Research Article

ABC_2-type short-wave infrared photodetector materials discovered via high-throughput screening and machine learning

Xiaoning Guan[1,2,#], Yanan Zhang[1,2,#], Suwen Han[1,2], Chunlian Xiong[3], Yuqing Yang[1,2], Changcheng Chen[3], Fan Zhang[1,4], Yanchao Zhang[2,4], Huachuan Gao[1,4], Feng Zhou[1,2,*], Pengfei Guan[5,*], Pengfei Lu[1,2,*]

[1]School of Integrated Circuits, Beijing University of Posts and Telecommunications, Beijing 100876, China.
[2]State Key Laboratory of Information Photonics and Optical Communications, Beijing University of Posts and Telecommunications, Beijing 100876, China.
[3]School of Science, Xi'an University of Architecture and Technology, Xi'an 710055, Shaanxi, China.
[4]Ningbo Weiyuan Optoelectronic Research Institute Co., Ltd., Ningbo 315200, China.
[5]Advanced Interdisciplinary Science Research (AiR) Center, Ningbo Institute of Materials Technology and Engineering, Chinese Academy of Sciences, Ningbo 315201, Zhejiang, China.

[#]Authors contributed equally.

[*]**Correspondence to:** Feng Zhou, Pengfei Lu, School of Integrated Circuits, Beijing University of Posts and Telecommunications, Beijing 100876, China; State Key Laboratory of Information Photonics and Optical Communications, Beijing University of Posts and Telecommunications, Beijing 100876, China. E-mail: zfsimon@163.com; photon.bupt@gmail.com; Pengfei Guan, Advanced Interdisciplinary Science Research (AiR) Center, Ningbo Institute of Materials Technology and Engineering, Chinese Academy of Sciences, Ningbo 315201, Zhejiang, China. E-mail: pguan@nimte.ac.cn

ORCID: Pengfei Guan (0000-0002-7679-6768)

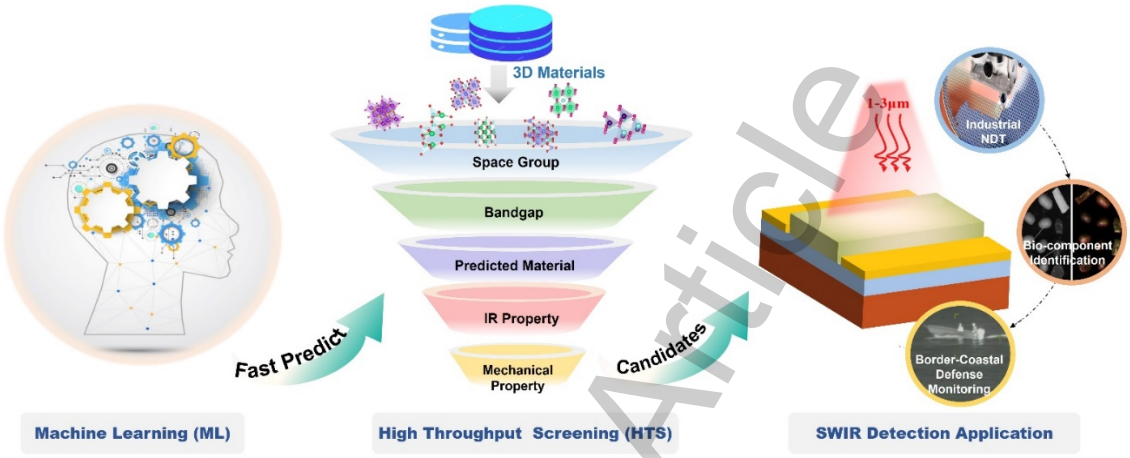
How to cite this article: Guan X, Zhang Y, Han S, Xiong C, Yang Y, Chen C, Zhang F, Zhang Y, Gao H, Zhou F, Guan P, Lu P. ABC_2-type short-wave infrared photodetector

Accepted Article

materials discovered via high-throughput screening and machine learning. *J Mater Inf* 2026;6:[Accept]. http://dx.doi.org/10.20517/jmi.2025.89

Received: 29 October 2025 | **Revised:** 25 February 2026 | **Accepted:** 26 February 2026

Abstract



Rapid discovery of short-wave infrared (SWIR) detection materials requires efficient strategies to identify candidates with suitable bandgaps, favorable carrier transport properties, and structural stability. Here, we propose a high-throughput screening (HTS) framework that integrates machine learning (ML) models with density functional theory (DFT) calculations to accelerate the prediction and validation of infrared-detection materials (see Figure 1). Using a curated dataset of 1327 I–X–VI chalcogenide compounds retrieved from the Materials Project database, we trained five regression models-random forest, gradient boosting, support vector regression, extreme gradient boosting, and decision tree-to predict electronic bandgaps with high accuracy and computational efficiency. The optimized extreme gradient boosting regression (XGBR) model delivers a test-set coefficient of determination (R^2) of 0.945, a mean absolute error (MAE) of 0.150 eV, and a mean squared error (MSE) of 0.056 eV, with a 5-fold cross-validation (R^2) of 0.927, verifying its robust prediction performance and generalization ability. This ML-guided screening highlights five promising chalcogenides: $KGaSe_2$, $KGaTe_2$, $KInSe_2$, $KInTe_2$, and $CsInTe_2$. These candidates were further evaluated using first-principles DFT calculations to assess their band structures, density of states, and carrier effective masses. Among them, $KGaSe_2$ exhibits a direct bandgap of ~0.8 eV, low effective mass, and excellent thermodynamic stability, making it a highly attractive

Accepted Article

candidate for SWIR detection. This work demonstrates the power of combining ML and DFT in accelerating the discovery of IR optoelectronic materials and provides a scalable, generalizable approach for next-generation photodetector design.

Keywords: High throughput screening, machine learning, DFT calculation, short-wave infrared detection, material prediction

INTRODUCTION

Short-wave infrared (SWIR) radiation, typically spanning wavelengths from approximately 1 to 3 μm, exhibits high detection efficiency and strong environmental penetration. Unlike other infrared regions, SWIR radiation is capable of detecting not only thermal emissions and atmospheric airglow but also reflected light from surrounding objects. These characteristics make it particularly advantageous for applications such as night vision, camouflage detection, and target recognition. In recent years, SWIR technologies have extended into civilian domains, including industrial Non-Destructive Testing (NDT), bio-component identification, and even border-coastal defense monitoring. The growing demand for high-performance and cost-effective SWIR detectors has consequently driven the development of novel materials and advanced device architectures[1-5].

Conventional materials such as mercury cadmium telluride (HgCdTe) and indium gallium arsenide (InGaAs) have long been used for SWIR detection due to their excellent optoelectronic properties. However, HgCdTe involves toxic elements and complex epitaxial growth, while InGaAs is costly and limited in detection range, restricting their broader application. Two-dimensional (2D) materials like black phosphorus, transition metal dichalcogenides (TMDs), and InSe offer high mobility and broadband absorption, but their instability, low monolayer absorption, and poor scalability hinder device-level integration (Xinyu J *et al.*, 2022)[6]. As a promising alternative, lead selenide (PbSe), an IV-VI semiconductor, offers a direct and narrow bandgap ideal for SWIR detection, along with relatively mature synthesis routes. In its colloidal quantum dot (CQD) form, PbSe enables solution processing, size-tunable bandgaps extending up to 3 μm, and low-cost fabrication. Although stability and surface defects remain challenges, recent advances have pushed the detectivity of PbSe QD-based devices beyond 10^{11} Jones and external quantum efficiency over 120% (Wu Z *et al.*, 2022)[7], making them increasingly viable

Accepted Article

for commercial SWIR applications.

The early discovery of infrared materials primarily depended on empirical approaches and labor-intensive experimental screening, which were often time-consuming and inefficient. For instance, HgCdTe, a widely used infrared detector material, was first identified in the late 1950s and subsequently optimized through iterative experimentation with techniques such as Bridgman growth, zone melting, and quench-anneal treatment (Schmit, 1983)[8]. Although the advent of first-principles simulations has enabled the theoretical prediction of key material properties-such as bandgap, thermal stability, and mechanical robustness-these methods remain computationally demanding, particularly when applied to large-scale screening tasks. To address these challenges, researchers have increasingly turned to more efficient strategies for materials discovery. High-throughput screening (HTS), supported by initiatives such as the Materials Genome Project and large-scale databases including the Materials Project[9-10], allows for the rapid evaluation of thousands of candidate compounds. When combined with machine learning (ML) techniques, this approach enables the prediction of critical properties-such as bandgap, thermodynamic stability, and carrier mobility-with significantly reduced computational costs. This data-driven methodology has proven effective in the discovery of functional materials for ultraviolet and mid-infrared applications, and holds substantial promise for accelerating the development of novel infrared detection materials. It thus represents a pivotal shift toward more scalable and intelligent material design.

Recent progresses in HTS computation and ML have significantly accelerated the discovery of novel materials, particularly in the field of optoelectronics and IR applications. HTS techniques enable the rapid evaluation of thousands of candidate compounds through automated workflows, facilitating efficient materials screening based on structural and electronic descriptors[11-14]. For instance, Wang *et al.* (2022)[14] performed large-scale HTS first-principles calculations to screen nearly 1000 monolayer structures, ultimately identifying 256 thermodynamically, mechanically, dynamically, and thermally stable 2D semiconductors with desirable electronic properties. The resulting database, termed the Two-Dimensional Semiconductor Database (2DSdb), provides a valuable resource for further theoretical modeling and device design. ML, on the other hand, has emerged as a powerful tool for uncovering complex structure-property relationships and accelerating predictive modeling[15-19]. Sun *et al.* (2019)[18] utilized ML-

Accepted Article

assisted synthesis to develop lead-free, perovskite-inspired materials with improved stability and optoelectronic performance. Similarly, Shaban Tameh *et al.* (2024)[19] employed ML models to explore organic molecules with long-wave infrared transparency, demonstrating the potential of data-driven approaches in guiding molecular design. More recently, integrated frameworks combining HTS, ML, and DFT have shown great promise for large-scale materials discovery[20-27]. Wu *et al.* (2023)[26] combined DFT and ML to identify calcium sulfides with optimal mid-infrared properties, while Shen *et al.* (2023)[27] conducted extensive screening of over 400,000 ternary chalcogenides by coupling HTS-DFT simulations with stacked ML models. This integrated approach enabled the identification of thermodynamically stable candidates with suitable bandgaps, low effective masses, and high theoretical solar conversion efficiencies (SLME > 32%), highlighting the power of HTS+ML+DFT pipelines in accelerating the discovery of next-generation IR materials.

As a promising material system for high-performance SWIR photodetectors, the I-X-VI chalcogenides—corresponding to the ABC_2 configuration (A=alkali metals (Li, Na, K, Rb, Cs), B=all main group and transition metal elements, C=chalcogen elements (S, Se, Te))—integrate intrinsic advantages that address core photodetection requirements: strong covalent bonding, tunable direct bandgaps (modulated via ordered-disorder transitions or structural distortion to match the SWIR range, as verified in $AgBiS_2$[28] and $CuTlSe_2$[29]), excellent p-type dopability, and robust mechanical/thermodynamic stability (confirmed by first-principles calculations[30]). Composed of earth-abundant, low-toxicity elements, these compounds also exhibit high carrier mobility, suppressed non-radiative recombination, and promising optoelectronic performance (e.g., EQE exceeding 13% in relevant systems[31]), while their structural flexibility and compatibility with thin-film growth further enhance practical applicability. Collectively overcoming limitations of traditional materials (fixed bandgaps, toxicity, poor stability), the integration of high-throughput screening (HTS) and machine learning (ML)-driven selection with targeted exploration of I-X-VI (ABC_2-type) chalcogenides holds great promise for the rational design of efficient, scalable, and environmentally friendly SWIR optoelectronic materials.

To efficiently explore this chemically diverse space, we developed a HTS framework that integrates machine learning ML with DFT calculations. A curated database of I-X-VI selenide compounds was constructed, and five regression models-random forest, gradient

Accepted Article

boosting, support vector regression, extreme gradient boosting, and decision tree-were trained to predict their electronic bandgaps. This data-driven approach led to the identification of five promising IR candidates: $KGaSe_2$, $KGaTe_2$, $KInSe_2$, and $CsInTe_2$. Among them, $KGaSe_2$ was found to possess an optimal bandgap (~0.8 eV), favorable carrier transport properties, and excellent thermodynamic stability, as confirmed by detailed DFT analysis.

These findings demonstrate the potential of integrating ML and first-principles methods for the accelerated discovery of high-performance IR detection materials. Compared with previous work (Kim *et al.*, 2018)[32], this method not only improved the accuracy and comprehensiveness of candidate material screening, but also revealed the orbital essence of performance differences through electronic structure analysis, providing a generalizable example for the design of SWIR materials.

MATERIALS AND METHODS

Overview of computational methods

This study aims to rapidly predict and screen SWIR materials with excellent absorption properties by integrating ML with HTS, and ultimately validate the screening results using first-principles calculations based on DFT. The research workflow mainly consists of three steps: HTS screening, ML, and DFT validation, as shown **Figure 1**. First, a material dataset covering infrared absorption properties is constructed, key descriptors are extracted, and a predictive model is established using ML algorithms. The trained model is then used to predict the infrared absorption properties of a large number of as-yet unsynthesized materials. After obtaining the prediction results from the ML model, HTS is applied to filter candidate materials. The selection criteria are primarily based on predicted bandgap ranges and other physicochemical parameters to ensure that the selected materials have potential advantages for target applications. Finally, to ensure that the screened candidate materials exhibit reliable infrared absorption properties, first-principles calculations based on DFT are performed for validation.

Accepted Article

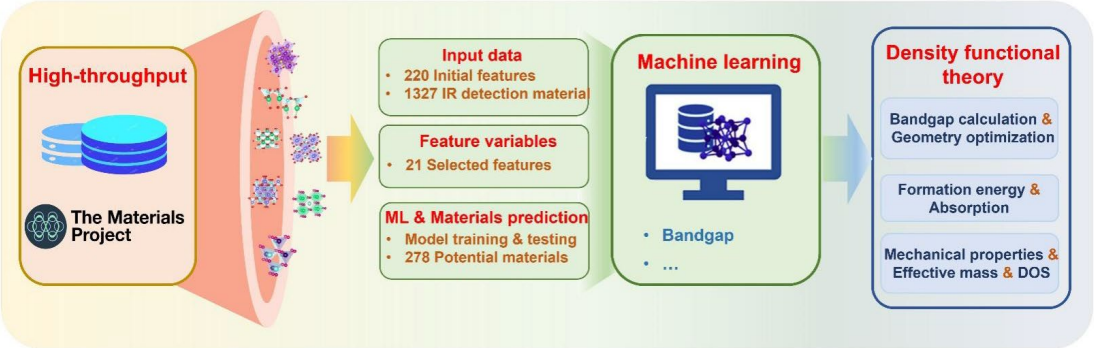


Figure 1. Flowchart for screening promising SWIR photodetector materials via an integrated high-throughput screening, ML, and DFT approach. Starting from ternary compounds retrieved from The Materials Project (MP) database, this workflow first implements high-throughput screening incorporating seven specific space groups. Through predefined selection criteria, the pool is refined to yield infrared detection material candidates and corresponding initial feature parameters. The resulting dataset is then processed via ML for feature optimization to identify key representative features, followed by ML-driven material prediction to obtain potential infrared detection candidates. Finally, comprehensive DFT calculations—including bandgap computation, structural optimization, formation energy analysis, optical absorption, mechanical properties, effective mass, and density of states (DOS) calculations—are performed to complete the final screening and identify the most promising SWIR photodetector materials.

Accepted Article

High-throughput screening framework

To identify high-performance SWIR photodetector materials, a multi-step high-throughput screening framework was developed. Starting from ~30,000 ternary compounds retrieved from the Materials Project database, sequential filters were applied based on symmetry, composition, stoichiometry (ABC_2-type[33,34]) and atomic complexity, reducing the pool to 1327 non-toxic, structurally feasible candidates. The initial band gap values used for screening were obtained directly from the Materials Project database. A machine learning model was then trained to predict infrared bandgaps, and 278 compounds meeting the target bandgap features were obtained, including those in the SWIR range (0.41–0.89 eV)[35]. Subsequent constraints on bandgap, mechanical and dynamical stability, as well as optical and transport properties, led to a final shortlist of robust candidates with strong potential for SWIR optoelectronic applications.

Machine learning methods

In this study, we integrated ML with DFT calculations to screen potential three-dimensional (3D) materials for SWIR photodetector and structural stability, targeting bandgap width as the key predictive property for ML modeling. Based on elemental composition and spatial configurations, through preliminary screening and literature research, a ML database consisting of 1327 materials and 220 initial characteristic variables was finally constructed. Five ML algorithms-RFR, GBR, SVR, XGBR, and DT[36][37] were implemented using the Scikit-Learn library and compiled via Jupyter Notebook[38][39]. The dataset was randomly partitioned into an 80% training set and a 20% testing set for model training and performance validation.

DFT calculations

Density functional theory (DFT) calculations were performed using the **Vienna *Ab initio* Simulation Package (VASP)** based on the **projector augmented wave (PAW)** method. The process also made use of plane-wave pseudo potentials[40][41]. To represent electron wave functions, a plane-wave basis set with a uniform kinetic energy cutoff energy of 500 eV for all calculations. The exchange–correlation interactions were treated within the generalized gradient approximation (GGA) using the Perdew–Burke–Ernzerhof (PBE) functional. Structural optimizations were carried out until the total energy convergence reached 10^{-5} eV, and the residual Hellmann–Feynman forces on each atom were smaller than 0.01 eV $Å^{-1}$. To accurately characterize the band structure features of the material, k-point sampling is carried out along the high-symmetry points of the Brillouin zone during the calculation process, in order to reveal the distribution laws of electronic states in different momentum space regions[42].

To maintain consistent computational accuracy across diverse crystal systems, Γ-centered k-point meshes were chosen to deliver comparable resolution in reciprocal space. The convergence of key properties with respect to k-point density—along with the associated k-point spacings for representative structures—is documented in the Supporting Information. Specific mesh dimensions were tailored according to each system's symmetry and lattice parameters, exemplified by a 4 × 4 × 2 mesh for space group $I4_1/amd$, 3 × 3 × 3 for $I\bar{4}2d$, and 9 × 9 × 1 for $R\bar{3}m$. The high-symmetry locations for the band structures were found to be, respectively, G (0 0 0), F (0 0.5 0), Q (0 0.5 0.5), Z (0 0 0.5)

Accepted Article

and G (0 0 0). K-point convergence tests were performed for representative $I4_1/amd$ and $R\bar{3}m$ structures. Total energy, band gap convergence, and estimated k-point spacings are summarized in Tables S1 and S2 of the Supporting Information. And the crystal structures were displayed using VESTA. The crystal structures were displayed using VESTA.

RESULTS AND DISCUSSION

High-throughput calculation

We start with all possible ternary chalcogenide structures from the Materials Project database[43]. Considering the difficulties in synthesizing multicomponent materials, we focus on ternary compounds, which account for approximately 30,000 materials. Further, we introduce a series of screening criteria in order to identify the most promising material candidates for SWIR photodetector potential candidates. And the candidate space groups include 7 types, namely *C2/c*, $I\bar{4}2d$, *I4/mcm*, $R\bar{3}m$, $Pna2_1$, $I4_1/amd$, and *P4/mmm*. The selection of these seven space groups is not arbitrary but based on their statistical prevalence in stable ABC_2-type (I-X-VI) chalcogenides, structural compatibility with SWIR photodetection performance, and verified synthetic feasibility[44-46]. Our screening workflow is schematically presented in **Figure 2(a)** and illustrated as follows.

Among them, *C2/c* belongs to the monoclinic crystal system and is a stable structure that can adapt to distortion effects. Materials with this structure have good structural stability under external field effects. $I\bar{4}2d$, *I4/mcm* and $I4_1/amd$ all belong to the tetragonal crystal system. $I\bar{4}2d$ has non-centrosymmetric characteristics and is suitable for infrared detection. *I4/mcm* has rotational symmetry and translational symmetry and is suitable for optoelectronic materials. $I4_1/amd$ originates from the rock salt structure's six-coordination structure and has potential application value in narrow bandgap semiconductor research. $R\bar{3}m$ belongs to the trigonal crystal system and is also a six-coordination structure derived from the rock salt structure. Its high symmetry helps stabilize the electronic structure of the material and improve the carrier transport capability, making it important in infrared detection materials. $Pna2_1$ belongs to the orthorhombic crystal system and is commonly found in polar materials, with wide applications in optoelectronic functional materials. *P4/mmm*, as the most symmetrical structure in the tetragonal crystal system, its high symmetry contributes to the isotropic optoelectronic properties of the material—which is beneficial for large-area, uniform SWIR detector fabrication.

Accepted Article

It is important to clarify that excluding other space groups does not mean they are inherently unstable or unable to adapt to distortion, but rather they fail to meet the comprehensive requirements of SWIR photodetectors (synergy of bandgap matching, high carrier mobility, and synthetic feasibility). For example, some high-symmetry space groups (e.g., Fm-3m) are stable but have fixed bandgaps outside the SWIR range, while low-symmetry ones (e.g., P1) suffer from insufficient carrier mobility. Thus, the seven selected space groups represent a rational trade-off for target applications.

As shown **Figure 2(a)**, we introduced a series of screening criteria to identify promising SWIR detection materials. The overall screening workflow is presented as following:

✧ **Step 1**: Starting from ~30,000 ternary compounds in the Materials Project database, we applied initial filters to narrow down candidates: (i)inclusion in one of 7 representative space groups, (ii) I-X-VI family, (iii) no rare earth or O and (iv) ABC_2-type stoichiometry constraint. These criteria reduced the dataset to 1900 compounds.

✧ **Step 2:** To further narrow down the candidates, compounds with a non-zero bandgap (Eg > 0 eV) as reported in the Materials Project database, a total number of atoms per unit cell ≤ 32, and composed only of non-toxic elements were selected. This refinement yielded 1327 compounds.

✧ **Step 3:** A machine learning model was trained using 1327 high-throughput-screened ABC_2-type compounds as the training dataset to learn structure–infrared bandgap relationships. Based on the trained model and literature-guided targeted element substitutions[44-46], a prediction space was constructed for bandgap inference of unexplored ternary compounds. The trained model was then applied to this space, yielding 278 predicted sets that meet the bandgap requirements for functional infrared materials.

✧ **Step 4:** In this section, a bandgap filter in the range of 0.41–0.89 eV, corresponding to the short-wave infrared (SWIR) region, was imposed. These combined criteria reduced the candidate list to 27 compounds.

✧ **Step 5:** Optical screening was performed by analyzing the absorption spectra.

Accepted Article

Materials that exhibited absorption coefficients below 10^4 cm^{-1} in the SWIR range were excluded. This filtering step reduced the candidate list to 5 candidates with strong SWIR optical absorption.

✧ **Step 6:** The remaining 5 compounds were subjected to comprehensive first-principles calculations to evaluate their mechanical properties and dynamical stability. Elastic constants were computed to assess mechanical robustness, and ab initio molecular dynamics (AIMD) simulations were conducted to verify thermal stability at finite temperature. All 5 materials were confirmed to be stable under both criteria.

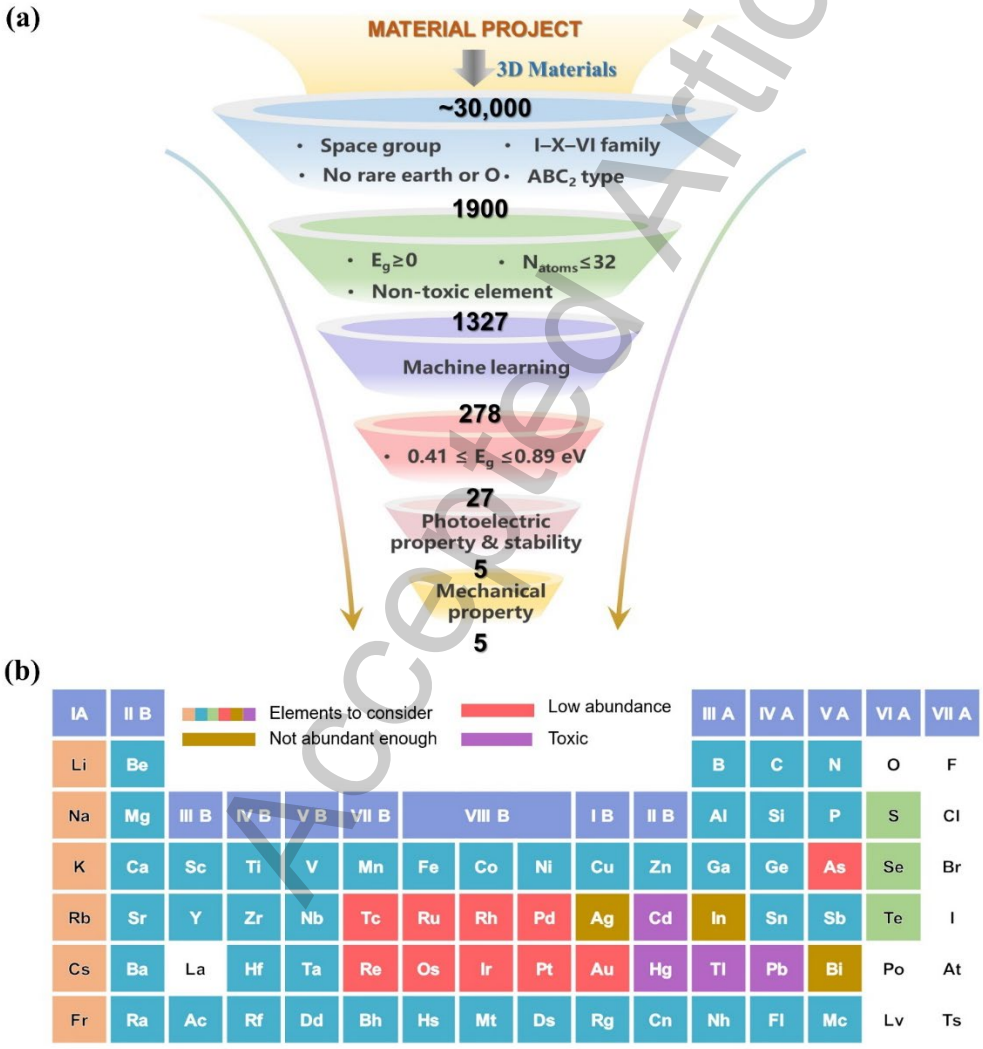


Figure 2. (a) HTS workflow for the computational screening of promising infrared detection materials. (b) The database of candidate materials, with the constituent elements shown in the figure.

Accepted Article

Machine learning screening

Based on the chemical composition and structural characteristics of ternary compound, a machine learning (ML) database was constructed using elements from group. As shown **Figure 2(b)**, the first-site elements were selected from group I elements (Li, Na, K, Rb, Cs) and the last-site elements from group VI elements (S, Se, Te). This selection resulted in a dataset comprising 1327 compounds, sourced from the Materials Project. The bandgap distribution of these 1327 ternary compounds spans a range from 0 to 7 eV, encompassing the primary bandgap range relevant for IR optoelectronic detection materials, as illustrated in **Figure 3(a)**. For the development and evaluation of the ML model, 1061 materials were selected as the training set, and 266 materials were chosen as the test set. The selection of feature variables for different ML prediction tasks varies; ideal feature variables can significantly enhance the predictive ability and generalization capability of the ML model. Therefore, the selected feature variables should be as rich as possible to effectively define the physical and chemical properties of the materials.

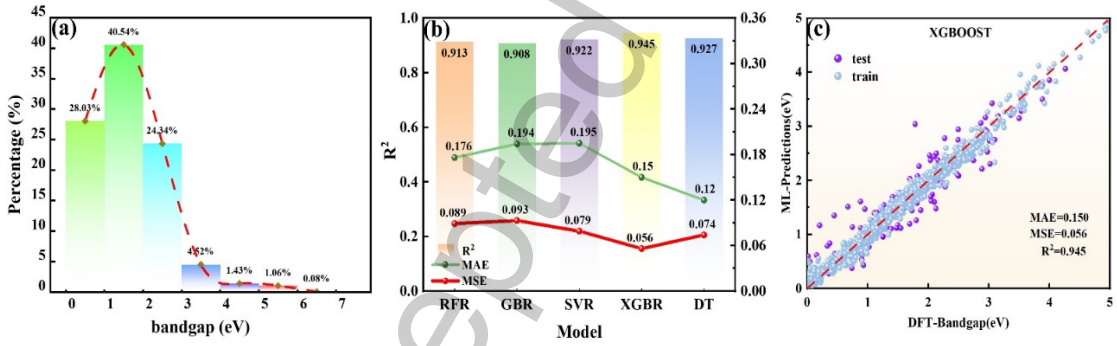


Figure 3. ML model training and prediction results. (a) The bandgap distribution of the 1327 materials. (b) Test set performance (R^2, MAE, MSE) for the five base ML models. (c) Fitting plot of the training and test set. prediction results for the best model, XGBR.

When selecting feature variables, two fundamental principles should be followed: features must be easy to obtain, and the selected features must be sufficiently comprehensive and complete. The number of feature descriptors should be significantly smaller than the number of materials in the database, as too many features can reduce the predictive power and accuracy of ML models. To enable the ML model to better capture the relationship between feature variables and bandgaps, this study selected over 30 core initial features for model training, which included ion radius, valence electron count,

Accepted Article

tolerance factor, electronegativity, Fermi level, etc. However, excessive features can lead to data redundancy, thereby affecting the model's accuracy and training speed. On the other hand, too few features may cause underfitting, preventing the model from identifying the correlation between features and bandgaps. Therefore, this study used Pearson correlation coefficient plots and importance rankings from the best training model XGBR for feature selection. Researchers first evaluated features with a Pearson correlation coefficient less than 0.7 using the correlation matrix (**Figure S1**): for each pair of highly correlated features, one was removed. Then, based on feature importance ranking, the top 21 features were selected, including the bandgap of ternary compounds, the electron affinity of elements, the Fermi level, the Mendeleev number of B-site elements, and the solid-phase average density, among others. These 21 feature descriptors are detailed in **Table S3**.

We utilized five base models: RFR, SVR, GBR, DT, and XGBR. These five ML models were trained and tested using the established database of 1327 materials. Additionally, hyperparameter optimization was performed using the grid search algorithm. By splitting the dataset into multiple training and validation sets, we reduced errors caused by randomness, ensuring the stability of the model across different data subsets.

The XGBR model demonstrated superior training performance compared to five other models based on the evaluation metrics of R^2, MAE, and MSE. On the test set, it achieved an R^2 of 0.945, MAE of 0.150, and MSE of 0.056, as shown in **Figure 3(b).**

As illustrated in **Figure 3(c)**, the predicted bandgap values (0~5 eV) of the XGBR model exhibited a high degree of consistency with the actual values across both training and test sets, confirming its robust generalization capability. The generalization ability of the hyperparameters was evaluated using K-fold cross-validation. **Table S4** summarizes the hyperparameter configurations of the five base models, with the optimized hyperparameters of the XGBR model presented immediately below the table. And **Table S5** provides the band gaps and formation energies of the 27 candidate materials, where the formation energies were calculated based on total energies of fully optimized structures relative to their elemental reference phases (see Supporting Information for computational details), thereby supporting feasibility evaluation and experimental selection.

Accepted Article

The optimal XGBR model was trained using 1327 high-quality ABC_2-type compounds (refined via high-throughput screening) as the training set, whose training performance is illustrated in **Figure 3(c)**, while the training outcomes of the other four supplementary models are presented in **Figure S2** for comprehensive validation of model reliability. To expand the candidate pool for SWIR photodetection, we constructed a prediction database based on well-established infrared detection materials reported in existing literature—such as the ABC_2-type $AgBiS_2$, a typical high-performance near-infrared detection material with tailored structural and electronic properties[47-51]. Specifically, we performed targeted element substitutions on the A, B, and C sites of these literature-validated materials, including substitutions of elements from the same group, as well as constrained substitutions within the same period only when the substituted elements exhibit effectively identical valence-electron configurations in the relevant bonding environment. In addition, substitutions were limited to elements with similar electronic orbital characteristics, defined here as having the same dominant valence-orbital types with effectively identical electronic configurations near the band edges without introducing additional strongly correlated d or f states. This strategy has been proven effective for tuning bandgaps within the SWIR range while retaining core material stability. These rational substitutions ensured the derived compounds retained core electronic properties consistent with the 1327 high-quality training samples, thus maintaining structural and chemical compatibility with the ABC_2-type system[52, 53]. Subsequently, we applied the trained XGBR model to predict the bandgaps of compounds in this constructed database, and guided by the fundamental screening criteria for SWIR detection materials[54], 278 promising candidates were identified from the prediction results. All predicted band gaps are based on the PBE benchmark, and the actual values may be systematically higher by 30–50%. To ensure the generalizability of our machine learning model for ABC_2-type ternary compounds, we employed k-fold cross-validation to select the best XGBoost configuration (**Figure S3**). We compare the predicted bandgaps against the true values, with 95% confidence intervals included, in **Figure S4**, and summarize the 95% confidence intervals for quantifying bandgap prediction uncertainty separately in **Table S6**.

After machine learning identified 278 promising candidates for SWIR applications, we further refined the candidate set based on their predicted bandgap values. Specifically,

Accepted Article

compounds with bandgaps falling within the SWIR range of 0.41–0.89 eV were selected, in accordance with the fundamental bandgap requirements for SWIR photodetector materials. Based on this bandgap criterion, 27 compounds were retained from the initial 278 candidates. This selection ensures that the final candidates exhibit bandgaps well suited for SWIR detection and therefore possess strong potential for infrared optoelectronic applications. These 27 materials were subsequently chosen for detailed first-principles DFT calculations.

Crystal stability and photoelectric performance

To further assess these 27 ABC_2-type candidates, we performed formation energy calculations[54] based on total energies of fully optimized structures relative to their elemental reference phases, and computed their electronic band gaps using the PBE functional. Concurrently, their optical absorption spectra were evaluated to determine suitability for infrared applications. This screening identified five compounds exhibiting strong absorption within the SWIR spectral range, as illustrated in Figure 4. These materials exhibit a favorable combination of low formation energies and infrared-relevant optical responses, highlighting them as promising candidates for further exploration toward SWIR photodetector–related applications.

As shown **Figure 4(a).** $KGaSe_2$-*I4₁/amd* exhibits the most pronounced narrow-band response characteristics in SWIR region, with its main absorption peak located at 0.45 eV (2.75 μm) and an absorption coefficient reaching 1.8×10^5 cm^{-1}, surpassing that of similar materials such as $KInSe_2$-*I4₁/amd* (1.5×10^5 cm^{-1}).Notably, $KGaTe_2$-*I4₁/amd*, despite having its main absorption peak at 0.65 eV (1.91 μm), extends its absorption edge up to 1.2 eV (1.03 μm), maintaining an effective absorption coefficient of $>5\times10^4 cm^{-1}$ in the SWIR-LWIR transition region. The high-symmetry *I4₁/amd* structure (e.g., $KGaSe_2$) enhances direct bandgap transition probabilities, leading to a full width at half maximum (FWHM) of only 0.12 eV, which significantly improves spectral selectivity. Meanwhile, the layered I4₂d structure of $CsInTe_2$, due to quantum confinement effects, sustains a high absorption coefficient of 1.2×10^5 cm^{-1} in the 1.0-1.2 eV (1.03-1.24 μm) range. This characteristic is closely associated with the strong spin-orbit coupling (SOC) effects of Te atoms. Among the five IR detection materials, the absorption spectra consistently show strong absorption edges. These properties indicate excellent optoelectronic response capabilities in IR detection materials, making these materials suitable for designing high-

Accepted Article

efficiency IR detection.

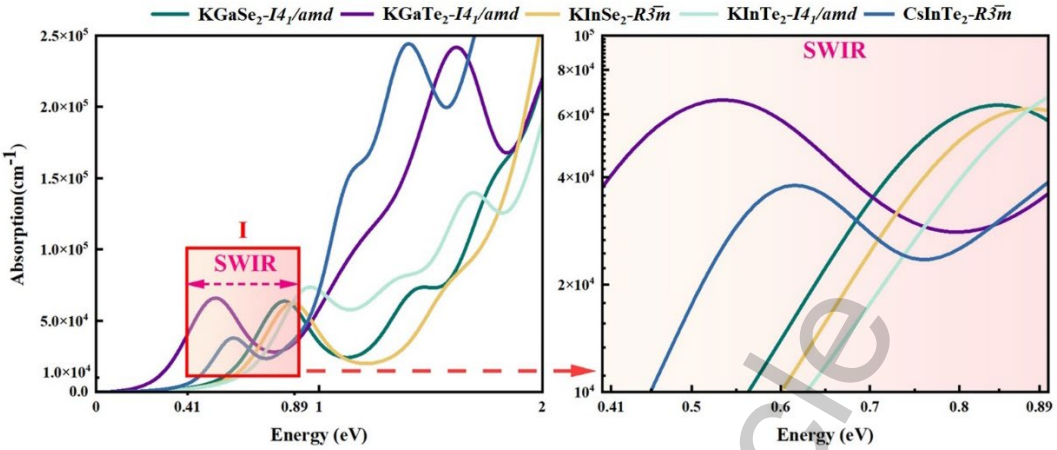


Figure 4. (a)Absorption for the six systems $KGaSe_2$-*I*4$_1$/*amd*, KGaTe2-*I*4$_1$/*amd*, $KInSe_2$-*I*4$_1$/*amd*，$KInTe_2$-*I*4$_1$/*amd*, $CsInTe_2$-*R*$\bar{3}$*m*. (b) An enlarged view of Area I in figure 7(a).

From previous literature, we know that in direct bandgap materials, the conduction band minimum (CBM) and valence band maximum (VBM) have the same momentum at the Gamma point, allowing electrons to transition directly from the CBM to the VBM without the need for lattice vibrations (phonons). The electron-hole pairs can directly recombine and emit photons[56-59]. This not only eliminates the need for complex methods such as heating, doping, quantum structure control, surface passivation, and electric field modulation, but also results in higher radiative recombination rates and photon generation efficiencies in direct bandgap materials. This characteristic significantly enhances the brightness and efficiency of devices, making candidate materials highly suitable for infrared photodetector with high recombination rates[60].

Additionally, the mechanical properties and flexibility of IR detection materials were calculated via the Young's modulus (E). The E reflects the material's ability to elastically deform under external force, while the shear modulus (G) characterizes the material's resistance to shear deformation. Generally, softer materials exhibit lower E and shear modulus. Based on the calculated elastic constants (C_{ij})[61], the mechanical parameters such as bulk modulus (B), shear modulus (G), E, and Poisson's ratio (σ) were calculated using the Voigt-Reuss-Hill (VRH) approximation[62]. These mechanical parameters are important for evaluating the material's mechanical stability, ductility, brittleness, and interactions, and provide a foundation for further calculations of thermodynamic

Accepted Article

parameters such as melting temperature and Debye temperature. For cubic structure materials, the elastic matrix typically consists of three independent constants: C_{11}, C_{12}, and C_{44}. To assess the mechanical stability of the materials, the Born mechanical stability conditions for cubic crystals at zero pressure are considered. Accordingly, the elastic constants (C_{11}, C_{12}, and C_{44}) should satisfy the criteria $C_{11} - C_{12} > 0$, $C_{11} + 2C_{12} > 0$, and $C_{44} > 0$[63]. The quantity C_{12} - C_{44} is not a strict stability requirement but is commonly discussed as a qualitative indicator of elastic anisotropy and bonding characteristics. In addition, the condition $C_{12} < B < C_{11}$ reflects the relationship between bulk and elastic moduli rather than serving as an independent mechanical stability criterion.

Mechanical stability is a critical parameter in evaluating the long-term reliability of SWIR detector materials. According to the elastic constants presented in **Table 1**, $KGaTe_2$-$I4_1/amd$, and $CsInTe_2$-$R\bar{3}m$ satisfy the generalized Born stability criteria, confirming their mechanical robustness.

Table 1. The elastic constants C_{ij} (GPa) and Cauchy pressure (GPa), calculated bulk modulus B (GPa), Young's modulus E (GPa), shear modulus G (GPa), Poisson's ratio σ, Pugh's ratio B/G, Debye temperature θ_D, and Zener anisotropy factor A of $KGaSe_2$-$I4_1/amd$, $KGaTe_2$-$I4_1/amd$, $KInSe_2$-$I4_1/amd$, and $CsInTe_2$-$R\bar{3}m$

Compounds	C_{11}	C_{12}	C_{44}	C_{12}-C_{44}
$KGaSe_2$-$I4_1/amd$	30.76	5.03	10.29	-5.27
$KGaTe_2$-$I4_1/amd$	32.75	6.00	3.75	2.25
$KInSe_2$-$I4_1/amd$	212.93	12.10	12.42	-0.32
$KInTe_2$-$R\bar{3}m$	44.09	20.61	12.90	7.71
$CsInTe_2$-$R\bar{3}m$	37.87	20.90	13.88	7.03

Compounds	B	E	G	σ	B/G	θ_D	A
$KGaSe_2$-$I4_1/amd$	19.22	27.92	11.10	0.26	1.73	183.50	0.80
$KGaTe_2$-$I4_1/amd$	17.57	19.29	7.32	0.32	2.40	132.60	0.28
$KInSe_2$-$I4_1/amd$	84.95	80.79	30.11	0.34	2.82	271.00	0.12
$KInTe_2$-$R\bar{3}m$	25.22	32.79	12.77	0.28	1.97	165.20	1.10
$CsInTe_2$-$R\bar{3}m$	25.95	29.56	11.28	0.31	2.30	142.20	1.64

Accepted Article

$KGaTe_2$-$I4_1/amd$ exhibits a positive $C_{12}-C_{44}$ value, indicating enhanced mechanical stability, which is commonly discussed as a qualitative indicator of elastic anisotropy and bonding characteristics rather than a strict measure of mechanical stability. This feature suggests a relatively balanced elastic response under combined normal and shear deformations, which may be favorable from a mechanical behavior perspective. $CsInTe_2$-$R\bar{3}m$, shows a higher C_{11}, indicating a stronger resistance to uniaxial compressive deformation along the longitudinal direction. Although its relatively large C_{12} implies an enhanced coupling between normal stress and lateral strain, the elastic constants remain within a reasonable range, and its moderate C_{44} reflects a balanced response to shear deformation, collectively suggesting mechanically robust elastic characteristics. In contrast, $KGaSe_2$-$I4_1/amd$ and $KInSe_2$-$I4_1/amd$ both exhibit lower C_{44} values and negative $C_{12}-C_{44}$, which are often associated with increased elastic anisotropy and a softer shear response. These elastic features indicate a greater sensitivity to shear-related distortions and highlight potential mechanical limitations that should be considered when evaluating their structural robustness under external perturbations.

Among the studied materials, $CsInTe_2$ shows favorable mechanical behavior in $R\bar{3}m$ phases. The $R\bar{3}m$ structure, displays higher C_{11} and C_{44} values, reflecting superior rigidity and resistance to external mechanical loads. These attributes enhance structural integrity and environmental adaptability, reducing the risk of degradation due to thermal or mechanical stress. In summary, $KGaTe_2$-$I4_1/amd$ and $CsInTe_2$-$R\bar{3}m$ are mechanically stable and better suited for long-term IR detection applications. Among them, $CsInTe_2$-$R\bar{3}m$ stands out for its excellent mechanical balance and should be prioritized for further development as a promising IR detector material.

Fig 5. (a)-(c) present the compressive, tensile, and bending strengths of the studied materials. $KGaSe_2$-$I4_1/amd$ and $KInSe_2$-$I4_1/amd$ exhibit high bulk moduli (19.22 and 84.95 GPa, respectively), indicating strong compressive resistance suitable for long-term stable operation and physical shock absorption in optoelectronic devices.

Accepted Article

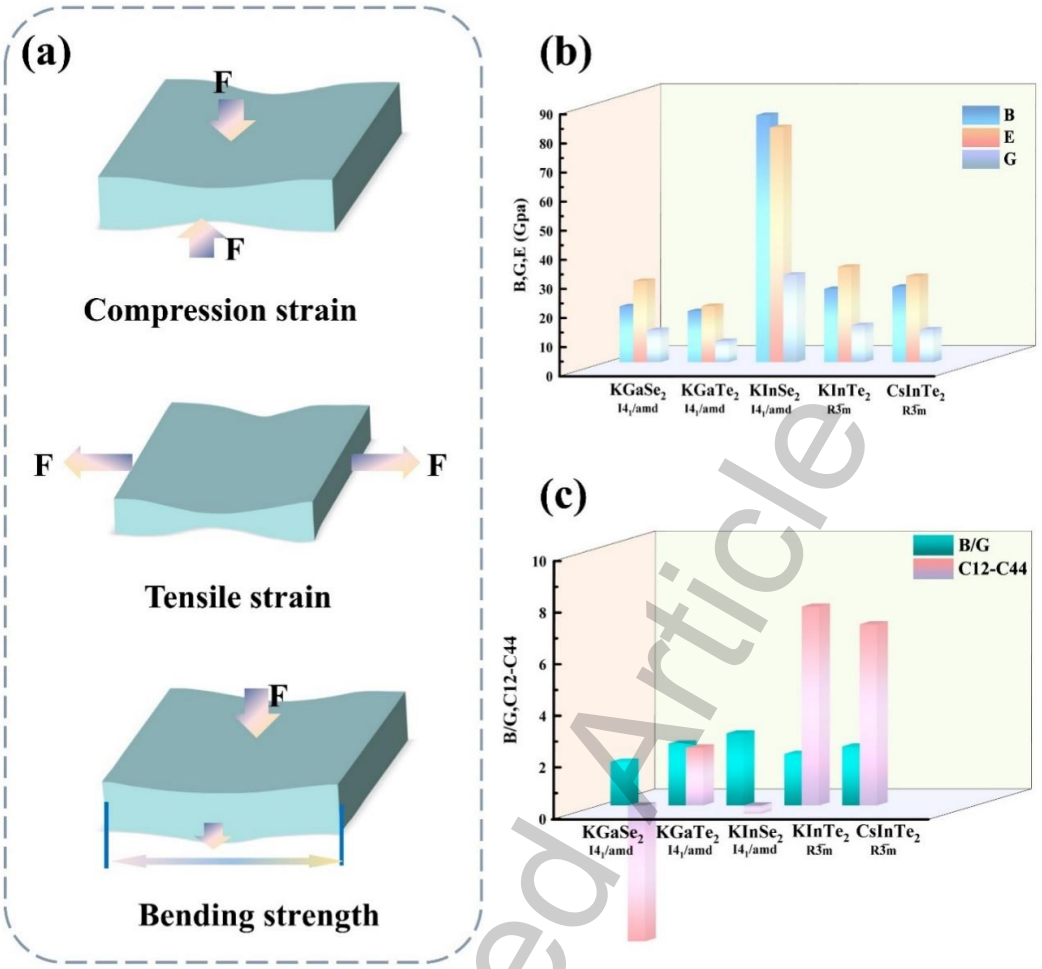


Figure 5. (a)-(c) Schematic diagrams of compressive strength, tensile strength, and bending strength. The elastic moduli *B*, *G*, and Young's modulus *E* are plotted in **(b)**. **(c)** The Pugh's ratio *(B/G)* and Cauchy pressure (C_{12}-C_{44}) of 5 materials.

The Young's modulus, a measure of stiffness, is critical for assessing rigidity. $KInSe_2$-*I4₁/amd* (80.79 GPa) and $CsInTe_2$-*R3̄m* (29.56 GPa) demonstrates strong rigidity, helping maintain structural stability under mechanical and thermal stress. Similarly, the shear modulus follows the same trend: $KInSe_2$- *I4₁/amd* (30.11 GPa) and $CsInTe_2$-*R3̄m* (11.28 GPa) offers good resistance to mechanical stress. According to the Pugh ratio (B/G), which indicates ductility (with values >1.75 denoting ductile behavior), all materials exceed this threshold. $KGaTe_2$-*I4₁/amd* and $CsInTe_2$-*R3̄m* (B/G = 2.40 and 2.99, respectively) strike a better balance between ductility and stability, making them promising for durable optoelectronic applications. The Cauchy pressure (C_{12}-C_{44}), an indicator of bond character and elastic behavior, is positive for all materials, implying ionic bonding and further supporting their ductile nature. Poisson's ratio (σ), which

Accepted Article

measures lateral deformation, ranges between 0.32 and 0.36 for these materials. $KGaTe_2$-$I4_1/amd$ and $CsInTe_2$-$R\bar{3}m$, with slightly lower σ, offer better deformation resistance, making them more suitable for flexible detection applications. The Debye temperature (θ_D), reflecting thermal stability, is highest for $KGaTe_2$-$I4_1/amd$ (132.60 K), suggesting they can better withstand temperature fluctuations and thermal expansion-desirable traits for high-power devices. Conversely, $CsInTe_2$-$R\bar{3}m$ (94.00 K) may experience thermal degradation over extended use, limiting its high-power IR photodetector applications.

Elastic anisotropy, characterized by the Zener anisotropy factor (A), reveals that none of the materials are perfectly isotropic ($A \neq 1$). **Figure 6** shows anisotropy in the Young's modulus distribution, with $KInSe_2$-$I4_1/amd$ exhibiting the greatest deviation. For infrared detection applications, moderate anisotropy-as seen in $CsInTe_2$-$R\bar{3}m$ is favorable, as it ensures more uniform mechanical behavior under stress. In contrast, high anisotropy in $KInSe_2$-$I4_1/amd$ demands additional thermal and mechanical management. $CsInTe_2$-$R\bar{3}m$, with their moderate anisotropy, good ductility, and acceptable thermal and mechanical properties, are well-suited for flexible detection applications. $KInSe_2$-$I4_1/amd$, despite its high rigidity, requires further optimization in terms of anisotropy and deformation resistance for stable application in flexible infrared detection systems.

Accepted Article

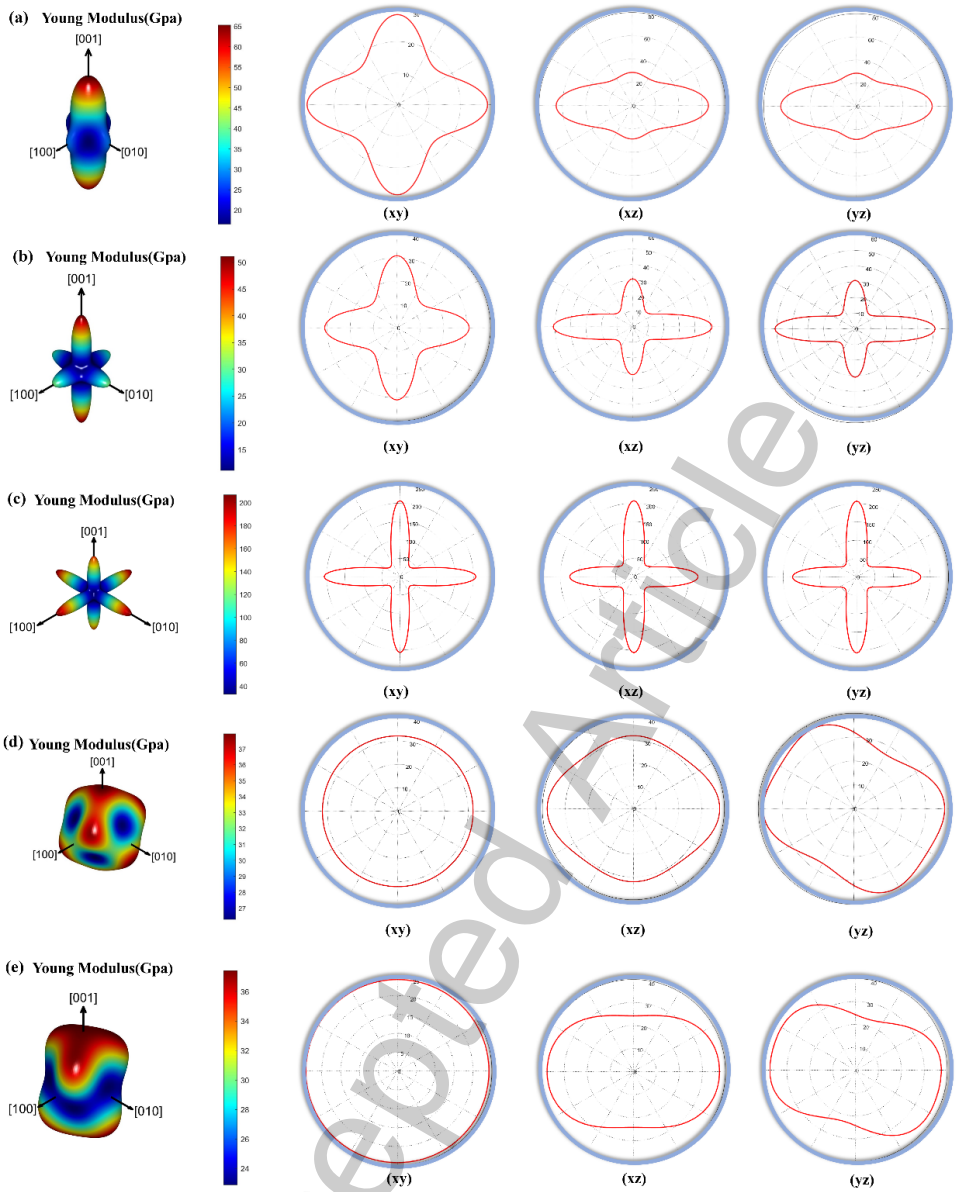


Figure 6. The directional dependence of Young's Modulus (*E*): **(a)** $KGaSe_2$-*I*4$_1$/*amd*, **(b)** $KGaTe_2$-*I*4$_1$/*amd*, **(c)** $KInSe_2$-*I*4$_1$/*amd*, **(d)**$KInTe_2$-*R*$\bar{3}$*m*, **(e)**$CsInTe_2$-*R*$\bar{3}$*m*. And the construction of the corresponding system on 2D *xy*, *xz*, and *yz* projection planes.

After mechanical property screening, currently there are 6 potential infrared detection materials. In high-performance electronic devices, a smaller effective mass enhances electron mobility, thereby improving the switching speed and overall device performance. In optoelectronic materials, higher carrier mobility facilitates the efficient collection of photo-generated carriers, thereby enhancing the material's photoresponse capability. The carrier mobility (μ) depends on the effective mass (m) and the relaxation time (τ), expressed as[64-67]:

Accepted Article

$$\mu = \frac{e}{m}\tau \quad (6)$$

Therefore, in the performance evaluation of infrared absorption materials, the effective mass (m^*) of electrons and holes is a crucial physical parameter, as it directly affects carrier mobility and the optical response characteristics of the material. Specifically, the effective mass of electrons (m_{e^*}) and holes (m_{h^*}) can be calculated using either the curvature approximation of the parabolic band structure or a numerical fitting approach[67]. The expression is given as follows:

$$m^* = \hbar^2(\frac{\partial^2 E}{\partial k^2})^{-1} \quad (7)$$

where $\hbar$ is the reduced Planck's constant, E is the energy of the electronic state, and k is the wave vector in reciprocal space.

For a single isotropic and parabolic conduction band, the effective mass can be obtained from the curvature of the conduction band[68]. Among the five candidate materials, $KGaSe_2$- *I4₁/amd*, $KGaTe_2$- *I4₁/amd*, and $KInSe_2$- *I4₁/amd* exhibit relatively low electron mobility, as shown in the table, with a decreasing trend in mobility. This also indicates that the In-Se bond in $KInSe_2$- *I4₁/amd* exhibits stronger covalent characteristics compared to the Ga-Se bond in $KGaTe_2$-*I4₁/amd*. Notably, $KInSe_2$-*I4₁/amd* has both hole and electron effective masses smaller than 1, which is advantageous for infrared photodetectors. A smaller hole mass implies a longer carrier diffusion length, reducing non-radiative recombination losses and thereby enhancing the photoelectric conversion efficiency. Notably, as shown in **Table 2**, $KInSe_2$-*I4₁/amd* exhibits both electron and hole effective masses below 1 m_0, indicating a favorable balance for infrared photodetector applications. A smaller electron effective mass enhances electron transport, allowing photo-generated electrons to quickly reach the collecting electrode [69-71]. Meanwhile, a lower hole effective mass implies a longer carrier diffusion length, reducing non-radiative recombination losses and thereby improving photoelectric conversion efficiency.

Table 2. Electronic effective mass, hole effective mass, and exciton binding energy for $KGaSe_2$-*I4₁/amd*， $KGaTe_2$-*I4₁/amd*，$KInSe_2$-*I4₁/amd*，$KInTe_2$-*R3̄m* ，

Accepted Article

$CsInTe_2$-$R\bar{3}m$

Compounds		electron effective mass(m_e^*)	hole effective mass(m_h^*)
$KGaSe_2$-$I4_1/amd$	**K − Γ**	0.25	1.86
	K − M	0.25	3.31
$KGaTe_2$-$I4_1/amd$	**K − Γ**	0.23	1.71
	K − M	0.23	1.82
$KInSe_2$-$I4_1/amd$	**K − Γ**	0.18	0.19
	K − M	1.5	0.51
$KInTe_2$-$R\bar{3}m$	**K − Γ**	0.25	0.39
	K − M	0.74	4.13
$CsInTe_2$-$R\bar{3}m$	**K − Γ**	0.48	0.23
	K − M	0.50	3.73

Furthermore, in optoelectronic devices such as photodetectors, maintaining effective mass values within an optimal range is essential. If the electron effective mass is too small (e.g., below 0.1 m_0), the resulting excessively high mobility may cause electrons to transit the active region too rapidly, reducing their probability of recombining with holes and thereby suppressing radiative recombination efficiency. Consequently, moderate and well-balanced effective masses for both electrons and holes are generally preferred, as they can simultaneously support efficient carrier transport and favorable recombination dynamics, which are critical for achieving high sensitivity and efficiency in infrared photodetection.

It should be emphasized that, in the present work, the calculated effective masses are not intended to provide a quantitative evaluation of carrier mobility, which would additionally require knowledge of carrier relaxation times and scattering mechanisms. Instead, the effective-mass analysis is used as an indicator of carrier transport potential, suggesting that the selected materials possess favorable transport characteristics under appropriate conditions. A comprehensive assessment of carrier mobility, including explicit relaxation-time calculations, is left for future studies.

Accepted Article

As shown **Figure 6,** $KGaSe_2$-*I4₁/amd* and $KInSe_2$-*I4₁/amd* exhibit high bulk moduli (19.22 and 84.95 GPa, respectively), indicating strong resistance to compression and potential for long-term operational stability in optoelectronic devices. $KInSe_2$-*I4₁/amd* (E = 80.79 GPa, G = 30.11 GPa) and $CsInTe_2$-*R3̄m* (E = 29.560 GPa, G = 11.280 GPa) show superior stiffness and resistance to deformation.

According to the Pugh criterion (B/G > 1.75), all three materials are ductile. Although this implies excellent ductility, its low bulk and Young's moduli suggest compromised mechanical stability. In contrast, $KGaTe_2$-*I4₁/amd* and $CsInTe_2$-*R3̄m* (B/G = 2.40 and 2.99, respectively) provide a more favorable trade-off between ductility and structural robustness. Positive Cauchy pressures confirm the dominance of ionic bonding in all structures, supporting their ductile behavior. Poisson's ratios for the materials range from 0.32 to 0.36, consistent with typical values for semiconductors. The Debye temperatures of $KGaTe_2$-*I4₁/amd* (132.60 K) indicate strong thermal stability, beneficial for high-power optoelectronic applications. $CsInTe_2$-*R3̄m*, with a lower Debye temperature (94.00 K), may experience more significant thermal expansion, limiting its use under high-temperature conditions. Elastic anisotropy, evaluated via the Zener factor and illustrated in **Figure 6,** reveals that none of the materials are elastically isotropic. $KInSe_2$-*I4₁/amd* shows the most pronounced anisotropy, which may lead to direction-dependent mechanical performance. In contrast, $CsInTe_2$-*R3̄m* exhibits moderate anisotropy, favorable for uniform mechanical response in flexible devices.

In summary, $CsInTe_2$-*R3̄m* offers a desirable combination of ductility, thermal stability, and manageable anisotropy, making them promising candidates for flexible applications. $KInSe_2$-*I4₁/amd*, while exhibiting excellent rigidity, requires further mechanical and thermal optimization for reliable use in flexible infrared detection systems.

In the band analysis section，5 of 6 materials were found to exhibit direct bandgaps, which has better electronic transition efficiency compared with indirect bandgap materials, and all the bandgap values are shown in **Figure 7**. The following sections will provide a detailed investigation of the optical properties of $KGaSe_2$-*I4₁/amd*, $KGaTe_2$-*I4₁/amd*, $KInSe_2$-*I4₁/amd*, $CsInSe_2$-*I4₁/amd* and $CsInTe_2$-*R3̄m*.

Accepted Article

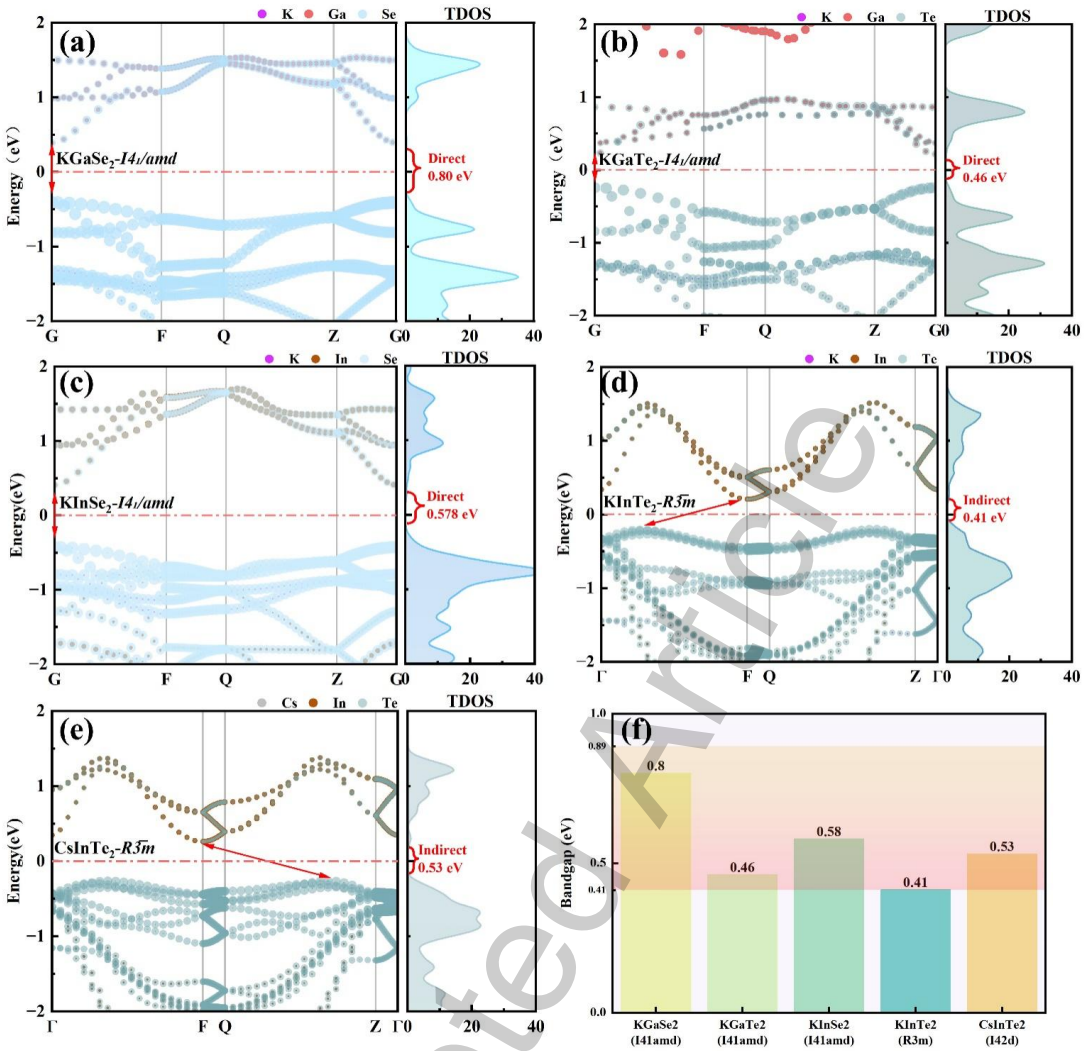


Figure 7. The band structures and total density of states (TDOS) of the (a) $KGaSe_2$-*I4₁/amd*, (b) $KGaTe_2$-*I4₁/amd*, (c) $KInSe_2$-*I4₁/amd*, (d) $KInTe_2$-*R3̄m* and (e) Cs_2InTe_2-*R3̄m*, (f)Bandgap data summary chart. The Fermi energy is set to zero.

The analysis of the band structure and TDOS in the images reveals that different space groups have a significant impact on the band properties and electronic state distribution of materials. In the tetragonal crystal system, $KGaSe_2$, $KGaTe_2$, and $KInSe_2$ with the *I4₁/amd* space group are all direct bandgap semiconductors, with the band extremum appearing at the same momentum space point, which is conducive to light absorption and photoelectric conversion. Among them, $KGaSe_2$ has the highest bandgap (0.80 eV), while $KGaTe_2$ has the lowest bandgap (0.46 eV), indicating that selenium (Se) has a greater influence on the band structure. Its TDOS shows that the electronic state distribution is relatively concentrated, which is suitable for infrared detection applications. Moreover, its TDOS shows that the electronic state distribution is relatively dispersed, which may

Accepted Article

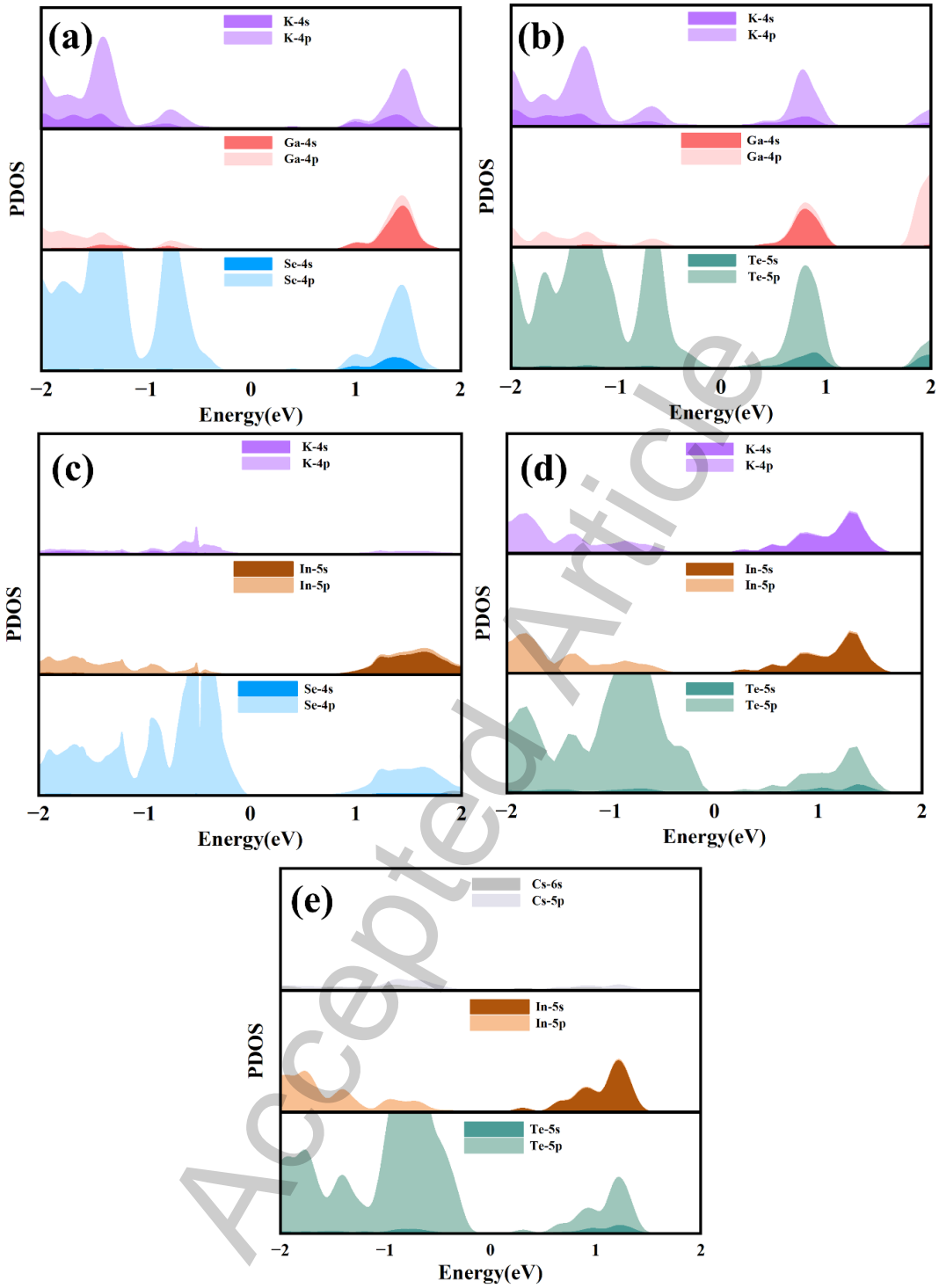
improve the transport performance of photo-generated carriers.

The $CsInTe_2$ and the $KInTe_2$ of the $R\bar{3}m$ space group exhibits an indirect bandgap (0.525 eV, 0.409eV), and the band extremum does not match in the momentum space. This will affect the optical transition and the efficiency of photoelectric conversion, but the longer carrier lifetime makes it advantageous in certain photoelectric applications. Its TDOS shows a gentler distribution of electronic states, suggesting better electronic transport performance. Overall, the $I4_1/amd$ structure tends to form a direct bandgap, suitable for efficient photoelectric conversion; while the $R\bar{3}m$ structure may introduce an indirect bandgap, suitable for infrared detection applications with long carrier lifetime. Reasonable selection of the space group structure can optimize the band properties of the material, thereby enhancing the performance of optoelectronic devices.

A comparison between structures with the same chemical composition ($CsInTe_2$) but different space groups highlight significant variations in bandgaps and physical properties. To gain microscopic insights, partial density of states (PDOS) was conducted for the five selected materials.

From the **Figure 8**, it can be observed that the chalcogen atoms (Se and Te) contribute dominantly to the outermost orbital states, while the A-site atoms, including K and Cs, contribute the least. In particular, Cs exhibits an almost negligible contribution. At the valence band maximum (VBM), the *p* orbitals of C-site atoms play a prominent role. Since the VBM is the starting point for electronic transitions in semiconductors and represents the location where photo-generated electron-hole pairs form, this indicates that the C-site atoms have a significant impact on hole effective mass and carrier mobility. Furthermore, in terms of bonding, the C-site atoms form strongly polar covalent bonds with the B-site atoms. Notably, in the high-energy region of the conduction band, different bonding characteristics can be observed. In $KGaSe_2$-$I4_1/amd$, the conduction band is primarily formed by the overlap between Ga *4p* orbitals and Se *4s* orbitals. In contrast, in $KInTe_2$-$R\bar{3}m$, and $CsInTe_2$-$R\bar{3}m$, the conduction band is mainly composed of the overlap between In *5p* orbitals and Te *5s* orbitals. This strong orbital overlap further confirms the covalent nature of the bonding between B-site and C-site atoms.

Accepted Article



Accepted Article

Figure 8. PDOS for six systems: (a) $KGaSe_2$-*I*4$_1$/*amd*, (b) $KGaTe_2$-*I*4$_1$/*amd*, (c) $KInSe_2$-*I*4$_1$/*amd*, (d) $KInTe_2$-*R*$\bar{3}$*m*, and (e) $CsInTe_2$-*R*$\bar{3}$*m*. In the calculations, the Fermi level (E_F) is set to 0 eV. The abscissa represents the energy relative to the Fermi level (unit: eV), and the ordinate denotes the DOS (unit: states/eV). Different lines in the figure correspond to the orbital contributions of different elements (e.g., Se 4p, Ga 4p and so on).

To address the systematic band gap underestimation of the PBE functional, we further validated the band gaps of the five final candidates using the HSE06 hybrid functional (a standard approach for accurate semiconductor band gap prediction). The results are summarized in **Table S7**: $KGaSe_2$- *I4₁/amd* (1.59 eV), $KGaTe_2$-*I4₁/amd* (1.08 eV), $KInSe_2$-*I4₁/amd* (1.55 eV), $KInTe_2$- *R3̄m* (0.92 eV), and $CsInTe_2$- *R3̄m* (1.03 eV). Importantly, HSE06 calculations confirm that the bandgap characteristics of all five candidates remain unchanged compared to the PBE results, and all HSE06-corrected band gaps fall within or near the SWIR-relevant range (0.4–1.55 eV, 0.8–3.0 μm), confirming the suitability of these candidates for SWIR detection. Specifically, $KGaTe_2$-*I4₁/amd*, $KInTe_2$-*R3̄m*, and $CsInTe_2$-*R3̄m* are within the conventional SWIR range (0.92–1.08 eV), while $KGaSe_2$- *I4₁/amd* and $KInSe_2$-*I4₁/amd* lie in the NIR-SWIR transition region (~0.78–0.80 μm), which is widely included in broadband SWIR detection systems[72]. As a representative example, the HSE06 band structure of $KGaSe_2$ is shown in **Fig. S5**, which explicitly demonstrates that the direct bandgap character at the Γ point is preserved at the hybrid-functional level. The remaining four candidates exhibit the same qualitative behavior in their HSE06 band-edge characteristics; therefore, only one representative band structure is presented for clarity.

To ensure that the 5 candidates identified are dynamically stable and are viable to be experimentally synthesized, we also calculate AIMD simulations specifically, as shown in **Figure 4(b)**. The energy fluctuations of the material remained within a narrow range, indicating its thermodynamic stability at 300 K. **Figure 9(a)** shows that the formation energies of all candidate materials are negative, indicating their thermodynamic stability; the AIMD simulation results in **Figure 9(b)** demonstrate small energy fluctuations of the materials at 300 K, further verifying their thermal stability. The results from formation energy analysis and molecular dynamics simulations further confirm that this material possesses excellent stability and manufacturability.

Accepted Article

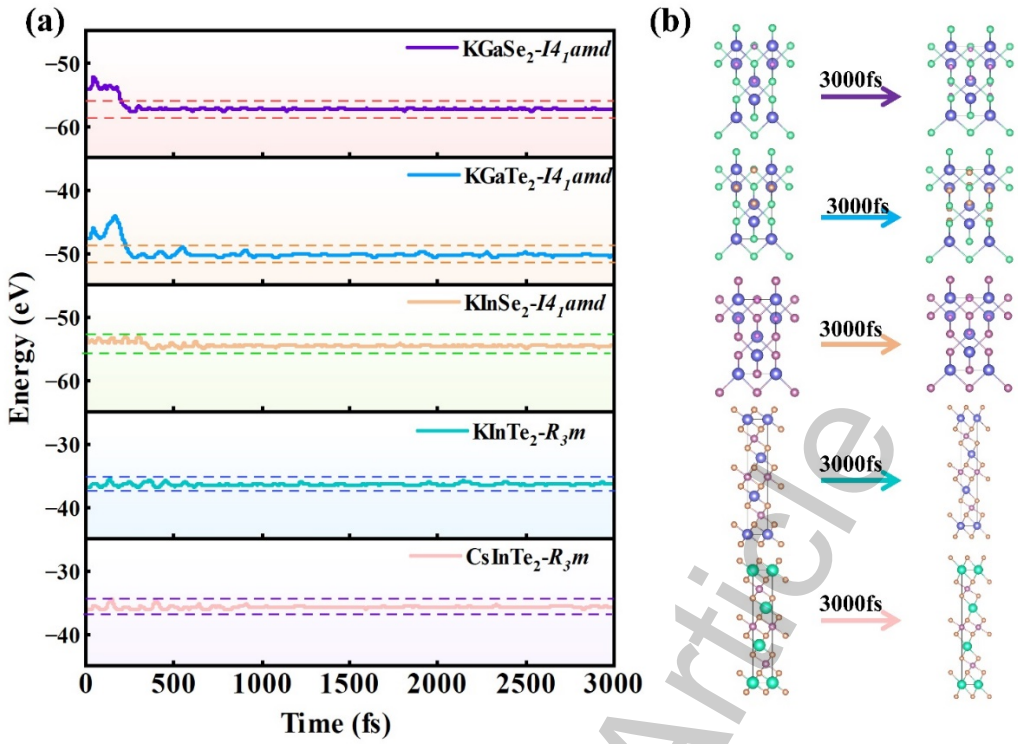


Figure 9. (a) Total energy evolution as a function of simulation time obtained from AIMD simulations performed at finite temperature for the corresponding candidate structures. (b) Representative atomic configurations of the candidate structures before and after 3000 fs of AIMD simulation. The AIMD simulations are used to assess short-time local structural robustness rather than to establish definitive thermodynamic stability.

Note that PBE was used only for initial high-throughput screening (rapidly excluding metallic/wide-gap systems) and relative optical property analysis (reliable for trend prediction[73], while the final candidate validation relies on HSE06 band gaps, AIMD-based dynamical stability, and elastic constant-verified mechanical stability.

CONCLUSIONS

In summary, this work establishes an integrated HTS framework that synergizes ML with DFT to expedite the discovery of SWIR detection materials. By leveraging a curated database of I–X–VI chalcogenides and optimizing ML regression models, the framework efficiently identifies promising candidates with tailored bandgaps and desirable electronic properties, with XGBR proving to be the most robust predictor. DFT validation confirms the reliability of the screening pipeline and highlights $KGaSe_2$ as a particularly promising candidate due to its optimal bandgap and superior carrier transport characteristics. This full-chain design paradigm offers a scalable and interpretable alternative to conventional

Accepted Article

discovery methods. Future research will expand this framework to broader chemical spaces and advanced ML architectures, facilitating the rapid development of functional materials for next-generation optoelectronic devices.

DECLARATIONS

Acknowledgements

We are thankful for the helpful discussion with the computational support from the Beijing Computational Science Research Center (CSRC).

Authors' contributions

Xiaoning Guan and Yanan Zhang contributed equally to this work.

Xiaoning Guan: Validation, Supervision, Methodology, Funding acquisition, Data curation, Conceptualization.

Yanan Zhang: Writing – review & editing, Writing – original draft, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization.

Suwen Han: Writing – review & editing, Writing – original draft, Investigation, Data curation.

Chunlian Xiong: Writing – review & editing, Data curation.

Yuqing Yang: Writing – review & editing.

Changcheng Chen: Writing – review & editing, Methodology, Formal analysis, Data curation.

Fan Zhang: Validation, Methodology, Formal analysis, Conceptualization.

Yanchao Zhang: Validation, Methodology, Formal analysis, Conceptualization.

Huachuan Gao: Validation, Methodology, Formal analysis, Conceptualization.

Feng Zhou: Writing – review & editing, Supervision, Methodology.

Pengfei Guan: Writing – review & editing, Supervision, Methodology.

Pengfei Lu: Writing – review & editing, Supervision, Methodology.

Availability of data and materials

The original contributions presented in this study are included in the supplementary material. Further inquiries can be directed to the corresponding authors.

Financial support and sponsorship

This work was supported financially by the National Key Research and Development

Accepted Article

Program Projects of China (No. 2022YFB3204203) and Infrared Detection National Key Laboratory Open Project (No. IRDT-24-02).

Conflicts of interest

Fan Zhang, Yanchao Zhang, and Huachuan Gao are affiliated with Ningbo Weiyuan Optoelectronic Research Institute Co., Ltd. The other authors declare that they have no competing interests.

Ethical approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Copyright

© The Author(s) 2026.

REFERENCES

1. Cao, F.; Liu, L.; Li, L. Short-wave infrared photodetector. Mater. Today 2023, 62, 327-349. DOI:10.1016/j.mattod.2022.11.003

2. Fu, L.; He, Y.; Zheng, J.; Hu, Y.; Xue, J.; Li, S. et al. $TexSe_{1-x}$ photodiode shortwave infrared detection and imaging. Adv. Mater. 2023, 35(24), 2211522. DOI:10.1002/adma.202211522

3. Amani, M.; Tan, C.; Zhang, G.; Zhao, C.; Bullock, J.; Song, X. et al. Solution-synthesized high-mobility tellurium nanoflakes for short-wave infrared photodetectors. ACS Nano 2018, 12(7), 7253–7263. DOI:10.1021/acsnano.8b03424

4. Guan, X.; Yu, X.; Periyanagounder, D.; Benzigar, M. R.; Huang, J. K.; Lin, C. H. et al. Recent progress in short-to long-wave infrared photodetection using 2D materials and heterostructures. Adv. Opt. Mater. 2021, 9(4), 2001708. DOI:10.1002/adom.202001708

5. Zulfiqar, M. W.; Nisar, S.; Dastgeer, G. et al. 2D material-based infrared photodetectors: recent progress, challenges, and perspectives. Nanoscale 2025, 17(31), 17881-17918. DOI:10.1039/d5nr01920f

6. Ji, X.; Lu, C.; Cheng, L. Recent advances in two-dimensional materials in infrared photodetectors (invited). Infrared Laser Eng. 2022, 51(7). DOI:10.3788/irla20220065

Accepted Article

7. Wu, Z.; Ou, Y.; Cai, M.; Wang, Y.; Tang, R.; Xia, Y. Short-wave infrared photodetectors and imaging sensors based on lead chalcogenide colloidal quantum dots. Adv. Opt. Mater. 2022, 11(1). DOI:10.1002/adom.202201577

8. Schmit, J. L. Growth, properties and applications of HgCdTe. J. Cryst. Growth 1983, 65(1), 249–261. DOI:10.1016/0022-0248(83)90057-x

9. Kirklin, S.; Saal, J.; Meredig, B. et al. The Open Quantum Materials Database (OQMD): assessing the accuracy of DFT formation energies. npj Comput. Mater. 2015, 1, 15010. DOI:https://doi.org/10.1038/npjcompumats.2015.10

10. Curtarolo, S.; Setyawan, W. et al. AFLOWLIB.ORG: A distributed materials properties repository from high-throughput ab initio calculations. Comput. Mater. Sci. 2012, 58, 227–235. DOI:10.1016/j.commatsci.2012.02.002

11. Jain, A.; Voznyy, O.; Sargent, E. H. High-throughput screening of lead-free perovskite-like materials for optoelectronic applications. J. Phys. Chem. C 2017, 121(13), 7183–7187. DOI:10.1021/acs.jpcc.7b02221

12. Li, Y.; Yang, K. High-throughput computational design of organic–inorganic hybrid halide semiconductors beyond perovskites for optoelectronics. Energy Environ. Sci. 2019, 12(7), 2233–2243. DOI:10.1039/c9ee01371g

13. Jiang, X.; Yin, W. High-throughput computational screening of oxide double perovskites for optoelectronic and photocatalysis applications. J. Energy Chem. 2021, 57, 351–358. DOI:10.1016/j.jechem.2020.09.014

14. Wang, V.; Tang, G.; Liu, Y.; Wang, R.; Mizuseki, H.; Kawazoe, Y.; Nara, J.; Geng, W. T. High-throughput computational screening of two-dimensional semiconductors. J. Phys. Chem. Lett. 2022, 13(50), 11581–11594. DOI:10.1021/acs.jpclett.2c02972

15. Jacobs, R.; Liu, J.; Abernathy, H.; Morgan, D. Machine learning design of perovskite catalytic properties. Adv. Energy Mater. 2024, 14(12), 2303684. DOI:10.1002/aenm.202303684

16. Ding, R.; Chen, J.; Chen, Y.; Liu, J.; Bando, Y.; Wang, X. Unlocking the potential: machine learning applications in electrocatalyst design for electrochemical hydrogen energy transformation. Chem. Soc. Rev. 2024, 53(23), 11390–11461. DOI:10.1039/d4cs00844h

17. Lu, B.; Xia, Y.; Ren, Y.; Xie, M.; Zhou, L.; Vinai, G.; Morton, S. A.; Wee, A. T. S.; van der Wiel, W. G.; Zhang, W.; Wong, P. K. J. When machine learning meets 2D materials: A review. Adv. Sci. 2024, 11(13), 2305277. DOI:10.1002/advs.202305277

18. Tameh, M. S.; Coropceanu, V.; Coen, A. G.; Pyun, J.; Lichtenberger, D. L.; Brédas, J.

Accepted Article

High-throughput computational screening of hydrocarbon molecules for long-wavelength infrared imaging. ACS Mater. Lett. 2024, 6(9), 4371–4378. DOI:10.1021/acsmaterialslett.4c01037.s001

19. Sun, S.; Hartono, N. T. P.; Ren, Z. D.; Oviedo, F.; Buscemi, A. M.; Layurova, M.; Chen, D. X.; Ogunfunmi, T.; Thapa, J.; Ramasamy, S. Accelerated development of perovskite-inspired materials via high-throughput synthesis and machine-learning diagnosis. Joule 2019, 3(6), 1437–1451. DOI:10.1016/j.joule.2019.05.014

20. Wang, R.; Yu, H.; Zhong, Y.; Xiang, H. Identifying direct bandgap silicon structures with high-throughput search and machine learning methods. J. Phys. Chem. C 2024, 128(30), 12677–12685. DOI:10.1021/acs.jpcc.4c02967.s001

21. Hong, W.; Yuan, J.; Mao, Y. Janus halogenated silicene for photocatalytic and photovoltaic applications: a machine learning combined with first-principles study. J. Phys. Chem. C 2024, 128(34), 14282–14293. DOI:10.1021/acs.jpcc.4c02939.s001

22. Sa, B.; Hu, R.; Zheng, Z.; Xiong, R.; Zhang, Y.; Wen, C.; Zhou, J.; Sun, Z. High-throughput computational screening and machine learning modeling of Janus 2D III–VI van der Waals heterostructures for solar energy applications. Chem. Mater. 2022, 34(15), 6687–6701. DOI:10.1021/acs.chemmater.2c00226

23. Li, Y.; Yang, J.; Zhao, R.; Zhang, Y.; Wang, X.; He, X.; Fu, Y.; Zhang, L. Design of organic–inorganic hybrid heterostructured semiconductors via high-throughput materials screening for optoelectronic applications. J. Am. Chem. Soc. 2022, 144(36), 16656–16666. DOI:10.1021/jacs.2c07434

24. Hu, W.; Zhang, L.; Pan, Z. Designing two-dimensional halide perovskites based on high-throughput calculations and machine learning. ACS Appl. Mater. Interfaces 2022, 14(18), 21596–21604. DOI:10.1021/acsami.2c00564

25. Ma, X.; Lewis, J. P.; Yan, Q.; Su, G. Accelerated discovery of two-dimensional optoelectronic octahedral oxyhalides via high-throughput ab initio calculations and machine learning. J. Phys. Chem. Lett. 2019, 10(21), 6734–6740. DOI:10.1021/acs.jpclett.9b02420.s001

26. Wu, Q.; Kang, L.; Lin, Z. A machine learning study on high thermal conductivity assisted to discover chalcogenides with balanced infrared nonlinear optical performance. Adv. Mater. 2023, 36(6). DOI:10.1002/adma.202309675

27. Shen, Y.; Wang, X.; Chen, X.; Wang, Y.; Zhang, J. Accelerated screening of ternary chalcogenides for high-performance optoelectronic materials. npj Comput. Mater. 2023, 9(1), 124. URL: https://arxiv.org/abs/2305.02634

Accepted Article

28. Teng, H.-Y.; Yang, J.-X. Optical Response Modulation of AgBiS2 Through Order-Disordered Transition. Technol. Dev. Chem. Ind. 2023, 52(8), 1-4+13. URL: https://navi.cnki.net/knavi/JournalDetail?pcode=CJFD&pykm=HGXB

29. Ruan, C.; Yang, S.-Y.; Gu, Z.-C.; Sun, J.-Z.; Zhang, Y.; Zhang, C.; Li, Y.; Wang, Y.; Wang, L.; Liu, Q. Ideal Weyl Semimetal State in ABC2-Type Chalcopyrite Compounds. Phys. Rev. Lett. 2016, 116(22), 226801. DOI:10.1103/PhysRevLett.116.226801

30. Daoud, S. Proposal of New Stable ABC2 Type Ternary Semiconductor Pnictides K3Cu3P2 and K3Ni3P2: First-Principles Calculations and Prospects for Thermophysical and Optoelectronic Applications. Eng. Rep. 2020, 2(12), e5992. DOI:10.1002/eng2.12222

31. PubMed. Efficient Large-Area (81 cm²) Ternary Copper Halides Light-Emitting Diodes with External Quantum Efficiency Exceeding 13% via Host-Guest Strategy. Natl. Center Biotechnol. Inf. 2025, PMID: 38693828. URL: https://www.ncbi.nlm.nih.gov/pubmed/38693828

32. Kim, K.; Ward, L.; He, J.; Krishna, A.; Agrawal, A.; Wolverton, C. Machine-learning-accelerated high-throughput materials screening: Discovery of novel quaternary Heusler compounds. Phys. Rev. Mater. 2018, 2, 123801. DOI:10.1103/physrevmaterials.2.123801

33. Wada, T. CuInSe2 and related I–III–VI2 chalcopyrite compounds for photovoltaic application. Jpn. J. Appl. Phys. 2021, 60(8), 080101. DOI:10.35848/1347-4065/ac08ac

34. Zhang, T.; Yang, Y.; Liu, D.; et al. High efficiency solution-processed thin-film Cu(In,Ga)(Se,S)2 solar cells. Energy Environ. Sci. 2016, 9(12), 3674-3681. DOI:10.1039/C6EE02352E

35. García‐Hemme, E.; Caudevilla, D.; Algaidy, S.; et al. On the optoelectronic mechanisms ruling Ti‐hyperdoped Si photodiodes. Adv. Electron. Mater. 2022, 8(2), 2100788. DOI:10.1002/aelm.202100788

36. Obulesu, O.; Mahendra, M.; Thrillokreddy, M. Machine learning techniques and tools: a survey. In 2018 International Conference on Inventive Research in Computing Applications (ICIRCA); IEEE, 2018; pp 605–611. DOI:10.1109/ICIRCA.2018.8597302

37. El Mrabet, M. A.; El Makkaoui, K.; Faize, A. Supervised machine learning: a survey. In 2021 4th International Conference on Advanced Communication Technologies and Networking (CommNet); IEEE, 2021; pp 1–10. DOI:10.1109/CommNet52204.2021.9641998

38. Smajić, A.; Grandits, M.; Ecker, G. F. Using Jupyter Notebooks for re-training

Accepted Article

machine learning models. J. Cheminf. 2022, 14(1), 54. DOI:10.1186/s13321-022-00635-2

39. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; Vanderplas, J.; Passos, A.; Cournapeau, D.; Brucher, M.; Perrot, M.; Duchesnay, É. Scikit-learn: machine learning in Python. J. Mach. Learn. Res. 2011, 12, 2825–2830. DOI:10.5281/zenodo.13749328

40. Kresse, G.; Furthmüller, J. Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set. Phys. Rev. B 1996, 54(16), 11169–11186. DOI:10.1103/PhysRevB.54.11169

41. Segall, M. D.; Lindan, P. J. D.; Probert, M. J.; Pickard, C. J.; Hasnip, P. J.; Clark, S. J.; Payne, M. C. First-principles simulation: ideas, illustrations and the CASTEP code. J. Phys.: Condens. Matter 2002, 14(11), 2717–2744. DOI:10.1088/0953-8984/14/11/301

42. Setyawan, W.; Curtarolo, S. High-throughput electronic band structure calculations: challenges and tools. Comput. Mater. Sci. 2010, 49(2), 299–312. DOI:10.1016/j.commatsci.2010.05.010

43. Huck, P.; Jain, A.; Gunter, D.; Winston, D.; Persson, K. A community contribution framework for sharing materials data with materials project. In 2015 IEEE 11th International Conference on e-Science; IEEE, 2015; pp 535–541. DOI:10.1109/eScience.2015.75

44. Cheng, J.; et al. Growth and characterization of AgBiS2 bulk single crystals for near-infrared detectors. Cryst. Growth Des. 2024, 24(10), 5364–5372. DOI:10.1021/acs.cgd.4c00970

45. Woods-Robinson, R.; Han, Y.; Zhang, H.; Ablekim, T.; Khan, I.; Persson, K. A.; Zakutayev, A. Wide band gap chalcogenide semiconductors. Chem. Rev. 2020, 120(9), 4007–4055. DOI:10.1021/acs.chemrev.9b00600

46. Fatmi, M.; Bouferrache, K.; Ghebouli, M. A.; Ghebouli, B.; Alomairy, S.; Alanazi, F. K. Investigation of structural elastic electronic optical and thermoelectric properties of LiInS2 and LiInTe2 for optoelectronic and energy conversion. Sci. Rep. 2025, 15, 27859. DOI:10.1038/s41598-025-97851-z

47. Khalfin, S.; Bekenstein, Y. Advances in lead-free double perovskite nanocrystals, engineering band-gaps and enhancing stability through composition tunability. Nanoscale 2019, 11(18), 8665–8679. DOI:10.1039/c9nr01031a

48. Miah, M. H.; Khandaker, M. U.; Rahman, M. B.; et al. Band gap tuning of perovskite solar cells for enhancing the efficiency and stability: issues and prospects. RSC Adv. 2024,

Accepted Article

14(23), 15876–15906. DOI:10.1039/d4ra01640h

49. Yang, J. N.; Chen, T.; Ge, J.; et al. High color purity and efficient green light-emitting diode using perovskite nanocrystals with the size overly exceeding Bohr exciton diameter. J. Am. Chem. Soc. 2021, 143(47), 19928–19937. DOI:10.1021/jacs.1c09948

50. Lin, K.; Xing, J.; Quan, L. N.; et al. Perovskite light-emitting diodes with external quantum efficiency exceeding 20 per cent. Nature 2018, 562(7726), 245–248. DOI:10.1038/s41586-018-0575-3

51. Yang, X.; Zhang, X.; Deng, J.; et al. Efficient green light-emitting diodes based on quasi-two-dimensional composition and phase engineered perovskite with surface passivation. Nat. Commun. 2018, 9(1), 570. DOI:10.1038/s41467-018-02978-7

52. Hoang, K.; Mahanti, S. D. Atomic and electronic structures of I–V–VI2 ternary chalcogenides. J. Sci. Adv. Mater. Devices 2016, 1(1), 51–56. DOI:10.1016/j.jsamd.2016.04.004

53. Tan, C.; et al. Evaporated Se_xTe_{1-x} thin films with tunable bandgaps for short-wave infrared photodetectors. Adv. Mater. 2020, 32(38), 2001329. DOI:10.1002/adma.202001329

54. Wu, Z.; et al. Ultranarrow-bandgap small-molecule acceptor enables sensitive SWIR detection and dynamic upconversion imaging. Sci. Adv. 2024, 10(23), eadm9631. DOI:10.1126/sciadv.adm9631

55. Bartel, C. J.; Trewartha, A.; Wang, Q.; et al. A critical examination of compound stability predictions from machine-learned formation energies. npj Comput. Mater. 2020, 6(1), 97. DOI:10.1038/s41524-020-00362-y

56. Wang, Q.; Xu, B.; Sun, J.; et al. Direct band gap silicon allotropes. J. Am. Chem. Soc. 2014, 136(28), 9826–9829. DOI:10.1021/ja5035792

57. Agrawal, G. P.; Dutta, N. K. Recombination mechanisms in semiconductors. In Semiconductor Lasers; Agrawal, G. P., Dutta, N. K., Eds.; Springer US: Boston, MA, 1993; pp 74–146. DOI:10.1007/978-1-4613-0481-4_3

58. Wang, S.; Zhang, Z.; Chen, C.; et al. Achieve near-infrared absorption in Cs3Sb2Br9 through 3d orbital energy level splitting to construct high-performance intermediate band solar cells. Chem. Eng. J. 2024, 499, 158638. DOI:10.1016/j.cej.2024.158638

59. Wang, S.; Chen, C.; Shi, S.; et al. Flexibility potential of Cs2BX6 (B = Hf, Sn, Pt, Zr, Ti; X = I, Br, Cl) with application in photovoltaic devices and radiation detectors. J. Energy Chem. 2024, 95, 271–287. DOI:10.1016/j.jechem.2024.03.036

60. Fadaly, E. M. T.; Dijkstra, A.; Suckert, J. R.; et al. Direct-bandgap emission from

Accepted Article

hexagonal Ge and SiGe alloys. Nature 2020, 580(7802), 205–209. DOI:10.1038/s41586-020-2150-y

61. Sarritzu, V.; Sestu, N.; Marongiu, D.; et al. Direct or indirect bandgap in hybrid lead halide perovskites. Adv. Opt. Mater. 2018, 6(10), 1701254. DOI:10.1002/adom.201701254

62. Khandy, S. A.; Yousuf, S.; Gupta, D. C. Structural, Magneto‐electronic, Mechanical, and Thermophysical Properties of Double Perovskite Ba2ZnReO6. Phys. Status Solidi B 2019, 256(10), 1800625. DOI:10.1002/pssb.201800625

63. Yang, K.; Wan, R.; Zhang, Z.; et al. First-principle investigation on the thermoelectric and electronic properties of HfCoX (X = As, Sb, Bi) half-Heusler compounds. J. Solid State Chem. 2022, 314, 123386. DOI:10.1016/j.jssc.2022.123386

64. Poncé, S.; Li, W.; Reichardt, S.; Marom, N.; Arias, T. A. First-principles calculations of charge carrier mobility and conductivity in bulk semiconductors and two-dimensional materials. Rep. Prog. Phys. 2020, 83(3), 036501. DOI:10.1088/1361-6633/ab6a43

65. Su, L.; Shi, H.; Wang, S.; et al. Enhancing carrier mobility and Seebeck coefficient by modifying scattering factor. Adv. Energy Mater. 2023, 13(18), 2300312. DOI:10.1002/aenm.202300312

66. Qian, X.; Zhang, X.; Guo, H.; et al. Enhancing thermoelectric performance of n-type AgBi3S5 through synergistically optimizing the effective mass and carrier mobility. J. Materiomics 2023, 9(5), 874–881. DOI:10.1016/j.jmat.2023.02.010

67. Xiao, Y.; Liu, W.; Zhang, Y.; et al. Rationally optimized carrier effective mass and carrier density leads to high average ZT value in n-type PbSe. J. Mater. Chem. A 2021, 9(40), 23011–23018. DOI:10.1039/D1TA07039H

68. Hautier, G.; Miglio, A.; Waroquiers, D.; et al. How does chemistry influence electron effective mass in oxides? A high-throughput computational analysis. Chem. Mater. 2014, 26(19), 5447-5458. DOI:10.1021/cm404079a

69. Sun, Q.; Jiang, T.; Ou, Q.; et al. Influence of intermolecular packing on light emitting efficiency and carrier‐mobility of organic semiconductors: theoretical descriptor for molecular design. Adv. Opt. Mater. 2023, 11(5), 2202621. DOI:10.1002/adom.202202621

70. Xie, M.; Li, T.; Liu, C.; et al. Realizing highly efficient blue electrofluorescence by optimized hybridized local and charge transfer state and balanced carrier mobilities. Chem. Eng. J. 2023, 472, 144950. DOI:10.1016/j.cej.2023.144950

Accepted Article

71. Ye, G. D.; Ding, R.; Li, S. H.; et al. Single-crystalline hole-transporting layers for efficient and stable organic light-emitting devices. Light Sci. Appl. 2024, 13(1), 136. DOI:10.1038/s41377-024-01484-4
72. Liu, X.; et al. Broadband Short-Wave Infrared Photodetectors: Materials, Mechanisms, and Applications. Adv. Opt. Mater. 2024, 12, 2302215. DOI:10.1002/adom.202302215
73. Da Silva, E. K.; et al. Selected I-III-VI2 Semiconductors: Synthesis, Properties and Applications in Photovoltaic Cells. Int. J. Mol. Sci. 2025, 26(14), 4344. DOI:10.3390/ijms26144344

Accepted Article

Supplementary Materials

The supporting information includes the following content, all available for free: Specific names and abbreviations of 21 characteristic variables. Training results of Random Forest Regression (RFR), Support Vector Regression (SVR), Gradient Boosting Regression (GBR), Decision Tree (DT), and Extreme Gradient Boosting Regression (XGBR) models. Best Model XGBoost K-fold Cross validation Results. XGBoost model predicts bandgap values of 4 infrared materials. Predict the bandgap of materials between 0 eV and 7 eV.

Accepted Article

Table. S1 Convergence of total energy and PBE band gap with respect to k-point sampling for representative crystal systems.

Material	Space group	k-point grid	Total energy (eV)	ΔE* (meV/atom)	E_{g_PBE}(eV)	ΔEg(eV)	k-point grid	Total energy (eV)
$KGaSe_2$	$I4_1/amd$	2 × 2 × 1	−54.0375	+12.1	0.7490	−0.0448	2 × 2 × 1	−54.0375
		4 × 4 × 2	**−54.2312**	**0.0**	**0.7986**	**+0.0048**	**4 × 4 × 2**	**−54.2312**
		6 × 6 × 3	−54.2289	+0.14	0.7938	0.0	6 × 6 × 3	−54.2289
$CsInTe_2$	$R\bar{3}m$	7 × 7 × 1	−36.1596	+0.94	0.5158	−0.0105	7 × 7 × 1	−36.1596
		9 × 9 × 1	**−36.1670**	**+0.09**	**0.5251**	**−0.0012**	**9 × 9 × 1**	**−36.1670**
		11 × 11 × 1	−36.1683	0.0	0.5263	0.0	11 × 11 × 1	−36.1683

* ΔE is referenced to the most converged k-point grid for each material and normalized per atom.

To ensure the reliability of the calculations, we verified the convergence of k-point grids. For the $I4_1/amd$ phase, the selected k-point grid is 4×4×2, while for the $R\bar{3}m$ phase it is 9×9×1. The total energy difference between these selected grids and denser meshes is below 1 meV/atom, and the corresponding variation in PBE band gap is less than 0.005 eV—confirming excellent convergence of the calculation parameters. These k-point grids were therefore adopted consistently in all calculations presented in the manuscripts.

Accepted Article

Table. S2 k-point grids and corresponding reciprocal-space spacing (Δk) for representative crystal systems.

Material	Crystal system	Lattice parameters (Å)	k-point grid	$\Delta k_a/\Delta k\_b$ ($Å^{-1}$)	Δk_c ($Å^{-1}$)
$KGaSe_2$	Tetragonal ($I4_1/amd$)	a = 6.3954, c = 11.2442	4 × 4 × 2	0.034	0.036
$CsInTe_2$	Layered rhombohedral ($R\bar{3}m$)	a = 4.4910, c = 26.4392	9 × 9 × 1	0.035	0.237

To ensure the accuracy and comparability of computational results, we determined the k-point spacing via the formula

$$\Delta k_i = \frac{2\pi}{a_i \cdot n_i}.$$

For both target materials, we maintained a uniform reciprocal-space resolution of $\Delta k \approx 0.035$ $Å^{-1}$ in the primary periodic directions, thus guaranteeing a rigorous and fair comparison of their electronic properties. Regarding the layered $CsInTe_2$ structure, we only employed a single k point along the long c axis. This parameter setting follows the widely accepted standard protocol for layered materials, since the weak interlayer coupling does not require high-density k-point sampling in this specific direction.

Accepted Article

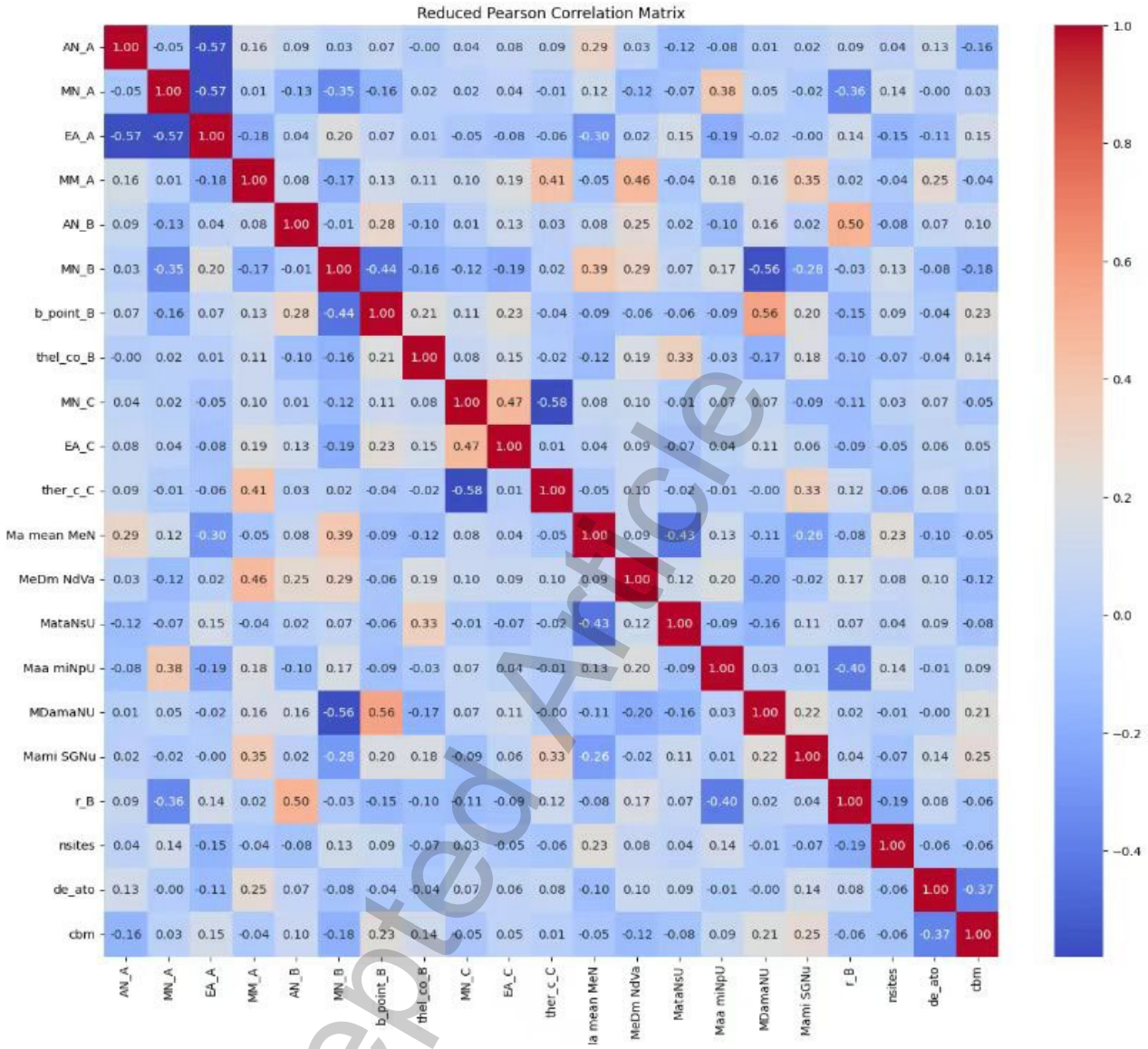


Fig. S1 ML model training and prediction results. Pearson correlation coefficient plot for the top 21 feature descriptors.

Accepted Article

Table. S3 Specific names and abbreviations of 21 characteristic variables.

Feature Variables	Explanation
AN_A	AN_A refers to the atomic number of A-site atoms in ternary compounds
MN_A	Relates to "Mean Neighbor A Site," representing the average properties of atoms neighboring the A site in a ternary compound
EA_A	electron affinity of A-site element
MM_A	MM_A refers to the Mendeleev number of A-site atoms in ternary compounds
AN_B	AN_B refers to the atomic number of B-site atoms in ternary compounds.
MN_B	Relates to "Mean Neighbor B Site," representing the average properties of atoms neighboring the B site in a ternary compound
b_point_B	Boiling point of B-site atoms
thel_co_B	Thermal conductivity of B-site atoms
MN_C	Mendeleev number
EA_C	electron affinity of C-site element
ther_c_C	Thermal conductivity of C-site atoms
Ma mean MeN	MagpieData mean Mendeleev Number of the system
MeDm NdVa	MagpieData maximum NdValence
MataNsU	MagpieData mode NsUnfilled
Maa miNpu	MagpieData maximum NpUnfilled
MDamaNu	MagpieData maximum NUnfilled
Mami SGNu	MagpieData minimum Space Group Number
r_B	r_B refers to the atomic radius of B-site atoms in ternary compounds
nsites	Number of atomic sites in the ternary compound system
de_ato	Atomic density
cbm	Conduction Band Minimum of the material

Accepted Article

Table. S4 Hyperparameter settings for the five base machine learning models.

ML models	Hyperparameters
RFR	bootstrap=True, max_depth=45, max_features= 'sqrt', min_samples_leaf=1, min_samples_split=4, n_estimators=900
GBR	learning_rate=0.05, max_depth=4, max_features='sqrt', min_samples_leaf=5, min_samples_split=10, n_estimators=300, subsample=0.8, andom_state=42
SVR	kernel='rbf', gamma= 0.1, epsilon=0.2, C=15
XGBR	colsample_bytree=0.6, gamma=0.1, learning_rate=0.1, max_depth=4, min_child_weight=3, n_estimators=400, subsample=0.8, reg_alpha=0.1, reg_lambda=1
DT	max_depth=15, min_samples_leaf=3, min_samples_split=3

Supplementary details on hyperparameter optimization of the XGBR model

To ensure the generalization performance of the XGBR model, its hyperparameters were optimized via five-fold cross-validation (n_splits=5, shuffle=True, random state=42) for global optimization. The final optimized hyperparameters of the XGBR model are as follows: learning rate=0.05, max depth=4, max features='sqrt', min samples leaf=5, min samples split=10, n estimators=300, subsample=0.80, random state=42.

Accepted Article

Table. S5 Table of band gaps and formation energies for 27 candidate materials

No.	Material Formula	Space groups	Formation energy(eV)	Bandgap(eV)
1	$LiGaTe_2$	*I4/mcm*	-0.60	0.68
2	$LiInTe_2$	*I4/mcm*	-0.64	0.44
3	$LiInSe_2$	*R3̄m*	-1.89	0.82
4	$RbAlTe_2$	*R3̄m*	-1.29	0.66
5	$RbGaTe_2$	*I4/mcm*	-0.66	0.88
6	$RbGaTe_2$	*I4₁/amd*	-0.36	0.49
7	$RbInTe_2$	*R3̄m*	-0.58	0.44
8	$KAlTe_2$	*R3̄m*	-1.31	0.63
9	$KAlTe_2$	*I4₁/amd*	-1.15	0.70
10	$KGaSe_2$	*R3̄m*	-0.71	0.80
11	$KGaSe_2$	*I4₁/amd*	-0.62	0.80
12	$KGaTe_2$	*I4/mcm*	-0.66	0.79
13	$KGaSe_2$	*I4₁/amd*	-0.39	0.80
14	$KInSe_2$	*I4₁/amd*	-0.71	0.58
15	$KInTe_2$	*R3̄m*	-0.59	0.41
16	$KInTe_2$	*Pna2₁*	-1.09	0.67
17	$CsAlTe_2$	*R3̄m*	-1.29	0.64
18	$CsAlTe_2$	*I4₁/amd*	-1.09	0.79
19	$CsGaSe_2$	*R3̄m*	-0.66	0.85
20	$CsGaTe_2$	*I4₁/amd*	-0.38	0.84
21	$CsInTe_2$	*R3̄m*	-0.59	0.49
22	$CsInTe_2$	*I4₁/amd*	-0.47	0.66
23	$NaAlTe_2$	*I4₁/amd*	-1.19	0.42
24	$NaGaSe_2$	*R3̄m*	-0.70	0.53
25	$NaGaSe_2$	*I4₁/amd*	-0.65	0.75
26	$NaInSe_2$	*I4₁/amd*	-0.76	0.87
27	$NaInTe_2$	*I4/mcm*	-0.11	0.72

The formation energy of each optimized structure was calculated based on total energies obtained from first-principles calculations. For a given compound, the formation energy is defined as:

$$E_{\text{form}} = E_{\text{tot}} - \sum_i n_i E_i,$$

where $E_{\text{tot}}$is the total energy of the fully relaxed compound with its optimized crystal structure, $n_i$is the number of atoms of element i in the chemical formula, and E_i is the total energy of element $i$in its most stable reference structure calculated using the same computational settings.

In this work, all structures were fully optimized prior to total energy evaluation. The elemental reference energies were obtained from the corresponding elemental phases, and all formation energies were derived consistently from the same first-principles dataset.

Accepted Article

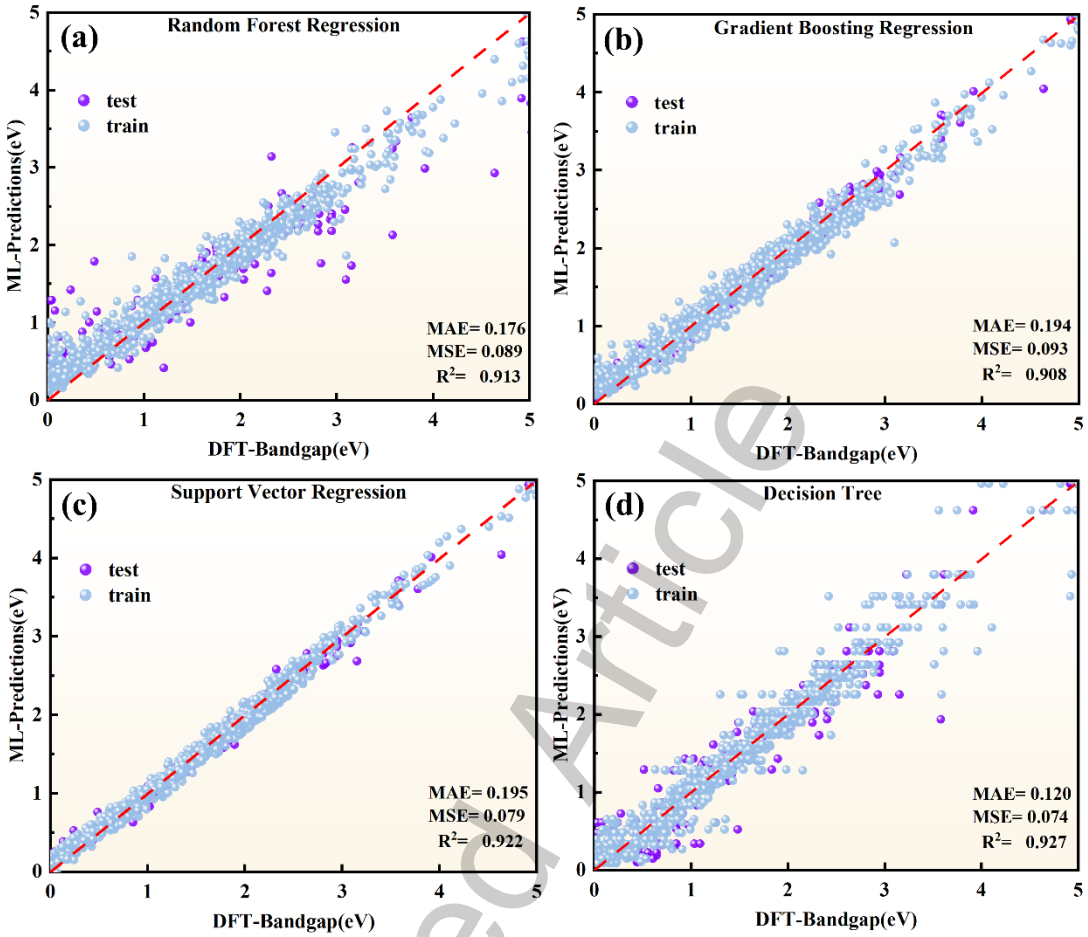


Fig. S2 Training results of Random Forest Regression (RFR), Support Vector Regression (SVR), Gradient Boosting Regression (GBR), and Decision Tree (DT)

Accepted Article

```
]: import xgboost as xgb

#xgboost = xgb.XGBRegressor(colsample_bytree=0.7, gamma =0.1, learning_rate=0.1, max_depth=3, min_child_weight=1, n_estimators=300, subsample
xgboost = xgb.XGBRegressor(colsample_bytree=0.6, gamma=0.1, learning_rate=0.05, max_depth=5,
                           min_child_weight=3, n_estimators=400, subsample=0.8, reg_alpha=0.2, reg_lambda=1)

xgboost.fit(X_train, y_train)


predictions_xgb = xgboost.predict(X_test)

r2_rf = r2_score(y_test, predictions_xgb)
mse = mean_squared_error(y_test, predictions_xgb)
mae = mean_absolute_error(y_test, predictions_xgb)


print('XGB Test MAE= %.3f' % mae)
print('XGB Test MSE= %.3f' % mse)
print('XGB Test R2= %.3f' % r2_rf)

plt.figure(figsize=(6, 6))
plt.scatter(y_test, predictions_xgb, color='blue')
plt.plot([y_test.min(), y_test.max()], [y_test.min(), y_test.max()], 'k--', lw=4)
plt.xlabel('Actual Values')
plt.ylabel('Predicted Values')
plt.title('Stacked Model Predicted vs Test Actual')
plt.show()
```

```
XGB Test MAE= 0.150
XGB Test MSE= 0.056
XGB Test R2= 0.945
```

```
import xgboost as xgb
from sklearn.model_selection import cross_val_score, KFold
from sklearn.metrics import mean_squared_error, mean_absolute_error, r2_score
import matplotlib.pyplot as plt


xgboost = xgb.XGBRegressor(
    colsample_bytree=0.6, gamma=0.1, learning_rate=0.05, max_depth=3,
    min_child_weight=5, n_estimators=300, subsample=0.8, reg_alpha=0.1, reg_lambda=1
)


kf = KFold(n_splits=10, shuffle=True, random_state=42)


cv_scores_mae = cross_val_score(xgboost, X_train, y_train, cv=kf, scoring='neg_mean_absolute_error')
cv_scores_mse = cross_val_score(xgboost, X_train, y_train, cv=kf, scoring='neg_mean_squared_error')
cv_scores_r2 = cross_val_score(xgboost, X_train, y_train, cv=kf, scoring='r2')


print("Cross-Validation MAE: ", -cv_scores_mae)
print("Average MAE: %.3f" % (-cv_scores_mae.mean()))
print("Cross-Validation MSE: ", -cv_scores_mse)
print("Average MSE: %.3f" % (-cv_scores_mse.mean()))
print("Cross-Validation R2: ", cv_scores_r2)
print("Average R2: %.3f" % cv_scores_r2.mean())

xgboost.fit(X_train, y_train)


predictions_xgb = xgboost.predict(X_test)


test_r2 = r2_score(y_test, predictions_xgb)
test_mse = mean_squared_error(y_test, predictions_xgb)
test_mae = mean_absolute_error(y_test, predictions_xgb)


print('XGB Test MAE: %.3f' % test_mae)
print('XGB Test MSE: %.3f' % test_mse)
print('XGB Test R2: %.3f' % test_r2)


plt.figure(figsize=(6, 6))
plt.scatter(y_test, predictions_xgb, color='blue')
plt.plot([y_test.min(), y_test.max()], [y_test.min(), y_test.max()], 'k--', lw=4)
plt.xlabel('Actual Values')
plt.ylabel('Predicted Values')
plt.title('XGB Predicted vs Test Actual')
plt.grid(True)
plt.show()
```

```
Cross-Validation MAE:  [0.13864205 0.12887342 0.19452578 0.13656334 0.15521336 0.14874709
 0.1598401  0.15408146 0.173895   0.19703285]
```

Accepted Article

Fig. S3 Best Model XGBOOST K-fold Cross validation Results.

Table. S6 95% Confidence Interval Results for XGBoost Model-Predicted Bandgaps

Sample_Index	True_Value	Predicted_Value	95%CI_Lower	95%CI_Upper	CI_Width
0	1.583	1.70599997	0	1.706	1.706
1	1.466	1.32700002	0	1.327	1.327
2	1.06	1.20399996	0	1.204	1.204
3	2.474	2.31599985	0	2.316	2.316
4	1.653	1.69400006	0	1.694	1.694
5	2.546	2.68600009	0	2.686	2.686
6	1.444	1.19000057	0	1.19	1.19
7	0.707	0.79299983	0	0.793	0.793
8	1.959	1.86800031	0	1.868	1.868
9	0.603	0.68500002	0	0.685	0.685
10	0.485	0.75	0	0.75	0.75
11	1.432	1.48599942	0	1.486	1.486
12	1.893	1.59700003	0	1.597	1.597
13	2.066	2.20399996	0	2.204	2.204
14	0.04	0.06899998	0	0.069	0.069
15	2.322	2.24699979	0	2.247	2.247
16	0.998	1.06299964	0	1.063	1.063
17	1.438	1.35699993	0	1.357	1.357
18	1.143	1.32200027	0	1.322	1.322
19	1.669	1.58099997	0	1.581	1.581

Accepted Article

20	2.605	2.647000074	0	2.647	2.647
21	0.029	0.708000004	0	0.708	0.708
22	0.117	0.222000003	0	0.222	0.222
23	2.18	2.165999889	0	2.166	2.166
24	1.929	1.930999994	0	1.931	1.931
25	2.805	2.319000006	0	2.319	2.319
26	0.898	0.772000015	0	0.772	0.772
27	1.108	1.133999944	0	1.134	1.134
28	1.938	2.059999943	0	2.06	2.06
29	1.189	1.225999951	0	1.226	1.226
30	1.686	1.585000038	0	1.585	1.585
31	0.872	1.182999969	0	1.183	1.183
32	1.032	0.657999992	0	0.658	0.658
33	0.64	0.301999986	0	0.302	0.302
34	2.413	2.361000061	0	2.361	2.361
35	0.04	0.685000002	0	0.685	0.685
36	2.522	2.420000076	0	2.42	2.42
37	3.084	2.625999928	0	2.626	2.626
38	1.172	1.225999951	0	1.226	1.226
39	0.832	0.759999999	0	0.76	0.76
40	2.288	2.710000038	0	2.71	2.71
41	0.51	0.818000019	0	0.818	0.818

Accepted Article

42	0.818	0.958999991	0	0.959	0.959
43	1.132	1.097000003	0	1.097	1.097
44	2.322	2.926000118	0	2.926	2.926
45	0.202	0.254000008	0	0.254	0.254
46	1.496	1.48300004	0	1.483	1.483
47	1.934	1.912999988	0	1.913	1.913
48	2.32	1.962000012	0	1.962	1.962
49	4.915	4.146999836	0	4.147	4.147
50	1.513	1.585000038	0	1.585	1.585
51	1.185	1.223999977	0	1.224	1.224
52	0.685	0.69599998	0	0.696	0.696
53	1.303	1.215999961	0	1.216	1.216
54	1.117	1.332000017	0	1.332	1.332
55	2.654	2.642999887	0	2.643	2.643
56	0.438	0.549000025	0	0.549	0.549
57	3.577	3.535000086	0	3.535	3.535
58	2.745	2.686000109	0	2.686	2.686
59	1.387	1.421000004	0	1.421	1.421
60	1.927	1.957000017	0	1.957	1.957
61	0.431	0.469999999	0	0.47	0.47
62	2.158	2.177999973	0	2.178	2.178
63	1.744	1.909999967	0	1.91	1.91

Accepted Article

64	1.284	1.182999969	0	1.183	1.183
65	0.549	0.685000002	0	0.685	0.685
66	0.694	0.75	0	0.75	0.75
67	0.604	0.469999999	0	0.47	0.47
68	0.934	0.966000021	0	0.966	0.966
69	1.833	1.133999944	0	1.134	1.134
70	1.292	1.327000022	0	1.327	1.327
71	2.634	2.686000109	0	2.686	2.686
72	0.453	0.428000003	0	0.428	0.428
73	2.03	1.868000031	0	1.868	1.868
74	2.816	2.482000113	0	2.482	2.482
75	2.383	2.480000019	0	2.48	2.48
76	2.698	2.674999952	0	2.675	2.675
77	1.724	1.725999951	0	1.726	1.726
78	2.929	2.710000038	0	2.71	2.71
79	0.846	0.830999997	0	0.831	0.831
80	0.285	0.291999996	0	0.292	0.292
81	0.076	0.131999999	0	0.132	0.132
82	0.41	0.216999993	0	0.217	0.217
83	2.244	2.073999882	0	2.074	2.074
84	1.416	1.297999978	0	1.298	1.298
85	1.479	1.085999966	0	1.086	1.086

Accepted Article

86	2.368	2.52699995	0	2.527	2.527
87	5.733	5.56500057	0	5.565	5.565
88	1.712	1.84099961	0	1.841	1.841
89	1.21	1.19099985	0	1.191	1.191
90	1.133	0.96600021	0	0.966	0.966
91	1.369	1.42100004	0	1.421	1.421
92	0.554	0.36300005	0	0.363	0.363
93	2.948	2.90599999	0	2.906	2.906
94	1.949	1.94200031	0	1.942	1.942
95	3.914	3.57200027	0	3.572	3.572
96	0.437	0.48100006	0	0.481	0.481
97	1.337	1.12800021	0	1.128	1.128
98	2.264	2.48099947	0	2.481	2.481
99	3.578	2.64299987	0	2.643	2.643
100	1.926	1.82200027	0	1.822	1.822
101	2.085	1.96200012	0	1.962	1.962
102	0.911	0.90600018	0	0.906	0.906
103	2.252	2.17000076	0	2.17	2.17
104	1.448	1.31099999	0	1.311	1.311
105	1.211	1.29799978	0	1.298	1.298
106	1.05	1.06799959	0	1.068	1.068
107	2.231	2.22099956	0	2.221	2.221

Accepted Article

108	1.533	1.485999942	0	1.486	1.486
109	1.252	1.294999957	0	1.295	1.295
110	5.011	4.146999836	0	4.147	4.147
111	1.79	1.822000027	0	1.822	1.822
112	1.663	1.735000014	0	1.735	1.735
113	0.987	0.84799999	0	0.848	0.848
114	2.944	2.772000074	0	2.772	2.772
115	1.316	1.203999996	0	1.204	1.204
116	0.359	0.703000009	0	0.703	0.703
117	2.406	2.480000019	0	2.48	2.48
118	0.846	0.772000015	0	0.772	0.772
119	0.003	0.43900001	0	0.439	0.439
120	1.718	1.718000054	0	1.718	1.718
121	3.612	3.862999916	0	3.863	3.863
122	2.13	2.211999893	0	2.212	2.212
123	1.473	1.613999963	0	1.614	1.614
124	1.489	1.694000006	0	1.694	1.694
125	2.008	2.073999882	0	2.074	2.074
126	2.215	2.246999979	0	2.247	2.247
127	0.258	0.48300001	0	0.483	0.483
128	1.596	1.705999997	0	1.706	1.706
129	0.657	0.600000024	0	0.6	0.6

Accepted Article

130	0.002	0.068000004	0	0.068	0.068
131	1.079	1.177000046	0	1.177	1.177
132	2.15	2.073999882	0	2.074	2.074
133	1.113	1.182999969	0	1.183	1.183
134	0.238	0.75	0	0.75	0.75
135	2.276	2.420000076	0	2.42	2.42
136	2.276	2.395999908	0	2.396	2.396
137	0.13	0.268999994	0	0.269	0.269
138	3.225	3.115999937	0	3.116	3.116
139	2.277	1.94599998	0	1.946	1.946
140	2.023	2.084000111	0	2.084	2.084
141	0.637	0.685000002	0	0.685	0.685
142	0.843	0.958999991	0	0.959	0.959
143	0.971	1.11500001	0	1.115	1.115
144	2.342	2.36500001	0	2.365	2.365
145	2.592	2.480000019	0	2.48	2.48
146	0.868	0.708000004	0	0.708	0.708
147	1.752	1.519999981	0	1.52	1.52
148	1.024	1.11500001	0	1.115	1.115
149	2.063	1.993999958	0	1.994	1.994
150	1.678	1.613999963	0	1.614	1.614
151	0.511	0.216999993	0	0.217	0.217

Accepted Article

152	2.051	2.059999943	0	2.06	2.06
153	2.458	2.295000076	0	2.295	2.295
154	3.778	3.812000036	0	3.812	3.812
155	1.204	0.842000008	0	0.842	0.842
156	2.916	2.926000118	0	2.926	2.926
157	2.485	2.674999952	0	2.675	2.675
158	0.048	0.090999998	0	0.091	0.091
159	2.944	2.642999887	0	2.643	2.643
160	1.812	1.822000027	0	1.822	1.822
161	2.413	2.480999947	0	2.481	2.481
162	1.693	1.585000038	0	1.585	1.585
163	5.251	4.997000217	0	4.997	4.997
164	0.871	0.83099997	0	0.831	0.831
165	1.018	0.975000024	0	0.975	0.975
166	4.637	3.762000084	0	3.762	3.762
167	0.028	0.549000025	0	0.549	0.549
168	1.542	1.585000038	0	1.585	1.585
169	0.012	0.101000004	0	0.101	0.101
170	1.117	1.223999977	0	1.224	1.224
171	1.9	1.932999969	0	1.933	1.933
172	1.468	1.501999974	0	1.502	1.502
173	1.262	1.215999961	0	1.216	1.216

Accepted Article

174	2.775	2.89199996	0	2.892	2.892
175	0.659	0.444000006	0	0.444	0.444
176	1.074	1.190999985	0	1.191	1.191
177	2.041	2.073999882	0	2.074	2.074
178	0.883	0.912999988	0	0.913	0.913
179	1.682	1.800999999	0	1.801	1.801
180	1.68	1.618999958	0	1.619	1.619
181	1.228	1.225999951	0	1.226	1.226
182	1.092	0.966000021	0	0.966	0.966
183	1.605	1.822000027	0	1.822	1.822
184	2.405	2.188999891	0	2.189	2.189
185	2.514	2.482000113	0	2.482	2.482
186	0.473	0.481000006	0	0.481	0.481
187	3.093	2.084000111	0	2.084	2.084
188	1.862	1.840999961	0	1.841	1.841
189	1.935	1.930999994	0	1.931	1.931
190	2.294	2.395999908	0	2.396	2.396
191	0.764	0.785000026	0	0.785	0.785
192	0.935	0.906000018	0	0.906	0.906
193	1.057	1.027999997	0	1.028	1.028
194	1.111	1.133999944	0	1.134	1.134
195	0.273	0.216999993	0	0.217	0.217

Accepted Article

196	1.192	1.128000021	0	1.128	1.128
197	2.64	2.686000109	0	2.686	2.686
198	4.926	4.997000217	0	4.997	4.997
199	1.423	1.190000057	0	1.19	1.19
200	0.848	0.661000013	0	0.661	0.661
201	1.952	2.084000111	0	2.084	2.084
202	1.902	1.909999967	0	1.91	1.91
203	0.06	0.068999998	0	0.069	0.069
204	2.835	2.075000048	0	2.075	2.075
205	1.252	0.898000002	0	0.898	0.898
206	1.995	2.084000111	0	2.084	2.084
207	1.584	1.534999967	0	1.535	1.535
208	1.691	1.70599997	0	1.706	1.706
209	1.97	1.912999988	0	1.913	1.913
210	2.806	2.480999947	0	2.481	2.481
211	1.396	1.325000048	0	1.325	1.325
212	1.831	1.962000012	0	1.962	1.962
213	2.525	2.487999916	0	2.488	2.488
214	1.601	1.713999987	0	1.714	1.714
215	2.427	2.625999928	0	2.626	2.626
216	0.972	0.906000018	0	0.906	0.906
217	1.416	1.513000011	0	1.513	1.513

Accepted Article

218	1.442	1.421000004	0	1.421	1.421
219	1.598	1.567999959	0	1.568	1.568
220	2.027	2.084000111	0	2.084	2.084
221	1.372	1.220999956	0	1.221	1.221
222	1.33	1.48300004	0	1.483	1.483
223	1.717	1.664000034	0	1.664	1.664
224	1.105	0.967999995	0	0.968	0.968
225	5.111	4.146999836	0	4.147	4.147
226	2.363	2.246999979	0	2.247	2.247
227	1.633	1.667000055	0	1.667	1.667
228	0.401	0.222000003	0	0.222	0.222
229	1.131	1.072999954	0	1.073	1.073
230	1.615	1.618999958	0	1.619	1.619
231	5.021	4.769999981	0	4.77	4.77
232	2	1.94599998	0	1.946	1.946
233	1.535	1.58099997	0	1.581	1.581
234	0.651	0.814000001	0	0.814	0.814
235	0.452	0.549000025	0	0.549	0.549
236	1.874	1.993999958	0	1.994	1.994
237	1.738	1.694000006	0	1.694	1.694
238	2.485	2.674999952	0	2.675	2.675
239	0.149	0.216999993	0	0.217	0.217

Accepted Article

240	1.669	1.713000059	0	1.713	1.713
241	2.035	1.667000055	0	1.667	1.667
242	0.633	0.768000007	0	0.768	0.768
243	2.064	1.993999958	0	1.994	1.994
244	0.836	0.772000015	0	0.772	0.772
245	1.647	1.735000014	0	1.735	1.735
246	0.697	0.657999992	0	0.658	0.658
247	1.759	1.713999987	0	1.714	1.714
248	3.158	3.259000063	0	3.259	3.259
249	2.568	2.647000074	0	2.647	2.647
250	2.101	1.840999961	0	1.841	1.841
251	0.164	0.254000008	0	0.254	0.254
252	1.641	1.713999987	0	1.714	1.714
253	1.873	1.713000059	0	1.713	1.713
254	1.257	1.332000017	0	1.332	1.332
255	1.4	1.404000044	0	1.404	1.404
256	1.65	1.712000012	0	1.712	1.712
257	1.419	1.421000004	0	1.421	1.421
258	1.776	1.613999963	0	1.614	1.614
259	2.284	2.177999973	0	2.178	2.178
260	0.642	0.226999998	0	0.227	0.227
261	1.63	1.534999967	0	1.535	1.535

Accepted Article

262	3.151	2.828999996	0	2.829	2.829
263	1.071	1.082000017	0	1.082	1.082
264	0.913	0.906000018	0	0.906	0.906
265	1.664	1.705999997	0	1.706	1.706

Column Descriptions:

- **Sample_Index**: Unique identifier of each material sample in the test set;
- **True_Value**: Reference true bandgap value of the material (unit: eV);
- **Predicted_Value**: Bandgap value of the material predicted by the XGBoost model (unit: eV);
- **95%CI_Lower**: Lower bound of the 95% confidence interval (unit: eV, set to 0 to comply with the physical constraint that bandgap ≥ 0);
- **95%CI_Upper**: Upper bound of the 95% confidence interval (unit: eV);
- **CI_Width**: Width of the 95% confidence interval (unit: eV, calculated as "95%CI_Upper - 95%CI_Lower"), which quantifies predictive uncertainty—larger values indicate higher uncertainty.

Accepted Article

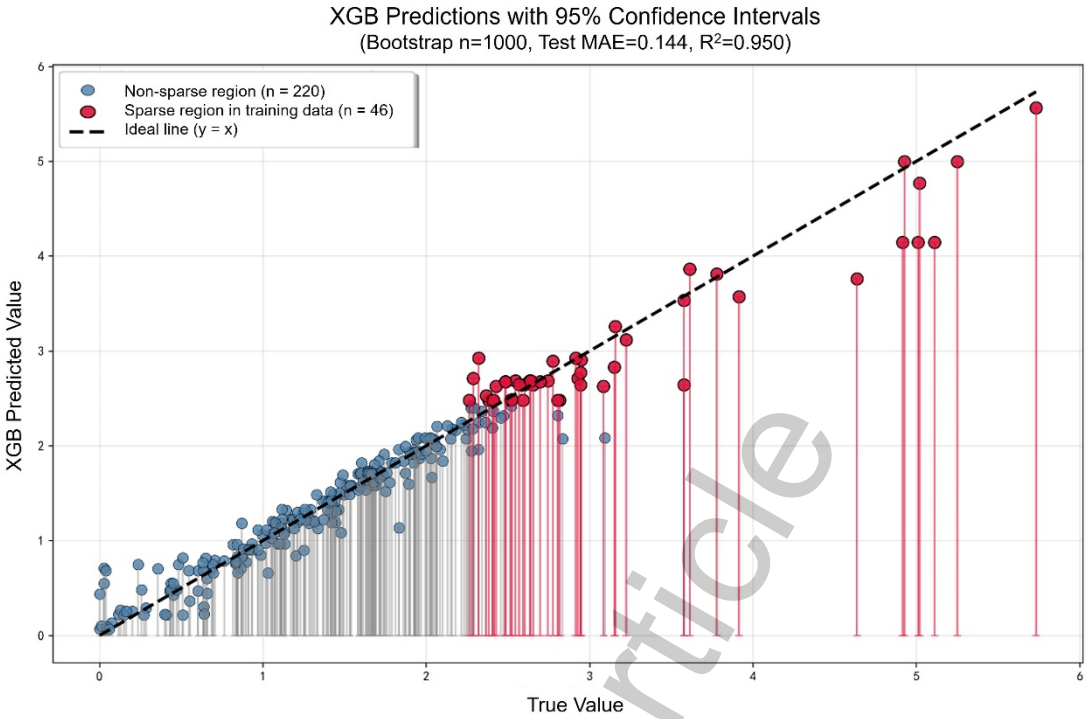


Fig. S4 XGBoost Model-Predicted Bandgaps vs. True Values with 95% Confidence Intervals. This figure shows the relationship between the true bandgap values (x-axis) and XGBoost-predicted bandgap values (y-axis) for 278 candidate materials in the test set. The black dashed line denotes the ideal "predicted = true" reference. Blue scatter points represent samples in non-sparse training data regions (CI width ≤ 1.5× the overall mean), with gray error bars indicating 95% confidence intervals (reflecting low predictive uncertainty). Red scatter points represent samples in sparse training data regions (CI width > 1.5× the overall mean), with red error bars indicating 95% confidence intervals (reflecting high predictive uncertainty).

Accepted Article

Table. S7 Band gap comparison of five final candidate materials calculated by the PBE functional and HSE06 hybrid functional. "Band gap PBE" denotes the band gap obtained from PBE functional calculations; "Band gap HSE" corresponds to the band gap calculated using the HSE06 hybrid functional (adopted to verify the systematic underestimation of band gaps by PBE); "ΔBand gap" represents the difference between the HSE06 and PBE band gaps. This table quantifies the systematic underestimation of band gaps by the PBE functional for these I-III-VI_2 candidate compounds, while confirming that the HSE06-corrected band gaps of the materials fall within or near the energy range relevant to short-wave infrared (SWIR) detection.

Material	Band_gap_PBE (eV)	Band_gap_HSE (eV)	ΔBand_gap (eV)
KGaSe2-I41amd	0.80	1.59	0.79
KGaTe2-I41amd	0.46	1.08	0.62
KInSe2-I41amd	0.58	1.55	0.97
KInTe2-R3m	0.41	0.92	0.51
CsInTe2-R3m	0.53	1.03	0.50

Accepted Article

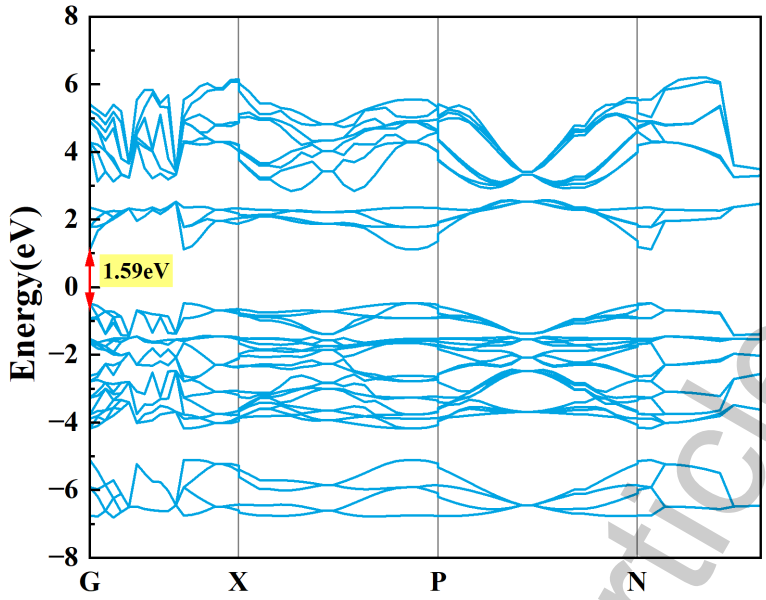


Fig. S5 HSE06 band structure of $KGaSe_2$ along high-symmetry directions. The valence band maximum and conduction band minimum are both located at the Γ point, confirming the preservation of the direct bandgap character under the hybrid-functional level.

Accepted Article