

HTSC-2025: A benchmark dataset of ambient-pressure high-temperature superconductors for AI-driven critical temperature prediction

Xiao-Qi Han(韩小琪), Ze-Feng Gao(高泽峰), Xin-De Wang(王馨德), Zhenfeng Ouyang(欧阳振峰), Peng-Jie Guo(郭朋杰), and Zhong-Yi Lu(卢仲毅)

Citation: Chin. Phys. B, 2025, 34 (**10**): 100301. DOI: 10.1088/1674-1056/adf042
http://cpb.iphy.ac.cn; http://iopscience.iop.org/cpb

What follows is a list of articles you may be interested in

HTSC-2025: A benchmark dataset of ambient-pressure high-temperature superconductors for AI-driven critical temperature prediction

Xiao-Qi Han(韩小琪), Ze-Feng Gao(高泽峰), Xin-De Wang(王馨德), Zhenfeng Ouyang(欧阳振峰), Peng-Jie Guo(郭朋杰), and Zhong-Yi Lu(卢仲毅)
Chin. Phys. B, 2025, 34 (**10**): 100301. DOI: 10.1088/1674-1056/adf042

Multicast-oriented key provision in hybrid DV/CV multi-domain quantum networks

Xinyu Chen(陈欣宇), Yuan Cao(曹原), Yuxiang Lu(陆宇翔), Yue Chen(陈越), Kunpeng Zheng(郑昆朋), Xiaosong Yu(郁小松), Yongli Zhao(赵永利), Jie Zhang(张杰), and Qin Wang(王琴)
Chin. Phys. B, 2025, 34 (**9**): 090301. DOI: 10.1088/1674-1056/adecf8

Exploring Majorana zero modes in iron-based superconductors

Geng Li(李更), Shiyu Zhu(朱诗雨), Peng Fan(范朋), Lu Cao(曹路), and Hong-Jun Gao(高鸿钧)
Chin. Phys. B, 2022, 31 (**8**): 080301. DOI: 10.1088/1674-1056/ac70c3

Average fidelity estimation of twirled noisy quantum channel using unitary 2*t*-design

Linxi Zhang(张林曦), Changhua Zhu(朱畅华), Changxing Pei(裴昌幸)
Chin. Phys. B, 2019, 28 (**1**): 010304. DOI: 10.1088/1674-1056/28/1/010304

Momentum distribution and non-local high order correlation functions of 1D strongly interacting Bose gas

EJKP Nandani, Xi-Wen Guan(管习文)
Chin. Phys. B, 2018, 27 (**7**): 070306. DOI: 10.1088/1674-1056/27/7/070306

DATA PAPER

HTSC-2025: A benchmark dataset of ambient-pressure high-temperature superconductors for AI-driven critical temperature prediction

Xiao-Qi Han(韩小琪)[1,2], Ze-Feng Gao(高泽峰)[1,2,†], Xin-De Wang(王馨德)[1,2], Zhenfeng Ouyang(欧阳振峰)[1,2], Peng-Jie Guo(郭朋杰)[1,2], and Zhong-Yi Lu(卢仲毅)[1,2,3]

[1] *School of Physics and Beijing Key Laboratory of Opto-electronic Functional Materials & Micro-nano Devices, Renmin University of China, Beijing* 100872, *China*

[2] *Key Laboratory of Quantum State Construction and Manipulation (Ministry of Education), Renmin University of China, Beijing* 100872, *China*

[3] *Hefei National Laboratory, Hefei* 230088, *China*

(Received 10 June 2025; revised manuscript received 14 July 2025; accepted manuscript online 16 July 2025)

The discovery of high-temperature superconducting materials holds great significance for human industry and daily life. In recent years, research on predicting superconducting transition temperatures using artificial intelligence (AI) has gained popularity, with most of these tools claiming to achieve remarkable accuracy. However, the lack of widely accepted benchmark datasets in this field has severely hindered fair comparisons between different AI algorithms and impeded further advancement of these methods. In this work, we present HTSC-2025, an ambient-pressure high-temperature superconducting benchmark dataset. This comprehensive compilation encompasses theoretically predicted superconducting materials discovered by theoretical physicists from 2023 to 2025 based on BCS superconductivity theory, including the renowned X_2YH_6 system, perovskite MXH_3 system, M_3XH_8 system, cage-like BCN-doped metal atomic systems derived from LaH_{10} structural evolution, and two-dimensional honeycomb-structured systems evolving from MgB_2. In addition, we note a range of approaches inspired by physical intuition for designing high-temperature superconductors, such as hole doping, the introduction of light elements to form strong covalent bonds, and the tuning of spin–orbit coupling. The dataset presented in this paper is openly available at ScienceDB. The HTSC-2025 benchmark has been open-sourced on Hugging Face at https://huggingface.co/datasets/xiao-qi/HTSC-2025 and will be continuously updated, while the Electronic Laboratory for Material Science platform is available at https://in.iphy.ac.cn/eln/link.html#/124/V2s4.

Keywords: benchmark, superconductors, artificial intelligence

PACS: 03.75.Lm, 07.05.Mh, 84.35.+i, 05.70.Jk

DOI: 10.1088/1674-1056/adf042 **CSTR:** 32038.14.CPB.adf042

1. Introduction

Superconducting materials, characterized by their exceptional property of zero electrical resistance, have remained a central research focus in condensed matter physics since their discovery in 1911 (when mercury exhibited superconductivity at 4.2 K[1]). As evidenced by the arXiv[2] preprint repository, scholarly publications containing the keyword "Superconductivity" have now surpassed 60000 entries to date, underscoring the sustained scientific significance of this field. High-temperature superconducting materials serve as key components for energy-efficient transmission,[3] magnetic resonance imaging (MRI),[4] nuclear magnetic resonance (NMR),[5] and fusion devices.[6] Therefore, the discovery of new high-temperature, or even room-temperature, superconductors holds great significance for human production and daily life.

With technological advancements, AI has become an essential tool for discovering superconducting materials.[7] Kamal Choudhary *et al.*[8] performed large-scale calculations on crystal materials databases based on atomistic line graph neural network (ALIGNN)[9] predictions and identified 105 dynamically stable superconductors with $T_c > 5$ K. Similarly, Marques *et al.*[10] applied this approach to hydrogen-based systems and discovered 59 superconducting materials, among which Li_2CuH_6 exhibits a superconducting transition temperature of 86 K. InvDesFlow[11] explored new chemical spaces through a crystal generative model, demonstrating the feasibility of generative models in superconducting material discovery. In its latest application, InvDesFlow identified Li_2AuH_6[12] with a superconducting transition temperature of 140 K, which significantly exceeds the McMillan limit and even surpasses the liquid nitrogen temperature regime.

Recently, numerous AI tools have been employed to predict superconducting transition temperatures. The ALIGNN model has achieved a mean absolute error (MAE) of less than 2 K in its predictions.[8] The bootstrapped ensemble of tem-

†Corresponding author. E-mail: zfgao@ruc.edu.cn

© 2025 Chinese Physical Society and IOP Publishing Ltd. All rights, including for text and data mining, AI training, and similar technologies, are reserved.

pered equivariant graph neural networks (BETE-NET)[13] predicts T_c using three moments (λ, $\langle\omega\rangle$, and ω^2) of the spectral function $\alpha^2F(\omega)$, reducing the MAE to 2.1 K. The BANS[14] model, based on a deep learning framework that incorporates a 3D vision transformer (3D-ViT) architecture and attention mechanisms, predicts T_c by analyzing electronic band structures. For systems with $T_c < 10$ K, the prediction error is less than 2 K, while for high-T_c superconductors, the error remains below 25 K. The InvDesFlow-AL[15] proposes the Supercon-GNN model, which employs an equivariant graph neural network based on spherical harmonics to predict superconducting transition temperatures. This framework achieves a prediction error of less than 2 K for conventional BCS superconductors as well. However, these models lack a unified standard for evaluating their performance. Developing standardized benchmarks is crucial for the advancement of algorithms. For example, in protein structure prediction, the PoseBusters benchmark introduced by AlphaFold3[16] has been widely adopted for assessing various AI tools.

In this work, we compile superconducting materials newly reported during the 2024–2025 period, including X_2YH_6 systems, perovskite-type MXH_3 structures, fluorite-type M_3XH_8 compounds, cage-like BCN-based materials evolved from the LaH_{10} structure with metal atom doping, and two-dimensional honeycomb structures derived from MgB_2. Additionally, we highlight several approaches based on physical intuition for the inverse design of high-T_c superconductors. A retrospective analysis of these strategies offers valuable insights for future efforts in discovering materials with even higher superconducting transition temperatures.

2. Methods

In this section, we present the establishment process of HTSC-2025. As illustrated in Fig. 1(a), our construction process initiates with an extensive literature review of theoretically predicted superconductors. Considering the essential requirements of timeliness, high application value, and rapid validation for benchmark systems, we specifically focus on high-temperature superconducting materials discovered since 2024 that exhibit ambient-pressure superconductivity. To enable efficient verification of AI performance, we prioritize materials identified through BCS superconducting mechanisms, as this selection ensures that AI predictions can be rapidly validated through density functional theory (DFT) calculations. Subsequently, we systematically collect crystal structure information (CIF files) and corresponding theoretically predicted superconducting transition temperatures through multiple approaches: expert manual construction, automated script replacement, and direct communication with original authors via email. Finally, through systematic literature analysis, we categorize these materials into specific crystalline

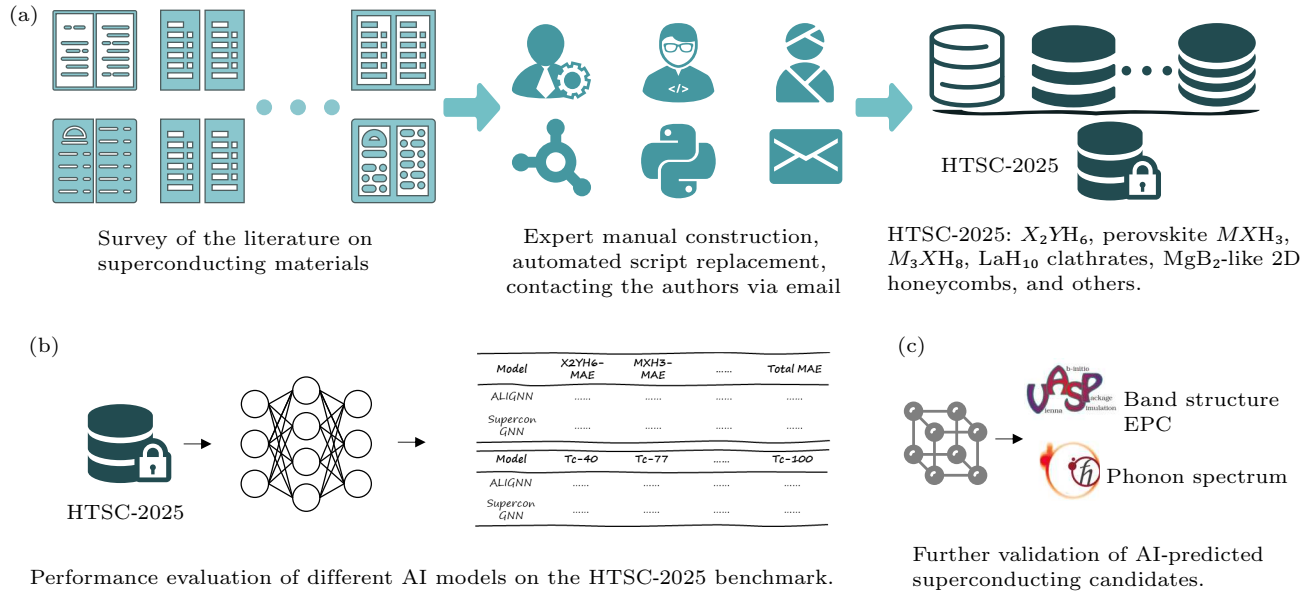


Fig. 1. Construction and application of the HTSC-2025 benchmark. (a) Workflow of HTSC-2025 benchmark dataset construction, starting from a literature review of theoretically predicted ambient-pressure high-T_c superconductors (since 2024), focusing on BCS-type candidates for rapid DFT validation. Crystal structures and T_c values are collected via expert curation, scripts, and author correspondence, and categorized into representative structural types for AI evaluation. (b) Evaluation pipeline for AI models on HTSC-2025: Training data exclusion prevents leakage, T_c prediction from crystal structures, MAE analysis (system-specific/overall), success rates across T_c intervals highlight high-T_c identification. (b) Evaluation pipeline for AI models on HTSC-2025: To prevent data leakage, candidate materials in the test set are rigorously excluded from training data based on structural similarity. AI models convert crystal structures into graph-based or structured representations and predict superconducting T_c values. Performance is evaluated using mean absolute error, both within individual crystal structure systems (e.g., X_2YH_6, MXH_3) and across the full dataset. Additionally, success rates are calculated for high-T_c predictions within key temperature intervals, particularly above the liquid nitrogen threshold ($T_c > 77$ K) and the current theoretical upper limit ($T_c > 160$ K), to assess the model's practical effectiveness in discovering technologically relevant superconductors. (c) First-principles calculations for validating AI-predicted new superconducting materials, including crystal structure optimization, phonon spectrum evaluation, and electron–phonon coupling analysis.

systems: X_2YH_6-type structures, perovskite MXH_3-type structures, M_3XH_8-type structures, LaH_8-type clathrates, MgB_2-like 2D honeycomb lattices, and other structures, thereby establishing a structured framework for subsequent AI performance evaluation.

3. Data records

In this section, we analyze the ambient-pressure superconducting materials and their mechanisms within HTSC-2025. Notably, HTSC-2025 does not include all literature-reported materials but instead retains approximately 50% of them as representative candidates. These were selected through targeted sampling based on the distribution of superconducting transition temperatures, aiming to ensure that the dataset captures the typical characteristics of high-T_c superconductors. This selective curation helps balance data coverage and quality, and is designed to enhance the generalization ability of AI models by avoiding redundancy and ensuring exposure to related materials beyond the benchmark itself.

Figure 2(a) presents a histogram of the T_c distribution, where the vertical axis indicates the number of materials within each T_c interval. The HTSC-2025 dataset contains a total of 155 samples, with an average T_c of 30.3 K. Notably, more than half of the crystals exhibit a T_c exceeding 20 K, highlighting the dataset's focus on evaluating materials with relatively high T_c values. Figure 2(b) shows a heatmap overlaid on the periodic table, in which the color intensity represents the frequency of each element appearing in HTSC-2025. Hydrogen is the most frequently occurring element, present in 79 compounds, underscoring the significant potential of hydrides in achieving high-temperature superconductivity. Figure 2(c) illustrates representative materials from different material systems. Table 1 presents the average T_c, maximum T_c, corresponding chemical formula, space group, and the number of materials for each material system.

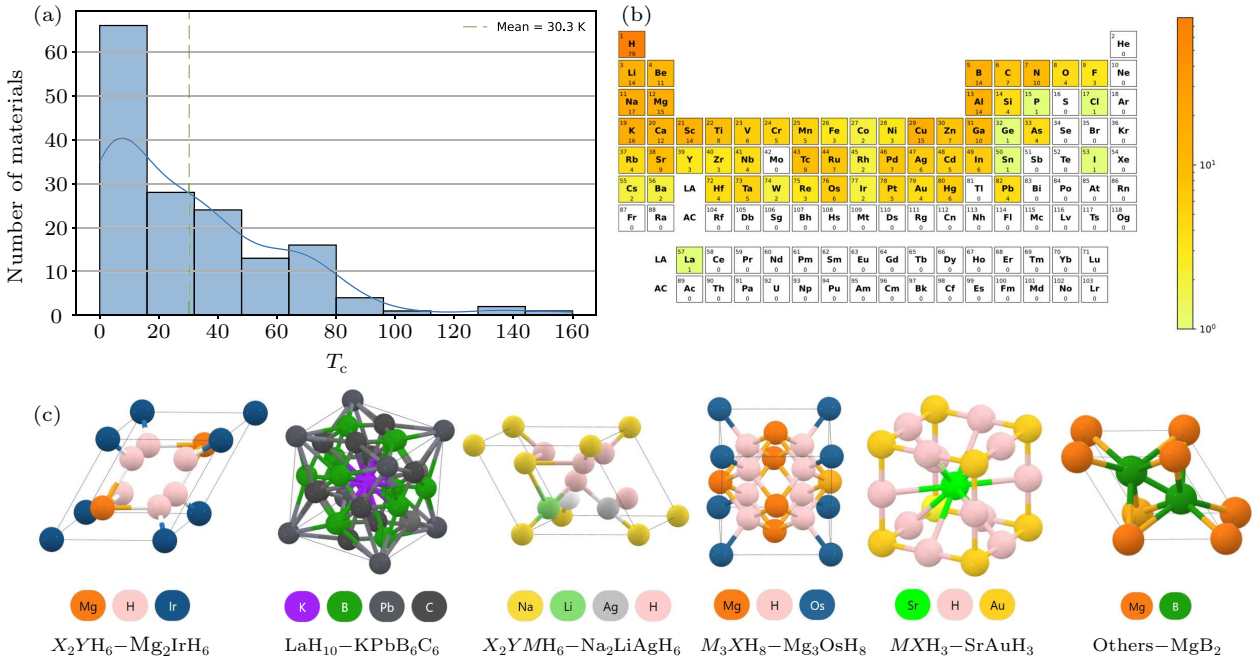


Fig. 2. Statistics and elemental composition of the HTSC-2025 benchmark. (a) Histogram of superconducting transition temperatures (T_c) for 155 materials in HTSC-2025, with an average T_c of 30.3 K. (b) Elemental heatmap on the periodic table based on occurrence frequency; hydrogen appears most frequently, highlighting the prominence of hydrides. (c) Representative high-T_c materials from different material systems.

Table 1. Crystalline materials included in the HTSC-2025 benchmark.

Material class	Average T_c (K)	Max T_c formula	Space group	Max T_c (K)	Number
X_2YH_6	55.4	Mg_2IrH_6	$Fm\bar{3}m$	160	19
LaH_{10}	53.0	$KPbB_6C_6$	$Pm\bar{3}$	88	12
X_2YMH_6	35.5	Na_2LiAgH_6	$Fm\bar{3}m$	86	23
MXH_3	35.3	$SrAuH_3$	$Pm\bar{3}m$	132	15
LiMBC	58.0	$Li_3KB_6C_6$	$P6_3/mmc$	73	15
M_3XH_8	20.4	Mg_3OsH_8	$Pm\bar{3}m$	73	18
Others	8.0	MgB_2	$P6_3/mmm$	40.0	53
Total	30.3	Mg_2IrH_6	$Fm\bar{3}m$	160	155

The primary objective of HTSC-2025 is to advance the development of AI algorithms for predicting superconducting transition temperatures, thereby accelerating the discovery of novel superconducting materials. Summarizing the commonly used design principles in recent theoretical high-T_c predictions helps build physical intuition and provides useful guidance for future AI-driven exploration. As illustrated in Fig. 2(c), the discovery of Mg_2IrH_6 (T_c = 160 K)[17] has inspired researchers to conduct more detailed investigations into X_2YH_6 systems. Many systems with superconducting T_c exceeding 20 K — and even reaching up to 70 K — have been dis-

covered. Notable examples include Li_2CuH_6 (T_c = 86 K), Mg_2PtH_6 (T_c = 78 K), Mg_2PdH_6 (T_c = 63 K), Mg_2RhH_6 (T_c = 53 K),[10] as well as the recently AI-discovered Li_2AuH_6 (T_c = 140 K).[12] The superconducting mechanisms of these structures appear to be similar. The electronic structure of X_2YH_6 compounds reveals the presence of van Hove singularities near the Fermi level, which may contribute to their high T_c. In addition, the significant contribution of hydrogen 1s orbitals to the electronic density of states enables strong coupling between high-frequency phonon modes and electrons near the Fermi level, thereby enhancing the overall electron–phonon coupling strength.

Inspired by studies on X_2YH_6 compounds, fluorite-type M_3XH_8 compounds (Fig. 2(c)) with the $Pm\bar{3}m$ space group, where M = Li, Na, Mg, Al, K, Ca, Ga, Rb, Sr, and In, and X denotes 3d, 4d, or 5d transition metals, have also attracted considerable attention. A total of 29 dynamically stable compounds have been identified. The band structure of Mg_3OsH_8 (T_c = 73 K)[18] shows three bands crossing the Fermi level, indicating its metallic nature. The electronic density of states of hydrogen is uniformly distributed near the Fermi level, facilitating electron–phonon coupling. Furthermore, the significant phonon density of states contribution from hydrogen atoms in the high-frequency phonon region further enhances this coupling.

The high-throughput computational discovery of $ZnHCr_3$ (T_c = 30 K), $ZnHAl_3$ (T_c = 80 K),[19] and $MgHCu_3$ (T_c = 42 K)[20] has sparked increasing interest in the superconducting properties of cubic perovskite MXH_3 systems.[21] Miguel *et al.* have also identified a series of superconducting materials[10,22,23] within the MXH_3 family, such as $KInH_3$ (T_c = 73 K), $KAlH_3$ (T_c = 52 K), $AlHgH_3$ (T_c = 28 K), $PbHgH_3$ (T_c = 25 K), $KCaH_3$ (T_c = 23 K), and $PbOsH_3$ (T_c = 23 K). Recently, Li *et al.* systematically explored perovskite hydrides by selecting alkali and post-transition metals as the M-site elements and 3d, 4d, and 5d transition metals as the X-site elements. Through high-throughput calculations, they identified 17 dynamically stable perovskite hydrides.[24] Notably, $SrAuH_3$ (T_c = 132 K) and $SrZnH_3$ (T_c = 107 K) exhibit exceptionally high superconducting transition temperatures, highlighting the great potential of this family of materials for achieving high-T_c superconductivity. Multiple electronic bands cross the Fermi level. The 5d orbitals of Au contribute significantly near the Fermi level, forming a pronounced van Hove singularity, which may enhance the density of states at the Fermi level and thereby promote superconductivity. More recently, additional superconducting materials within this system have continued to be discovered,[25] such as $KScH_3$ (T_c = 40 K), $RbScH_3$ (T_c = 32 K), and $CsScH_3$ (T_c = 18 K).

The dynamical stability of cage-like hydrides such as CaH_6,[26,27] YH_6,[28,29] CeH_{10},[30] and LaH_{10}[31,32] requires high-pressure conditions, which limits their synthesis and practical applications. However, the conventional superconducting mechanisms observed in these materials have inspired physicists to explore the possibility of achieving high-temperature superconductivity within BCS-type materials. Given that B and C are the lightest elements capable of forming strong covalent bonds, and considering the compositional diversity of guest metal atoms, significant progress has been made in substituting hydrogen in cage-like hydride structures with these elements. The hexagonal cage-structured XB_8C compounds (X = Ca, Sr, Ba) exhibit superconductivity with T_c = 77.1 K, 64.4 K, and 53.2 K, respectively.[33] First-principles calculations reveal a metallic ground state characterized by multiple bands crossing the Fermi level. The coexistence of flat and dispersive bands near Fermi level leads to an enhanced electron–phonon coupling strength, providing a mechanism for the observed elevated T_c values. Phonon spectrum calculations show that the electron–phonon coupling is mainly concentrated in the high-frequency optical modes.

$La(BN)_5$ and $Y(BN)_5$[34] adopt cage-like crystal structures analogous to that of LaH_{10}, in which boron and nitrogen atoms substitute for hydrogen to reduce the high pressure typically required for maintaining dynamical stability. Their superconducting transition temperatures are 69 K and 59 K, respectively. A similar structure, AlB_6N_6, exhibits a predicted T_c of 47 K.[35] Quaternary superconductors, including $RbBaB_6C_6$ (T_c = 68 K)[36] and $KPbB_6C_6$ (T_c = 88 K),[37] exhibit crystal structures primarily composed of a covalent B–C framework, with metal atoms doped as guest species. Recent studies[38] have shown that a reinsertion strategy involving metals such as Na, K, Mg, and Ca to construct quaternary Li_xM_yBC systems can effectively stabilize layered structures, suppress defects, and enhance electron–phonon coupling. Theoretically, the superconducting transition temperature can reach as high as 73 K, as predicted for compounds such as $Li_3KB_6C_6$.

By doping metals into the boron-nitride cage $(BN)_5$, four stable superconductors $M(BN)_5$ (M = Na, Al, In, Tl) are obtained, with corresponding T_c of 8 K, 22 K, 15 K, and 15 K, respectively. The electronic band structure of $Al(BN)_5$[39] reveals multiple bands crossing the Fermi level, giving rise to both electron-type and hole-type Fermi surfaces. The electronic states at the Fermi surface are primarily derived from the p orbitals of B and N atoms, as well as the p orbitals of the Al atom, which is favorable for Cooper pair formation and enhances electron–phonon coupling. Atomic doping is an effective approach to enhance the superconducting transition temperature. The transition from insulator to metal and the induction of high-temperature superconductivity are achieved by doping guest atoms (such as Li and Mg) into zinc blende-

type structures (XY_4Z_4, where $Y_4Z_4 = B_4N_4$, Si_4C_4, B_4P_4).[39] Notably, LiB_4N_4 and MgB_4P_4 exhibit superconducting transition temperatures as high as 67 K and 45 K, respectively, surpassing that of the conventional high-temperature superconductor MgB_2. The elevated T_c values in these compounds are attributed to a high electronic density of states at the Fermi level and the softening of low-frequency acoustic phonon modes, which significantly enhance the electron–phonon coupling strength. By introducing potassium atoms into the metal borohydride $Ca(BH_4)_2$, strong electron–phonon coupling between B–H σ molecular orbitals and bond-stretching phonon modes can be achieved. At a doping concentration of 0.10, the transition temperature reaches a maximum value of 130 K.[41]

In addition to the material systems mentioned above, we also observe progress in physics-inspired strategies for discovering high-temperature superconductors, such as hole doping, introducing light elements to form strong covalent bonds, and tuning spin–orbit coupling. The compound $(BN)_5$ exhibits insulating behavior in its pristine state. However, hole doping induces a downward shift of the Fermi level, leading to the metallization of the sp^3 hybridized σ-bonding band. This electronic transition is accompanied by the emergence of high-temperature superconductivity with a T_c reaching 96 K.[42] With increasing hole concentration, although the overall electronic structure remains largely unchanged, strong coupling arises between mid-frequency optical phonon modes and the σ-electrons, contributing significantly to the total electron–phonon coupling.

At ambient pressure, MgB_2C_2 and hole-doped NaBC exhibit potential for high-temperature superconductivity.[43] The hole concentration can be tuned via thermal deintercalation methods. For instance, low-temperature sodium deintercalation in NaBC effectively suppresses the formation of defects in BC layers and preserves strong electron–phonon coupling. The deintercalated layered structures (e.g., $Na_{3/4}BC$ and $Mg_{2/3}B_2C_2$) remain stable under kinetic constraints, and the planarity of the BC layers is crucial to superconductivity. Excessive buckling, such as in $Na_{2/3}BC$ where $d_{avg} = 0.12$ Å, disrupts the coupling between σ-band electrons and high-frequency bond-stretching phonons ($\tilde{6}5$ meV), significantly reducing the T_c. These hole-doped materials exhibit strong electron–phonon coupling strengths ($\lambda \sim 0.95$–1.32) and a characteristic two-gap superconducting behavior similar to MgB_2, with T_c values ranging from 43 K to 88 K, the highest being observed in $Na_{7/8}BC$.

The two-dimensional honeycomb structure KB_2C_2 ($T_c = 153$ K)[44] is designed based on the AlB_2-type structure of MgB_2.[45] By introducing carbon atoms to form strong covalent B–C σ bonds, a strong coupling between electrons and high-frequency in-plane phonon vibrations is achieved. Its multigap nature (coexistence of σ and π states) further facilitates Cooper pair formation. Biaxial tensile strain induces phonon mode softening and significantly enhances the electron–phonon coupling constant λ, far exceeding that of conventional two-dimensional superconductors. Similarly, by replacing the B–B surface layer of MgB_2 with a B–C layer, the two-dimensional material $Mg_2B_4C_2$ exhibits high-temperature superconductivity, with a predicted T_c estimated to be around 47–48 K.[46]

At ambient pressure, $I4mm$-Mg_2BN, Cm-Mg_2BN, $Cmmm$-MgB_2N, and $R3m$-Mg_2BN exhibit superconducting transition temperatures T_c of 31 K, 19 K, 11 K, and 4.5 K,[47] respectively. The low-frequency lattice vibrations primarily contributed by Mg and B atoms play a dominant role in electron–phonon coupling, serving as a key factor for superconductivity. Among these phases, $I4mm$-Mg_2BN exhibits the strongest EPC ($\lambda = 1.33$), corresponding to its highest T_c (31 K), whereas $R3m$-Mg_2BN shows the weakest EPC ($\lambda = 0.38$).

The kagome metals $Rh_3M_2S_2$ (M = Pb, In, Tl) exhibit the coexistence of superconductivity and topological states.[48] These materials are weak superconductors, with superconducting transition temperatures T_c of 1.03 K (Pb), 2.31 K (In), and 5.39 K (Tl), respectively. Notably, spin–orbit coupling has a significant influence on T_c.

Most of AI-based studies on superconducting T_c prediction have claimed significant improvements in accuracy.[8,13–15] However, the field still lacks a unified, open, and extensible benchmark, which hinders fair comparisons between models and limits further development. In this work, we propose HTSC-2025, a benchmark dataset for high-T_c superconductors under ambient pressure. It features timely coverage (materials predicted from 2023 to May 2025), practical value (ambient pressure and high T_c focus), quick verifiability (compatible with BCS theory and DFT), and broad scope (including X_2YH_6, perovskite-type MXH_3, M_3XH_8 systems, etc.). We also review a series of physically inspired design strategies, highlighting that hole doping, the introduction of light atoms forming strong covalent bonds, and spin–orbit coupling engineering can enhance T_c. HTSC-2025 is publicly released and will be continuously maintained. It provides a reproducible and quantitative basis for cross-model evaluation and offers physically interpretable templates to support inverse design and the discovery of new high-T_c superconductors.

Recently, AI has driven significant progress in the discovery of high-pressure superconductors. For example, Jiang *et al.*[49] developed a T_c prediction model under high pressure and identified 14 new clathrate hydride prototypes. Among them, 11 ternary clathrate structures were predicted to exhibit T_c values above 250 K at 300 GPa, with Li_2NaH_{17} and ThY_2H_{24} approaching room-temperature T_c (297 K and 303 K, respectively) at lower pressures. Wang *et al.*[50]

also identified 144 high-pressure superconductors using AI tools. However, HTSC-2025 does not include these materials. This reflects a balance between practical and academic values: while high-pressure superconductors offer insights into achieving room-temperature T_c, their extreme pressure conditions limit real-world applications. HTSC-2025 also excludes unconventional superconductors, as the lack of a unified theoretical framework makes it difficult to verify AI-predicted candidates or to assess benchmark quality. We hope to extend HTSC-2025 to these domains as scientific understanding advances.

In the future, we will continuously expand and update the HTSC-2025 benchmark, develop related AI algorithms, and conduct extensive evaluations on this benchmark to facilitate the discovery of high-temperature superconductors. In addition, we will also explore physics-inspired approaches to discover new high-T_c superconducting materials.

4. Usage notes

As illustrated in Fig. 1(b), prior to evaluating AI models on the HTSC-2025 benchmark, the training dataset must rigorously exclude these materials to prevent data leakage. The AI models then convert crystal data into graph representations or other structured formats for input, subsequently predicting the superconducting T_c of the materials. To quantify predictive performance, we calculate the mean absolute error (MAE) between model predictions and DFT-computed results. For comprehensive evaluation, the test set is partitioned according to the aforementioned crystalline systems (X_2YH_6, MXH_3, etc.), enabling both system-specific MAE calculations and total MAE. Furthermore, given the greater application value of high-temperature superconductors compared to their low-temperature counterparts, we introduce an additional evaluation metric: the prediction success rate across different critical temperature intervals. This stratified analysis aims to quantify the performance of AI in identifying materials within specific T_c ranges, with particular focus on technologically critical regimes such as the McMillan limit and the liquid nitrogen temperature range.

4.1. Data validation

As shown in Fig. 1(c), the performance of AI models is often further validated by their ability to discover new high-temperature superconducting materials. For AI-recommended candidate superconductors, the BCS theory provides a convenient framework for verification using first-principles methods, such as those implemented in VASP[51] and Quantum Espresso.[52] According to BCS theory, superconductivity arises from the formation of Cooper pairs–electron pairs bound via phonon-mediated interactions that overcome Coulomb repulsion. The collective condensation of these Cooper pairs near the Fermi surface opens a superconducting energy gap, resulting in the characteristic zero-resistance state. To verify whether a candidate material follows the BCS superconducting mechanism, a systematic first-principles computational protocol is employed, including crystal structure optimization, phonon spectrum analysis, and electron–phonon coupling calculations. Specifically, for validating superconducting properties, a computational scheme based on density functional theory and density functional perturbation theory can be employed. The exchange–correlation functional is treated using the PBE-GGA form,[53] with optimized norm-conserving Vanderbilt pseudopotentials.[54] Taking the X_2YH_6 system as an example, the kinetic energy cutoff is set to 80 Ry, and the charge density cutoff to 320 Ry. Brillouin zone sampling is performed using an unshifted $16 \times 16 \times 16$ $\boldsymbol{k}$-mesh. Electronic occupations are treated with Methfessel–Paxton smearing of 0.02 Ry width.[55] Phonon properties are calculated using DFPT on a $4 \times 4 \times 4$ $\boldsymbol{q}$-mesh.[56]

Data availability statement

The data that support the findings of this study are openly available in Science Data Bank at the following link https://doi.org/10.57760/sciencedb.j00113.00259.

To support the discovery of superconducting materials and the evaluation of AI tools, we open-source this benchmark on Hugging Face at https://huggingface.co/datasets/xiaoqi/HTSC-2025.

Acknowledgments

Project supported by the National Natural Science Foundation of China (Grant Nos. 62476278, 12434009, and 12204533), the National Key R&D Program of China (Grant No. 2024YFA1408601), and the Innovation Program for Quantum Science and Technology (Grant No. 2021ZD0302402). Computational resources have been provided by the Physical Laboratory of High Performance Computing at Renmin University of China.

References

[1] Onnes H K 1911 *Commun. Theor. Phys.* **120**
[2] arXiv https://arxiv.org/
[3] Coombs T A, Wang Q, Shah A, Hu J, Hao L, Patel I, Wei H, Wu Y, Coombs T and Wang W 2024 *Nat. Rev. Electr. Eng.* **1** 788
[4] Yao C and Ma Y 2021 *iScience* **24** 102541
[5] Pustogow A, Luo Y, Chronister A, *et al.* 2019 *Nature* **574** 72
[6] Bruzzone P, Fietz W H, Minervini J V, Novikov M, Yanagi N, Zhai Y and Zheng J 2018 *Nucl. Fusion* **58** 103001
[7] Han X Q, Wang X D, Xu M Y, Feng Z, Yao B W, Guo P J, Gao Z F and Lu Z Y 2025 *Chin. Phys. Lett.* **42** 027403
[8] Choudhary K and Garrity K 2022 *NPJ Comput. Mater.* **8** 244
[9] Choudhary K and DeCost B 2021 *NPJ Comput. Mater.* **7** 185
[10] Cerqueira T F T, Fang Y W, Errea I, Sanna A and Marques M A L 2024 *Adv. Funct. Mater.* **34** 2404043
[11] Han X Q, Ouyang Z F, Guo P J, Sun H, Gao Z F and Lu Z Y 2024 *Chin. Phys. Lett.* **42** 047301

[12] Ouyang Z F, Yao B W, Han X Q, Guo P J, Gao Z F and Lu Z Y 2025 *Phys. Rev. B* **111** L140501
[13] Gibson J B, Hire A C, Dee P M, Barrera O, Geisler B, Hirschfeld P J and Hennig R G 2025 *npj Comput. Mater.* **11** 7
[14] Li J, Fang W Q, Jin S J, Zhang T D, Wu Y L, Xu X D, Liu Y and Yao D X 2025 *AI Sci.* **1** 015001
[15] Han X Q, Guo P J, Gao Z F, Sun H and Lu Z Y 2025 arXiv:2505.09203 [cond-mat.mtrl-sci]
[16] Abramson J, Adler J and Dunger J 2024 *Nature* **630** 493
[17] Dolui K, Conway L J, Heil C, Strobel T A, Prasankumar R P and Pickard C J 2024 *Phys. Rev. Lett.* **132** 166001
[18] Li B, Fan Y X, Zhai J J, Cao Z S, Zhu C, Cheng J, Liu S L and Shi Z X 2025 *J. Phys. Chem. C* **129** 7098
[19] Liu S M, Shi J J, He Y, Tian C, Zhu Y H, Wang X Q and Zhong H X 2024 *Adv. Funct. Mater.* **34** 2315386
[20] Tian C, He Y, Zhu Y H, Du J, Liu S M, Guo W H, Zhong H X, Lu J, Wang X Q and Shi J J 2024 *Adv. Funct. Mater.* **34** 2304919
[21] He Y, Lu J, Wang X Q and Shi J J 2023 *Phys. Rev. B* **108** 054515
[22] Cerqueira T F T, Sanna A and Marques M A L 2024 *Adv. Mater.* **36** 2307085
[23] Wines D and Choudhary K 2024 *Mater. Futures* **3** 025602
[24] Li B, Zhu C, Zhai J J, Yin C H, Fan Y X, Cheng J, Liu S L and Shi Z X 2024 *Phys. Rev. B* **110** 214504
[25] Quan H, Shi X B, Han Y L, Zhang P and Wang B T 2025 *Phys. Rev. B* **111** 134509
[26] Wang H, Tse J S, Tanaka K, Iitaka T and Ma Y M 2012 *Proc. Natl. Acad. Sci. USA* **109** 6463
[27] Ma L, Wang K, Xie Y, Yang X, Wang Y Y, Zhou M, Liu H Y, Yu X H, Zhao Y S, Wang H B, Liu G T and Ma Y M 2022 *Phys. Rev. Lett.* **128** 167001
[28] Li Y W, Hao J, Liu H Y, Tse J S, Wang Y C and Ma Y M 2015 *Sci. Rep.* **5** 9948
[29] Troyan I A, Semenok D V, Kvashnin A G, Sadakov A V, Sobolevskiy O A, Pudalov V M and Oganov A R 2021 *Adv. Mater.* **33** 2006832
[30] Chen W H, Semenok D V, Huang X L, Shu H Y, Li X, Duan D F, Cui T and Oganov A R 2021 *Phys. Rev. Lett.* **127** 117001
[31] Somayazulu M, Ahart M, Mishra A K, Geballe Z M, Baldini M, Meng Y, Struzhkin V V and Hemley R J 2019 *Phys. Rev. Lett.* **122** 027001
[32] Drozdov A P, Kong P P, Minkov V S, Besedin S P, Kuzovnikov M A, Mozaffari S, Balicas L, Balakirev F F, Graf D E, Prakapenka V B, Greenberg E, Knyazev D A, Tkacz M and Eremets M I 2019 *Nature* **569** 528
[33] Li B, Fan Y X, Yin C H, Zhai J J, Zhu C, Cao Z S, Cheng J and Liu S L 2025 *Phys. Rev. B* **111** 014510
[34] Ding H B, Feng Y J, Jiang M J, Tian H L, Zhong G H, Yang C L, Chen X J and Lin H Q 2022 *Phys. Rev. B* **106** 104508
[35] Li X, Yong X, Wu M, Lu S Y, Liu H Y, Meng S, Tse J S and Li Y W 2019 *J. Phys. Chem. Lett.* **10** 2554
[36] Duan Q Z, Zhan L H, Shen J Y, Zhong X and Lu C 2024 *Phys. Rev. B* **109** 054505
[37] Geng N, Hilleke K P, Zhu L, Wang X Y, Strobel T A and Zurek E 2023 *J. Am. Chem. Soc.* **145** 1696
[38] Tomassetti C R, Kafle G P, Marcial E T, Margine E R and Kolmogorov A N 2024 *J. Mater. Chem. C* **12** 4870
[39] Chen C, Miao J Y, Zuo J N, He S and Lu C 2024 *Phys. Rev. B* **110** 174514
[40] Jiang Q W, Chen L, Ma H, Li C D, Duan D F and Cui T 2025 *Inorg. Chem.* **64** 400
[41] Di Cataldo S and Boeri L 2023 *Phys. Rev. B* **107** L060501
[42] Han Y L, Jiang K Y, Wang B T, Zhang P and Lu H Y 2025 *Phys. Rev. B* **111** 094520
[43] Tomassetti C R, Gochitashvili D, Renskers C, Margine E R and Kolmogorov A N 2024 *Phys. Rev. Mater.* **8** 114801
[44] Liu H D, Fu X P, Fu Z G, Lu H Y and Zhang P 2025 *Phys. Rev. B* **111** 184502
[45] Nagamatsu J, Nakagawa N, Muranaka T, Zenitani Y and Akimitsu J 2001 *Nature* **410** 63
[46] Singh S, Romero A H, Mella J D, Eremeev V, Muñoz E, Alexandrova A N, Rabe K M, Vanderbilt D and Muñoz F 2022 *npj Quantum Mater.* **7** 37
[47] Jiang J J, Xue Y M, Zha L L, Yao S W, Wang B, Hu W J, Peng L, Shi T T, Chen J, Liu X L and Lin J 2025 *J. Mater. Chem. C* **13** 9799
[48] Liu W Y, Wang X J, Li Y Q, Wei Y H, Zhong M M and Kuang M Q 2025 *Supercond. Sci. Technol.* **38** 045024
[49] Jiang B, Luo X S, Sun Y, Zhong X, Lv J, Xie Y, Ma Y M and Liu H Y 2025 *Phys. Rev. B* **111** 054505
[50] Wang X Y, Zhang C Q, Wang Z Y, Liu H Y, Lv J, Wang H, E W N and Ma Y M 2025 arXiv:2502.16558 [cond-mat.supr-con]
[51] Wang V, Xu N, Liu J C, Tang G and Geng W T 2021 *Comput. Phys. Commun.* **267** 108033
[52] Quantum ESPRESSO https://www.quantum-espresso.org/
[53] Perdew J P, Burke K and Ernzerhof M 1996 *Phys. Rev. Lett.* **77** 3865
[54] Hamann D R 2013 *Phys. Rev. B* **88** 085117
[55] Methfessel M and Paxton A T 1989 *Phys. Rev. B* **40** 3616
[56] Baroni S, de Gironcoli S, Dal Corso A and Giannozzi P 2001 *Rev. Mod. Phys.* **73** 515