A deep learning approach to search for superconductors from electronic bands

To cite this article: Jun Li *et al* 2025 *AI Sci.* **1** 015001

OPEN ACCESS
IOP Publishing | Songshan Lake Materials Laboratory
AI FOR SCIENCE
AI Sci. 1 (2025) 015001 (13pp)
https://doi.org/10.1088/3050-287X/adf6cb

A deep learning approach to search for superconductors from electronic bands

Jun Li[1,*], Wenqi Fang[3], Shangjian Jin[2], Chenyu Suo[1], Tengdong Zhang[1], Yanling Wu[1], Xiaodan Xu[1], Yong Liu[1,*] and Dao-Xin Yao[2,*]

[1] State Key Laboratory of Metastable Materials Science and Technology & Hebei Key Laboratory of Microstructural Material Physics, School of Science, Yanshan University, Qinhuangdao, 066004, People's Republic of China
[2] State Key Laboratory of Optoelectronic Materials and Technologies, Guangdong Provincial Key Laboratory of Magnetoelectric Physics and Devices, School of Physics, Sun Yat-sen University, Guangzhou, 510275, People's Republic of China
[3] Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen, Guangdong, 518055, People's Republic of China

E-mail: ljcj007@ysu.edu.cn, yongliu@ysu.edu.cn and yaodaox@mail.sysu.edu.cn

Received 2 April 2025, revised 15 July 2025
Accepted for publication 31 July 2025
Published 12 August 2025

Abstract

The prediction of superconducting transition temperatures (T_c) from electronic band structures remains challenging due to the complex interplay of electronic correlations, lattice dynamics, and symmetry. Here, we develop **DeeperBand**, a symmetry-aware 3D Vision Transformer model, to decode band structure features linked to superconductivity. By integrating density functional theory calibrated bands with an attention mechanism, **DeeperBand** identifies steep Fermi-level density-of-states slopes and flat bands as key indicators, aligning with Bardeen, Cooper, and Schrieffer and van Hove singularity paradigms. Trained on 2474 samples with symmetry-informed augmentation (1362 superconductors and 1112 non-superconductors), the model achieves $R^2 = 0.887$ for T_c prediction. Crucially, independent validation using 23 newly synthesized superconductors (published post-training) demonstrates robust generalization across diverse material classes. High-throughput screening of 46 442 materials predicts 15 026 potential superconductors ($T_c > 2$ K), including 1502 candidates ($T_c > 10$ K) such as noble-metal oxides (e.g. $BaAgO_2$) with cuprate-like band features. The **DeeperBand** model can help identifying band structure features linked to superconductivity through interpretable attention maps. This work bridges machine learning with superconductivity theory, establishing data-driven band structure analysis as a powerful paradigm for targeted material discovery with preserved physical interpretability.

Supplementary material for this article is available online

Keywords: deep learning, electronic bands, superconductor

* Authors to whom any correspondence should be addressed.

1. Introduction

Superconductivity, a quantum mechanical phenomenon characterized by zero electrical resistance and the expulsion of magnetic fields, has captivated researchers for over a century. Since the pioneering work of Bardeen, Cooper, and Schrieffer (BCS) [1], which established the microscopic theory of conventional superconductivity via electron–phonon coupling, the field has expanded to encompass unconventional superconductors such as cuprates [2], iron-based compounds [3], and nickelates [4]. These materials often defy BCS paradigms, driven instead by strong electronic correlations [5], spin fluctuations [6], or exotic pairing mechanisms. Despite decades of exploration, the discovery of high-temperature superconductors remains largely serendipitous, relying on experimental trial-and-error rather than predictive theoretical frameworks.

Traditional approaches to superconductivity research heavily rely on density functional theory (DFT) to compute electronic band structures, which provide critical insights into phenomena such as Fermi surface nesting and van Hove singularities. DFT-derived parameters, including the electron–phonon coupling constant (λ) and the density of states (DOSs) near the Fermi level(E_F), are indispensable for understanding conventional superconductors [7]. As demonstrated by DFT calculations [8], this approach successfully describes the electron–phonon coupling in the BCS superconductor MgB_2, which exhibits a relatively high superconducting transition temperature (T_c). Similarly, DFT effectively characterizes hydrogen-rich superconductors under high pressure—recently setting new records T_c [9]—and two-dimensional (2D) carbon-based materials [10, 11].

Moreover, DFT has proven instrumental in elucidating complex phenomena within unconventional superconductors. Key examples include nonlinear lattice dynamics in $YBa_2Cu_3O_{6.5}$ [12], magnetic interactions [6], tight-binding model parameters [13], and electronic Coulomb correlations in iron-based superconductors [5]. DFT further illuminates critical features such as spin–orbit coupling strength in heavy fermion systems [14], interlayer interactions in twisted bilayer graphene [15], σ-bond metallization in nickel-based superconductors under high pressure [16], and superconducting pairing symmetry in bilayer silicene [17] and other 2D carbon-based materials [18, 19].

Recent advances in machine learning (ML) have introduced data-driven methods for predicting T_c. These encompass: (i) periodic table-embedded deep learning for interpretable material design [20]; (ii) hybrid CNN-LSTM algorithms for superconductivity analysis [21]; (iii) characterization of Eliashberg spectral functions [22] and optimized Allen–Dynes fitting using low-dimensional descriptors [23]; (iv) analyses leveraging atomic vectors/symmetries [24]; (v) investigations of the Debye temperature—T_c correlation [25]; (vi) structural search method identifying ambient-pressure BCN superconductors [26]; (vii) predicting high-temperature superconductivity in ternary hydrides under high pressure [27].

These studies demonstrate that deep learning not only accelerates material discovery but also reveals novel superconducting systems beyond the scope of conventional theories. However, prevailing studies predominantly focus on chemical composition or empirical descriptors (e.g. atomic radii, electronegativity), often overlooking the fundamental role of electronic band structures. This neglect risks conflating compositional similarity with genuine physical causality. For instance, isomers with identical stoichiometries can exhibit markedly different superconducting behaviors due to distinct band dispersions, whereas materials with disparate chemistries may share analogous band features conducive to superconductivity. These observations underscore the necessity for a band-centric ML framework that bridges *ab initio* calculations and experimental outcomes. The present work extends this direction by focusing on electronic band structures as key descriptors of superconductivity.

Here, we introduce **DeeperBand**, a deep learning model that correlates three-dimensional (3D) electronic band structures with T_c by integrating a Vision Transformer (ViT) architecture [28] and attention mechanisms. Departing from prior ML approaches, we construct a training dataset consisting of 3D electronic band structures of 1362 experimentally validated superconductors and 1112 non-superconducting materials. We overcome the limited availability of high-fidelity band structure data through symmetry-informed augmentation strategies (e.g. k-space permutations, atomic coordinate swaps and band structure folding), expanding the training dataset while preserving physical invariants. Our model identifies band structure features linked to superconductivity through interpretable attention maps. By screening 46 442 materials from the materials project [29], we predict 14 956 potential superconductors ($T_c > 2$ K), including 1502 candidates with $T_c > 10$ K. By embedding deep learning within the bedrock of band theory, **DeeperBand** offers a principled framework for accelerating superconductor discovery, harnessing the predictive power of modern artificial intelligence (AI) while respecting the quantum mechanical underpinnings of these materials.

2. Methods

2.1. Dataset generation

To construct ML-compatible electronic band structures, we curate the Superband Dataset [30] comprising T_c values and DFT-calculated band structures for 1362 superconductors and 1112 non-superconducting compounds. For model training, we select the reported T_c and 18 electronic bands near the E_F. Each electronic band is discretized on a $32 \times 32 \times 32\mathbf{k}$ grid with the energies defined related to the E_F, resulting in band data of one material with dimensions C (channel) $\times$ L (length) $\times$ H (height) $\times$ W (width) $= 18 \times 32 \times 32 \times 32$. Given the exponential decay in superconductor count with increasing T_c, we apply a logarithmic transformation $\log(T_c + 1)$ (T_c in Kelvin) to

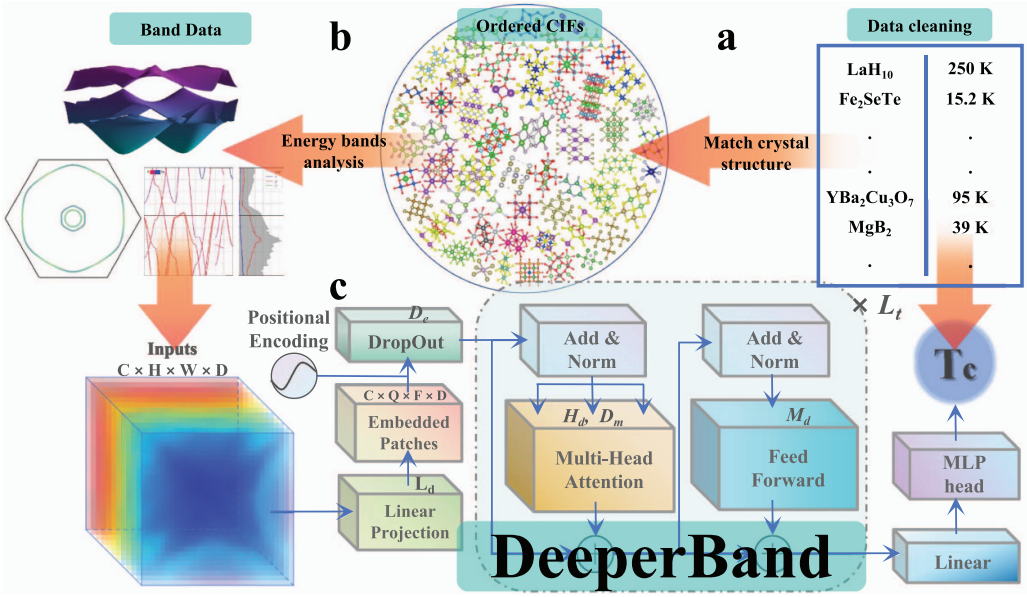

Figure 1. DeeperBand workflow for superconductivity prediction. (a) Data acquisition & cleaning: superconductor formulas and T_c from SuperCon [31] → Remove duplicates. CIF files from MP [29]/OQMD [32] → Disorder removal via 3DSC [33] (lowest Ewald energy). (b) Band structure calculation: high-throughput DFT (Atomate [34]) with MIT parameters [35], automated by FireWorks [36]. Band extraction via Pymatgen [29, 37] and Ifermi [38] → Superband dataset [30] ($18 \times 32 \times 32 \times 32$). (c) Deep learning architecture: 3D-ViT [28] processes band data with optimal hyperparameters: patch size $(P \times Q \times F \times D) = 18 \times 8 \times 8 \times 8$; latent dim $(L_d) = 534$; attention heads $(H_d) = 64$; transformer layers $(L_t) = 3$.

mitigate data skewness during training. The $\log(T_c + 1)$ transformation counters the exponential decay of superconductor counts with rising T_c [30]. While inducing higher absolute errors at high T_c, it ensures balanced learning across the T_c spectrum—critical for predicting unconventional high-temperature superconductivity.

The initial training set of 2474 is insufficient for deep learning. To prevent overfitting, we implement symmetry-preserving data augmentation:

(i) k-space permutation: interchange (k_x, k_y, k_z) axes to exploit rotational symmetry of the Brillouin zone;
(ii) Sign reversal: apply transformations (e.g. $k_x \rightarrow -k_x$) preserving time-reversal symmetry.
(iii) Band folding: repeated select sampling of bands near the Fermi level enables effective simulation of band structure folding in supercell.

These operations retain electronic structure invariants while expanding the dataset. The curated dataset of 2474 materials reflects the current limit of experimentally validated, DFT-computable band structures. While ViT demand large datasets, our symmetry-informed augmentation generates 39 600 physically consistent samples—a critical step to mitigate overfitting given material scarcity.

2.2. Deep learning algorithms

We develop **DeeperBand** (figure 1), a deep learning model based on the 3D-ViT architecture [28]. This design leverages the inherent structural similarities of 3D electronic band representations. The processing pipeline begins with a linear projection layer that standardizes input band data into L_d dimensional embeddings. Subsequent patch embedding restructures the data into 18 channels $\times$ 8 $\times$ 8 $\times$ 8 patches ($Q \times F \times D$ spatial dimensions). To regularize the model, a dropout layer D_e ranging between 0 and 1, stochastically zeroes intermediate features during training.

As shown in the light blue box of figure 1, the core transformer module contains L_t layers. Each layer includes a multi-head attention mechanism, with the number of attention heads controlled by H_d and the dropout rate of the attention output controlled by D_m. A feedforward neural network, with its dimension controlled by M_d, provides additional processing for the attention mechanism. Here, H_d, D_m, and M_d are adjustable positive integers. The network concludes with a linear layer and a multilayer perceptron head as the output layers, producing a one-dimensional tensor as the final output, which is the predicted T_c.

Model training employed the Optuna framework [39] for hyperparameter optimization. We minimize the mean squared error (MSE) between predicted values and rescaled targets $\log(T_c + 1)$ across 100 epoch searches. The optimal configuration (L_d = 534, D_e = 0.127, H_d = 64, D_m = 0.197, M_d = 1038, and L_t = 3, see appendix A) is identified via 8:1:1 training-validation-test splitting with cross-validation. We use 1980 samples to generate 39 600 physically consistent training samples through systematic augmentation, with 247 validation and test samples. We also do 5-fold cross-validation to make sure no overfitting. This procedure yielded a robustly optimized model.

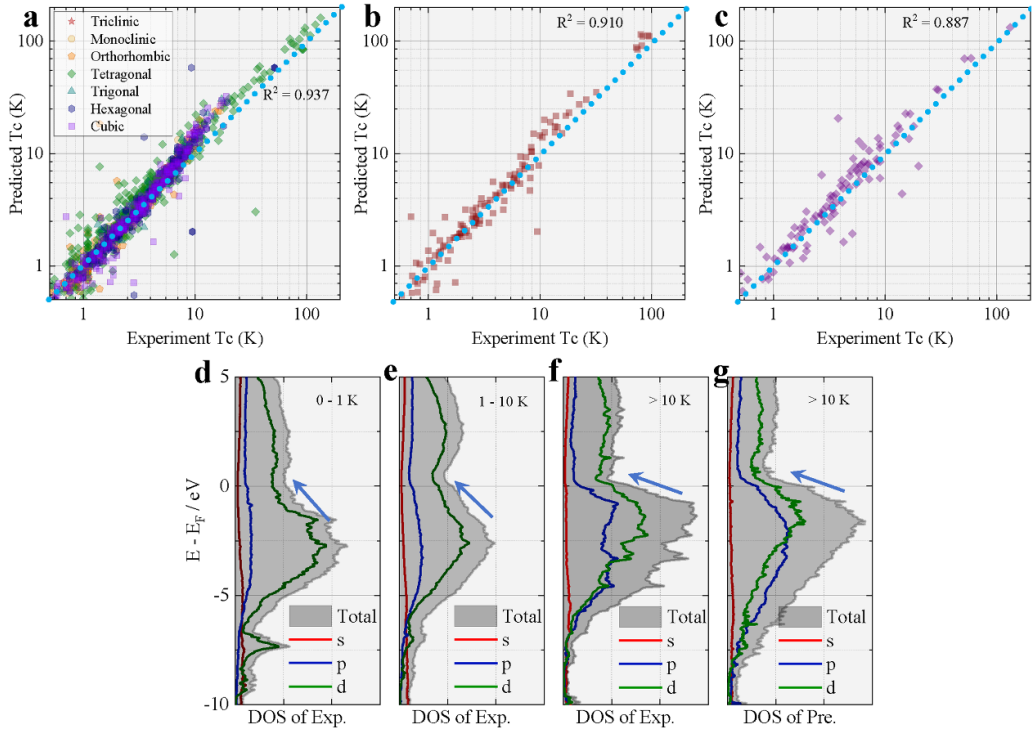

Figure 2. Prediction accuracy and DOSs analysis of DeeperBand. (a) Training set T_c predictions ($R^2 = 0.937$). (b) Validation set T_c predictions ($R^2 = 0.910$). (c) Test set T_c predictions ($R^2 = 0.887$). (d)–(f) Ensemble-averaged DOS for experimentally confirmed superconductors: (d) $T_c < 1$ K (206 materials); (e) $1 \leqslant T_c \leqslant 10$ K (967 materials); (f) $T_c > 10$ K (189 materials). *Dashed lines* indicate E_F; *shaded regions* show $\pm 1\sigma$ variance. (g) Predicted DOS for 1502 high-T_c candidates ($T_c > 10$ K). *Blue arrows* mark characteristic 'peak-valley-peak' signature near E_F. *Note*: non-superconductors ($T_c = 0$) excluded from regression analysis. DOS units: states/eV/unit cell.

3. Results and discussion

3.1. Effectiveness of ***DeeperBand***

To assess model performance, we compare predicted T_c with the experimental values for the known superconductors (figure 2). We employ the coefficient of determination, R^2, defined as:

$$R^2 = 1 - \frac{S_{\mathrm{Res}}}{S_{\mathrm{Tot}}} = 1 - \frac{\sum_i \left(T_i - \hat{T}_i\right)^2}{\sum_i \left(T_i - \bar{T}\right)^2}, \qquad (1)$$

where T_i represents the predicted T_c, $\hat{T}_i$ denotes the average of predicted T_c, and $\bar{T}$ denotes averaged T_c. While MSE serves as an effective loss function during training, its interpretation becomes ambiguous when applied to transformed targets like $\log(T_c + 1)$. As a scale-invariant metric, R^2 quantifies the proportion of variance explained by the model, providing an intuitive and standardized measure of predictive accuracy relative to experimental data.

On the training set, the model achieves $R^2 = 0.937$. For validation set (not used in update model weights), the model's predictions provide good agreement with the experimental superconductors, giving an $R^2 = 0.910$. For test set (not used in model training), the model achieves $R^2 = 0.887$. Accuracy is particularly high for cubic crystal systems. Most of cubic superconductors are conventional (BCS-type), show lower prediction uncertainty. Predictions for hexagonal and tetragonal superconductors exhibit deviations of approximately 20% above experimental values, a discrepancy not attributable to insufficient training (loss convergence is demonstrated in appendix A). This suggests that optimized sample preparation (e.g. doping, pressure) may yield higher T_c in these materials than currently reported. Collectively, these results demonstrate that electronic band structures alone effectively correlate with superconductivity and serve as robust predictors of T_c.

Table 1. Benchmark of predicted T_c vs experimental T_c of four pairs of non-superconducting and superconducting compounds with similar chemical formula.

Structure	System	T_c^{Exp} (K)	T_c^{Pre} (K)
$Ba_2Cu_4PrO_8$ [40]	Cmmm	—	<0.01
$Ba_2Cu_4TmO_8$ [41]	Ammm	80	107.9
$LaYAs_2Fe_2O_2$ [42]	P4mm	—	<0.01
LaAsFeO [43]	P4/nmm	10.2	11.9
$La_3Ni_2O_7$ [4]	Amam	—	<0.01
$La_3Ni_2O_7$ (29.5 GPa) [4]	Fmmm	80	111.6
IrSiTb [44]	Pnma	—	<0.01
IrSiZr [45]	Pnma	2.04	1.80

To further validate **DeeperBand**, we examine four representative material pairs (table 1):

Cuprates: doping critically determines superconductivity. $Ba_2Cu_4PrO_8$ [40] (non-superconducting) and $Ba_2Cu_4TmO_8$

Table 2. Benchmark of predicted T_c vs experimental T_c of on 23 newly synthesized superconductors published after model training completion.

Structure	System	T_c^{Exp} (K)	T_c^{Pre} (K)
$Re_{0.1}Mo_{0.9}B_2$ (44 GPa) [46]	P6/mmm	20	6.246
$RhBi_2$ [47]	—	5.1	2.376
β-$IrSn_4$ [48]	$I4_1/acd$	0.9	0.899
V_xTaS_2 [49]	—	2.7	0.983
β-Bi_2Pd [49]	—	5.4	7.666
MnB_4 (31.4 GPa) [50]	$P2_1/c$	2	3.301
$Pd_3Bi_2Se_2$ [51]	C2/m	0.8	1.098
$TiNb_2Mo$ [52]	—	3.22	7.21
Pb_7Bi_3 [53]	—	9	10.42
TaSeS [54]	$P6_3m$	4.15	3.123
$LaCoSi_2$ [55]	Cmcm	1.2	1.358
$Y_3Rh_4Ge_{13}$ [56]	$I2_13$	1.45	2.203
$ZrOs_2$ [57]	—	3.23	2.507
$Ca_2Pd_3Sb_4$ [58]	I4/mmm	1.3	0.257
$TaGe_2$ [59]	P6222	1.9	2.049
Ni_2NbSn [60]	$Fm\bar{3}m$	3	3.363
$PbMo_6Se_8$ [61]	$R\bar{3}$	3.8	8.170
ZrB_{12} [62]	$Fm\bar{3}m$	5.2	7.215
RuN [63]	R3m	1.29	<0.01
$CuRh_2Se_4$ [64]	$Fd\bar{3}m$	3.4	4.190
TaC [65]	$Fm\bar{3}m$	9	4.161
TiS_2 (136.6 GPa) [66]	C2/m	4.1	3.087
Ta_2PdS_6 (82 GPa) [67]	C2/m	5.2	<0.01

[41] ($T_c = 80$ K) exhibit nearly identical band structures, a distinction accurately captured by our model.

Iron-based systems: magnetism-induced band splitting poses prediction challenges. The model correctly differentiates superconducting $LaYAs_2Fe_2O_2$ [42] from non-superconducting LaAsFeO [43], despite both showing magnetic ordering.

Nickelates: pressure-dependent superconductivity is verified. Our model confirms the pressure-induced superconductivity in $La_3Ni_2O_7$ [4] and its absence at ambient pressure.

Alloys: complex band structures are resolved. The model distinguishes superconducting IrSiTb [44] and non-superconducting IrSiZr [45], despite structural similarity (Pnma space group) and comparable band features.

To rigorously evaluate the predictive generalizability of **DeeperBand**, we tested the model on 23 newly synthesized superconductors published after model training completion (table 2). These materials, spanning binary alloys (e.g. $RhBi_2$ [47]), high-pressure phases(MnF_4 [50]) and complex chalcogenides ($Pd_3Bi_2Se_2$ [51]), were completely absent from both training and validation datasets. While this temporally out-of-sample set provides critical validation of model generalizability across diverse crystal systems, its limited size ($N = 23$) precludes definitive statistical conclusions about global performance. Consequently, we emphasize the methodological value of this test, demonstrating the framework's capability to identify plausible superconducting candidates, with relatively predictive accuracy. However, larger errors in high-pressure systems suggest that extreme condition physics may require specialized augmentation. We commit to updating these benchmarks in superband dataset [30] as new experimental superconductors emerge.

3.2. Attention mechanism associated to superconductivity

We now illustrate the attention mechanism in our deep learning model. Using the attention rollout method [68] within the 3D-ViT framework, we identify regions of the electronic band structure most strongly associated to superconductivity. We demonstrate this with the example of KFe_2Se_2 [69]. Figure 3 shows the electronic band structure near the Fermi surface of KFe_2Se_2, along with the corresponding attention mask. KFe_2Se_2 crystallizes in the space group I4/mmm, which leads to C_4 symmetry observed in the k_x–k_y plane of the electronic band and a tilt C_2 symmetry in the k_z-direction. The attention mask for the k_x–k_y plane similarly exhibits C_4 symmetry, with regions near $(0, \frac{4\pi}{3})$ identified by the model as contributing significantly to superconductivity. In the z-direction, prominent regions appear around $k_z = \frac{4\pi}{5}$ and $\frac{6\pi}{5}$.

Despite the model's initial randomization of weights without explicit consideration of material symmetry, the trained deep learning model effectively captures these symmetry-related features. This underscores the potential of deep learning to reveal meaningful insights from complex 3D electronic band structures. Further details on the attention mechanism and additional examples are provided in appendix B.

3.3. High-throughput searching potential superconductivity

Beyond experimental validation, we deploy **DeeperBand** to predict superconductivity in unexplored materials with DFT-calculated band structures. While traditional superconductors emerge from conductive metals/alloys, unconventional variants often arise from strongly correlated nominal insulators (e.g. Mott systems) exhibiting vanishing band gaps. To probe this space, we screen 46 442 materials project compounds [29], selecting those with band gap <0.2 eV, a threshold capturing correlated metallic states. We identify 14 956 potential candidates with an estimated T_c above 2 K, including 1502 materials that could exhibit T_c values exceeding 10 K. Further details are provided in appendix C.

We further correlate predicted T_c normal-state DOS features. Figure 2(c)–(e) shows the averaged DOS for known superconductors, revealing:

(i) Minimal s-orbital contributions to superconductivity;
(ii) Dominant p-/d-orbital roles, with d-orbitals critical for $T_c > 10$ K (consistent with transition-metal prevalence);
(iii) A universal DOS signature for high-T_c systems: depressed DOS near Fermi surface, flanked by sharp peaks, indicating steep DOS gradients and flat bands near the E_F.

Remarkably, this signature recurs in the ensemble DOS of predicted high-T_c candidates for $T_c > 10$ K (figure 2(f)). These consistent patterns validate our model's physical

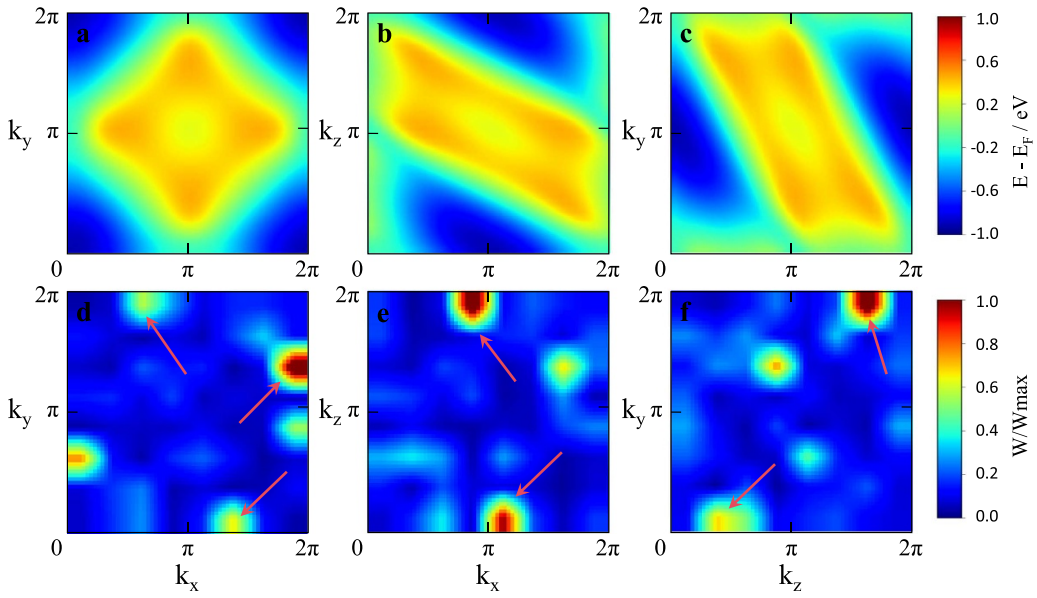

Figure 3. Attention mechanism analysis for KFe_2Se_2. (a) Electronic band crossing E_F in k_x–k_y plane (C_4 symmetry axis indicated). (b) k_x–k_z plane showing tilted C_2 symmetry. (c) k_z–k_y projection. Bottom panels (d)–(f) show attention masks obtained via attention rollout [68], with **red arrows** marking high-attention regions. These masks indicate which regions of the electronic bands are most relevant to superconductivity. In the visualizations, red areas signify regions with a strong contribution to superconductivity, while blue areas denote regions with little to no impact.

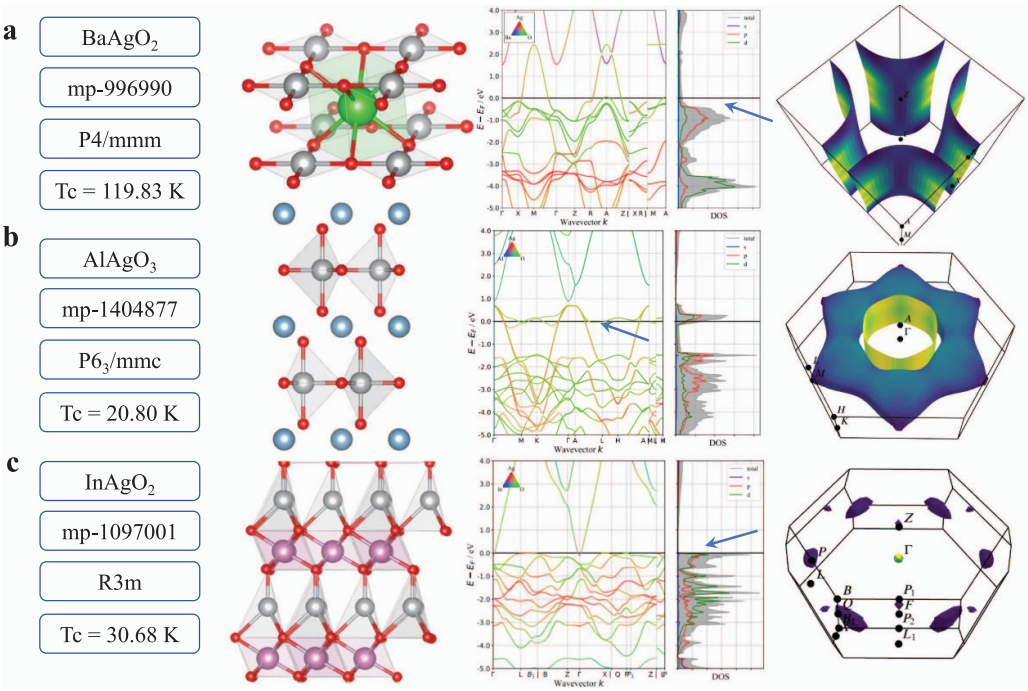

Figure 4. Predicted noble-metal high-temperature superconductor candidates, where data from materials project [29]. (a) $BaAgO_2$: cuprate-like quasi-circular Fermi surface. DOS minimum at E_F with peak 0.2 eV below; (b) $AlAgO_3$: near-flat band (±0.2 eV) and high DOS at E_F. Fermi surface nesting indicated by blue arrows; (c) $InAgO_2$: sharp DOS peak <0.1 eV below E_F. Band gap closure under strain suggests tunable superconductivity.

interpretability and its utility for accelerated superconductor discovery.

3.4. Superconductor Candidates

Among the predicted candidates, we highlight three noble-metal-containing systems exhibiting promising superconducting signatures:

$BaAgO_2$ (mp-996 990) exhibits a DOS minimum near the Fermi surface, with a maximum $\sim$0.2 eV below the E_F. It features a pronounced (π, π) band extremum, yielding a quasi-circular Fermi surface reminiscent of cuprate superconductors. Although $BaAgO_2$ has yet to be experimentally synthesized, theoretical studies suggest its kinetic and thermodynamic stability [70], making it a candidate high-temperature superconductors.

$AlAgO_3$ (mp-1404 877) features an almost flat band near the Fermi surface ($\pm$0.2 eV) and a very high DOS at the E_F. Its Fermi surface also shows nesting, which is favorable for unconventional superconductivity.

$InAgO_2$ (mp-1097 001) does not exhibit significant band overlap at the Fermi surface, it has a high DOS value just below the E_F ($<$0.1 eV), indicating potential superconductivity.

Crucially, replacing Ag with Au preserves these electronic traits (e.g. $BaAuO_2$ mp-1147 676), opening avenues for discovering new unconventional superconductors in noble-metal oxides. More potential superconductors are shown in appendix C, and the data of all predicted superconductor candidates is provided at supplementary information.

4. Conclusion

This study establishes electronic band structures as fundamental descriptors of superconductivity through a deep learning framework. By developing **DeeperBand**, a 3D-ViT model with symmetry-informed augmentation, we achieve accurate T_c prediction ($R^2 > 0.91$) across diverse materials, from cuprates to alloys. Crucially, attention mechanisms identify momentum-space hotspots linked to superconductivity, advancing mechanistic understanding beyond empirical correlations.

Our neural architecture, while relatively simple, effectively predicts superconducting properties. However, prediction deviations ($\sim$20% error) persist in hexagonal/tetragonal systems due to intrinsic lattice complexity. The blind test performance demonstrates effective generalization despite limited original data. However, larger errors in high-pressure systems suggest that extreme condition physics may require specialized augmentation. Future work will integrate multi-fidelity data from high-pressure systems. While symmetry-informed augmentation expanded our physically consistent training set to 59 376 samples, we note that the original dataset of 2474 materials remains limitation for transformer architectures. Though regularization techniques (dropout, weight decay) mitigate overfitting, performance on radically novel material classes may require future supplement to the dataset.

This study presents a novel approach to electronic band analysis and proposes a new framework for predicting the electrical properties of materials. As the developing of AI and high-throughput DFT technologies, the difficulties related to neural network model limitations and training dataset bottlenecks discussed in this paper are likely to be overcome. The integration of ML with electronic band analysis, as introduced here, will be a possible new direction for AI in materials.

Data availability statement

Code and data are available free of charge. The data of all predicted superconductor candidates is provided at supplementary information. We publicly provide the full SuperBand dataset on Science Data Bank [71]. The **DeeperBand** code for accessing and reading the SuperBand dataset file, and neural network model capable of training this dataset provided on the our Github repository (https://github.com/ljcj007/DeeperBand). The other data that support the findings of this study are available from the corresponding author upon request.

Author contributions

J Li: Writing—review & editing, Writing—original draft, Conceptualization, Visualization, Validation, Software, Project administration, Methodology, Data curation. W Fang: Writing—review & editing, Writing—original draft, Validation, Software, Methodology. S Jin: Writing—review & editing, Validation, Formal analysis, Methodology. C Suo: Writing—review & editing, Formal analysis. T Zhang: Writing—review & editing, Formal analysis. Y Wu: Writing—review & editing, Visualization. X Xu: Writing—review & editing, Visualization. Y Liu: Writing—review & editing, Resources, Software, Funding acquisition. D X Yao: Writing—review & editing, Conceptualization, Supervision, Methodology, Data curation, Funding acquisition.

Conflict of interests

The authors declare no competing interests.

Acknowledgments

This work is supported by the National Natural Science Foundation of China (Grant Nos. 12204400, 12494591, 12401676, 92165204), Natural Science Foundation of Hebei Province (Grant Nos. A2022203010, A2024203011), Innovation Capability Improvement Project of Hebei province (Grant No. 22567605 H), National Key R&D Program of China (2022YFA1402802), Guangdong Fundamental Research Center for Magnetoelectric Physics, Guangdong Fundamental Research Center for Magnetoelectric Physics (2024B0303390001), and Guangdong Provincial Quantum Science Strategic Initiative (GDZX2401010). J L, W F and S J contributed equally to this work.

Table A1. The deep learning performance of hyperparameters on the MSE validation loss after 100 epochs of training.

Validation loss	L_t	L_d	H_d	M_d	D_e	M_d
0.324	7	734	24	2154	0.457	0.458
0.240	4	82	32	246	0.283	0.360
0.408	8	618	8	2346	0.204	0.266
0.346	8	468	16	2290	0.412	0.223
0.313	6	962	64	2796	0.461	0.438
0.248	7	120	36	3810	0.415	0.294
0.239	5	754	16	1372	0.268	0.451
0.308	7	188	22	3298	0.161	0.189
0.363	3	114	46	226	0.172	0.241
0.266	4	912	24	880	0.149	0.327
0.313	6	562	48	482	0.273	0.141
0.340	7	558	64	306	0.492	0.240
0.277	4	334	14	3478	0.289	0.157
0.228	3	534	64	1038	0.127	0.197

Appendix A. Model training

Studies on both conventional and unconventional superconductors indicate that the electronic band density near the Fermi surface greatly influences superconductivity, whereas bands far from the Fermi surface contribute little. However, the number of electronic bands near the Fermi surface varies across different superconductors. While ML can indeed take the superconductor with the highest number of electronic bands as a reference and pad the bands of superconductors with fewer electronic bands with zero tensors to form a training set, this approach leads to a waste of computational resources and introduces additional uncertainties in deep learning.

Based on our analysis, the average energy of the 18 bands around to the Fermi surface for most superconductors is typically within the range of -3 eV to 3 eV. Consequently, we uniformly select 18 bands near the Fermi surface for our study. Concerning the grid resolution of the band data, a denser spatial grid can certainly yield more accurate results but at the expense of increased storage and computational resources. In DFT calculations, the typical reciprocal space grid resolution for band structures is generally below $12 \times 12 \times 12$. Expanding to a higher grid resolution during data augmentation could potentially introduce data distortion issues. Therefore, after careful consideration, we select an $18 \times 32 \times 32 \times 32$ grid as the dimensionality for the band data, aiming to find a balance between accuracy and computational efficiency.

It is well known that ViTs typically utilize a patch size of 16×16 when processing 2D images [28]. For our 3D data, we adapt a resolution $8 \times 8 \times 8$ of for each image patch, resulting in a total of $4 \times 4 \times 4$ patches. Additionally, we employ the Optuna software [39] to optimize the hyperparameters of our model, using validation loss as the optimization criterion. The performance of selected hyperparameters, following 100 epochs of training, is summarized in table A1.

Interestingly, our observations indicate that increasing the number of deep learning parameters does not necessarily improve the results. For example, a lower transformer layer count (L_t) yields superior training outcomes compared to models with higher L_t values. We suppose that an excessive number of parameters might lead to overfitting, which explains why the hyperparameters set with the lowest validation loss, as presented in table A1, is selected.

To train the model, we utilize stochastic gradient descent with a learning rate of 0.001, momentum of 0.9, weight decay of 10^{-5}, and batch size of 32. The training process is depicted in figure A1(a), where we observe that although the training loss continued to decrease slightly up to 400 epochs, the validation loss plateaued at around 0.06. Therefore, we extend the training to 700 epochs to finalize the deep learning model.

We employ 5-fold cross-validation to assess model performance and conduct parameter fine-tuning. The dataset is first uniformly partitioned into 5 folds. In each round, one fold (20%) is held out as an independent test set, while the remaining 80% constituted the combined training and validation set. This combined set is further split into a training subset (72%, augmented for dynamic spatial augmentation techniques preserving physical symmetries, such as k-space permutations and sign reversals) and a validation subset (8%, consisting of original, unaugmented data). Training for each fold commenced from the same pre-trained model initialization. The results aggregated across all five folds are presented in figure A1(b), demonstrating minimal deviation from the performance metrics obtained using the previous training.

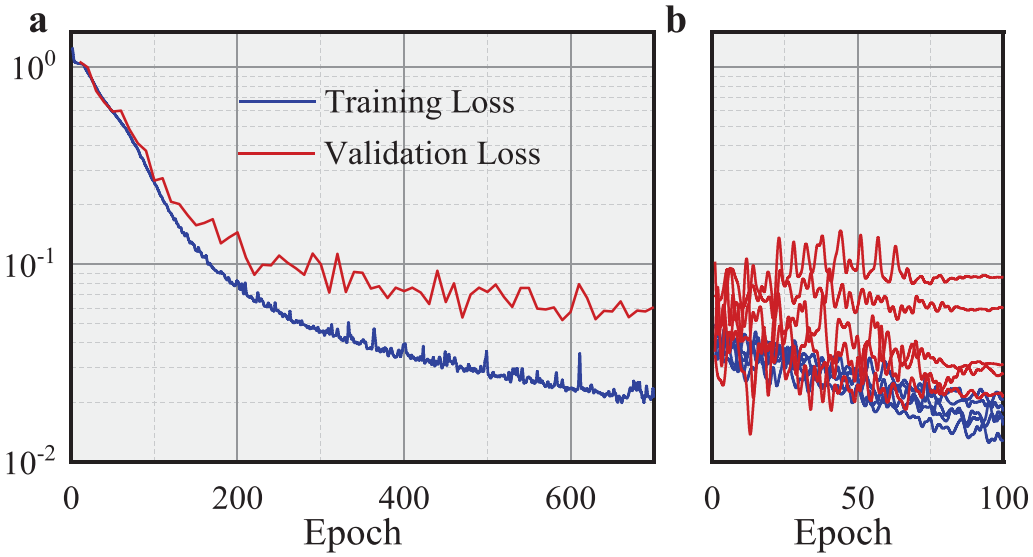

Figure A1. The training loss and validation loss During the actual training process. (a) Main training process with 700 epochs. (b) Five-fold cross-validation to assess model performance and conduct parameter fine-tuning with 100 epochs.

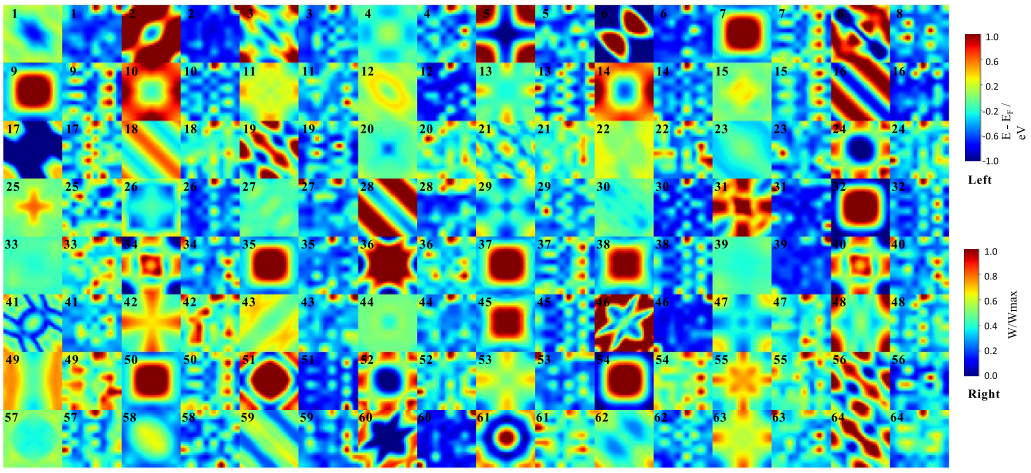

Figure B1. Further example attention weights as in figure 3(a), (d), we have only selected the k_x–k_y plane, as the z-axis is typically set as the principal axis, and the k_x–k_y plane tends to exhibit the highest symmetry. The examples shown here are randomly selected from those with T_c above 5 K. The left graph of each material represents the band closest to the E_F, while the right graph represents the attention mask.

Appendix B. Attention mechanism

To generate attention maps from the output token to the input space, we employ the attention rollout method [68]. In summary, we average the attention weights across all heads in the 3D-ViT model and recursively multiply the weight matrices across all layers. This method effectively captures the cumulative effect of attention across tokens throughout the entire network, providing a comprehensive view of how information is propagated and emphasized from input to output.

The attention rollout method offers an intuitive approach to trace the flow of information as it propagates from the input layer through the higher layers of a neural network. By systematically aggregating and multiplying attention weights across layers, this technique enables us to visualize how different parts of the input influence the model's decision-making process at various stages. Attention Rollout provides a detailed representation on how the model integrates and refines information as it progresses through each layer, ultimately leading to the final output. This makes it an invaluable tool for dissecting the inner workings of 3D-ViT models, offering insights into how these models interpret and prioritize various aspects of the input data.

Further examples of attention weights are shown in figure B1(random selection within the k_x–k_y plane). While some images may appear blurred, the majority of attention weights successfully captures the crystalline symmetry of the materials. We believe that as this technology advances, these attention weights will help identify the key features of electronic bands that contribute to superconductivity, providing a reliable reference for addressing the challenges of high-temperature superconductivity mechanisms. Additionally, our model can also use theoretical lattice models, such as tight-binding models, as inputs, aiding in determining whether these

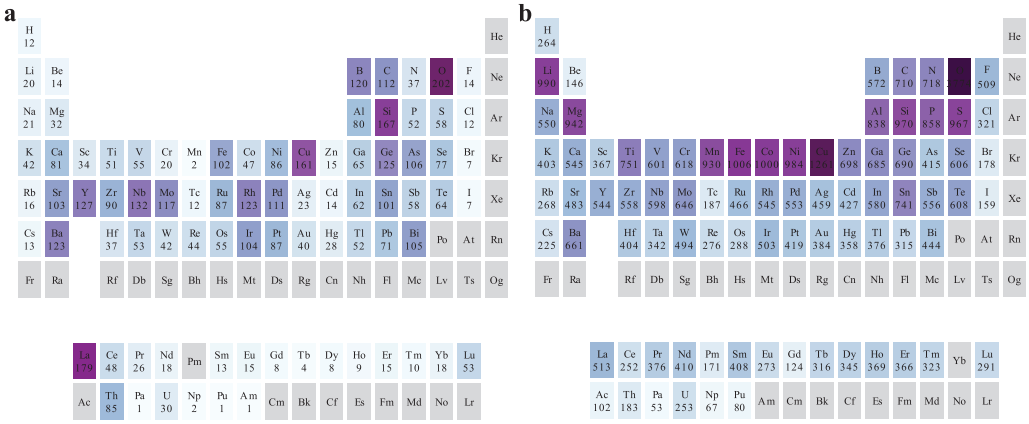

Figure C1. The elemental distribution of (a) superconductors found experimentally [30] and (b) predicted superconductors. For any superconductor containing element A, the count for element A is incremented by 1. By following this process, we obtain the elemental distribution across all superconductors. It is important to note that this analysis encompasses all superconductors, without distinguishing between conventional and unconventional ones. The abundance of metallic elements primarily originates from early alloy-based superconductors. Additionally, oxygen is present not only in cuprate superconductors but also in many low-temperature superconductors. Adapted from [30], with permission from Springer Nature.

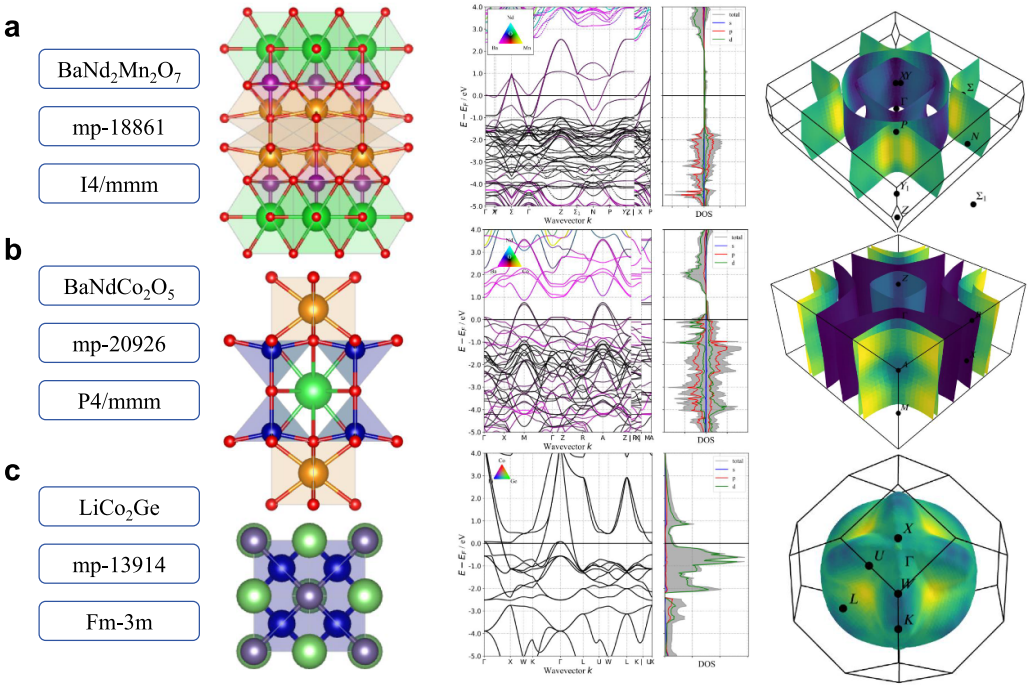

Figure C2. Further example as in the left of figure 4 with three possible manganese, cobalt and lithium-based superconductor candidates.

lattice models contain critical factors conducive to the formation of superconductivity.

Appendix C. Superconductor candidates

Early superconductors were primarily derived from metals or alloys, which inherently possess conductive properties. In contrast, unconventional superconductors, apart from those that naturally exhibit conductivity, often emerge from strongly correlated systems like Mott insulators, with no band gap. Therefore, when identifying potential superconductor candidates, our focus is primarily on materials with no band gap or possess a slightly doped gap. To this end, we systematically screened 46 442 materials in MP database [29] with band gaps up to 0.2 eV to find possible superconductor candidates.

We employ the trained deep learning model to conduct high-throughput computational analysis of the band structures of these materials, ultimately identifying 14 956 potential superconductors. The elemental distribution of these candidates is depicted in figure C1(b), and is compared to the distribution of experimentally confirmed superconductors in figure C1(a).

We observe that among the predicted superconductors, the fourth-period transition metals—manganese (Mn), iron (Fe), cobalt (Co), nickel (Ni), and copper (Cu)—appear with notable frequency, all of which are adjacent in the periodic table. This is particularly significant given that the three known unconventional superconductors with relatively high T_c involve iron, nickel, and copper. This observation suggests the possibility of discovering new unconventional superconductors based on manganese and cobalt. To explore this further, we have identified two potential manganese- and cobalt-based superconductors, with their electronic band structures illustrated in the top and middle panels of figure C2.

In addition to these transition metals, lithium (Li) and magnesium (Mg) also show strong potential for becoming superconductors. We speculate that these elements are more likely to be conventional superconductors when combined with nonmetallic elements, similar to the well-known MgB_2. A promising candidate superconductor involving lithium has been identified, and its electronic band structure is depicted in the bottom panel of figure C2.

ORCID iDs

Jun Li 0000-0001-8450-3702
Wenqi Fang 0000-0003-3347-0511
Shangjian Jin 0000-0002-6686-3365
Yanling Wu 0000-0001-5126-6501
Xiaodan Xu 0000-0003-1701-6044
Yong Liu 0000-0002-5435-9217
Dao-Xin Yao 0000-0003-1097-3802

References

[1] Bardeen J, Cooper L N and Schrieffer J R 1957 Theory of superconductivity *Phys. Rev.* **108** 1175–204
[2] Bednorz J G and Muller K A 1986 Possible high T_c superconductivity in the Ba-La-Cu-O system *Z. Phys. B* **64** 189–93
[3] Stewart G R 2011 Superconductivity in iron compounds *Rev. Mod. Phys.* **83** 1589–652
[4] Sun H *et al* 2023 Signatures of superconductivity near 80 K in a nickelate under high pressure *Nature* **621** 493–8
[5] Aichhorn M, Biermann S, Miyake T, Georges A and Imada M 2010 Theoretical evidence for strong correlations and incoherent metallic state in FeSe *Phys. Rev. B* **82** 064504
[6] Graser S, Kemper A F, Maier T A, Cheng H-P, Hirschfeld P J and Scalapino D J 2010 Spin fluctuations and superconductivity in a three-dimensional tight-binding model for $BaFe_2As_2$ *Phys. Rev. B* **81** 214503
[7] Giustino F 2017 Electron-phonon interactions from first principles *Rev. Mod. Phys.* **89** 015003
[8] Bohnen K-P, Heid R and Renker B 2001 Phonon dispersion and electron-phonon coupling in MgB_2 and AlB_2 *Phys. Rev. Lett.* **86** 5771–4
[9] Drozdov A P *et al* 2019 Superconductivity at 250 K in lanthanum hydride under high pressures *Nature* **569** 528–31
[10] Jun Li and Yao D-X 2022 Superconductivity in octagraphene *Chin. Phys. B* **31** 017403
[11] Chen S, Liu Z, Duan W and Liu F 2013 First-principles calculations on the effect of doping and biaxial tensile strain on electron-phonon coupling in graphene *Phys. Rev. Lett.* **111** 196802
[12] Mankowsky R *et al* 2014 Nonlinear lattice dynamics as a basis for enhanced superconductivity in $YBa_2Cu_3O_{6.5}$ *Nature* **516** 71–73
[13] Chao Cao P J H and Cheng H-P 2008 Proximity of antiferromagnetism and superconductivity in $LaFeAsO_{1-x}F_x$: Effective Hamiltonian from *ab initio* studies *Phys. Rev. B* **77** 220506
[14] Samokhin K V, Zijlstra E S and Bose S K 2004 $CePt_3Si$ an unconventional superconductor without inversion center *Phys. Rev. B* **69** 094514
[15] Carr S, Fang S, Jarillo-Herrero P and Kaxiras E 2018 Pressure dependence of the magic twist angle in graphene superlattices *Phys. Rev. B* **98** 085144
[16] Luo Z, Xunwu H, Wang M, Wei W and Yao D-X 2023 Bilayer two-orbital model of $La_3Ni_2O_7$ under pressure *Phys. Rev. Lett.* **131** 126001
[17] Liu F, Liu C-C, Kehui W, Yang F and Yao Y 2013 $d+id'$ chiral superconductivity in bilayer silicene *Phys. Rev. Lett.* **111** 066804
[18] Jun Li, Jin S, Yang F and Yao D-X 2020 Electronic structure, magnetism and high-temperature superconductivity in multilayer octagraphene and octagraphite *Phys. Rev. B* **102** 174509
[19] Jiacheng Y, Jun Li, Zhong D and Yao D-X 2023 Possible superconductivity in Biphenylene *Chin. Phys. Lett.* **40** 077401
[20] Konno T, Kurokawa H, Nabeshima F, Sakishita Y, Ogawa R, Hosako I and Maeda A 2021 Deep learning model for finding new superconductors *Phys. Rev. B* **103** 014509
[21] Shaobo Li, Dan Y, Xiang Li, Tiantian H, Dong R, Cao Z and Jianjun H 2020 Critical temperature prediction of superconductors based on atomic vectors and deep learning *Symmetry* **12** 262
[22] Xie S R *et al* 2022 Machine learning of superconducting critical temperature from Eliashberg theory *npj Comput. Mater.* **8** 1–8
[23] Xie S R, Stewart G R, Hamlin J J, Hirschfeld P J and Hennig R G 2019 Functional form of the superconducting critical temperature from machine learning *Phys. Rev. B* **100** 174513
[24] Zhang J, Zhu Z, Xiang X-D, Zhang K, Huang S, Zhong C, Qiu H-J, Kailong H and Lin Xi 2022 Machine learning prediction of superconducting critical temperature through the structural descriptor *J. Phys. Chem. C* **126** 8922–7
[25] Smith A D, Harris S B, Camata R P, Yan D and Chen C-C 2023 Machine learning the relationship between Debye temperature and superconducting transition temperature *Phys. Rev. B* **108** 174514
[26] Han X-Q, Ouyang Z, Guo P-J, Sun H, Gao Z-F and Lu Z-Y 2024 AI-accelerated discovery of high critical temperature superconductors (arXiv:2409.08065)
[27] Wang X, Zhang C, Wang Z, Liu H, Jian Lv, Wang H, Weinan E and Yanming M 2025 Discovery of high-temperature superconducting ternary hydrides via deep learning (arXiv:2502.16558)
[28] Dosovitskiy A, Beyer L, Kolesnikov A, Weissenborn D, Zhai X, Unterthiner T, Dehghani M, Minderer M, Heigold G, Gelly S, Uszkoreit J and Houlsby N 2021 An image is worth 16×16 words: transformers for image recognition at scale (arXiv:2010.11929)
[29] Jain A *et al* 2013 Commentary: the materials project: a materials genome approach to accelerating materials innovation *APL Mater.* **1** 011002
[30] Zhang T, Suo C, Yanling W, Xiaodan X, Liu Y, Yao D-X and Jun Li 2025 SuperBand: an electronic-band and Fermi surface structure database of superconductors *Sci. Data* **12** 744
[31] Center for Basic Research on Materials 2024 Mdr supercon datasheet ver. 240322

[32] Kirklin S, Saal J E, Meredig B, Thompson A, Doak J W, Aykol M, Rühl S and Wolverton C 2015 The open quantum materials database (OQMD): assessing the accuracy of DFT formation energies *npj Comput. Mater.* **1** 1–15

[33] Sommer T, Willa R, Schmalian J and Friederich P 2023 3DSC - a dataset of superconductors including crystal structures *Sci. Data* **10** 1–13

[34] Mathew K *et al* 2017 Atomate: A high-level interface to generate, execute and analyze computational materials science workflows *Comput. Mater. Sci.* **139** 140–52

[35] Jain A, Hautier G, Moore C J, Ong S P, Fischer C C, Mueller T, Persson K A and Ceder G 2011 A high-throughput infrastructure for density functional theory calculations *Comput. Mater. Sci.* **50** 2295–310

[36] Jain A *et al* 2015 Fireworks: a dynamic workflow system designed for high-throughput applications *Concurr. Comput. Pract. Exp.* **27** 5037–59

[37] Ong S P, Cholia S, Jain A, Brafman M, Gunter D, Ceder G and Persson K A 2015 The materials application programming interface (API): a simple, flexible and efficient API for materials data based on representational state transfer (REST) principles *Comput. Mater. Sci.* **97** 209–15

[38] Ganose A M, Searle A, Jain A and Griffin S M 2021 IFermi: a python library for Fermi surface generation and analysis *J. Open Source Softw.* **6** 3089

[39] Akiba T, Sano S, Yanase T, Ohta T and Koyama M 2019 Optuna: a next-generation hyperparameter optimization framework (arXiv:1907.10902)

[40] Yuh Yamada S H, Yamada N, Guo Z, Kodama Y, Kawamoto K, Mizutani U and Hirabayashi I 1994 High oxygen pressure synthesis of the $PrBa_2Cu_4O_8$ compound *Physica C* **231** 131–6

[41] Hijar C A, Stern C L, Poeppelmeier K R, Rogacki K, Chen Z and Dabrowski B 1995 Single-crystal growth and characterization of $REBa_2Cu_4O_8$ and $Y_{1-x}CaxBa_2Cu_4O_8$ *Physica C* **252** 13–21

[42] Tropeano M, Fanciulli C, Canepa F, Cimberle M R, Ferdeghini C, Lamura G, Martinelli A, Putti M, Vignolo M and Palenzona A 2009 Effect of chemical pressure on spin density wave and superconductivity in undoped and 15% F-doped $La_{1-y}Y_yFeAsO$ compounds *Phys. Rev. B* **79** 174523

[43] Okada H, Igawa K, Takahashi H, Kamihara Y, Hirano M, Hosono H, Matsubayashi K and Uwatoko Y 2008 Superconductivity under high pressure in LaFeAsO *J. Phys. Soc. Japan* **77** 113712

[44] Szytuła A, Hofmann M, Leciejewicz J, Penc B and Zygmunt A 2001 Magnetic properties and magnetic structures of RIrSi (R = Tb-Er) series of compounds *J. Alloys Compd.* **316** 58–63

[45] Suzuki H, Kase N, Nakano T and Takeda N 2016 Superconducting and normal state properties in the ternary silicide NbIrSi, TaIrSi and NbPtSi *Phys. Proc.* **81** 57–60

[46] Sinha S *et al* 2024 Superconductivity in pressurized $Re_{0.10}Mo_{0.90}B_2$ (arXiv:2408.17416)

[47] KeYuan M, Roychowdhury S, Noky J, Borrmann H, Schnelle W, Shekhar C, Medvedev S A and Felser C 2024 Pressure-induced superconductivity in monoclinic $RhBi_2$ (arXiv:2409.06358)

[48] Ahmad N, Shimada S, Hasegawa T, Suzuki H, Afzal Md A, Nakamura N, Higashinaka R, Matsuda T D and Aoki Y 2024 Linear magnetoresistance and type-i superconductivity in (β)-$IrSn_4$ *J. Phys. Soc. Japan* **93** 044706

[49] Trivini S, Ortuzar J, Zaldivar J, Herrera E, Guillamón I, Suderow H, Sebastian Bergeret F and Pascual J I 2024 Diluted Yu-Shiba-Rusinov arrays on the β-Bi_2Pd anisotropic superconductor (arXiv:2408.11704)

[50] Xiang Z-N, Zhang Y-J, Lu Q, Li Q, Li Y, Huang T, Zhu Y, Yongze Y, Sun J and Wen H-H 2025 Superconductivity up to 14.2 K in MnB_4 under pressure *Adv. Mater.* **37** 2416882

[51] Chapai R, Peterson G, Smylie M P, Chen X, Jiang J S, Graf D, Mitchell J F and Welp U 2024 Fermi surface topology and magneto-transport properties of superconducting $Pd_3Bi_2Se_2$ (arXiv:2408.06298)

[52] Kuan Li, Chen C-Q, Zeng L, Longfu Li, Chen R, Peifeng Y, Wang K, Xiang Z, Yao D-X and Luo H 2025 Superconductivity in medium-entropy alloys: Nb_2TiW and Nb_2TiMo *Chin. Phys. Lett.* **42** 017401

[53] Karn N K, Kumar K, Kumar N, Kumar Y, Sharma M M, Hu J and Awana V P S 2024 Type-II superconductivity at 9k in Pb-Bi alloy *Solid State Commun.* **391** 115639

[54] Yadav K, Lamba M, Singh M, Yadav M, Kumar A and Patnaik S 2024 Order parameter symmetry in superconducting 2H-TaSeS *J. Phys.: Condens. Matter* **37** 075602

[55] Ito F, Kabeya N and Kimura N 2025 Weakly coupled type-II superconductivity in $LaCoSi_2$ single crystal *Physica C* **630** 1354655

[56] Afzal Md A, Higashinaka R, Iwasa K, Ahmed N, Tsubota R, Nakamura N, Matsuda T D and Aoki Y 2024 Superconductivity in chiral cubic $Y_3Rh_4Ge_{13}$ *J. Alloys Compd.* **978** 172914

[57] Meena P K, Mandal M, Manna P, Srivastava S, Sharma S, Mishra P and Singh R P 2024 Superconductivity in breathing kagome-structured C14 Laves phase XOs_2(X = Zr, Hf) *Supercond. Sci. Technol.* **37** 075004

[58] Xiaoping M *et al* 2024 Superconductivity and signatures of nontrivial topology in the layered compound $Ca_2Pd_3Sb_4$ *Chem. Mater.* **36** 2363–9

[59] Li Z *et al* 2024 Type-I superconductivity in the Weyl semimetal $TaGe_2$ with chiral structure *Phys. Rev. B* **110** 174506

[60] Nalevanko S *et al* 2024 Ni_2NbSn Heusler ferromagnetic superconductor: magneto-structural and transport characterisation *Intermetallics* **167** 108231

[61] Banerjee S, Ghosh S, Kataria A and Sundaresan A 2024 Evidence of unconventional vortex states in the Chevrel phase superconductor $PbMo_6Se_8$ *Phys. Rev. B* **110** 014512

[62] Zhang Z, Zheng X, Luo H, Gao C, Xue X, Zhu J, Ruobin Li, Jin C and Xiaohui Y 2024 Superconductivity in ZrB_{12} under high pressure *Metals* **14** 1082

[63] Ilin A S, Strugova A O, Cohn I A, Pavlovskiy V V, Zaitsev-Zotov S V, Sadakov A V, Sobolevskiy O A, Morgun L A, Matrovitskii V P and Rybalchenko G V 2024 Superconductivity in thin films of RuN *Phys. Rev. Mater.* **8** 074801

[64] Kuan Li, Zeng L, Longfu Li, Chen R, Peifeng Y, Wang K, Zhang C, Xiang Z and Luo H 2024 Structural and superconducting properties in the Te-doped spinel $CuRh_2Se_4$ *J. Alloys Compd.* **995** 174756

[65] Castro A A, Ascencio F and Escudero R 2024 Superconductivity in centrosymmetric TaC_x with x = 0.0, 1.25 and 1.50 *J. Supercond. Novel Magn.* **37** 509–14

[66] Zhang H, Zhong W, Yue B, Xiaohui Y and Hong F 2024 Anomalous metal-semiconductor transition and superconductivity in van der Waals TiS_2 under pressure *Phys. Rev. B* **110** 134525

[67] Liu W, Feng J, Hou Q, Bin Li, Wang K, Chen B, Sheng Li and Shi Z 2024 Pressure-induced superconductivity and isosymmetric structural transition in quasi-one-dimensional Ta_2PdS_6 *Phys. Rev. B* **109** 054513

[68] Abnar S and Zuidema W 2020 Quantifying attention flow in transformers (arXiv:2005.00928)
[69] Ying T P, Chen X L, Wang G, Jin S F, Zhou T T, Lai X F, Zhang H and Wang W Y 2012 Observation of superconductivity at 30–46 K in $A_xFe_2Se_2$(A = Li, Na, Ba, Sr, Ca, Yb and Eu) *Sci. Rep.* **2** 426
[70] Cerqueira T F T, Lin S, Amsler M, Goedecker S, Botti S and Miguel A L M 2015 Identification of Novel Cu, Ag and Au ternary oxides from global structural prediction *Chem. Mater.* **27** 4562–73
[71] Jun Li, Zhang T and Suo C 2024 SuperBand: superconductor's energy band *Science Data Bank* (https://doi.org/10.57760/sciencedb.16728)