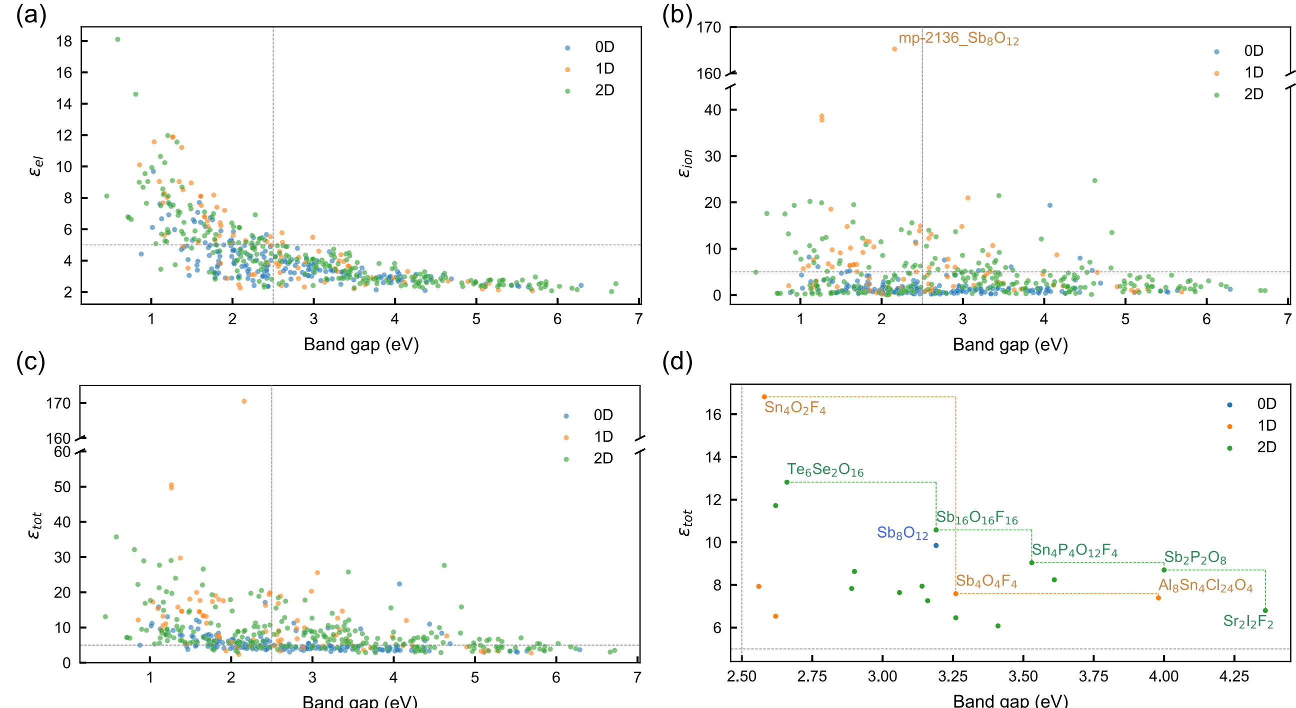

Fig. 2 | Dielectric constant vs. band gap plots. a Electronic contribution to dielectric constant ($\varepsilon_{\mathrm{el}}$) vs. band gap, **b** Ionic contribution to dielectric constant ($\varepsilon_{\mathrm{ion}}$) vs. band gap, **c** Total dielectric constant ($\varepsilon_{\mathrm{tot}}$) vs. band gap, and **d** $\varepsilon_{\mathrm{tot}}$ vs. band gap plot for 20 materials with band gap larger than 2.5 eV, $\varepsilon_{\mathrm{tot}}$ larger than 5, and exfoliation energy less than 30 meV/Å^2 after checking their suitability for field-effect transistor (FET) applications. Here, the blue, orange, and green points represent 0D, 1D, and 2D vdW materials, respectively. The horizontal and vertical gray dashed lines, respectively, mark the positions with a dielectric value of 5 and a bandgap value of 2.5 eV. The Pareto fronts for 1D and 2D candidates are plotted in **d** with orange and green dashed lines, respectively. The Pareto front of 0D materials is mp-1999_Sb_8O_{12}, which has been experimentally reported as a molecular crystal dielectric[12]. These materials on the Pareto fronts are labeled with their full chemical formula and color-coded to match their corresponding data points, whose numerical values are given in Table 1.

gradient approximation (GGA) function typically underestimates bandgap values, our calculated bandgaps are approximately 1 eV lower than the experimental values. However, bandgaps obtained under hybrid functionals or GW approximation closely align with experimental values. These results affirm the reliability of our calculated data. The calculated results are shown in Fig. 2. Since the dielectric response of materials to an external electric field involves both electronic polarization and ionic polarization, we presented the distributions of the electronic contribution to the dielectric constant ($\varepsilon_{\mathrm{el}}$) and ionic contribution to the dielectric constant ($\varepsilon_{\mathrm{ion}}$) with respect to band gap in Fig. 2a, b, respectively. It's worth noting that $\varepsilon_{\mathrm{el}}$ and $\varepsilon_{\mathrm{ion}}$ values discussed here are extracted from corresponding dielectric matrixes along the vdW direction (see Methods section). Notably, $\varepsilon_{\mathrm{el}}$ exhibits an inverse relationship with the band gap value, while $\varepsilon_{\mathrm{ion}}$ exhibits no obvious correlation with the band gap value, consistent with previous observations in 3D materials[27,28]. This observation is reasonable according to the calculation formulas for the dielectric matrix. The electronic contribution to the dielectric matrix can be expressed as[29],

$$\varepsilon_{el,\alpha\beta}^{real} = 1 + \frac{2}{\pi} P \int_0^\infty \frac{\varepsilon_{el,\alpha\beta}^{imag}(\omega')}{\omega'} d\omega' \quad (1)$$

where α and β $(=x, y, z)$ are Cartesian directions. $\varepsilon_{el,\alpha\beta}^{imag}(\omega)$ is the imaginary part of the frequency-dependent dielectric function,

$$\varepsilon_{el,\alpha\beta}^{imag}(\omega) = \frac{4\pi e^2}{\Omega} \lim_{q\to 0} \frac{1}{q^2} \sum_{c,v} 2 \int \delta(\epsilon_{c\mathbf{k}} - \epsilon_{v\mathbf{k}} - \omega) \times \left\langle u_{c\mathbf{k}+\mathbf{e}_\alpha q} | u_{v\mathbf{k}} \right\rangle \left\langle u_{v\mathbf{k}} | u_{c\mathbf{k}+\mathbf{e}_\beta q} \right\rangle d\mathbf{k} \quad (2)$$

$\epsilon_{c\mathbf{k}}$ and $\epsilon_{v\mathbf{k}}$ are the energies of conduction and valence band at the wavevector $\mathbf{k}$, respectively. $u_{c\mathbf{k}}$ and $u_{v\mathbf{k}}$ are the periodic parts of their corresponding wavefunctions. From Eq. 1, we knew that the electronic static dielectric matrix mainly depends on the values of $\varepsilon_{el,\alpha\beta}^{imag}(\omega)$ function in the low-frequency integrating region. $\varepsilon_{el,\alpha\beta}^{imag}(\omega)$ function is related to the electron transition from valence band to conduction band, with its first peak position determined by the minimum of direct band gap. Therefore, the electronic contribution to the dielectric matrix is approximately inversely proportional to bandgap value. The ionic contribution to dielectric matrix can be expressed as[30],

$$\varepsilon_{ion,\alpha\beta} = \frac{4\pi}{\Omega} \sum_m \frac{\left(\sum_{\kappa\alpha'} Z_{\kappa,\alpha\alpha'} U_m(\kappa\alpha')\right)\left(\sum_{\kappa\beta'} Z_{\kappa,\beta\beta'} U_m(\kappa\beta')\right)}{\omega_m^2} \quad (3)$$

where, $Z_{\kappa,\alpha\beta}$ is the Born effective charge tensor of atom κ, U_m and ω_m are the atomic displacement (eigenvector) and frequency of phonon mode m. $\sum_{\kappa\alpha'} Z_{\kappa,\alpha\alpha'} U_m(\kappa\alpha')$ is the total ion dipole moment along α direction under the vibration of phonon mode m. Since the denominator of Eq. 3 is the square of phonon vibration frequency, the ionic contribution to dielectric matrix primarily depends on the infrared (IR) active phonon vibration at low-frequency region, rather than the band gap.

As can be observed, the range of $\varepsilon_{\mathrm{el}}$ value is an order of magnitude smaller than that of $\varepsilon_{\mathrm{ion}}$ value for 522 vdW materials. Notably, the 1D vdW crystal Sb_8O_{12} (material ID in MP: mp-2136) stands out with the highest $\varepsilon_{\mathrm{ion}}$ value up to 165.24, alongside $\varepsilon_{\mathrm{el}}$ and bandgap values of 5.27 and 2.16 eV, respectively. In comparison, the experimentally reported

Table 1 | Details of 9 dielectric candidates on the Pareto fronts

Material ID in MP	Full Formula	Bandgap (eV)	ε_{el}	ε_{ion}	ε_{tot}	Space Group	E_{exf} (meV/Å^2)	D
mp-1999	Sb_8O_{12}	3.19	4.56	5.29	9.85	$P\bar{4}3m$	24.52	0
mp-27480	$Sn_4O_2F_4$	2.58	4.57	12.25	16.82	P2/m	17.59	1
mp-7609	$Sb_4O_4F_4$	3.26	3.50	4.09	7.59	$P2_1/m$	28.04	1
mp-560633	$Al_8Sn_4Cl_{24}O_4$	3.98	2.90	4.49	7.39	$P\bar{1}$	11.74	1
mp-29320	$Te_6Se_2O_{16}$	2.66	4.89	7.92	12.82	$P\bar{1}$	21.15	2
mp-561533	$Sb_{16}O_{16}F_{16}$	3.19	3.67	6.91	10.58	$P2_1/c$	29.90	2
mp-555838	$Sn_4P_4O_{12}F_4$	3.53	3.09	5.95	9.04	Pc	15.85	2
mp-3439	$Sb_2P_2O_8$	4.00	3.54	5.15	8.70	$P2_1/m$	27.96	2
mp-23046	$Sr_2I_2F_2$	4.36	3.29	3.51	6.80	P4/nmm	15.69	2

The table includes Material ID in the MP database, full formula, bandgap, ε_{el}, ε_{ion}, ε_{tot}, space group of clusters, exfoliation energy per area (E_{exf}), and cluster dimension (D).

0D vdW dielectric Sb_8O_{12}[12] (material ID in MP: mp-1999) has a ε_{ion} value of 5.29, a ε_{el} value of 4.56 and a bandgap value of 3.19 eV. We carefully compared the difference between mp-1999_Sb_8O_{12} and mp-2136_Sb_8O_{12} from several aspects, including lattice structure, electronic structure, bonding information, and IR active phonon vibration intensity, as shown in Supplementary Note 1. The analysis suggests that mp-2136_Sb_8O_{12} features weaker Sb-O bond strength within the cluster and denser cluster stacking, resulting in a smaller bandgap. Besides, given their similar orbital components around the Fermi level (Supplementary Fig. 3b, e), mp-2136_Sb_8O_{12} thus has larger ε_{el} according to Eq. 1. The IR intensity spectra in Supplementary Fig. 3c, f present that mp-2136_Sb_8O_{12} has much lower-frequency IR active phonon mode, explaining its significantly larger ε_{ion} value of mp-2136_Sb_8O_{12}, as per Eq. 3. Nevertheless, predicting the infrared (IR) intensity spectrum remains challenging due to its dependence on atomic structure. Generally, weaker bond strength and heavier atom mass are more likely to induce softer phonon mode, hence lower-frequency IR active phonon mode. The comparison between mp-1999_Sb_8O_{12} and mp-2136_Sb_8O_{12} suggests that materials with readily excited electron-hole pairs tend to exhibit larger electronic dielectric values. Furthermore, materials with easily excited IR active phonon modes are likely to have larger ionic dielectric values.

Candidates on the Pareto fronts

Figure 2c shows the distribution of the total dielectric constant (ε_{tot}, the sum of ε_{el} and ε_{ion}) with respect to the band gap for 522 vdW materials. In the pursuit of promising dielectrics with substantial capacitance and minimal tunneling current, we mainly focused on materials with ε_{tot} surpassing 5 and bandgap exceeding 2.5 eV (under GGA functional), resulting in a total of 124 materials. To address the demands of device miniaturization, we calculated the exfoliation energy for these 124 materials (see Methods section for details). Materials with lower exfoliation energy exhibit weaker interlayer binding, thereby facilitating the formation of stable few-layer structures. Consequently, we identified 81 materials with exfoliation energy below 30 meV/Å^2, comprising 11 0D, 16 1D, and 54 2D vdW materials. Given that some of these materials are volatile, deliquescent, or toxic and therefore unsuitable for 2D FET applications, we further investigated the physical and chemical properties of the 81 materials through literature and web searches (Supplementary Table 3), and finally remain 20 materials. We then employed Pareto front analysis, a method commonly used for multi-objective optimization, to identify desired materials with both a large dielectric value and a large bandgap[31]. Accordingly, we selected 9 dielectric candidates situated on the Pareto fronts across materials of different dimensions, as marked in Fig. 2d and listed in Table 1. Among them, 0D vdW material mp-1999_Sb_8O_{12} has been experimentally confirmed as a molecular crystal dielectric.

Subsequently, we investigated the band alignments between the 9 dielectric candidates and a common channel material 2D MoS_2. The candidates are cleaved along the vdW direction with the largest dielectric response. Minimum slab models of candidates are constructed to extract their band edge positions using HSE06 hybrid functional[32]. Their projected density of states (PDOS) are shown in Supplementary Fig. 5. The layered structures, composed of clusters with different dimensions, are shown in Fig. 3b, while their band alignments with MoS_2 monolayer are shown in Fig. 3a. The energy ranges within 1 eV above the conduction band minimum (CBM) and below the valence band maximum (VBM) of MoS_2 are shaded in light cyan. When 2D MoS_2 serves as the channel material, layered $Sb_4O_4F_4$ and $Al_8Sn_4Cl_{24}O_4$ dielectrics are well-suited for p-type field-effect transistors (FETs), due to the high tunnel barrier for hole carriers. On the other hand, layered Sb_8O_{12}, $Sn_4O_2F_4$, $Te_6Se_2O_{16}$, and $Sb_{16}O_{16}F_{16}$ dielectrics are suitable for n-type FETs, given the high tunnel barrier for electron carriers. Layered $Sn_4P_4O_{12}F_4$, $Sb_2P_2O_8$, and $Sr_2I_2F_2$ are suitable for both n-type and p-type FETs.

It's worth noting that, when accounting for environmental dielectric screening, the actual valence and conduction band offsets between 2D semiconductor channels and vdW dielectrics will be increased[33]. Therefore, some vdW dielectrics initially excluded due to their small band offsets with 2D semiconductor channels may also prove to be promising upon further investigation. Interestingly, several dielectric candidates are characterized by buckled geometric structure, such as $Sn_4O_2F_4$, $Sb_4O_4F_4$, $Sb_{16}O_{16}F_{16}$, and $Sn_4P_4O_{12}F_4$. While the vacuum between the 2D layer and the channel could potentially degrade the uniform dielectric environment, they can impart anisotropic conductance to isotropic semiconductor channels, as demonstrated in 2D SiP_2-gated MoS_2 FET[16], and enable unique device functionalities, such as anisotropic digital inverter[34], anisotropic memorizers[35], artificial synaptic devices[36], and so on.

Statistical analysis of dimension and element preference

To identify characteristics of promising dielectric candidates, we analyzed the distributions of dimension and element for 522 vdW materials. These vdW materials are categorized into four regions based on their dielectric and bandgap values. As shown in Fig. 4a, materials in region 1 have a dielectric value larger than 5 and bandgap less than 2.5 eV, those in region 2 have a dielectric value less than 5 and bandgap less than 2.5 eV, region 3 comprises materials with a dielectric value larger than 5 and bandgap larger than 2.5 eV, and region 4 consists of materials with a dielectric value less than 5 and bandgap larger than 2.5 eV. As shown in Fig. 4b, there are 192, 57, 124, and 149 vdW materials in the four regions, respectively. The prevalence of materials in regions 1 and 4 is attributed to the inverse relation between ε_{el} and bandgap (Fig. 2a). Materials in region 3 have a slightly larger ε_{ion}, enhancing their total dielectric value compared to materials in region 4. Therefore,

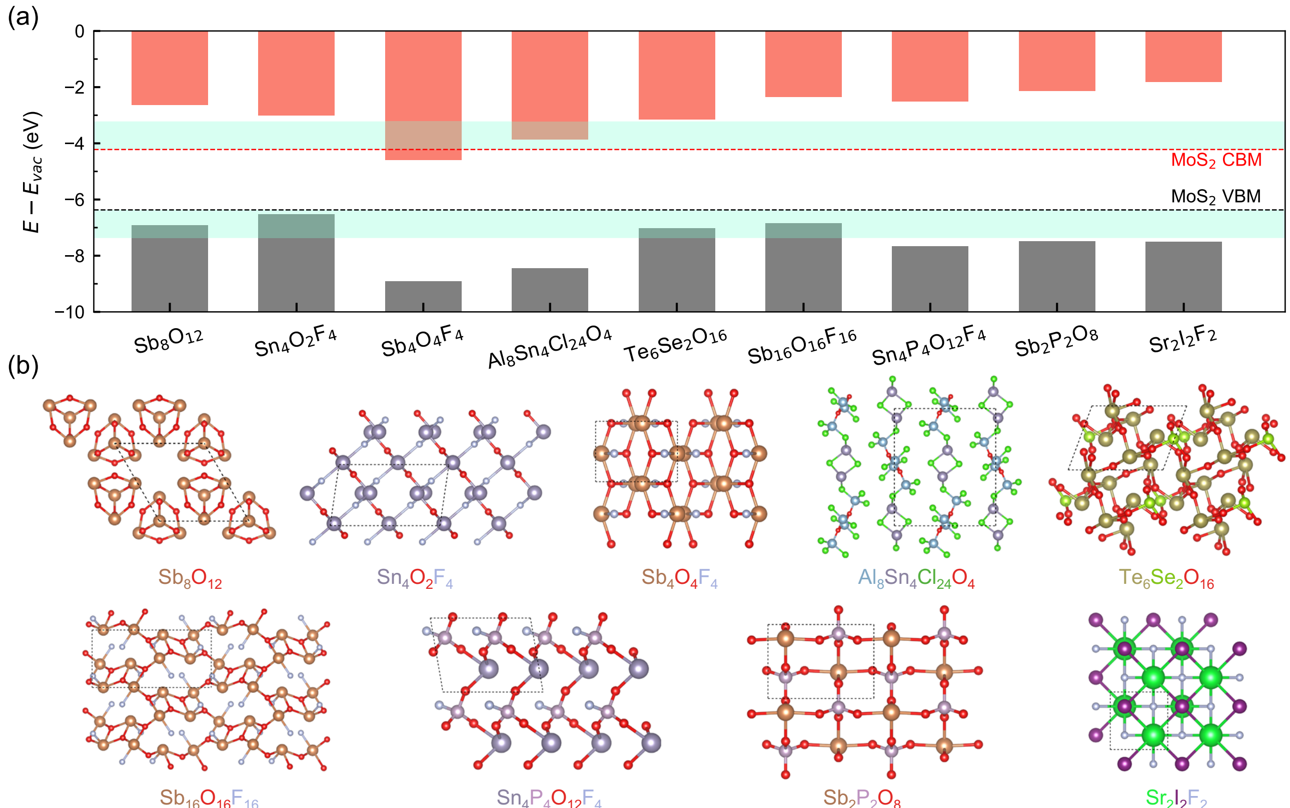

Fig. 3 | Band edges and atomic structures for the layered dielectric candidates on the Pareto fronts. a The band alignments between layered candidates and 2D MoS_2 calculated with HSE06 functional. Here, the gray and salmon bars show the valence band maximum (VBM) and conduction band minimum (CBM) of candidates, respectively. The black and red dashed lines show the VBM and CBM of MoS_2, respectively. The light cyan regions present the energy ranges of 1 eV above the MoS_2 CBM and beneath the MoS_2 VBM. **b** Top views of monolayer candidates. Sb_8O_{12} is composed of 0D clusters. $Sn_4O_2F_4$, $Sb_4O_4F_4$, and $Al_8Sn_4Cl_{24}O_4$ are composed of 1D clusters. $Te_6Se_2O_{16}$, $Sb_{16}O_{16}F_{16}$, $Sn_4P_4O_{12}F_4$, $Sb_2P_2O_8$, and $Sr_2I_2F_2$ are 2D clusters. Their side views are shown in Supplementary Fig. 4. The dashed lines mark the single unit cell of each material.

materials in region 3, where ε_{ion} is critical, align with our objective. Region 2 contains the fewest materials, characterized by both small ε_{el} and ε_{ion}, with bandgap values clustering around 2 eV (Fig. 2c). Figure 2a reveals that ε_{el} is not strictly inversely proportional to bandgap. There is a spread for ε_{el} value when given a bandgap, especially for a small bandgap. Some materials thus have small ε_{el}, even though their bandgap is small. The reason for this phenomenon is that the electron transition possibility depends not only on bandgap, but also the electronic structure and density of state around Fermi energy (Eq. 2)[37].

As shown in Fig. 4b, 0D vdW materials (blue pies) are less prevalent in region 3, while 1D (orange pies) and 2D vdW materials (green pies) are underrepresented in region 2. This suggests that 1D and 2D vdW materials are more likely to possess a large ε_{ion} value. Supplementary Fig. 6 confirms that only a small fraction of 0D vdW materials have ε_{ion} values larger than 5, whereas a substantial portion of 1D and 2D vdW materials do. To further analyze the distribution of anions and cations in each region, Supplementary Fig. 7 presents the prevalence of anions and cations in a histogram for each region, while Fig. 4d, e illustrates this distribution on the periodic table in a heatmap. Figure 4c displays the top five prevalent anions and cations in each region, excluding the hydrogen cation, due to its tendency to form hydroxyl anions when bonding with oxygen. By comparing the top five prevalent anions across different regions (gray bars in Fig. 4c), it is evident that materials with bandgap larger than 2.5 eV (regions 3 and 4) predominantly utilize oxygen, nitrogen, chlorine, fluorine, and bromine as anions, with most of these anions exhibiting strong electronegativity. Conversely, materials with bandgap less than 2.5 eV (regions 1 and 2) mainly utilize sulfur, oxygen, iodine, selenium, and bromine as anions, with most of these anions having weak electronegativity. As shown in Fig. 4d, elements with strong electronegativity (O, F, and Cl) are predominantly found in materials with large bandgaps (region 3 and 4, right sides of rectangles), whereas elements with weak electronegativity (S, Se, Te and I) mainly distribute in materials with small bandgaps (region 1 and 2, left sides of rectangles).

We further analyzed the distribution of cations across different regions. As shown in the salmon bars in Fig. 4c, materials with ε_{tot} larger than 5 (regions 1 and 3) predominantly take phosphorus, antimony, tin, bismuth, and lead as cations, with most of these cations having large atomic masses. In contrast, materials with ε_{tot} less than 5 (regions 2 and 4) predominantly take phosphorus, sulfur, silicon, carbon, and boron as cations, with most of these cations having small atomic mass. Figure 4e illustrates that elements in periods 4, 5, and 6 with larger atomic mass (e.g., As, Se, Sn, Sb, Te, Pb and Bi) are primarily found in materials with large ε_{tot} (regions 1 and 3, upper portion of rectangles), whereas elements in period 2 and 3 with small atomic mass (e.g., B, C, Si, P and S) are mainly distributed in materials with small ε_{tot} (regions 2 and 4, lower portion of rectangles). Notably, alkali metal and alkaline earth metal cations, characterized by very weak electronegativity, are primarily observed in materials with large bandgaps (regions 3 and 4, right sides of rectangles) in Fig. 4e.

In summary, our analysis suggests that materials with large bandgaps typically incorporate strongly electronegative elements (O, F, and Cl) as anions, or very weakly electronegative elements (alkali metals and alkaline earth metals) as cations. Conversely, materials with small bandgap typically incorporate elements with weak electronegativity (e.g., S, Se, Te, and I) as anions. Furthermore, materials with

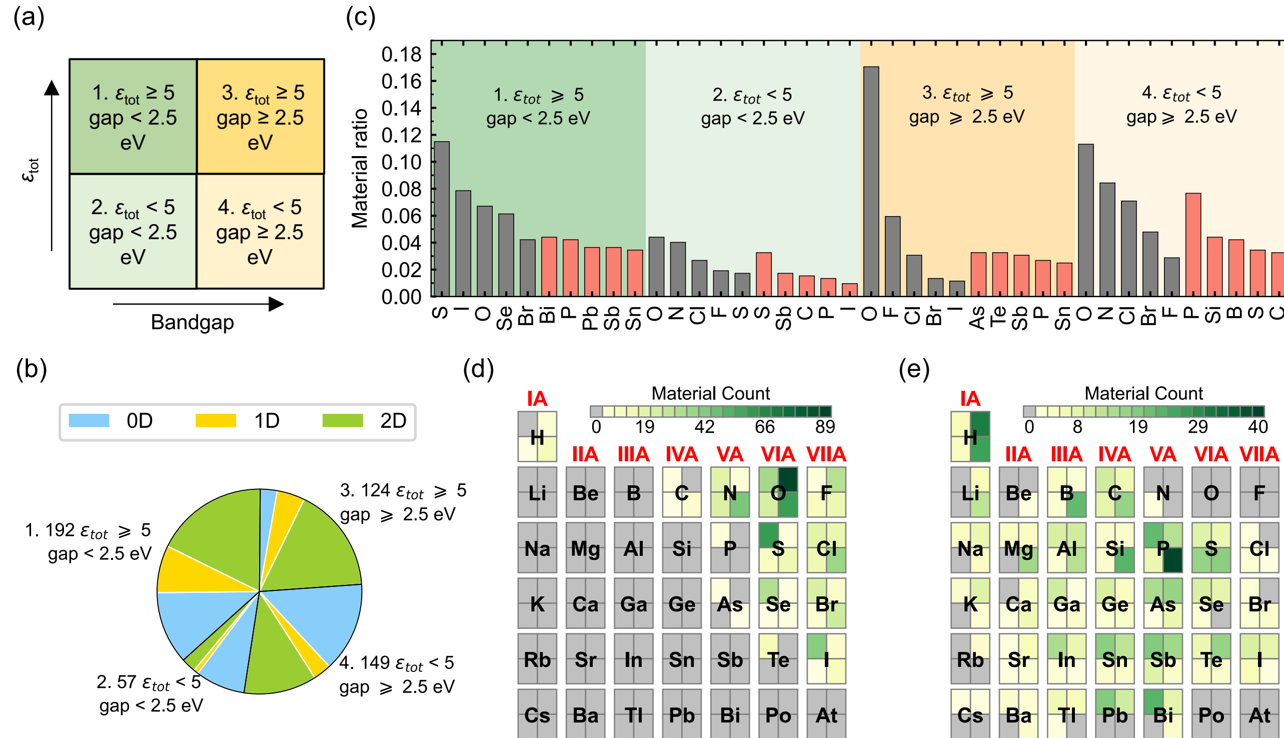

Fig. 4 | Statistical analysis of 522 calculated materials. a Division of materials into four regions according to their band gap and dielectric values in order to investigate the dimension and element preference in each region, as shown in **b**–**e**. **b** Distribution of material dimension in each region. The blue, yellow, and green pies represent the percentages of 0D, 1D, and 2D vdW materials in each region, respectively. **c** Statistic of top five prevalent anion and cation (excluding hydrogen cation) in each region, shown as a histogram. The gray and salmon bars represent the material ratio containing certain anion or cation, respectively. **d** Statistic of anion prevalence and **e** cation prevalence in each region, shown as a heatmap on a periodic table. Here, every element rectangle is divided into four parts, corresponding to four regions in **a**.

large ε_{tot} tend to incorporate heavier elements (e.g., As, Se, Sn, Sb, Te, Pb, and Bi) as cations, while those with small ε_{tot} generally involve light elements (e.g., B, C, Si, P, and S) as cations. These observations are understandable and reasonable.

Firstly, the bandgap value hinges on the position of the band edges. The positions of VBM and CBM are generally determined by the energy levels of anionic and cationic valence electrons, respectively, according to molecular orbital theory[38,39]. Introducing non-oxide anion with lower electronegativity into oxides has been utilized to narrow the bandgap for visible light applications[40,41]. Various semiempirical correlations between bandgap and element electronegativity have been proposed over the last century[42,43], including a relationship where the bandgap value of oxides is proportion to the square of the difference between average cationic and anionic electronegativity[43]. Secondly, materials with heavier cations often exhibit softer phonon mode, making them more likely to generate lower-frequency IR active phonon mode, thereby increasing the ε_{ion} and ε_{tot}. Furthermore, heavier cations typically have larger atomic radii, leading to increased atomic polarization and longer bond lengths. These factors are likely to contribute to higher ε_{el} and ε_{ion}, which have been extensively studied and applied in the design of polymer dielectrics[44–46]. Hence, in the pursuit of promising dielectrics with both large bandgaps and high dielectric constants, materials featuring strongly electronegative anions (e.g., O, F, and Cl) and heavy cations (e.g., As, Se, Sn, Sb, Te, Pb and Bi) should be prioritized.

In addition, it's noteworthy to consider why some materials with bandgap less than 2.5 eV (in regions 1 and 2) also contain anions with strong electronegativity (e.g., O anion). This can be attributed to several factors. Firstly, most of them have bandgap values close to 2.5 eV, especially those in region 2. Secondly, certain materials contain multiple anions (e.g., mp-23072_$Bi_2Br_2O_2$ in region 1, whose VBM is mainly contributed by Br anions, shown in Supplementary Fig. 8). Thirdly, additional factors beyond anionic electronegativity, such as cation type and crystal structure, also influence bandgap values. As for why some materials with ε_{tot} exceeding 5 (in regions 1 and 3) also contain light cations (e.g., P cation). Firstly, for materials with a gap less than 2.5 eV (in region 1), most of them have ε_{el} values larger than 5. Hence, the value of ε_{ion} is not crucial to make their ε_{tot} larger than 5 (Fig. 2a). However, those with a bandgap larger than 2.5 eV and ε_{el} less than 5 (in region 3), generally contain multiple cations besides light cations (e.g., mp-581609_$Bi_2P_2Pb_8O_{16}$ in region 3, whose cations include Bi, Pb, and P, shown in Supplementary Fig. 8). Lastly, other factors such as crystal structure and anion type, also influence ε_{ion} values beyond cationic atomic mass.

Two-step ML screening classifier

According to the above analysis, materials with strongly electronegative anions and heavy cations are more likely to be promising dielectrics. To efficiently screen such materials, we developed a quantitative machine learning (ML) model comprising two classifiers, as illustrated in Fig. 5a. The first classifier aims to identify materials with bandgap larger than 2.5 eV, while the second classifier is further applied to screen materials with ε_{tot} larger than 5.

The dataset for our ML models is derived from our high-throughput first-principles calculations. For the bandgap classifier, all 522 calculated data are utilized, including 249 materials with a bandgap less than 2.5 eV (negative class '0') and 273 materials with a bandgap larger than 2.5 eV (positive class '1'). For the ε_{tot} classifier, only the subset of 273 materials with a bandgap larger than 2.5 eV is used. This subset includes 149 materials with ε_{tot} less than 5 (negative class '0') and 124 materials with ε_{tot} larger than 5 (positive class '1').

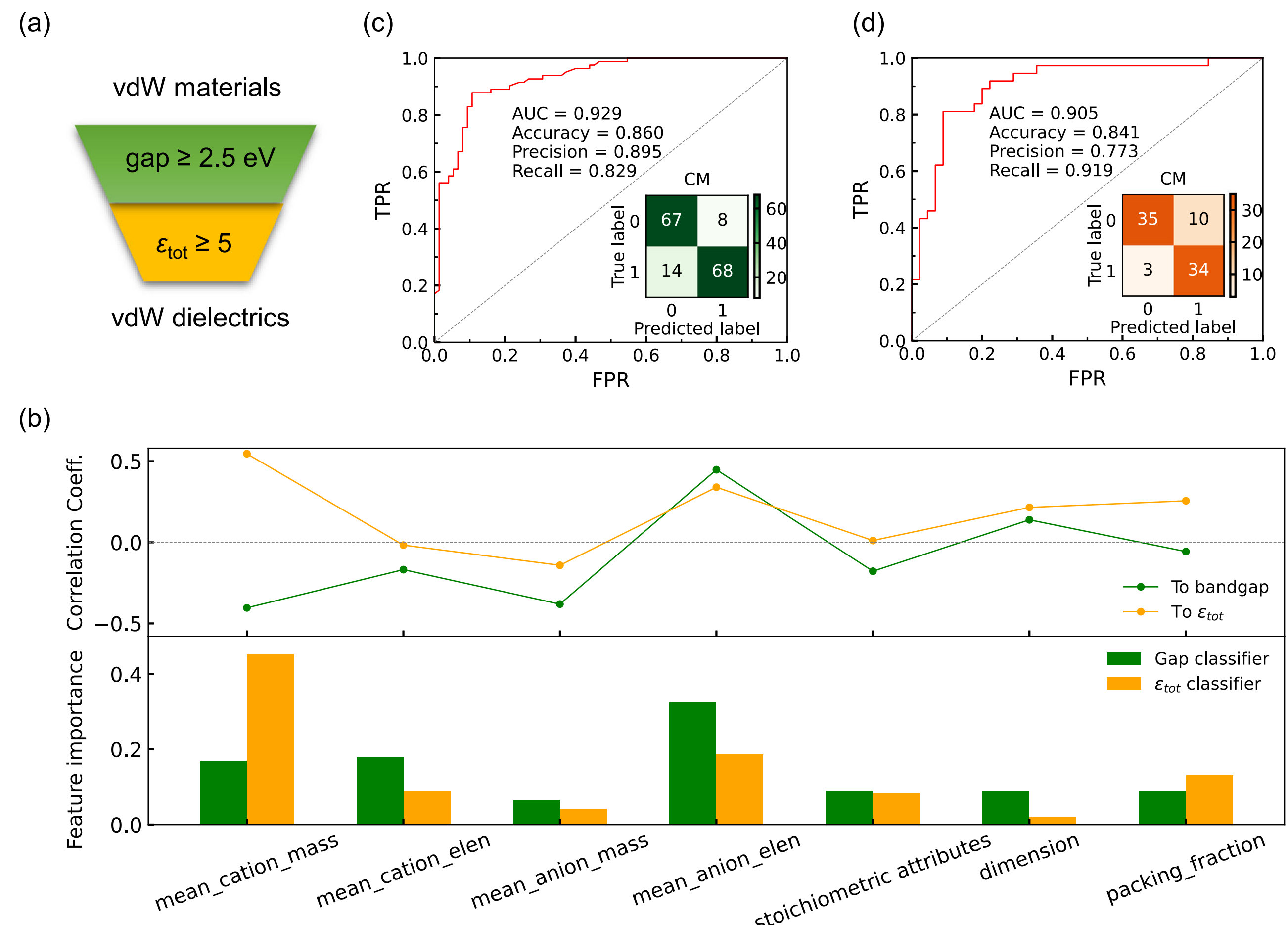

Fig. 5 | Two-step machine learning (ML) screening classifier. a Schematic workflow for screening materials with a band gap larger than 2.5 eV and ε_{tot} larger than 5. The first step involves using a bandgap classifier to screen materials with gap larger than 2.5 eV, the second step utilizes a ε_{tot} classifier to identify materials with ε_{tot} larger than 5. **b** Pearson correlation coefficients (upper panel) between the 7 features and band gap (green line), and between the 7 features and ε_{tot} (orange line). The feature importance of the two classifiers is depicted in the bottom panel, with green and orange bars representing the feature importance for gap classifier and ε_{tot} classifier, respectively. **c** Receiver operating characteristic (ROC) curve (red line) of gap classifier in the test set. The insert subplot is confusion matrix on the test set, where the labels '0' and '1' represent materials with a gap less than 2.5 eV and materials with gap larger than 2.5 eV, respectively. **d** ROC curve (red line) of ε_{tot} classifier in the test set. The insert subplot is confusion matrix on the test set, where the labels '0' and '1' represent materials with ε_{tot} less than 5 and materials with ε_{tot} larger than 5, respectively. The gray dashed lines in panels **c**, **d** represent the ROC curves of random classifiers. The model performance indicators, including the area under the ROC curve (AUC), accuracy, precision, and recall, are listed in panels **c**, **d**.

Next, we selected 7 features based on the previous statistics analysis, including 4 elemental attributes, 1 stoichiometric attribute, and 2 structural attributes. Elemental attributes contain the mean atomic mass and electronegativity of cations (Mean_mass_cations, Mean_elen_cations) and mean atomic mass and electronegativity of anions (Mean_mass_anions, Mean_elen_anions). Stoichiometric attribute is the L^2 norm of elemental fractions[47]. Structural attributes contain cluster dimensionality (Dimension) and packing fraction (Packing_fraction). By computing the Pearson correlation coefficients among 7 features, as shown in Supplementary Figs. 9 and 10, the absolute values of Pearson correlation coefficients among 7 features are below 0.65, indicating minimal feature redundancy[48]. Besides, the upper panel of Fig. 5b shows the Pearson correlation coefficients between these features and bandgap, as well as ε_{tot}. Several features, particularly elemental attributes, show correlation coefficients with absolute values exceeding 0.3, demonstrating clear relevance to bandgap and ε_{tot}. Hence, our selection of features is deemed reasonable. Moreover, material bandgap shows a positive correlation with anionic electronegativity and a negative correlation with anionic atomic mass, cationic atomic mass, and cationic electronegativity. ε_{tot} primarily exhibits a positive correlation with cationic atomic mass, anionic electronegativity, and packing fraction.

Following data collection and featurization, we then trained the classification models using the XGBoost algorithm, which is based on the gradient boosting tree method[49]. The selection of model parameters is summarized in the Methods section. Figure 5c, d, respectively, show the final model performance of the bandgap classifier and ε_{tot} classifier, presenting the receiver operating characteristic (ROC) curve, the area under ROC (AUC), accuracy, precision, recall, and confusion matrix. Both classifiers achieve accuracies exceeding 80%, indicating good prediction performance. We further analyzed the feature importance for these two classifiers, as shown in the bottom panel of Fig. 5b. For the bandgap classifier, the mean electronegativity of anions is the most important factor in determining bandgap value. As for the ε_{tot} classifier, the mean atomic mass of cations is the most crucial feature among the seven considered. These results are consistent with our statistics analysis, indicating that the classification models are reasonable.

To demonstrate the practicality of our two-step classification models, we further applied them to identify additional promising vdW dielectrics. Following the screening process for vdW materials

depicted in Supplementary Fig. 1, we first filtered out 199 vdW materials without ICSD ID numbers from the Materials Project database. Given the significant differences in data distributions between this 199-materials dataset and our original 522-materials dataset, as illustrated in Supplementary Fig. 11, we adopted active learning to mitigate potential model errors arising from data bias. Upon completion of active learning iterations, 72 materials are predicted to be promising vdW dielectrics. Among them, 49 materials are verified to possess a bandgap larger than 2.5 eV and a dielectric constant exceeding 5. Detailed information on all 60 promising vdW dielectrics is listed in Supplementary Table 7, while their ε_{tot} vs. bandgap plot is shown in Supplementary Fig. 12. We found that all dielectric candidates located on the Pareto fronts are successfully predicted, including 0D $Sb_8O_8F_8$ (material ID: mp-753868; $E_g = 3.57$ eV; $\varepsilon_{tot} = 9.33$), 1D $Bi_4O_4F_4$ (material ID: mp-753160; $E_g = 3.33$ eV; $\varepsilon_{tot} = 29.16$) and 2D $Bi_4O_2F_8$ (material ID: mp-753309; $E_g = 4.00$ eV; $\varepsilon_{tot} = 28.38$).

Besides, it's important to note that while our dataset is limited to vdW materials, the analysis of influencing factors on bandgap and ε_{tot} is also applicable to non-vdW materials. As calculated by Petousis et al.[50], most reported high-κ dielectrics feature both heavy cations and strongly electronegative anions, such as BiF_3, Tl_2SnF_6, and $AlTlF_4$. This observation indicates that our selected features can also be adapted to build ML classifiers for screening promising 3D non-vdW dielectrics.

Discussion

In summary, the bandgap and dielectric values of 189 0D vdW materials, 81 1D vdW materials, and 252 2D vdW materials are obtained via high-throughput first-principles calculations. They are all synthesized in experimental and screened out from the Materials Project database. Among them, 9 highly promising dielectric candidates are identified as suitable for MoS_2-based FET due to their large band offsets. Statistical analysis of the dimensionality and element prevalence among the 522 vdW materials reveal that materials containing strongly electronegative anions and heavy cations are more likely to be promising dielectrics, and 1D and 2D vdW materials are more likely to possess large dielectric constants. Upon these findings, we developed a two-step ML classifier to screen promising candidates. The model incorporates seven carefully selected feature descriptors and achieves accuracies exceeding 80%. The feature importance analysis indicates that the mean electronegativity of anions and the mean atomic mass of cations are most relevant to bandgap and dielectric values, respectively. Implementing an active learning framework, we successfully predicted 49 additional promising vdW dielectrics. Overall, our work provides a rich candidate list of vdW dielectrics and a high-accuracy ML classification model for screening promising dielectrics, which will facilitate the development and application of future 2D FETs.

Methods

Dimensionality identification

The process follows a topology-scaling algorithm[21], as illustrated in Supplementary Fig. 1. First, identify the number of covalently bonded clusters in the unit cell, denoted as n. Then, create a $2 \times 2 \times 2$ supercell and count the clusters, denoted as N. Calculate the ratio N/n: if $N/n = 8$, the material is composed of 0D clusters; if $N/n = 4$, it is composed of 1D clusters; if $N/n = 2$, it is composed of 2D clusters; and if $N/n = 1$, it is 3D. Non-integer ratios indicate that the material is composed of mix-dimensional clusters, which we ignored.

To identify covalently bonded clusters, we used the geometric criterion,

$$d_{ij} \leq k(r_i^{\mathrm{cov}} + r_j^{\mathrm{cov}}) \tag{4}$$

where d_{ij} is the distance between atom i and j, and r_i^{cov} and r_j^{cov} are their covalent radii. The scaling factor k affects cluster identification: larger k values identify larger clusters, indicating higher dimensions, while smaller k values identify smaller clusters, indicating lower dimensions. We tested multiple k values (1.05, 1.1, 1.15, 1.2, and 1.3). Results from these multiple k values are collected, and manually corrected in case of misidentifications.

Calculation details

First-principles calculations are performed using the Vienna Ab initio Simulation Package (VASP), based on Kohn-Sham density functional theory[51,52]. The projector augmented-wave (PAW) pseudopotentials[53] are used to describe the valence electron wavefunctions with a plane-wave cutoff of 520 eV. The Perdew-Burke-Ernzerhof generalized gradient approximation (PBE-GGA)[54] is applied for electron exchange-correlation energy. For vdW systems, the DFT-D3 method of Grimme[55] corrects the vdW interactions. The k-point sampling density is set to $0.02 \times 2\pi$ Å^{-1}. Considering that some vdW materials contain heavy elements such as Pb, Tl, and Bi, we also examined the effects of spin-orbit coupling (SOC) on key properties, including the bandgap (E_g), the electronic and ionic contributions to the dielectric constant (ε_{el} and ε_{ion}). The results, summarized in Supplementary Table 8, indicate that the inclusion of SOC typically reduces the bandgap while increasing ε_{el}, consistent with the inverse relationship between E_g and ε_{el}. No clear trend was observed for ε_{ion}. Although there are slight differences between the SOC and non-SOC results, the relative differences among vdW materials remain consistent.

Dielectric constant along the vdW direction

The dielectric matrix is determined using density functional perturbation theory (DFPT)[30]. From this matrix, the orientationally averaged dielectric constant along the vdW direction is extracted. For 0D vdW materials with three independent vdW directions, the spherically averaged dielectric constant is:

$$\begin{aligned}\varepsilon_{\mathrm{vdW}}^{\mathrm{0D}} &= \frac{\int_0^{\pi} d\theta \int_0^{2\pi} \sin\theta d\phi (\sin\theta\cos\phi\boldsymbol{\alpha} + \sin\theta\sin\phi\boldsymbol{\beta} + \cos\theta\boldsymbol{\gamma})^T \mathcal{E} (\sin\theta\cos\phi\boldsymbol{\alpha} + \sin\theta\sin\phi\boldsymbol{\beta} + \cos\theta\boldsymbol{\gamma})}{4\pi} \\ &= \frac{\lambda_1 + \lambda_2 + \lambda_3}{3}\end{aligned} \tag{5}$$

Here, $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, and $\boldsymbol{\gamma}$ are the three independent eigenvectors corresponding to the eigenvalues λ_1, λ_2 and λ_3 of dielectric matrix $\mathcal{E}$. For 1D vdW materials with two independent vdW directions ($\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$), the circularly averaged dielectric constant is:

$$\varepsilon_{\mathrm{vdW}}^{\mathrm{1D}} = \frac{\int_0^{2\pi} d\theta (\sin\theta\boldsymbol{\alpha} + \cos\theta\boldsymbol{\beta})^{\mathrm{T}} \mathcal{E} (\sin\theta\boldsymbol{\alpha} + \cos\theta\boldsymbol{\beta})}{2\pi} = \frac{\boldsymbol{\alpha}^{\mathrm{T}}\mathcal{E}\boldsymbol{\alpha} + \boldsymbol{\beta}^{\mathrm{T}}\mathcal{E}\boldsymbol{\beta}}{2} \tag{6}$$

In this case, the numerator is the trace of 2D matrix $\begin{bmatrix} \boldsymbol{\alpha}^{\mathrm{T}}\mathcal{E}\boldsymbol{\alpha} & \boldsymbol{\alpha}^{\mathrm{T}}\mathcal{E}\boldsymbol{\beta} \\ \boldsymbol{\beta}^{\mathrm{T}}\mathcal{E}\boldsymbol{\alpha} & \boldsymbol{\beta}^{\mathrm{T}}\mathcal{E}\boldsymbol{\beta} \end{bmatrix}$, which equals the sum of the eigenvalues ($\lambda_1 + \lambda_2$). For 2D vdW materials with one vdW direction ($\boldsymbol{\gamma}$),

$$\varepsilon_{\mathrm{vdW}}^{\mathrm{2D}} = \boldsymbol{\gamma}^{\mathrm{T}}\mathcal{E}\boldsymbol{\gamma} \tag{7}$$

Calculation of exfoliation energy

The exfoliation energy per unit area is defined as[56]

$$E_{exf} = \frac{E_{iso} - E_{bulk}/n}{A} \tag{8}$$

where E_{iso} is the energy of an isolate cluster in vacuum. E_{bulk} is the energy of bulk composed of n clusters, and A is half the surface area of

every cluster in bulk. The surface area A is calculated as[57]

$$A = \begin{cases} 2\pi\left(\frac{3V}{4\pi n}\right)^{2/3} & for\ 0\mathrm{D}\ materials \\ \left(\frac{\pi V h}{n}\right)^{1/2} & for\ 1\mathrm{D}\ materials \\ |\mathbf{a}\times\mathbf{b}| & for\ 2\mathrm{D}\ materials \end{cases} \quad (9)$$

Where V is the volume of bulk composed of n clusters, h is the period length along the 1D chain direction, and $\mathbf{a}$ and $\mathbf{b}$ are the lattice vectors in the 2D plane. Examples illustrating the calculation of cluster surface area are shown in Supplementary Fig. 2.

Machine learning model

The two classification models are implemented using XGBoost, a highly efficient, flexible, and portable distributed gradient boosting library[49]. The decision tree model is utilized as the base learner, and hyperparameter optimization is conducted using the evolutionary algorithm of Neural Network Intelligence (NNI) with 10-fold cross-validation. The dataset is divided into a 70% training set and a 30% test set. Data bias between the 522-material and 199-material datasets is observed using the Uniform Manifold Approximation and Projection (UMAP) algorithm[58]. To address the data bias, an active learning framework is employed, wherein the top 10 materials with predictive probabilities close to 0.5 are fed back into the training set. More details can be found in Supplementary Note 2.

Data availability

Relevant data supporting the key findings of this study are available within the article and the Supplementary Information files. All raw data generated during the current study are available from the corresponding authors upon request.

Code availability

The central code used in this paper is VASP. Detailed information related to the license and user guide are available at https://www.vasp.at. Other codes used in the study are available from the corresponding authors upon request.

References

1. Chhowalla, M., Jena, D. & Zhang, H. Two-dimensional semiconductors for transistors. *Nat. Rev. Mater.* **1**, 16052 (2016).
2. Liu, C. et al. Two-dimensional materials for next-generation computing technologies. *Nat. Nanotechnol.* **15**, 545–557 (2020).
3. Zhang, X. et al. Two-dimensional transition metal dichalcogenides for post-silicon electronics. *Natl. Sci. Open* **2**, 20230015 (2023).
4. Desai, S. B. et al. MoS_2 transistors with 1-nanometer gate lengths. *Science* **354**, 99–102 (2016).
5. Li, L. et al. Black phosphorus field-effect transistors. *Nat. Nanotechnol.* **9**, 372–377 (2014).
6. Bandurin, D. A. et al. High electron mobility, quantum hall effect and anomalous optical response in atomically thin InSe. *Nat. Nanotechnol.* **12**, 223–227 (2016).
7. Guo, Y. et al. Charge trapping at the MoS_2-SiO_2 interface and its effects on the characteristics of MoS_2 metal-oxide-semiconductor field effect transistors. *Appl. Phys. Lett.* **106**, 103109 (2015).
8. Illarionov, Y. Y. et al. Insulators for 2D nanoelectronics: the gap to bridge. *Nat. Commun.* **11**, 3385 (2020).
9. Liu, K., Liu, L. & Zhai, T. Emerging two-dimensional inorganic molecular crystals: the concept and beyond. *J. Phys. Chem. Lett.* **13**, 2173–2179 (2022).
10. Lee, G.-H. et al. Flexible and transparent MoS_2 field-effect transistors on hexagonal boron nitride-graphene heterostructures. *ACS Nano* **7**, 7931–7936 (2013).
11. Dean, C. R. et al. Boron nitride substrates for high-quality graphene electronics. *Nat. Nanotechnol.* **5**, 722–726 (2010).
12. Liu, K. et al. A wafer-scale van der Waals dielectric made from an inorganic molecular crystal film. *Nat. Electron.* **4**, 906–913 (2021).
13. Holler, B. A., Crowley, K., Berger, M. H. & Gao, X. P. A. 2D semiconductor transistors with van der Waals oxide MoO_3 as integrated high-κ gate dielectric. *Adv. Electron. Mater.* **6**, 2000635 (2020).
14. Li, T. et al. A native oxide high-κ gate dielectric for two-dimensional electronics. *Nat. Electron.* **3**, 473–478 (2020).
15. Zhang, C. et al. Single-crystalline van der Waals layered dielectric with high dielectric constant. *Nat. Mater.* **22**, 832–837 (2023).
16. Li, Z. et al. An anisotropic van der Waals dielectric for symmetry engineering in functionalized heterointerfaces. *Nat. Commun.* **14**, 5568 (2023).
17. Sharma, V. et al. Rational design of all organic polymer dielectrics. *Nat. Commun.* **5**, 4845 (2014).
18. Osanloo, M. R., Van de Put, M. L., Saadat, A. & Vandenberghe, W. G. Identification of two-dimensional layered dielectrics from first principles. *Nat. Commun.* **12**, 5051 (2021).
19. Bagheri, M. & Komsa, H. P. Screening 0D materials for 2D nanoelectronics applications. *Adv. Electron. Mater.* **9**, 2200393 (2022).
20. Knobloch, T. et al. The performance limits of hexagonal boron nitride as an insulator for scaled cmos devices based on two-dimensional materials. *Nat. Electron.* **4**, 98–108 (2021).
21. Ashton, M., Paul, J., Sinnott, S. B. & Hennig, R. G. Topology-scaling identification of layered solids and stable exfoliated 2D materials. *Phy. Rev. Lett.* **118**, 106101 (2017).
22. Jain, A. et al. Commentary: The materials project: a materials genome approach to accelerating materials innovation. *APL Mater* **1**, 011002 (2013).
23. Jariwala, D., Marks, T. J. & Hersam, M. C. Mixed-dimensional van der Waals heterostructures. *Nat. Mater.* **16**, 170–181 (2017).
24. Liu, Y., Huang, Y. & Duan, X. Van der Waals integration before and beyond two-dimensional materials. *Nature* **567**, 323–333 (2019).
25. You, J. et al. Epitaxial growth of 1D Te/2D $MoSe_2$ mixed-dimensional heterostructures for high-efficient self-powered photodetector. *Adv. Funct. Mater.* **34**, 2311134 (2023).
26. Deng, J., Pan, J. B., Zhang, Y. F. & Du, S. X. Database construction of two-dimensional charged building blocks for functional-oriented material design. *Nano Lett* **23**, 4634–4641 (2023).
27. Takahashi, A., Kumagai, Y., Miyamoto, J., Mochizuki, Y. & Oba, F. Machine learning models for predicting the dielectric constants of oxides based on high-throughput first-principles calculations. *Phys. Rev. Mater.* **4**, 103801 (2020).
28. Lee, M., Youn, Y., Yim, K. & Han, S. High-throughput ab initio calculations on dielectric constant and band gap of non-oxide dielectrics. *Sci. Rep.* **8**, 14794 (2018).
29. Gajdoš, M., Hummer, K., Kresse, G., Furthmüller, J. & Bechstedt, F. Linear optical properties in the projector-augmented wave methodology. *Phys. Rev. B* **73**, 045112 (2006).
30. Gonze, X. & Lee, C. Dynamical matrices, born effective charges, dielectric permittivity tensors, and interatomic force constants from density-functional perturbation theory. *Phys. Rev. B* **55**, 10355–10368 (1997).
31. Gopakumar, A., Pal, K. & Wolverton, C. Identification of high-dielectric constant compounds from statistical design. *NPJ Comput. Mater.* **8**, 146 (2022).
32. Heyd, J., Scuseria, G. E. & Ernzerhof, M. Hybrid functionals based on a screened coulomb potential. *J. Chem. Phys.* **118**, 8207–8215 (2003).
33. Ryou, J., Kim, Y.-S., Kc, S. & Cho, K. Monolayer MoS_2 bandgap modulation by dielectric environments and tunable bandgap transistors. *Sci. Rep.* **6**, 29184 (2016).
34. Liu, E. et al. Integrated digital inverters based on two-dimensional anisotropic ReS_2 field-effect transistors. *Nat. Commun.* **6**, 6991 (2015).

35. Wang, H. et al. Gate tunable giant anisotropic resistance in ultra-thin GaTe. *Nat. Commun.* **10**, 2302 (2019).
36. Tian, H. et al. Anisotropic black phosphorus synaptic device for neuromorphic applications. *Adv. Mater.* **28**, 4991–4997 (2016).
37. Naccarato, F. et al. Searching for materials with high refractive index and wide band gap: A first-principles high-throughput study. *Phys. Rev. Mater.* **3**, 044602 (2019).
38. Tao, S. et al. Absolute energy level positions in tin- and lead-based halide perovskites. *Nat. Commun.* **10**, 2560 (2019).
39. Zeier, W. G. et al. Thinking like a chemist: intuition in thermoelectric materials. *Angew. Chem. Int. Ed.* **55**, 6826–6841 (2016).
40. Kageyama, H. et al. Expanding frontiers in materials chemistry and physics with multiple anions. *Nat. Commun.* **9**, 772 (2018).
41. Harada, J. K., Charles, N., Poeppelmeier, K. R. & Rondinelli, J. M. Heteroanionic materials by design: Progress toward targeted properties. *Adv. Mater.* **31**, e1805295 (2019).
42. Vijh, A. K. Chemical approaches to the approximate prediction of band gaps of binary semiconductors and insulators. *J. Electrochem. Soc.* **117**, 173C (1970).
43. Di Quarto, F. et al. Semiempirical correlation between optical band gap values of oxides and the difference of electronegativity of the elements. Its importance for a quantitative use of photocurrent spectroscopy in corrosion studies. *J. Phys. Chem. B* **101**, 2519–2525 (1997).
44. Pilania, G. et al. New group IV chemical motifs for improved dielectric permittivity of polyethylene. *J. Chem. Inf. Model.* **53**, 879–886 (2013).
45. Wang, C. C., Pilania, G. & Ramprasad, R. Dielectric properties of carbon-, silicon-, and germanium-based polymers: a first-principles study. *Phys. Rev. B* **87**, 035103 (2013).
46. Chen, J. et al. Engineering the dielectric constants of polymers: From molecular to mesoscopic scales. *Adv. Mater.* https://doi.org/10.1002/adma.202308670 (2023).
47. Ward, L., Agrawal, A., Choudhary, A. & Wolverton, C. A general-purpose machine learning framework for predicting properties of inorganic materials. *NPJ Comput. Mater.* **2**, 16028 (2016).
48. Lu, S., Zhou, Q., Ma, L., Guo, Y. & Wang, J. Rapid discovery of ferroelectric photovoltaic perovskites and material descriptors via machine learning. *Small Methods* **3**, 1900360 (2019).
49. Chen, T. & Guestrin, C. In *Proc. 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 785–794 (Association for Computing Machinery, San Francisco, California, USA, 2016).
50. Petousis, I. et al. High-throughput screening of inorganic compounds for the discovery of novel dielectric and optical materials. *Sci. Data* **4**, 160134 (2017).
51. Kresse, G. & Furthmuller, J. Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set. *Phys. Rev. B* **54**, 11169–11186 (1996).
52. Kresse, G. & Furthmuller, J. Efficiency of ab-initio total energy calculations for metals and semiconductors using a plane-wave basis set. *Comp. Mater. Sci.* **6**, 15–50 (1996).
53. Blochl, P. E. Projector augmented-wave method. *Phys. Rev. B* **50**, 17953–17979 (1994).
54. Perdew, J. P., Burke, K. & Ernzerhof, M. Generalized gradient approximation made simple. *Phy. Rev. Lett.* **77**, 3865–3868 (1996).
55. Grimme, S. Semiempirical GGA-type density functional constructed with a long-range dispersion correction. *J. Comput. Chem.* **27**, 1787–1799 (2006).
56. Jung, J. H., Park, C.-H. & Ihm, J. A rigorous method of calculating exfoliation energies from first principles. *Nano Lett* **18**, 2759–2765 (2018).
57. Li, S. et al. Graph-based discovery and analysis of atomic-scale one-dimensional materials. *Natl. Sci. Rev.* **9**, nwac028 (2022).
58. McInnes, L., Healy, J. & Melville, J. Umap: Uniform manifold approximation and projection for dimension reduction. Preprint at https://arxiv.org/abs/1802.03426 (2018).

Acknowledgements

We thank Fang Han Lim for the helpful discussions. This work was supported by funds from the National Natural Science Foundation of China (52272172, J.P. and 61888102, S.D.), the Major Program of the National Natural Science Foundation of China (92163206, J.P.), the National Key Research and Development Program of China (2022YFA1204100, S.D. and 2021YFA1201501, J.P.). Computational resources were provided by the National Supercomputing Center in Tianjin.

Author contributions

S.D. and J.P. conceived and supervised the research project. Y.L. and G.W. contributed equally to this work. Y.L. performed the first-principles calculations and data analysis. G.W. built the machine learning models. Y.L. and G.W. co-wrote the initial manuscript. Y.Z., J.Y., and Y.-F.Z. participated in discussing the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41467-024-53864-4.

Correspondence and requests for materials should be addressed to Jinbo Pan or Shixuan Du.

Peer review information *Nature Communications* thanks William Vandenberghe and the other, anonymous, reviewer for their contribution to the peer review of this work. A peer review file is available.

Reprints and permissions information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.

© The Author(s) 2024