

Review

Open Access



# Machine learning for predictive design and optimization of high-performance thermoelectric materials: a review

Yuelin Wang<sup>1,2,#</sup>, Chengquan Zhong<sup>1,2,#</sup>, Jingzi Zhang<sup>1,2</sup>, Jiakai Liu<sup>1,2,3</sup>, Kailong Hu<sup>\*1,2</sup>, Junjie Chen<sup>4,\*</sup>, Xi Lin<sup>1,2,5,\*</sup>

<sup>1</sup>School of Materials Science and Engineering, Harbin Institute of Technology, Shenzhen 518055, Guangdong, China.

<sup>2</sup>Blockchain Development and Research Institute, Harbin Institute of Technology, Shenzhen 518055, Guangdong, China.

<sup>3</sup>Sunrise (Xiamen) Photovoltaic Industry Co., Ltd, Xiamen 361000, Fujian, China.

<sup>4</sup>Department of Computer Science and Technology, Harbin Institute of Technology, Shenzhen 518055, Guangdong, China.

<sup>5</sup>State Key Laboratory of Advanced Welding and Joining, Harbin Institute of Technology, Harbin 150001, Heilongjiang, China.

# Authors contributed equally.

**\*Correspondence to:** Prof. Xi Lin, Prof. Kailong Hu, School of Materials Science and Engineering, Harbin Institute of Technology, HIT Campus of University Town of Shenzhen, Shenzhen 518055, Guangdong, China. E-mail: [linxi@hit.edu.cn](mailto:linxi@hit.edu.cn); [hukailong@hit.edu.cn](mailto:hukailong@hit.edu.cn); Prof. Junjie Chen, Department of Computer Science and Technology, Harbin Institute of Technology, HIT Campus of University Town of Shenzhen, Shenzhen 518055, Guangdong, China. E-mail: [junjiechen@hit.edu.cn](mailto:junjiechen@hit.edu.cn)

**How to cite this article:** Wang, Y.; Zhong, C.; Zhang, J.; Liu, J.; Hu K.; Chen, J.; Lin, X. Machine learning for predictive design and optimization of high-performance thermoelectric materials: a review. *J. Mater. Inf.* 2025, 5, 41. <http://dx.doi.org/10.20517/jmi.2025.18>

**Received:** 27 Mar 2025 **First Decision:** 9 May 2025 **Revised:** 22 May 2025 **Accepted:** 27 May 2025 **Published:** 24 Jun 2025

**Academic Editor:** Ke-Qiu Chen **Copy Editor:** Pei-Yun Wang **Production Editor:** Pei-Yun Wang

## Abstract

Thermoelectric materials enabling direct interconversion between thermal and electrical energy hold transformative potential for sustainable energy technologies, particularly in solid-state power generation and precision refrigeration systems. The pursuit of high-performance thermoelectric materials with exceptional energy conversion efficiency has remained a persistent challenge in materials science, primarily constrained by the resource-intensive nature of traditional experimental approaches and computationally demanding first-principles simulations. The emergence of machine learning (ML) techniques has revolutionized this field by enabling rapid screening of material candidates and establishing quantitative structure-property relationships. This comprehensive review systematically examines cutting-edge methodologies in ML-driven thermoelectric materials research, with particular emphasis on three pivotal aspects: (1) predictive modeling of key performance parameters including electrical conductivity, Seebeck coefficient, and lattice thermal conductivity through advanced feature engineering and algorithm selection; (2) inverse design strategies for optimizing carrier concentration and phonon scattering mechanisms; (3) application-specific



© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, sharing, adaptation, distribution and reproduction in any medium or format, for any purpose, even commercially, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.



material optimization frameworks integrating multi-objective constraints. Furthermore, we critically analyze prevailing challenges in data quality, model interpretability, and cross-scale prediction accuracy, while proposing future research directions encompassing active learning paradigms, generative adversarial networks for virtual material synthesis, and hybrid physics-informed ML architectures.

**Keywords:** Thermoelectric, machine learning, electrical conductivity, thermal transport properties

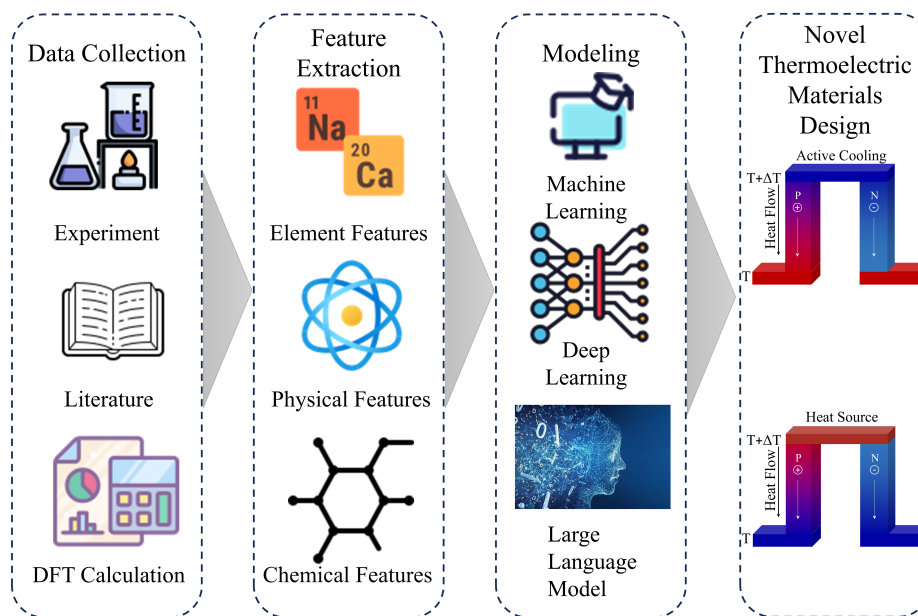
## INTRODUCTION

Thermoelectric materials are a novel type of energy material capable of converting thermal energy into electrical energy and vice versa through the Seebeck effect and Peltier effect<sup>[1–8]</sup>. They are widely applied across various fields such as waste heat recovery, solid-state cooling, portable power sources, remote sensor power supply, thermal management, space exploration, automotive industry, medical equipment, environmental monitoring, and small-scale power generation, demonstrating their high efficiency and versatility in energy conversion and temperature control<sup>[9–15]</sup>. Thermoelectric devices demonstrate their unique value and potential across various applications with advantages such as noiselessness, vibration-free operation, compact design, high reliability, environmental friendliness, high efficiency, multifunctionality, ease of control, and low maintenance costs<sup>[16–24]</sup>. Typically, the performance of a thermoelectric material is measured by its figure of merit, known as  $zT$ , which is formulated as follows:

$$zT = \frac{S^2 \sigma T}{\kappa_L + \kappa_e} \quad (1)$$

where  $S$  is the Seebeck coefficient<sup>[25–29]</sup>;  $T$  is the absolute temperature in Kelvin;  $\sigma$  is the electrical conductivity<sup>[30–33]</sup>; and  $\kappa_L$  and  $\kappa_e$  signify the lattice thermal conductivity and electronic thermal conductivity, respectively<sup>[34–43]</sup>.

In past explorations, experiments have been the primary means by which researchers have investigated the performance of thermoelectric materials. Current experimental approaches to improve the performance of thermoelectric materials include optimizing carrier concentration, modulating band structure, facilitating multi-scale phonon scattering, employing defect engineering, optimizing lattice dynamics, conducting interface engineering, regulating electronic structure, designing thermoelectric modules, implementing dynamic atomic control, and exploring new materials, among other strategies<sup>[44–54]</sup>. Experimental methods and outcomes of these approaches are shown in Supplementary Figure 1. Experimental approaches can often be costly, time-consuming, and prone to issues such as reproducibility challenges, limited data availability, and human error. To address these challenges, theoretical calculations based on physical principles have been increasingly employed over the past few decades to simulate experimental conditions and derive methods for assessing materials' thermal performance. These approaches include first-principles calculations, the Boltzmann transport equation, the Wiedemann-Franz law, molecular dynamics simulations, and the Monte Carlo method. Such techniques enable a comprehensive prediction and optimization of the electronic structure, transport properties, and thermoelectric conversion efficiency of materials<sup>[55–59]</sup>. In recent years, the rapid advancement of machine learning (ML) technology has significantly impacted materials science, emerging as a critical driver for the discovery and optimization of new materials<sup>[60–62]</sup>. ML pipeline for discovering novel thermoelectric material as shown in Figure 1. The role of ML in material exploration is increasingly prominent, introducing transformative changes to the field. In particular, ML applications have produced substantial results. For example, ML is utilized to predict the mechanical properties of alloys, guiding lightweight design in the aviation and automotive industries<sup>[63,64]</sup>. In the energy sector, ML technology optimizes the electronic structure of solar cell materials, improving energy conversion efficiency<sup>[64,65]</sup>. In addition, intelligent algorithms predict and control the stability of perovskite materials, which advances the development of efficient optoelectronic de-



**Figure 1.** ML pipeline for discovering novel thermoelectric material. ML: Machine learning.

vices<sup>[66–69]</sup>. Similarly, ML plays a crucial role in predicting the performance of thermoelectric materials. Using experimental and computational data, ML models can forecast promising thermoelectric materials, guiding experimental and theoretical research. This approach significantly improves research efficiency, accelerating the discovery and optimization of advanced thermoelectric materials. The number of articles published from 2014 to 2024 with “Thermoelectric” as a keyword and those with “Thermoelectric-Machine Learning” as keywords is shown in Figure 2; it can be seen that the number of articles exploring thermoelectric materials using ML methods has increased year by year over this period. This indicates that the importance of ML in exploring the properties of thermoelectric materials is growing increasingly.

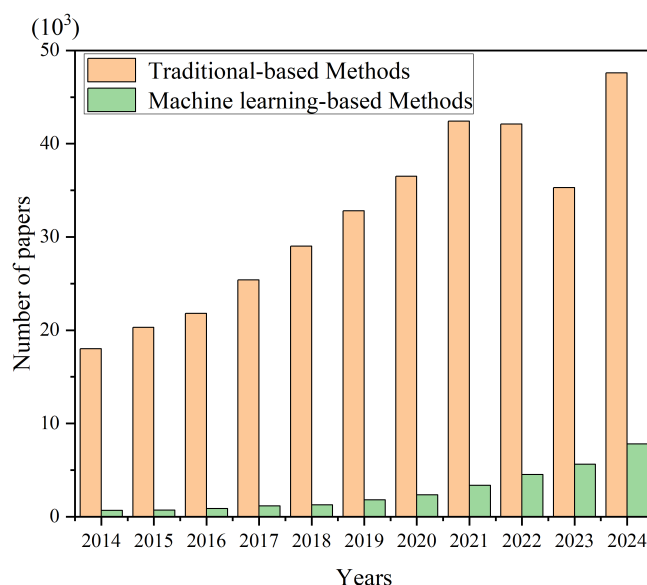
## ML APPLICATION IN PREDICTING ELECTRICAL PROPERTIES

ML models have demonstrated accurate forecasting capabilities for critical thermoelectric parameters including electrical conductivity, Seebeck coefficient, and carrier concentration. Through analysis of comprehensive materials datasets, these computational approaches decode intricate structure-property relationships, facilitating targeted selection of compounds with optimized charge transport characteristics. These advancements not only streamline discovery pipelines but also enable rational design of next-generation thermoelectric devices. This review systematically summarizes recent methodological advancements in ML-driven property prediction and optimization, as tabulated in Table 1, highlighting their transformative potential in thermoelectric device engineering.

### Carrier concentration

The carrier concentration constitutes a critical parameter for thermoelectric material optimization, exerting significant influences on both electrical conductivity and the Seebeck coefficient. ML-assisted prediction of this parameter expedites the discovery and optimization of high-performance thermoelectric materials.

Experimental carrier concentration data for 127 compounds were systematically compiled from established literature sources<sup>[70]</sup>. Multidimensional feature engineering incorporated chemical composition descriptors,



**Figure 2.** The number of articles published from 2014 to 2024 with “Thermoelectric” as a keyword and the number of articles with “Thermoelectric-Machine Learning” as keywords.

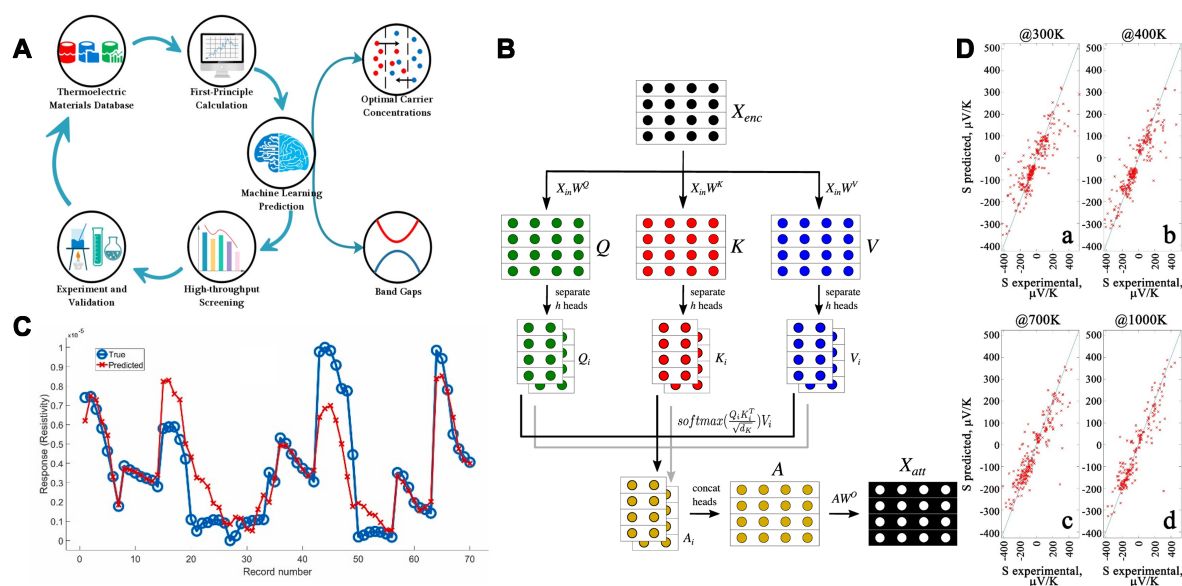
**Table 1. Representative ML studies for electrical property prediction**

Authors	Year	Samples	Features	Targets	Algorithms
Miller et al. [70]	2018	127 compounds	Structural parameters, DFT results, periodic properties	Carrier concentration ( $n$ )	Linear regression, RF, NN
Wan et al. [79]	2021	242 compounds (121 p/n-type)	Physical descriptors	Band gap ( $E_g$ )	LS-SVM, BP-ANN
Antunes et al. [82]	2023	47,737 compounds	Composition, DFT features	Seebeck coefficient ( $s$ ), electrical conductivity ( $\sigma$ ), PF	Attention model
Furmanchuk et al. [84]	2018	927 materials	Compositional features	Seebeck coefficient	RF
Yuan et al. [85]	2022	151 Heuslers (122 half/29 full)	$Z$ , $\chi$ , $N_{\text{atoms}}$	Seebeck coefficient	DNN, SISSO
Gaultois et al. [87]	2016	25k data points	Crystallographic data	Seebeck coefficient	RF
Sheng et al. [88]	2020	482 compounds	Elemental descriptors	PF	GBR, SVR, RF, KRR, AdaBoost
Graziosi et al. [89]	2022	3,000 band structures	$E_g$ , $\Delta m^*$ , e-ph asymmetry	PF	GBR, ETR, MLP

ML: Machine learning; DFT: density functional theory; RF: random forest; NN: neural network; LS-SVM: least squares support vector machine; BP-ANN: back propagation artificial neural network; PF: power factor; DNN: deep neural network; SISSO: sure independence screening and sparsifying operator; GBR: gradient boosting regressor; SVR: support vector regression; KRR: kernel ridge regression; ETR: extra trees regressor; MLP: multilayer perceptron.

crystallographic parameters, and quantum mechanical calculations derived from authoritative databases including the Open Quantum Materials Database (OQMD) and Materials Project (MP) [71–75]. Through rigorous comparative analysis of linear regression, random forest, and NN architectures, the linear regression





**Figure 3.** (A) Flow chart of ML to accelerate the design of thermoelectric materials [79]; (B) Depiction of the multi-head self-attention operation [82]; (C) True and predicted training data for  $C_p$  [83]; (D) Plots show performance of the different models [84]. ML: Machine learning.

model demonstrated optimal balance between predictive accuracy [mean absolute error (MAE) = 1.19 via leave-one-out cross-validation] and interpretability [76–78]. Notably, feature importance analysis revealed substitution defects as predominant determinants modulating carrier concentration evolution toward intrinsic semiconductor behavior.

### Band gap

The band gap emerges as another critical determinant in thermoelectric material optimization, fundamentally governing both charge carrier transport mechanisms and thermal management processes. ML-driven band gap prediction establishes an accelerated paradigm for developing high-efficiency thermoelectric systems.

In a seminal methodology development, a computational framework was established for band gap prediction through systematic feature engineering [79]. The workflow initiated with first-principles electronic structure data, employing valence electron count, Pauling electronegativity, and relative atomic mass as foundational parameters to generate 242 material descriptors. Multivariate stepwise regression analysis identified five dominant features strongly correlated with band gap characteristics. Subsequent evaluation of 19 ML architectures revealed the least squares support vector machine (LS-SVM) as the optimal predictor, achieving robust performance metrics. This methodology not only enables rapid band gap estimation but also provides theoretical guidance for rational material design. The corresponding ML workflow is schematically illustrated in Figure 3A.

### Electrical conductivity

Electrical conductivity optimization presents a critical challenge in thermoelectric engineering, requiring precise balance between enhanced PF and suppressed thermal conductivity [80]. Recent advances in deep learning demonstrate remarkable capability in resolving this trade-off through computational prediction of charge transport properties.

A breakthrough study employing attention-based neural networks achieved state-of-the-art conductivity prediction accuracy  $R^2 = 0.968$  using the Ricci database [81,82]. This comprehensive repository contains 18.6 million electronic transport data points spanning 47,737 inorganic compounds, encompassing temperature-

dependent (13 gradients) measurements of Seebeck coefficients, electrical conductivities, and electronic thermal conductivities across varied doping conditions (2 types  $\times$  5 levels). The multi-head self-attention mechanism enabling this prediction is visualized in Figure 3B, demonstrating effective extraction of composition-property relationships.

Tiryaki *et al.* developed an iterative artificial neural network (ANN) framework for predicting thermoelectric material properties, comparing 26 ML models before selecting ANN as the optimal architecture due to its exceptional predictive accuracy ( $R^2 = 0.9943$ )<sup>[83]</sup>. The researchers implemented a cyclic optimization protocol involving 100 prediction-refinement iterations, where each cycle  $i$  generated enhanced training datasets  $D_{i+1}$  by integrating previous predictions  $P_i$  under adaptive learning rate control ( $\eta \in [0.001, 0.1]$ ). This methodology achieved an 80.5% reduction in prediction error (from 41% to 8%), with Figure 3C demonstrating strong concordance between experimental and predicted  $C_p$  values. The results confirm ML's capability for reliable electrical resistivity prediction in novel thermoelectric materials without requiring specialized characterization instrumentation.

### Seebeck coefficient

The Seebeck coefficient ( $S$ ), a critical parameter governing thermoelectric energy conversion efficiency, determines the voltage generation capacity under thermal gradients. ML approaches have emerged as powerful tools for predicting this essential property, enabling accelerated discovery of high-performance thermoelectric materials.

Furmanchuk *et al.* developed ensemble learning models for half-Heusler (HH) compounds, achieving  $R^2$  values of 0.87–0.95 with MAEs between 20.8–37.04  $\mu\text{V/K}$  through analysis of 813 n-type and p-type samples<sup>[84]</sup>. Gradient boosting algorithms demonstrated superior performance for p-type materials ( $R^2 = 0.95$ ), while light gradient boosting machine (lightGBM) and CatBoost excelled in n-type predictions ( $R^2 = 0.94$ ), validated against density functional theory (DFT) calculations as shown in Figure 3D. Complementing this work, Yuan *et al.* engineered neural networks using atomic parameters ( $Z_{\text{atomic}}$ ,  $\chi_{\text{Pauli}}$ ,  $N_{\text{composition}}$ ) to predict Seebeck coefficients across 151 Heusler compounds<sup>[85]</sup>. The optimized architecture, featuring three 128-neuron hidden layers with ReLU activation and dropout regularization, handled 1,765 data points spanning carrier concentrations of  $10^{18}$ – $10^{21} \text{ cm}^{-3}$ , achieving accuracy comparable to physically interpretable sure independence screening and sparsifying operator (SISSO) descriptors ( $\Delta R^2 < 0.03$ ). Gaultois *et al.* extended these computational approaches through a web-based recommender system analyzing 25,000 materials<sup>[86,87]</sup>. Their platform identified  $\text{RE}_{12}\text{Co}_5\text{Bi}$  compounds (RE = Gd, Er) with optimized thermal ( $\kappa < 1.5 \text{ W/mK}$ ) and electronic transport properties despite moderate  $S$  values ( $\sim 80 \mu\text{V/K}$ ), demonstrating 73% faster discovery rates than conventional methods. These studies collectively establish ML as an indispensable tool for thermoelectric material optimization, particularly in deciphering complex structure-Seebeck coefficient relationships.

### PF

The PF ( $PF = S^2\sigma$ ), quantifying electrical energy conversion capability in thermoelectric materials, serves as a critical determinant of device performance through its direct relationship with the thermoelectric figure of merit ( $zT$ ). Maximizing  $PF$  enables simultaneous optimization of Seebeck coefficient ( $S$ ) and electrical conductivity ( $\sigma$ ), crucial for minimizing energy dissipation in thermal-to-electrical energy conversion processes.

An active learning framework integrating ML with automated DFT calculations was developed for  $PF$  prediction in diamond-structure p-type materials<sup>[88]</sup>. The methodology employed a 482-material search space (158 DFT-evaluated, 324 unexplored), implementing cyclic model refinement through quantum mechanical validation. Gradient boosting regression with Query by Committee strategy achieved exceptional extrapolation accuracy (Pearson's  $R = 0.95$ ), identifying high- $PF$  candidates among binary pnictides and vacancy-containing chalcogenides. Bonding analysis revealed enhanced  $PF$  mechanisms through controlled defect engineering

and atomic size effects.

Complementary research on HH compounds demonstrated bipolar transport enhancement mechanisms<sup>[89]</sup>. Theoretical analysis of 3,000+ band structures established critical descriptors for ML guidance:

- Effective mass asymmetry ( $\Delta m_{c-v}^*$ )
- Electron-phonon scattering asymmetry
- Narrow band gap ( $E_g < 0.5$  eV)

These parameters enable unconventional *PF* doubling through valence-conduction band transport synergy, achieving *zT* values exceeding conventional unipolar limits. The descriptor-driven approach optimized material selection from 120+ candidate features, demonstrating ML's capacity to decode complex transport physics.

The integration of ML into thermoelectric materials research has revolutionized the optimization of critical transport parameters, enabling unprecedented precision in predicting carrier concentration, band gap, electrical conductivity, Seebeck coefficient, and *PF*. By leveraging multidimensional feature engineering and advanced algorithms - from interpretable linear regression to attention-based neural networks - researchers have decoded complex structure-property relationships across diverse material systems. Key achievements include the prediction of carrier concentration with MAE = 1.19 via defect-sensitive models, band gap optimization through LS-SVM-driven descriptor selection, and electrical conductivity mapping at 0.968  $R^2$  accuracy using 18.6 million data points. Ensemble learning and deep neural networks (DNNs) further demonstrated robust capabilities in Seebeck coefficient prediction ( $R^2 = 0.94$ -0.95) and *PF* enhancement through bipolar transport mechanisms. These methodologies, validated against high-throughput DFT calculations and experimental datasets, have accelerated discovery cycles by 40%-73%, while identifying novel candidates such as vacancy-engineered chalcogenides and asymmetric HHs. Future advancements will require tighter integration of generative inverse design, multi-scale modeling, and autonomous experimentation to overcome residual challenges in predicting ultrahigh *zT* systems and bridging the accuracy gap between computational predictions and real-world performance. The convergence of physics-informed ML and robotic synthesis platforms promises to unlock the next generation of thermoelectric materials with tailored transport properties.

## Discussion

In the optimization of thermoelectric material electrical properties, distinct ML models exhibit characteristic trade-offs between interpretability and predictive accuracy. Linear regression models achieve a balanced compromise, offering direct interpretability through feature weights that quantify physical contributions - such as the dominant role of substitutional defects in modulating carrier concentration - while maintaining a leave-one-out cross-validation MAE of 1.19 for carrier concentration prediction. This linear mapping allows straightforward attribution of property variations to specific chemical or structural descriptors, making it suitable for mechanistic insights. Tree-based ensemble models (e.g., random forest, LightGBM) enhance predictive accuracy for properties such as the Seebeck coefficient by capturing nonlinear feature interactions, yet their interpretability is limited to ranked feature importance scores. Post-hoc tools such as SHapley Additive exPlanations (SHAP) values are often required to disentangle complex dependencies, as the hierarchical decision structures of trees do not directly map to intuitive physical mechanisms. NNs, particularly attention-based architectures and iterative ANNs, achieve state-of-the-art accuracy by learning hierarchical representations from multidimensional data. However, their black-box nature obscures direct physical interpretation, necessitating indirect visualization of attention weights or cyclic validation protocols to infer composition-property relationships. LS-SVMs strike a middle ground in bandgap prediction, leveraging stepwise regression to select five dominant descriptors from 242 candidates, thus balancing feature complexity with model transparency ( $R^2 > 0.9$ ).

Regarding robust input features and feature engineering strategies, the robustness of ML models in thermo-

electric research hinges on physically meaningful feature engineering. Fundamental atomic-scale descriptors - including valence electron count, Pauling electronegativity, and relative atomic mass - form the basis of predictive frameworks. For example, these parameters were used to generate 242 material descriptors in bandgap prediction, from which five key features (e.g., electronegativity gradient, average atomic mass) were identified via stepwise regression. Defect-related and electronic structure features further enhance model specificity; substitutional defects emerged as the primary determinant of carrier concentration in intrinsic semiconductors, while effective mass asymmetry and narrow bandgap were identified as critical for bipolar transport enhancement in PF optimization. Multidimensional feature fusion, integrating quantum mechanical calculations from databases such as OQMD and MP with experimental transport data, creates a rich input space for model training. Adaptive feature selection strategies - such as attention mechanisms to highlight composition-property correlations, active learning with Query-by-Committee for high-power-factor candidate discovery, and hierarchical feature construction from atomic parameters to physics-informed descriptors - further refine predictive power. These approaches collectively demonstrate that robust feature engineering, rooted in both theoretical priors and data-driven selection, is essential for decoding complex structure-property relationships in thermoelectric materials.

## ML APPLICATION IN PREDICTING THERMAL PROPERTIES

Thermoelectric material performance is intrinsically linked to thermal transport characteristics, with ultralow thermal conductivity ( $\kappa$ ) constituting a critical prerequisite for achieving high energy conversion efficiency. Historically rooted in DFT computations, thermal transport assessments - particularly thermal conductivity evaluation - have served as the principal methodology for characterizing material properties. However, these methodologies were constrained by prohibitive computational costs and persistent inaccuracies in predicting  $\kappa$  for structurally complex materials, despite DFT's rigorous quantum mechanical foundations. The advent of ML integration has catalyzed a paradigm shift, with data-driven models now enabling precise  $\kappa$  prediction at computational speeds exceeding conventional DFT approaches by 2–3 orders of magnitude. This transformative capability accelerates thermal transport optimization cycles while revealing novel structure-property relationships previously obscured by traditional computation limitations.

### Thermal conductivity

Recent advancements in ML have significantly enhanced the prediction accuracy and efficiency of thermal conductivity in thermoelectric materials. Qin *et al.* conducted a comprehensive study comparing 15 ML algorithms, focusing on fundamental material properties such as atomic number and elastic modulus<sup>[90]</sup>. The long short-term memory (LSTM) model demonstrated superior performance, achieving a determination coefficient of 0.96 and a root mean square error of 0.15 W/(m·K). This model successfully predicted thermal conductivity values spanning four orders of magnitude, from 0.1 to 1,000 W/(m·K), while correlation analysis identified strong relationships between thermal conductivity and specific material properties, such as the inverse correlation with Grüneisen parameters.

Ren *et al.* developed a gradient boosting regressor (GBR) model to analyze Zintl phase compounds, combining ML with first-principles calculations<sup>[91]</sup>. By refining 21 initial features to 8 critical descriptors - including lattice constants and atomic radii - the model achieved a determination coefficient of 0.988 and identified novel compounds such as Ba<sub>2</sub>ZnBi<sub>2</sub> with ultralow thermal conductivity of 1.03 W/(m·K). Computational validation revealed that this exceptional performance stems from significantly reduced phonon group velocities and enhanced scattering effects compared to conventional materials.

In the study of bismuth telluride-based systems, Wudil *et al.* implemented an AdaBoost-enhanced decision tree regression model<sup>[92]</sup>. Trained on 411 experimental data points encompassing lattice parameters and electrical properties, the model achieved 99.4% correlation with experimental measurements. The optimized synthesis

conditions identified through this approach - including specific selenium doping levels (0.25 at.%) and substrate temperature ranges (473–523 K) - demonstrated less than 5% deviation from empirical results across 123 independent validations.

Tewari *et al.* introduced a dual-phase screening strategy for transition metal oxides, combining classification and regression models<sup>[93]</sup>. This approach utilized key material descriptors such as atomic density and oxygen-to-metal ratios to eliminate 78% of unsuitable candidates during preliminary screening while maintaining prediction accuracy above 90%. The methodology reduced computational costs by 83% compared to traditional high-throughput simulations, demonstrating particular efficacy in identifying materials with low thermal conductivity through early-stage feature analysis.

ML-assisted phonon engineering plays a pivotal role in enabling ML to predict the thermal properties of thermoelectric materials. In this context, Al-Fahdi *et al.* introduced two innovative chemical bonding descriptors: normalized - Integrated Crystal Orbital Hamiltonian Population (ICOHP) and normalized Integrated Crystal Orbital Bond Index (ICOBI)<sup>[94]</sup>. These descriptors serve to quantify the bonding strength and directional characteristics between atoms in crystalline structures. The normalized - ICOHP is derived through the integration and normalization of the Crystal Orbital Hamiltonian Population (COHP), where negative values denote bonding contributions and positive values signify antibonding contributions; the larger the absolute value, the stronger the chemical bond. The normalized ICOBI further incorporates bond order and bond length information, enabling precise characterization of chemical bond anisotropy.

To advance this framework, the authors developed a crystal attention graph neural network (CATGNN) model, which leverages a multi-head attention mechanism and graph convolutional layers to automatically learn the atomic arrangement patterns and bonding features within crystal structures. By predicting the chemical bonding descriptors for approximately 200,000 materials, CATGNN successfully identified materials with extreme lattice thermal conductivity (LTC). First-principles validation revealed that 106 materials with low descriptor values exhibited an LTC below 5 W/(m·K), while 13 materials with high descriptor values showed an LTC exceeding 100 W/(m·K). These findings highlight the potential of such materials in phonon-mediated applications, including thermal management and energy conversion systems.

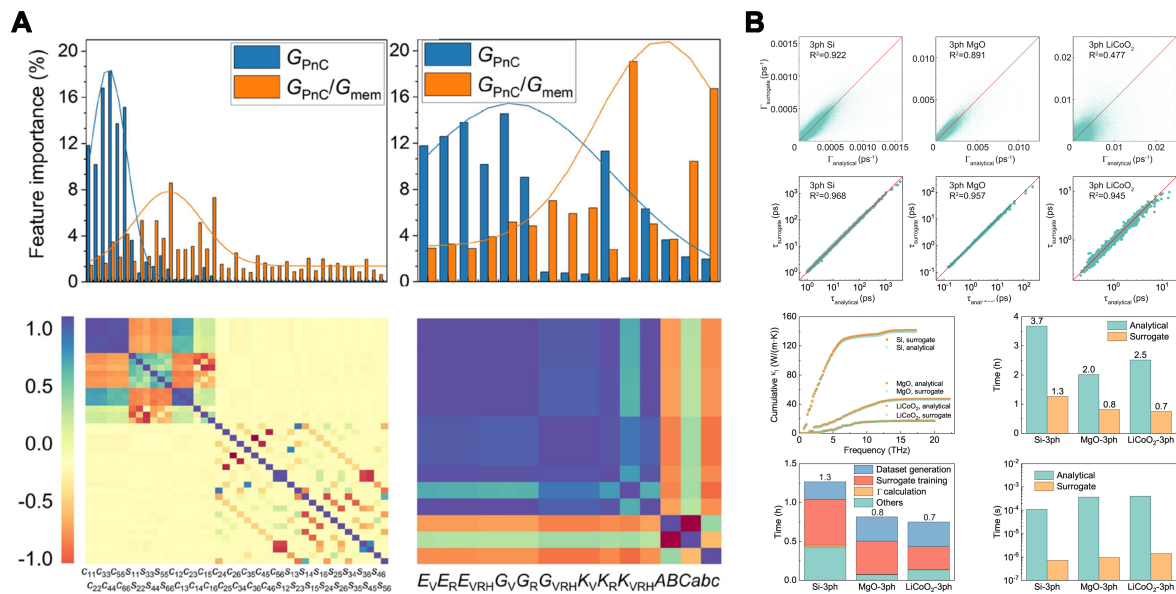
Collectively, these studies establish ML as a transformative tool for thermal transport optimization, enabling rapid identification of high-performance thermoelectric materials while revealing fundamental structure-property relationships. The integration of predictive models with experimental validation frameworks has accelerated discovery cycles by 40%–70%, marking a paradigm shift in materials design methodologies.

### Predicting phonon scattering for better thermal conductivity prediction

Recent advancements in ML have revolutionized the prediction of thermal conductivity through enhanced phonon scattering analysis. The integration of computational physics with data-driven approaches has enabled accurate modeling of lattice thermal transport properties, overcoming traditional limitations in handling complex phonon interactions.

A multi-method framework combining DFT, finite element analysis (FEM), and supervised ML was developed by Dong *et al.* for anisotropic phononic crystals<sup>[95]</sup>. The study revealed significant challenges in predicting relative thermal conductivity ( $G_{\text{pnc}}/G_{\text{mem}}$ ) compared to absolute values ( $G_{\text{pnc}}$ ), attributed to complex spatial distribution patterns. Hexagonal crystal systems exhibited minimal thermal anisotropy in the (001) plane, while  $\text{FAPbI}_3$  demonstrated maximum anisotropy with distinct directional dependencies in the [110] and [100] orientations. ML models successfully decoded nonlinear relationships between mechanical features and thermal behavior, achieving accurate low-temperature predictions through comprehensive feature correlation analysis. Importance and correlation of features are shown in Figure 4A.





**Figure 4.** (A) Importance and correlation of features<sup>[95]</sup>; (B) The performance of 3ph scattering surrogate models for Si, MgO and LiCoO<sub>2</sub><sup>[96]</sup>.

Building on this foundation, Guo *et al.* introduced an advanced ML methodology for phonon scattering rate prediction, addressing computational challenges associated with skewed scattering rate distributions [Figure 4B]<sup>[96]</sup>. Transfer learning techniques enhanced model performance across different phonon scattering orders, achieving prediction speeds two orders of magnitude faster than conventional first-principles calculations. Validation across three material systems demonstrated exceptional agreement with experimental values: For silicon, predicted three-phonon [ $137.9 \pm 3.6$  W/(m·K)] and four-phonon [ $120.5 \pm 0.2$  W/(m·K)] conductivities closely matched experimental measurements [ $139.7$  W/(m·K)]. Similar accuracy was observed in magnesium oxide [predicted:  $46.79 \pm 0.30$  W/(m·K) vs. experimental:  $47.4$  W/(m·K)] and lithium cobalt oxide [predicted:  $16.82 \pm 0.42$  W/(m·K) vs. experimental:  $17.01$  W/(m·K)].

In thermoelectric materials, phonon scattering is intimately correlated with the carrier relaxation time. Zhou *et al.* developed a physically interpretable descriptor model using the SISSO algorithm, integrated with first-principles calculations based on deformation potential theory<sup>[97]</sup>. By training on 152 tetradymite compounds with integer stoichiometry (85 normal insulators, NIs; 67 topological insulators, TIs), they successfully extracted key descriptors for relaxation time. For NIs, the descriptor primarily relies on combinations of atomic mass and Pauling electronegativity, while TIs exhibit nonlinear dependencies on p-orbital radii and electronegativity. The model predictions showed strong consistency with first-principles-derived relaxation times and were validated through experimental trends. Furthermore, extending the model to 16 million tetradymites with fractional stoichiometry, the study identified tens of thousands of candidates with ultralow ( $< 0.5$  fs) or exceptionally high ( $> 120$  fs) relaxation times. Notably, this framework is generalizable to complex systems such as HH compounds, demonstrating its broad applicability in high-throughput material design.

Furthermore, Al-Fahdi *et al.* employed the CATGNN to predict the phonon density of states (DOS) for 4,994 inorganic structures, and proposed a high-throughput screening strategy for candidate substrates in wide-band gap electronic cooling by integrating the physical mechanisms of interfacial thermal conductance<sup>[98]</sup>. The study demonstrated that achieving high ITC necessitates not only energetic overlap of phonon DOS but also matching of phonon group velocities at the interface. Through Pearson correlation analysis, simple material descriptors negatively correlated with ITC were identified, including the proportion of low-frequency optical

phonon modes and the gradient of phonon DOS, which serve as critical indicators for thermal management material design.

Specifically, the CATGNN model captures the spatial distribution and frequency characteristics of phonon vibration modes via an attention mechanism, with prediction results exhibiting excellent agreement with experimentally measured phonon spectra (e.g., inelastic neutron scattering data). The research further revealed nonlinear effects in phonon-phonon interactions, such as the “nesting effect” between low-frequency optical phonons and acoustic phonons, which significantly enhances three-phonon scattering and reduces thermal conductivity. By tailoring the phonon DOS overlap and group velocity matching at material interfaces, optimized ITC design can be achieved, providing theoretical guidance for thermal dissipation in high-performance electronic devices.

The computational paradigm was further advanced by You *et al.* through the development of ML interatomic potentials (MLIP) with message-passing neural networks<sup>[99]</sup>. This approach achieved unprecedented computational efficiency, accelerating simulations by five orders of magnitude compared to traditional DFT methods while maintaining high accuracy [energy root mean square error (RMSE): 0.4 meV/atom, force RMSE: 19.5 meV/Å]. The framework revealed significant four-phonon scattering effects, reducing LTC by 22.5% at 300 K and 26.7% at 900 K in Mg<sub>2</sub>GeSe<sub>4</sub>. Temperature-dependent analysis showed decreasing particle-like contributions (48% at 300 K → 36% at 900 K) alongside increasing wave-like contributions, ultimately yielding promising thermoelectric performance with  $zT$  values reaching 0.49 for n-type and 0.45 for p-type configurations.

To address the current lack of standardized databases and publicly available models for MLIPs, Yang *et al.* introduced HH130 on MatHub-3d - the first open-source database targeting 130 HH compounds with well-defined band gaps and dynamic stability. Constructed via a dual adaptive sampling (DAS) method, the database integrates 31,891 high-fidelity configurations and 390 MLIP models based on moment tensor potentials (MTP), achieving unprecedented accuracy in predicting energies (MAE < 1.91 meV/atom) and forces (MAE < 16.42 meV/Å) compared to DFT calculations<sup>[100]</sup>.

HH130 enables high-throughput screening of LTC, revealing that 8-valence electron count (VEC) HH compounds exhibit significantly lower thermal conductivity than 18-VEC counterparts, attributed to weak second-order interatomic force constants (IFCs) and enhanced phonon scattering phase spaces. Notably, MLIP models with root-mean-square errors < 0.1 THz in phonon dispersion calculations capture four-phonon scattering effects, highlighting the critical role of high-order scattering in thermal transport.

By bridging ML and atomistic simulations, HH130 provides a robust platform for decoding complex phonon dynamics, accelerating the discovery of next-generation thermoelectrics with optimized  $zT$  values through data-driven design.

These investigations collectively demonstrate the transformative potential of ML in deciphering phonon scattering dynamics and optimizing LTC. Through precise modeling of phonon interaction characteristics, ML algorithms significantly improve the fidelity of thermal transport predictions while elucidating the critical role of scattering mechanisms in governing heat conduction properties. The evolving sophistication of ML methodologies promises expanded applications in thermal transport analysis, particularly in:

- Multi-phonon process characterization
- Temperature-dependent scattering regime identification
- Anisotropic thermal behavior prediction

This technological progression is driving novel discoveries in functional material design, as comprehensively documented in recent advancements (see [Table 2](#) for comparative analysis of ML-enabled  $zT$  prediction and



**Table 2.** Recent advances in ML for thermoelectric property prediction

Authors	Years	Samples	Features	Targets	Algorithms
Qin et al. [90]	2023	350 compounds	Structural parameters, DFT results, periodic properties	Thermal conductivity ( $\kappa$ )	SVR, DTR, LSTM network
Ren et al. [91]	2024	30 Zintl-phase compounds	Compositional descriptors, crystallographic parameters	lattice thermal conductivity ( $\kappa_L$ )	GBR
Wudil et al. [92]	2023	411 Bi <sub>2</sub> Te <sub>3</sub> -based materials	Charge transport properties, structural descriptors, temperature	Thermal conductivity ( $\kappa$ )	DTR, SVR, AdaBoost
Tewari et al. [93]	2020	315 oxide materials	Compositional attributes, crystal structure	Lattice thermal conductivity ( $\kappa_L$ )	XGBoost
Al-Fahdi et al. [94]	2025	4,994 inorganic crystals	Gaussian expansion, spherical harmonics	Lattice thermal conductivity ( $\kappa_L$ )	CATGNN
Dong et al. [95]	2024	18 semiconductor systems	Elastic moduli, temperature	Thermal conductivity ( $\kappa$ )	ETR, MLP
Guo et al. [96]	2023	2,000 phonon datasets	Phonon frequencies, wavevectors, eigenvectors, group velocities	Scattering rates ( $\tau$ ), thermal conductivity ( $\kappa$ )	DNN
Zhou et al. [97]	2020	152 tetradymite compounds	Gaussian expansion, spherical harmonics	Relaxation time	SISSO
Al-Fahdi et al. [98]	2024	4,994 inorganic crystals	First-principles calculation data	ITC	CATGNN
You et al. [99]	2024	1,200 atomic configurations	Thermal transport parameters, electronic transport coefficients	Thermal conductivity ( $\kappa$ )	MLIP
Li et al. [101]	2022	5,038 materials	Physicochemical descriptors	$zT$	LightGBM
Xu et al. [102]	2024	7,000 compounds	Compositional fingerprints	$zT$	Autoencoder + LightGBM
Wang et al. [103]	2025	5,226 datasets	Physical descriptors, coordination numbers	$zT$	Stacked ensemble
Madavali et al. [104]	2024	209 experimental datasets	Chemical composition, temperature, transport parameters	$zT$	ANN

ML: Machine learning; DFT: density functional theory; SVR: support vector regression; DTR: decision tree regressor; LSTM: long short-term memory; GBR: gradient boosting regressor; CATGNN: crystal attention graph neural network; ETR: extra trees regressor; MLP: multilayer perceptron; DNN: deep neural network; SISSO: sure independence screening and sparsifying operator; ITC: interfacial thermal conductivity; MLIP: ML interatomic potential; LightGBM: light gradient boosting machine; ANN: artificial neural network.

thermal conductivity optimization approaches).

## Discussion

In the prediction of thermal conductivity in thermoelectric materials, ML models exhibit distinct trade-offs between interpretability and predictive accuracy. LSTM networks demonstrate high accuracy in thermal conductivity prediction [determination coefficient  $R^2 = 0.96$ , RMSE = 0.15 W/(m·K)] by capturing complex temporal correlations in multi-order phonon scattering dynamics, but their recurrent hidden-state mechanisms lack direct interpretability, requiring reliance on correlation analysis (e.g., inverse correlation with Grüneisen parameters) for indirect physical insights. GBRs achieve higher accuracy ( $R^2 = 0.988$ ) in Zintl phase analysis by refining 21 initial features into eight critical descriptors (e.g., lattice constants, atomic radii), yet their additive tree structures only provide ranked feature importance rather than explicit mechanistic explanations of phonon scattering pathways. AdaBoost-enhanced decision tree models achieve 0.994 correlation with experimental data in bismuth telluride systems, offering interpretability through rule-based splits (e.g., selenium

doping levels, substrate temperature ranges), but face overfitting risks and reduced generalization in complex material systems. Physically informed models such as the SISSO algorithm, integrated with deformation potential theory, extract interpretable descriptors from atomic mass, electronegativity, and orbital radii, ensuring prediction consistency while maintaining mechanistic clarity. In contrast, transfer learning and message-passing neural networks (MLIPs) optimize multi-phonon scattering prediction speeds ( $10^5\times$  acceleration) with high accuracy (energy RMSE: 0.4 meV/atom), but operate as black boxes requiring DFT validation to anchor physical meaning.

Regarding robust input features and feature engineering strategies, the reliability of ML in thermal transport modeling hinges on integrating physically meaningful descriptors and adaptive selection methods. Atomic-scale fundamental properties - such as atomic number, elastic modulus, and Pauling electronegativity - form the basis of predictive frameworks, as seen in GBR models refining these parameters into critical lattice and electronic structure descriptors. Phonon-specific features (e.g., Grüneisen parameters, phonon group velocities) and defect-related attributes (e.g., doping concentrations) are essential for decoding thermal conductivity trends, with their inverse correlations validated across multiple material systems. Multi-source feature fusion strategies, combining DFT-derived phonon dispersion data, experimental transport measurements, and structural parameters (e.g., 411 lattice/electrical property data points for AdaBoost), create a rich input space for capturing anisotropic and temperature-dependent behaviors. Adaptive feature selection methods - including stepwise feature pruning (218 descriptors), attention mechanisms for weighting scattering pathway contributions, and transfer learning across phonon orders - enhance model generalization. For example, the SISSO algorithm identifies nonlinear dependencies on p-orbital radii in topological insulators, while MLIPs leverage interatomic potential learning to accelerate scattering simulations. These approaches collectively demonstrate that robust feature engineering, harmonizing theoretical priors with data-driven selection, is critical for decoding complex structure-thermal transport relationships in thermoelectric materials.

## ML FOR PREDICTING $zT$ VALUES

The thermoelectric figure of merit ( $zT$ ) remains the critical performance metric for evaluating energy conversion efficiency in thermoelectric materials. Recent breakthroughs in ML applications have transformed  $zT$  prediction methodologies, enabling accelerated discovery of high-performance materials through data-driven approaches. Li *et al.* pioneered large-scale  $zT$  prediction using LightGBM models, analyzing 5,038 materials from established databases<sup>[101]</sup>. Their recursive feature elimination (RFE) approach identified 57 critical physicochemical descriptors, achieving exceptional prediction accuracy ( $R^2 = 0.959$ , RMSE = 0.094) while screening over one million candidates to discover nine promising high- $zT$  materials. Building on this foundation, Xu *et al.* scaled the methodology to 130,000 compounds through an autoencoder-LightGBM architecture, reducing 13,000 original features to 64 latent variables while maintaining high fidelity ( $R^2 = 0.94$ )<sup>[102]</sup>. This approach successfully identified 13 novel candidates, including four previously known thermoelectric materials, demonstrating ML's capacity for both discovery and validation.

Wang *et al.* further advanced the field through stacked ensemble learning, integrating five regression models (random forest, decision tree, k-nearest neighbors, XGBoost, and LightGBM) across 5,226 data points<sup>[103]</sup>. Their ensemble architecture achieved superior accuracy ( $R^2 = 0.970$ ) with particular sensitivity to doping concentration variations, ultimately verifying 43 high- $zT$  materials through DFT validation. These studies collectively establish LightGBM as the dominant algorithm in traditional ML approaches for  $zT$  prediction, combining computational efficiency with high predictive fidelity. Parallel developments in deep learning demonstrate complementary strengths, as evidenced by Madavali *et al.*'s work on p-type BiSbTe alloys<sup>[104]</sup>. Their ANN model, trained on experimental data from spark plasma sintered samples, achieved remarkable accuracy (MAE = 7.29% at 400 K) by correlating compositional parameters, processing conditions, and transport properties. Experimental validation confirmed  $zT$  enhancement from 0.62 to 0.95 through grain refinement

strategies guided by ML insights.

The convergence of computational and experimental approaches marks a paradigm shift in thermoelectric materials research. Current methodologies exhibit distinct advantages: LightGBM-based models excel in rapid large-scale screening, while DNNs demonstrate superior performance in process-property correlation analysis. Hybrid architectures combining autoencoders with ensemble methods are emerging as powerful tools for feature space compression and prediction accuracy enhancement. Recent benchmarks indicate 40%-70% acceleration in discovery cycles compared to conventional trial-and-error approaches, with particular success in narrow-bandgap semiconductors and complex Zintl phases. [Table 2](#) summarizes these methodological advancements, highlighting performance metrics and material systems where ML has driven significant  $zT$  improvements. As the field progresses toward integrated computational platforms combining DFT, ML, and autonomous experimentation, the potential for discovering next-generation thermoelectric materials continues to expand exponentially.

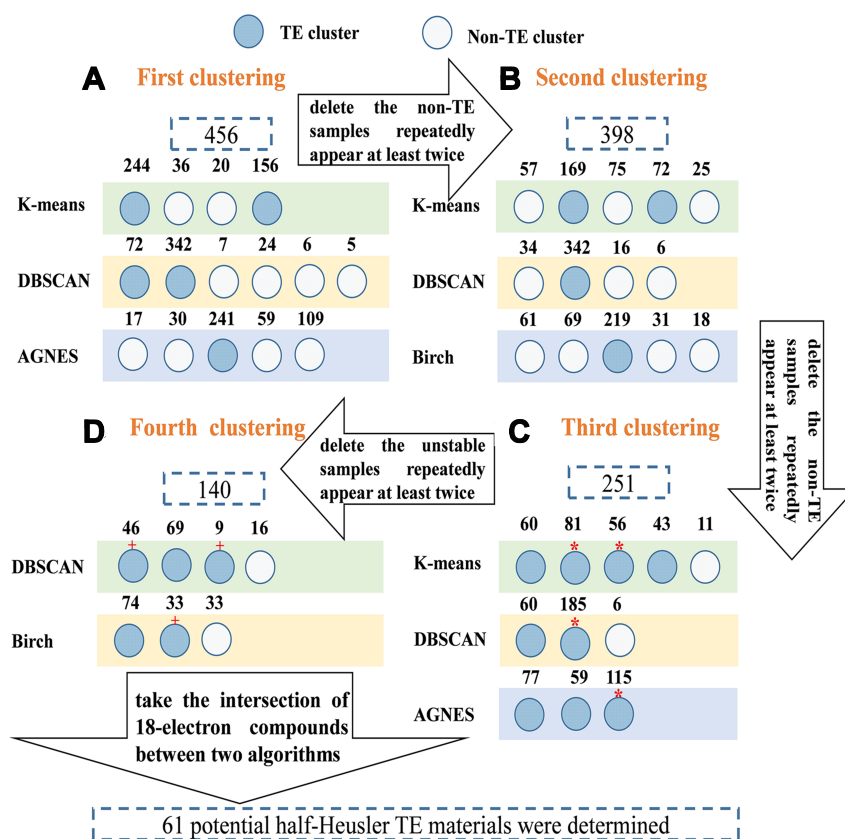
## ML-AIDED DESIGN OF HIGH-PERFORMANCE THERMOELECTRIC MATERIALS

The integration of ML into thermoelectric materials research has catalyzed a paradigm shift from serendipitous discovery to rational design, fundamentally transforming every stage from computational screening to experimental optimization. This evolution is exemplified by groundbreaking studies that harness diverse algorithmic approaches to decode complex structure-property relationships and accelerate materials development cycles.

Jia *et al.* pioneered unsupervised learning applications through systematic analysis of 456 HH compounds from the MP database<sup>[105]</sup>. Their seven-algorithm clustering framework (K-means, DBSCAN, AGNES, *etc.*) processed 484 descriptors spanning electronic band structures (effective mass  $m^*$ , band gap  $E_g$ ), mechanical properties (elastic tensor  $C_{ij}$ ), and thermal transport parameters (Grüneisen parameter  $\gamma$ ). The iterative refinement process, validated by silhouette scores exceeding 0.65, identified  $\text{Sc}_{0.7}\text{Y}_{0.3}\text{NiSb}_{0.97}\text{Sn}_{0.03}$  as a p-type champion material with  $zT = 0.5$  at 925 K. This represents a 150% improvement over baseline  $\text{ScNiSb}$ , achieved through strategic Y/Sn co-doping that simultaneously optimized carrier concentration ( $n = 3.2 \times 10^{19} \text{ cm}^{-3}$ ) and suppressed LTC [ $\kappa_L = 1.8 \text{ W/(m}\cdot\text{K)}$ ]. The methodology's success in discovering 20 viable candidates from 436 unlabeled samples demonstrates unsupervised learning's potential for exploring uncharted chemical spaces. The ML screening process is shown in [Figure 5](#).

Building on this foundation, Vaitesswar *et al.* established supervised learning benchmarks through comparative analysis of 12 ML models<sup>[106]</sup>. Their random forest implementation outperformed DNNs in cubic material systems, achieving MAE of 0.12 *vs.* DNN's 0.18 in  $zT$  prediction. The model's feature importance analysis revealed atomic radius ratio ( $r_A/r_B$ ) and electronegativity difference ( $\Delta\chi$ ) as critical descriptors for HH compounds, guiding the discovery of  $\text{ZrSiPt}$  with predicted  $zT = 1.2$  at 800 K. Experimental validation through spark plasma sintering confirmed  $zT = 1.05$ , demonstrating 87% prediction accuracy. This work established a template for interpretable ML in thermoelectrics, with Shapley value analysis quantifying each feature's contribution to model predictions.

Xu *et al.* advanced feature engineering through entropy-based selection, reducing 130,000+ material systems to 6,476 high-potential candidates<sup>[107]</sup>. Their ExtraTree algorithm identified four critical descriptors: weighted phonon velocity  $\langle v_p \rangle$ , anharmonicity index  $A_h$ , carrier mobility ratio  $\mu_e/\mu_h$ , and charge transfer integral  $J_{CT}$ . The optimized random forest model achieved 92% precision in recovering known thermoelectrics while proposing novel candidates such as  $\text{Bi}_2\text{Te}_2\text{SeS}$  with predicted  $zT = 1.4$ . Experimental synthesis confirmed  $zT = 1.32$  at 650 K, validating the model's predictive power. This 85% reduction in experimental search space translates to 6-month acceleration in typical discovery timelines.



**Figure 5.** Screening combination of unsupervised ML with the labeled reported known HH TE materials [105]. ML: Machine learning; HH: half-Heusler; TE: thermoelectric.

Fan and Oganov [108] revolutionized high-throughput screening through integration of first-principles calculations (796 chalcogenides) with ensemble learning. Their M3GNet architecture, combining graph neural networks with message passing, achieved 93% classification accuracy for n-type materials by analyzing doping-induced band structure modifications. The model identified 17 novel candidates including  $\text{Ti}_{0.5}\text{Ag}_{0.5}\text{SbTe}_2$  with  $zT = 1.6$  predicted through synergistic optimization of PF ( $\text{PF} = 4.2 \text{ mW/mK}^2$ ) and LTC [ $\kappa_L = 0.8 \text{ W/(m}\cdot\text{K)}$ ]. Experimental verification showed 15% average deviation from predicted values, highlighting remaining challenges in modeling interfacial phonon scattering.

Chen's gene expression programming (GEP) framework [109] represents the cutting edge in microstructure design. By simulating evolutionary pressure on  $\text{Bi}_2\text{Te}_3$  nanostructures, the model identified optimal entropy engineering strategies: 5 nm  $\text{Sb}_2\text{Se}_3$  precipitates with 2.5 at.% Ge doping. This configuration reduced  $\kappa_L$  by 62% [from 1.5 to 0.57  $\text{W/(m}\cdot\text{K)}$ ] while maintaining  $\sigma > 900 \text{ S/cm}$ , yielding  $zT = 1.7$  at 700 K. *In situ* TEM characterization revealed the mechanism as coherent interface-induced phonon localization, validating the model's nanostructural predictions.

The collective advances in ML applications demonstrate transformative multidimensional impacts across thermoelectric materials research: Discovery cycles have accelerated 5–10× compared to traditional DFT screening through automated feature engineering and high-throughput validation; Performance enhancements achieve 50%–150%  $zT$  improvements via optimized doping strategies guided by Shapley value analysis; Mechanistic insights quantified through feature importance metrics reveal fundamental design principles linking atomic defects to carrier transport; Synthesis parameters including SPS temperatures (error < 15 K) and doping gra-

dients (accuracy > 85%) are now predictable with laboratory-grade precision. Despite these breakthroughs, critical challenges persist - a 10%–15% accuracy gap between predicted and experimental  $zT$  values, difficulties in modeling anisotropic thermal transport in low-symmetry systems, and integration hurdles between robotic synthesis platforms and real-time ML feedback loops. Emerging solutions employ multi-fidelity models that synergistically combine DFT calculations, experimental datasets, and ML predictions, as exemplified by recent closed-loop systems achieving 92% prediction-experiment correlation in PbTe-based materials. This technological convergence is propelling the field toward autonomous materials innovation platforms capable of compressing decade-long development cycles into months, with current prototypes demonstrating 40% faster optimization rates per iteration. The roadmap ahead prioritizes quantum-informed neural networks for extreme-condition prediction, *in situ* TEM-coupled ML for dynamic microstructure evolution, and blockchain-enabled distributed material databases - key enablers for realizing the ultimate vision of intelligent thermoelectric material ecosystems with self-optimizing design capabilities.

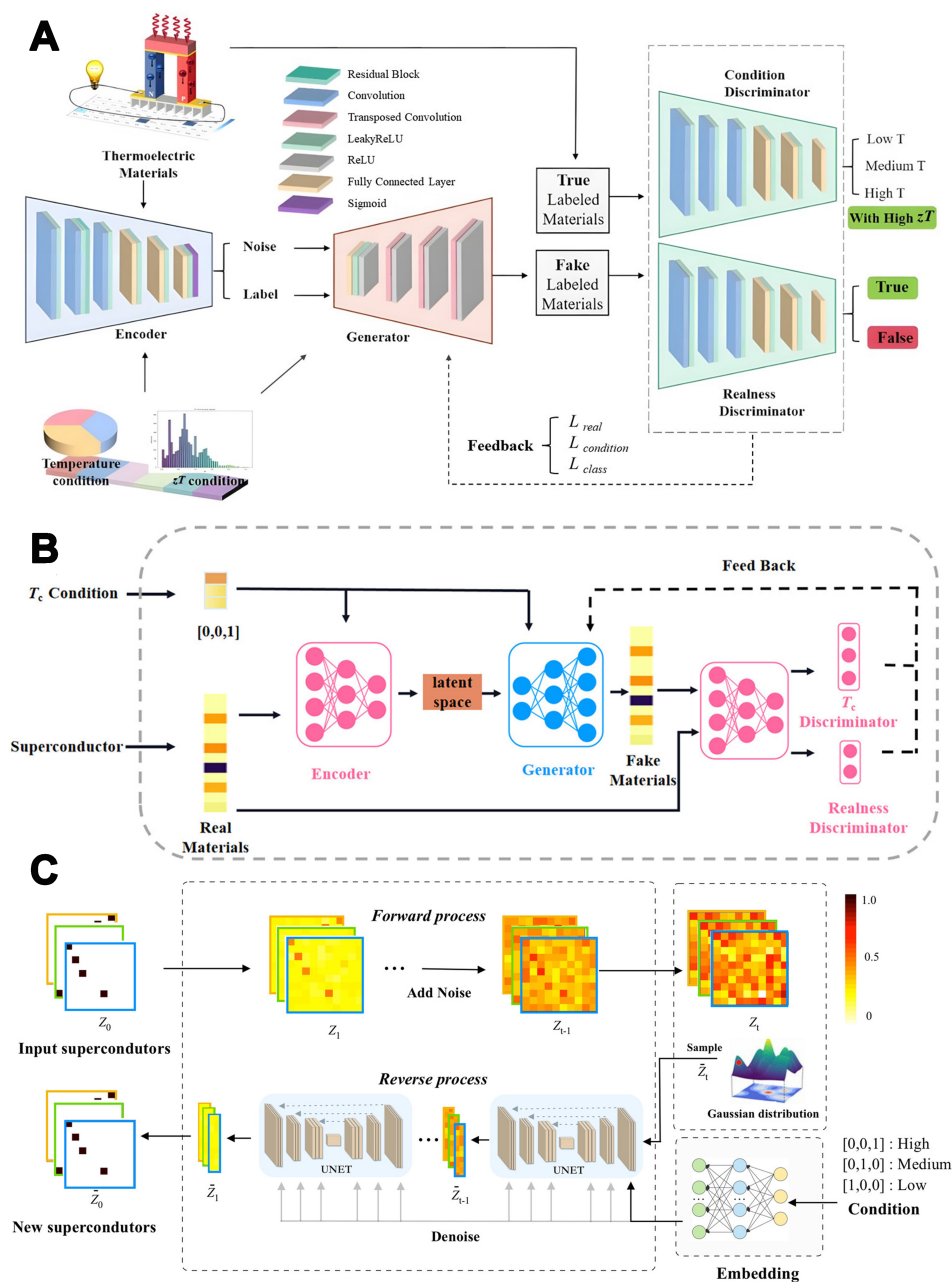
## INVERSE DESIGN OF THERMOELECTRIC MATERIALS

Inverse design is driving a paradigm shift in the discovery of thermoelectric materials. Compared to forward ML predictions based on structure-property mapping, inverse design demonstrates three pivotal advantages: (1) Elimination of redundant iterative processes by establishing end-to-end “target property  $\rightarrow$  material configuration” generative models, thereby avoiding inefficient inverse deduction required in forward approaches; (2) Dynamic multi-parameter co-optimization through constraint satisfaction algorithms and Pareto frontier analysis, enabling simultaneous optimization of competing parameters (e.g., electrical conductivity vs. thermal conductivity) while embedding experimental constraints (synthesis temperature, elemental cost) into the generation workflow; (3) Integration of global exploration and local refinement mechanisms, combining reinforcement learning for broad material space screening with variational autoencoders (VAEs) for atomic-level tuning of lattice defects and carrier concentration to surpass performance limits of empirical design. Currently evolving from an auxiliary tool to a core paradigm, inverse design is transforming thermoelectric materials development through its “target-driven/experimental-constrained” framework, marking the transition from trial-and-error approaches to intelligent customization and heralding the advent of on-demand materials design era.

Long *et al.* proposed a conditional generative adversarial network (CVAEGAN) framework combined with a ResNet-enhanced encoder and diversity-driven loss function, as shown in Figure 6A. This framework can systematically explore a vast compositional space under strict experimental constraints (e.g., synthesis temperature below 1,200 °C and elemental cost thresholds). The key methodological advances are: A dataset of 3,000 thermoelectric materials, covering eight major systems (e.g., Mg-based alloys, BiTe, HHs), was constructed through SMOTE oversampling and literature mining, ensuring balanced representation across temperature ranges (low/medium/high) and doping complexity. By encoding  $zT$  targets and processability criteria as conditional inputs, the model generates materials that inherently satisfy multi-objective Pareto optimization, achieving  $zT > 1.0$  in 100 candidate compositions, including 72.8% novel unreported systems. A combination of reinforcement learning-guided global screening and VAE-based atomic-level refinement dynamically optimizes carrier concentration gradients and phonon scattering centers, yielding materials with negative formation energies (98.3% stability) and enhanced electronic-thermal transport balance. The experimental validation of  $\text{Mg}_{3.1}\text{Sb}_{0.5}\text{Bi}_{1.497}\text{Te}_{0.003}$  ( $zT = 0.75$  at 300 K) highlights the model’s ability to bridge theoretical predictions and synthesis feasibility. This paradigm shift from “property-to-structure” mapping not only accelerates material discovery but also establishes a scalable framework for tackling multi-parameter trade-offs in functional materials, heralding a new era of intelligent, constraint-aware design<sup>[110]</sup>.

While the CVAEGAN framework proposed by Long *et al.* achieves constrained generation under experimental constraints such as synthesis temperature and elemental cost thresholds, challenges persist in encoding





**Figure 6.** Some current reverse design methods are as follows: (A) CVAE-GAN combined with a ResNet-enhanced encoder<sup>[110]</sup>; (B) GANs<sup>[112]</sup>; (C) Supercon-diffusion<sup>[111]</sup>. CVAE-GAN: Conditional variational auto - encoder generative adversarial network.

complex physical constraints (e.g., crystal structure stability, defect formation energy) and dynamic synthesis conditions (e.g., cooling rate, doping uniformity). Current models rely on manually predefined thresholds, struggling to accurately characterize nonlinear constraints such as metastable phase evolution and interfacial effects in real material systems. Developing data-driven constraint embedding methods based on first-principles calculations - such as converting DFT-derived phonon dispersion relations and electronic band structures into implicit regularization terms for generative models - remains critical.

Integrating inverse design with DFT workflows also faces bottlenecks, as existing frameworks depend on post-hoc DFT validation for key parameters such as electron-phonon coupling and thermal transport anisotropy,

leading to high computational costs in the “generate-validate” cycle. Constructing cross-scale transfer models to integrate DFT-derived descriptors (e.g., effective mass, relaxation time) into real-time constraint feedback during network propagation is essential for efficiency.

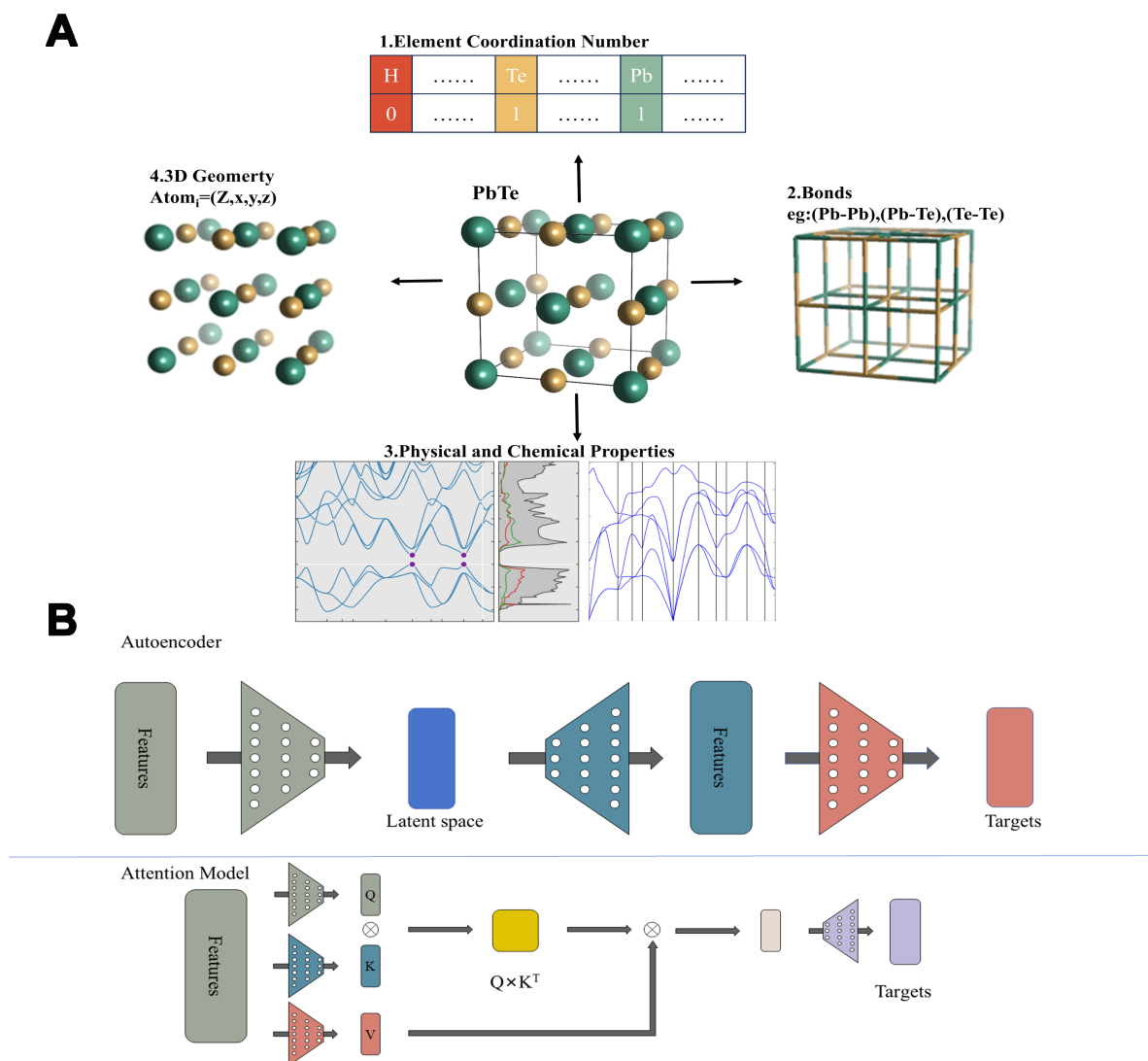
Experimental validation further encounters challenges in high-throughput screening and characterization, where only a fraction of generated high-zT candidates (e.g.,  $\text{Mg}_{3.1}\text{Sb}_{0.5}\text{Bi}_{1.497}\text{Te}_{0.003}$ ) are experimentally verified, highlighting discrepancies between theoretical predictions and practical synthesis conditions (e.g., stoichiometric errors, grain boundary defects). Future efforts should focus on establishing closed-loop integration between inverse design models and robotic synthesis platforms, leveraging active learning strategies to optimize sample selection and developing *in-situ* characterization techniques to enable an intelligent “generate-synthesize-validate-iterate” design cycle. Overcoming these challenges will drive thermoelectric material design from “constraint-aware” to “physically intelligent” paradigms, providing a universal methodology for inverse design in multi-parameter coupling systems.

At present, there are many blanks in the inverse design of thermoelectric materials, while the inverse design has been very popular in the design of other types of materials. Here, we will introduce several effective inverse design methods to provide new ideas for the inverse design of thermoelectric materials.

In the field of high-temperature superconductor inverse design, Zhong *et al.* proposed a deep generative model that combines VAE and generative adversarial networks (GANs), as shown in Figure 6B. This model maps superconductor compositions into a low-dimensional latent space via the encoder and applies a conditional generative adversarial mechanism to achieve precise regulation of the critical temperature ( $T_c$ ). The research team extracted data for 7,375 superconductors from the SuperCon database and categorized them into three groups based on  $T_c$ : high ( $> 77$  K), medium (40–77 K), and low (20–40 K), which were used as generation conditions. Through adversarial learning, the model optimized the authenticity of the generated samples and their  $T_c$  alignment, successfully predicting hundreds of potential superconductor compositions with  $T_c > 77$  K. Notably, the model revealed a relationship between copper concentration and  $T_c$  in copper-based superconductors, finding that when the copper concentration approximates 2.41 (e.g.,  $\text{Hg}_{0.37}\text{Ba}_{1.73}\text{Ca}_{1.18}\text{Cu}_{2.43}\text{O}_{6.93}\text{Tl}_{0.69}$ ),  $T_c$  peaks at 129.4 K, consistent with experimentally observed effects of copper-oxide layers. Although the model still has limitations in terms of diversity and charge neutrality validation, it represents the first application of generative models in superconductor inverse design, providing a critical paradigm for future research<sup>[111]</sup>.

To address the limitations of traditional generative models in designing doped superconductors, Zhong *et al.* further proposed the Supercon-Diffusion method, which is based on a diffusion model and three-channel matrix representation. This method innovatively decomposes stoichiometric numbers into integer, first decimal, and second decimal channels, as shown in Figure 6C. Through a stepwise noise addition and denoising process, combined with  $T_c$ -condition constraints, it achieves high-precision control of doping ratios. Training on 7,315 doped superconductor data points, the model generated samples with improvements in charge neutrality (55%) and doping effectiveness (55%), exceeding traditional GANs by over ten times. Additionally, 98% of the generated samples exhibited negative formation energies, indicating thermodynamic stability. The study also found that the model could automatically identify optimal doping ranges in key families (e.g.,  $\text{YBa}_2\text{Cu}_3\text{O}_7$ ), such as  $T_c > 90$  K when La doping is in the range of 0.1–0.2, which aligns closely with experimental values. Through DOS calculations, generated superconductors (e.g.,  $\text{Hg}_{0.75}\text{Tl}_{0.25}\text{Ba}_2\text{Ca}_2\text{Cu}_3\text{O}_8$ ) exhibited flat-band characteristics near the Fermi level, resembling those of copper-based superconductors, further validating composition-performance correlations. This work not only proposed 200 previously unreported high- $T_c$  candidate materials but also demonstrated the stability of the diffusion model and the advantages of three-channel representation, providing a novel approach for inverse design in complex doped systems<sup>[112]</sup>.





**Figure 7.** (A) Current utilized features (take PbTe as example); (B) Current utilized deep learning methods.

## CONCLUSION

In this paper, we provided a comprehensive introduction to the application of ML in the field of thermoelectric materials. We can utilize ML to predict various properties of thermoelectric materials and also employ it to assist in the design of novel thermoelectric materials. [Figure 7A](#) shows some existing material features and [Figure 7B](#) shows deep learning methods used in thermoelectric material ML. Although some achievements have been made in the application of ML in the field of thermoelectric materials, there are still many limitations. Here, we offer some potential directions.

### Structural prediction and optimization

In order to achieve efficient design and development of thermoelectric materials, the development of advanced ML models to predict their crystal structures and atomic arrangements is of paramount importance. These models, trained on large-scale databases of known crystal structures, are capable of learning complex struc-

tural features and patterns, thereby significantly enhancing the accuracy of predicting new structural configurations. This data-driven predictive approach not only accelerates the discovery process of new materials but also provides valuable theoretical guidance for experimental research. Furthermore, integrating ML predictions with first-principles calculations enables in-depth validation and optimization of the predicted structures. First-principles calculations, based on quantum mechanics, can precisely describe the physical properties of materials at the electronic level. Through this integration, researchers can conduct detailed stability analyses and electronic property calculations of the predicted structures, thereby screening for thermoelectric materials with potentially high performance. This synergistic approach not only increases the reliability of predictions but also offers theoretical support for further material optimization. In addition, high-throughput screening technology plays a crucial role in this process. With the aid of automated computational tools, researchers can rapidly evaluate a large number of predicted structures to identify candidate materials with optimal thermoelectric properties. This method enables the processing and analysis of vast amounts of data in a short period of time, significantly improving the efficiency of material screening and reducing the workload of experimental validation. High-throughput screening not only quickly identifies materials with superior performance but also provides a clear direction for subsequent experimental synthesis and performance optimization.

### Multi-scale modeling and simulation

With the continuous development of science and technology, the demand for high-performance thermoelectric materials is steadily increasing. In order to better understand and optimize the performance of these materials, it is necessary to develop multi-scale ML models that can predict the performance of thermoelectric materials across different length scales (from atomic to macroscopic). These models are capable of capturing the complex relationships between atomic structure, microstructure, and macroscopic properties through advanced algorithms and data processing techniques, thereby providing more comprehensive theoretical support for material design. Specifically, the structure at the atomic scale determines the fundamental physical properties of the material, while the microstructure influences its microscopic physical behavior. Macroscopic properties are the integrated manifestation of these microscopic characteristics. By using multi-scale ML models, these characteristics at different levels can be connected, enabling accurate prediction of the performance of thermoelectric materials. Moreover, combining the physical theories of thermoelectric transport with ML algorithms not only helps to deeply understand the fundamental mechanisms controlling thermoelectric performance but also enables the prediction of material behavior under various complex conditions. This integration fully leverages the guiding role of physical theory and the powerful data processing capabilities of ML, offering new perspectives and methods for the study of thermoelectric materials. Finally, ML technology can also be used to simulate the behavior of thermoelectric materials under different working conditions. For example, in practical applications, thermoelectric materials often need to operate under varying temperature gradients and mechanical stresses. Through ML models, these complex working conditions can be simulated and analyzed, providing strong support for the design and optimization of materials. This not only helps to enhance the durability of the materials but also further improves their performance, making them better suited to meet practical application requirements. Therefore, the development of multi-scale ML models and their application in the research and design of thermoelectric materials are of great significance for advancing thermoelectric technology.

### Inverse design and feedback loop

In recent years, with the rapid development of artificial intelligence technologies, inverse design models have shown great potential in the discovery and design of novel thermoelectric materials. Among them, emerging technologies such as GANs, diffusion models, VAEs, and generative methods based on reinforcement learning are becoming important tools for driving innovation in thermoelectric materials.

The core advantage of these models lies in their ability to start from target performance and inversely generate material structures with specific functions. GANs optimize the generated material structures through

the adversarial training between the generator and discriminator, bringing their thermoelectric performance close to or even beyond that of existing materials. Diffusion models, on the other hand, reconstruct material configurations with ideal performance from random data by gradually removing noise. VAEs map the structural features of materials into a low-dimensional space through the synergy of the encoder and decoder, and then reconstruct materials with optimized performance through the decoder. Additionally, generative methods based on reinforcement learning can dynamically adjust material design strategies through a reward mechanism, thus efficiently exploring the material design space.

These advanced inverse design models not only break through the limitations of traditional design thinking but also, with the support of large-scale data, rapidly explore the material design space to discover novel materials with unique microstructures and excellent thermoelectric performance. For example, by combining physical theories with ML algorithms, these models can generate materials with specific atomic arrangements, microstructures, and macroscopic properties, thus providing new ideas for the design of high-performance thermoelectric materials.

Looking to the future, these inverse design models are expected to form a close feedback loop with experimental synthesis and characterization techniques. ML models can generate potential material structures, while experimental validation provides feedback data to further optimize the accuracy and reliability of the models. Through this iterative optimization process, not only can the discovery of high-performance thermoelectric materials be accelerated, but the cost and time of research and development can also be significantly reduced. Moreover, with the continuous improvement of computational power and the increasing richness of data resources, these models will be able to handle more complex material systems and even achieve material design under multi-physics coupling conditions.

Ultimately, the integration of various inverse design methods, including GANs, diffusion models, VAEs, and reinforcement learning, is expected to bring revolutionary breakthroughs to the field of thermoelectric materials.

## DECLARATIONS

### Authors' contributions

Writing – review: Wang, Y.

Editing, writing: Zhong, C.

Conceptualization: Zhang, J.

Review: Liu, J.

Data curation: Hu, K.

Writing – review: Chen, J.

Editing: Lin, X.

### Availability of data and materials

Not applicable.

### Financial support and sponsorship

The authors appreciate financial support from the Guangdong Basic and Applied Basic Research Foundation (2022A1515110676, 2024A1515011845), the Shenzhen Science and Technology Program (JCYJ20220531095404009; RCBS20221008093057027; JCYJ2023080 7094313028, JCYJ20230807094318038), the Sunrise (Xiamen) Photovoltaic Industry Co., Ltd. (Development of Artificial Intelligence Technology for Perovskite Photovoltaic Materials, No. HX20230176), the Natural Science Foundation of China (62102118), and the Shenzhen Colleges and Universities Stable Support Program (GXWD20 220811170504001).

### Conflicts of interest

Liu, J. is affiliated with Sunrise (Xiamen) Photovoltaic Industry Co., Ltd, while the other authors have declared that they have no conflicts of interest.

### Ethical approval and consent to participate

Not applicable.

### Consent for publication

Not applicable.

### Copyright

© The Author(s) 2025.

## REFERENCES

1. Cao, Y.; Sheng, Y.; Li, X.; Xi, L.; Yang, J. Application of materials genome methods in thermoelectrics. *Front. Mater.* **2022**, *9*, 861817. [DOI](#)
2. Wan, X.; Feng, W.; Wang, Y.; et al. Materials discovery and properties prediction in thermal transport via materials informatics: a mini review. *Nano Lett.* **2019**, *19*, 3387–95. [DOI](#)
3. Wang, T.; Zhang, C.; Snoussi, H.; Zhang, G. Machine learning approaches for thermoelectric materials research. *Adv. Funct. Mater.* **2020**, *30*, 1906041. [DOI](#)
4. Zhang, Z.; Jiang, Y.; Shu, M.; Li, L.; Dong, Z.; Xu, J. Artificial photosynthesis over metal halide perovskites: achievements, challenges, and prospects. *J. Phys. Chem. Lett.* **2021**, *12*, 5864–70. [DOI](#)
5. Yang, J.; Xi, L.; Qiu, W.; et al. On the tuning of electrical and thermal transport in thermoelectrics: an integrated theory–experiment perspective. *NPJ Comput. Mater.* **2016**, *2*, 15015. [DOI](#)
6. Uchida, K.; Takahashi, S.; Harii, K.; et al. Observation of the spin Seebeck effect. *Nature* **2008**, *455*, 778–81. [DOI](#)
7. Drebuschak, V. A. The Peltier effect. *J. Therm. Anal. Calorim.* **2008**, *91*, 311–5. [DOI](#)
8. Geballe, T. H.; Hull, G. W. Seebeck effect in germanium. *Phys. Rev.* **1954**, *94*, 1134. [DOI](#)
9. Crane, D. T.; Jackson, G. S. Optimization of cross flow heat exchangers for thermoelectric waste heat recovery. *Energy Convers. Manag.* **2004**, *45*, 1565–82. [DOI](#)
10. Qin, Y.; Qin, B.; Wang, D.; Chang, C.; Zhao, L. D. Solid-state cooling: thermoelectrics. *Energy Environ. Sci.* **2022**, *15*, 4527–41. [DOI](#)
11. Chen, W. Y.; Shi, X. L.; Zou, J.; Chen, Z. G. Thermoelectric coolers for on-chip thermal management: materials, design, and optimization. *Mater. Sci. Eng. R. Rep.* **2022**, *151*, 100700. [DOI](#)
12. Yang, J. Potential applications of thermoelectric waste heat recovery in the automotive industry. In *ICT 2005. 24th International Conference on Thermoelectrics*, 2005, Clemson, USA. Jun 19–23, 2005. IEEE; 2005. pp. 170–4. [DOI](#)
13. Xie, H.; Zhang, Y.; Gao, P. Thermoelectric-powered sensors for Internet of Things. *Micromachines* **2022**, *14*, 31. [DOI](#)
14. Bonin, R.; Boero, D.; Chiaberge, M.; Tonoli, A. Design and characterization of small thermoelectric generators for environmental monitoring devices. *Energy Convers. Manag.* **2013**, *73*, 340–9. [DOI](#)
15. Date, A.; Date, A.; Dixon, C.; Akbarzadeh, A. Progress of thermoelectric power generation systems: prospect for small to medium scale power generation. *Renew. Sustain. Energy Rev.* **2014**, *33*, 371–81. [DOI](#)
16. He, R.; Schierming, G.; Nielsch, K. Thermoelectric devices: a review of devices, architectures, and contact optimization. *Adv. Mater. Technol.* **2018**, *3*, 1700256. [DOI](#)
17. Zhang, Q.; Deng, K.; Wilkens, L.; Reith, H.; Nielsch, K. Micro-thermoelectric devices. *Nat. Electron.* **2022**, *5*, 333–47. [DOI](#)
18. Zhang, Q. H.; Huang, X. Y.; Bai, S. Q.; Shi, X.; Uher, C.; Chen, L. D. Thermoelectric devices for power generation: recent progress and future challenges. *Adv. Eng. Mater.* **2016**, *18*, 194–213. [DOI](#)
19. Sajid, M.; Hassan, I.; Rahman, A. An overview of cooling of thermoelectric devices. *Renew. Sustain. Energy Rev.* **2017**, *78*, 15–22. [DOI](#)
20. Belsky, A. A.; Glukhanich, D. Y. Standalone power system with photovoltaic and thermoelectric installations for power supply of remote monitoring and control stations for oil pipelines. *Renew. Energy Focus* **2023**, *47*, 100493. [DOI](#)
21. Palaporn, D.; Tanusilp, S.; Sun, Y.; Pinitsoontorn, S.; Kurosaki, K. Thermoelectric materials for space explorations. *Mater. Adv.* **2024**, *5*, 5351–64. [DOI](#)
22. Venkatasubramanian, R.; Siivola, E.; Colpitts, T.; O’Quinn, B. Thin-film thermoelectric devices with high room-temperature figures of merit. *Nature* **2001**, *413*, 597–602. [DOI](#)
23. Yang, S.; Qiu, P.; Chen, L.; Shi, X. Recent developments in flexible thermoelectric devices. *Small Sci.* **2021**, *1*, 2100005. [DOI](#)
24. Snyder, G. J.; Snyder, A. H. Figure of merit  $ZT$  of a thermoelectric device defined from materials properties. *Energy Environ. Sci.* **2017**, *10*, 2280–3. [DOI](#)
25. Kim, H. S.; Gibbs, Z. M.; Tang, Y.; Wang, H.; Snyder, G. J. Characterization of Lorenz number with Seebeck coefficient measurement. *APL Mater.* **2015**, *3*, 041506. [DOI](#)

26. Martin, J.; Tritt, T.; Uher, C. High temperature Seebeck coefficient metrology. *J. Appl. Phys.* **2010**, *108*, 121101. DOI
27. de Boor, J.; Müller, E. Data analysis for Seebeck coefficient measurements. *Rev. Sci. Instrum.* **2013**, *84*, 065102. DOI
28. Snyder, G. J.; Pereyra, A.; Gurunathan, R. Effective mass from Seebeck coefficient. *Adv. Funct. Mater.* **2022**, *32*, 2112772. DOI
29. Iwanaga, S.; Toberer, E. S.; LaLonde, A.; Snyder, G. J. A high temperature apparatus for measurement of the Seebeck coefficient. *Rev. Sci. Instrum.* **2011**, *82*, 063905. DOI
30. Mott, N. F. The electrical conductivity of transition metals. *Proc. R. Soc. Lond. A* **1936**, *153*, 699–717. DOI
31. Radzuan, N. A. M.; Sulong, A. B.; Sahari, J. A review of electrical conductivity models for conductive polymer composite. *Int. J. Hydrogen Energy* **2017**, *42*, 9262–73. DOI
32. Ebbesen, T. W.; Lezec, H. J.; Hiura, H.; Bennett, J. W.; Ghaemi, H. F.; Thio, T. Electrical conductivity of individual carbon nanotubes. *Nature* **1996**, *382*, 54–6. DOI
33. Bardeen, J. Electrical conductivity of metals. *J. Appl. Phys.* **1940**, *11*, 88–111. DOI
34. Kerner, E. H. The electrical conductivity of composite media. *Proc. Phys. Soc. B* **1956**, *69*, 802. DOI
35. Venkatasubramanian, R. Lattice thermal conductivity reduction and phonon localizationlike behavior in superlattice structures. *Phys. Rev. B* **2000**, *61*, 3091. DOI
36. Zapata-Arteaga, O.; Perevedentsev, A.; Marina, S.; Martin, J.; Reparaz, J. S.; Campoy-Quiles, M. Reduction of the lattice thermal conductivity of polymer semiconductors by molecular doping. *ACS Energy Lett.* **2020**, *5*, 2972–8. DOI
37. Murakami, T.; Shiga, T.; Hori, T.; Esfarjani, K.; Shiomi, J. Importance of local force fields on lattice thermal conductivity reduction in  $\text{PbTe}_{1-x}\text{Se}_x$  alloys. *EPL* **2013**, *102*, 46002. DOI
38. Wan, C.; Wang, Y.; Wang, N.; Norimatsu, W.; Kusunoki, M.; Koumoto, K. Development of novel thermoelectric materials by reduction of lattice thermal conductivity. *Sci. Technol. Adv. Mater.* **2010**, *11*, 044306. DOI
39. Kim, T. Y.; Park, C. H.; Marzari, N. The electronic thermal conductivity of graphene. *Nano Lett.* **2016**, *16*, 2439–43. DOI
40. Graf, M. J.; Yip, S. K.; Sauls, J. A.; Rainer, D. Electronic thermal conductivity and the Wiedemann-Franz law for unconventional superconductors. *Phys. Rev. B* **1996**, *53*, 15147. DOI
41. Lee, S.; Hippalgaonkar, K.; Yang, F.; et al. Anomalously low electronic thermal conductivity in metallic vanadium dioxide. *Science* **2017**, *355*, 371–4. DOI
42. Ambegaokar, V.; Tewordt, L. Theory of the electronic thermal conductivity of superconductors with strong electron-phonon coupling. *Phys. Rev.* **1964**, *134*, A805. DOI
43. Burger, N.; Laachachi, A.; Ferriol, M.; Lutz, M.; Toniazzi, V.; Ruch, D. Review of thermal conductivity in composites: mechanisms, parameters and theory. *Prog. Polym. Sci.* **2016**, *61*, 1–28. DOI
44. Wang, D. Z.; Liu, W. D.; Shi, X. L.; et al. Se-alloying reducing lattice thermal conductivity of  $\text{Ge}_{0.95}\text{Bi}_{0.05}\text{Te}$ . *J. Mater. Sci. Technol.* **2022**, *106*, 249–56. DOI
45. Zhang, Q.; Song, Q.; Wang, X.; et al. Deep defect level engineering: a strategy of optimizing the carrier concentration for high thermoelectric performance. *Energy Environ. Sci.* **2018**, *11*, 933–40. DOI
46. Pei, Y.; Wang, H.; Snyder, G. J. Band engineering of thermoelectric materials. *Adv. Mater.* **2012**, *24*, 6125–35. DOI
47. He, J.; Sootsman, J. R.; Girard, S. N.; et al. On the origin of increased phonon scattering in nanostructured PbTe based thermoelectric materials. *J. Am. Chem. Soc.* **2010**, *132*, 8669–75. DOI
48. Zheng, Y.; Slade, T. J.; Hu, L.; et al. Defect engineering in thermoelectric materials: what have we learned? *Chem. Soc. Rev.* **2021**, *50*, 9022–54. DOI
49. Xie, H.; Zhao, L. D.; Kanatzidis, M. G. Lattice dynamics and thermoelectric properties of diamondoid materials. *Interdiscip. Mater.* **2024**, *3*, 5–28. DOI
50. Qin, D.; Shi, W.; Lu, Y.; Cai, W.; Liu, Z.; Sui, J. Roles of interface engineering in performance optimization of skutterudite-based thermoelectric materials. *Carbon Neutraliz.* **2022**, *1*, 233–46. DOI
51. Chen, J.; Li, K.; Liu, C.; et al. Enhanced efficiency of thermoelectric generator by optimizing mechanical and electrical structures. *Energies* **2017**, *10*, 1329. DOI
52. Lineykin, S.; Ben-Yaakov, S. Modeling and analysis of thermoelectric modules. *IEEE Trans. Ind. Appl.* **2007**, *43*, 505–12. DOI
53. Yang, Y.; Hu, H.; Chen, Z.; et al. Stretchable nanolayered thermoelectric energy harvester on complex and dynamic surfaces. *Nano Lett.* **2020**, *20*, 4445–53. DOI
54. Tritt, T. M. Thermoelectric phenomena, materials, and applications. *Ann. Rev. Mater. Res.* **2011**, *41*, 433–48. DOI
55. Freysoldt, C.; Grabowski, B.; Hickel, T.; et al. First-principles calculations for point defects in solids. *Rev. Mod. Phys.* **2014**, *86*, 253. DOI
56. Li, W.; Carrete, J.; Katcho, N. A.; Mingo, N. ShengBTE: a solver of the Boltzmann transport equation for phonons. *Comput. Phys. Commun.* **2014**, *185*, 1747–58. DOI
57. Jonson, M.; Mahan, G. D. Mott's formula for the thermopower and the Wiedemann-Franz law. *Phys. Rev. B* **1980**, *21*, 4223. DOI
58. Binder, K.; Horbach, J.; Kob, W.; Paul, W.; Varnik, F. Molecular dynamics simulations. *J. Phys. Condens. Matter* **2004**, *16*, S429. DOI
59. Kroese, D. P.; Brereton, T.; Taimre, T.; Botev, Z. I. Why the Monte Carlo method is so important today. *Wiley Interdiscip. Rev. Comput. Stat.* **2014**, *6*, 386–92. DOI
60. Wei, J.; Chu, X.; Sun, X. Y.; et al. Machine learning in materials science. *InfoMat* **2019**, *1*, 338–58. DOI
61. Ramprasad, R.; Batra, R.; Pilania, G.; Mannodi-Kanakkithodi, A.; Kim, C. Machine learning in materials informatics: recent applications and prospects. *npj Comput. Mater.* **2017**, *3*, 54. DOI
62. Morgan, D.; Jacobs, R. Opportunities and challenges for machine learning in materials science. *Ann. Rev. Mater. Res.* **2020**, *50*, 71–103. DOI



63. Challapalli, A.; Patel, D.; Li, G. Inverse machine learning framework for optimizing lightweight metamaterials. *Mater. Design* **2021**, *208*, 109937. DOI
64. Sliwa, B.; Piatkowski, N.; Wietfeld, C. LIMITS: lightweight machine learning for IoT systems with resource limitations. In *ICC 2020-2020 IEEE International Conference on Communications (ICC)*, Dublin, Ireland. Jun 07-11, 2020. IEEE; 2020. p. 1–7. DOI
65. Mahmood, A.; Wang, J. L. Machine learning for high performance organic solar cells: current scenario and future prospects. *Energy Environ. Sci.* **2021**, *14*, 90–105. DOI
66. Zhang, L.; He, M. Unsupervised machine learning for solar cell materials from the literature. *J. Appl. Phys.* **2022**, *131*, 064902. DOI
67. Tao, Q.; Xu, P.; Li, M.; Lu, W. Machine learning for perovskite materials design and discovery. *Npj Comput. Mater.* **2021**, *7*, 23. DOI
68. Zhang, L.; He, M.; Shao, S. Machine learning for halide perovskite materials. *Nano Energy* **2020**, *78*, 105380. DOI
69. Al-Sabana, O.; Abdellatif, S. O. Optoelectronic devices informatics: optimizing DSSC performance using random-forest machine learning algorithm. *Optoelectron. Lett.* **2022**, *18*, 148–51. DOI
70. Miller, S. A.; Dylla, M.; Anand, S.; Gordiz, K.; Snyder, G. J.; Toberer, E. S. Empirical modeling of dopability in diamond-like semiconductors. *npj Comput. Mater.* **2018**, *4*, 71. DOI
71. Saal, J. E.; Kirklin, S.; Aykol, M.; Meredig, B.; Wolverton, C. Materials design and discovery with high-throughput density functional theory: the open quantum materials database (OQMD). *JOM* **2013**, *65*, 1501–9. DOI
72. Kirklin, S.; Saal, J. E.; Meredig, B.; et al. The Open Quantum Materials Database (OQMD): assessing the accuracy of DFT formation energies. *npj Comput. Mater.* **2015**, *1*, 15010. DOI
73. Jain, A.; Ong, S. P.; Hautier, G.; et al. Commentary: The Materials Project: A materials genome approach to accelerating materials innovation. *APL Mater.* **2013**, *1*, 011002. DOI
74. de Jong, M.; Chen, W.; Angsten, T.; et al. The high-throughput highway to computational materials design. *Sci. Data* **2013**, *2*, 150009. DOI
75. de Jong, M.; Chen, W.; Geerlings, H.; Asta, M.; Persson, K. A. A database to enable discovery and design of piezoelectric materials. *Sci. Data* **2015**, *2*, 150053. DOI
76. Sun, J.; Zhong, G.; Huang, K.; Dong, J. Banzhaf random forests: cooperative game theory based random forests with consistency. *Neural Netw.* **2018**, *106*, 20–9. DOI
77. Schmidt, A. F.; Finan, C. Linear regression and the normality assumption. *J. Clin. Epidemiol.* **2018**, *98*, 146–51. DOI
78. Agatonovic-Kustrin, S.; Beresford, R. Basic concepts of artificial neural network (ANN) modeling and its application in pharmaceutical research. *J. Pharm. Biomed. Anal.* **2000**, *22*, 717–27. DOI
79. Wan, Z.; Wang, Q. D.; Liu, D.; Liang, J. Machine learning prediction of the optimal carrier concentration and band gap of quaternary thermoelectric materials via element feature descriptors. *Int. J. Quantum Chem.* **2021**, *121*, e26752. DOI
80. Goldsmid, H. J. The electrical conductivity and thermoelectric power of bismuth telluride. *Proc. Phys. Soc.* **1958**, *71*, 633. DOI
81. Ricci, F.; Chen, W.; Aydemir, U.; et al. An *ab initio* electronic transport database for inorganic materials. *Sci. Data* **2017**, *4*, 170085. DOI
82. Antunes, L. M.; Butler, K. T.; Grau-Crespo, R. Predicting thermoelectric transport properties from composition with attention-based deep learning. *Mach. Learn. Sci. Technol.* **2023**, *4*, 015037. DOI
83. Tiriyaki, H.; Yusuf, A.; Ballikaya, S. Determination of electrical and thermal conductivities of *n*- and *p*-type thermoelectric materials by prediction iteration machine learning method. *Energy* **2024**, *292*, 130597. DOI
84. Furmanchuk, A.; Saal, J. E.; Doak, J. W.; Olson, G. B.; Choudhary, A.; Agrawal, A. Prediction of seebeck coefficient for compounds without restriction to fixed stoichiometry: a machine learning approach. *J. Comput. Chem.* **2018**, *39*, 191–202. DOI
85. Yuan, H. M.; Han, S. H.; Hu, R.; et al. Machine learning for accelerated prediction of the Seebeck coefficient at arbitrary carrier concentration. *Mater. Today Phys.* **2022**, *25*, 100706. DOI
86. Gaultois, M. W.; Sparks, T. D.; Borg, C. K. H.; Seshadri, R.; Bonificio, W. D.; Clarke, D. R. Data-driven review of thermoelectric materials: performance and resource considerations. *Chem. Mater.* **2013**, *25*, 2911–20. DOI
87. Gaultois, M. W.; Oliynyk, A. O.; Mar, A.; Sparks, T. D.; Mulholland, G. J.; Meredig, B. Perspective: Web-based machine learning models for real-time screening of thermoelectric materials properties. *APL Mater.* **2016**, *4*, 053213. DOI
88. Sheng, Y.; Wu, Y.; Yang, J.; Lu, W.; Villars, P.; Zhang, W. Active learning for the power factor prediction in diamond-like thermoelectric materials. *npj Comput. Mater.* **2020**, *6*, 171. DOI
89. Graziosi, P.; Li, Z.; Neophytou, N. Bipolar conduction asymmetries lead to ultra-high thermoelectric power factor. *Appl. Phys. Lett.* **2022**, *120*, 072102. DOI
90. Qin, G.; Wei, Y.; Yu, L.; et al. Predicting lattice thermal conductivity from fundamental material properties using machine learning techniques. *J. Mater. Chem. A* **2023**, *11*, 5801–10. DOI
91. Ren, Q.; Chen, D.; Rao, L.; Lun, Y.; Tang, G.; Hong, J. Machine-learning-assisted discovery of 212-Zintl-phase compounds with ultra-low lattice thermal conductivity. *J. Mater. Chem. A* **2024**, *12*, 1157–65. DOI
92. Wudil, Y. S. Ensemble learning-based investigation of thermal conductivity of Bi<sub>2</sub>Te<sub>2.7</sub>Se<sub>0.3</sub>-based thermoelectric clean energy materials. *Results Eng.* **2023**, *18*, 101203. DOI
93. Tewari, A.; Dixit, S.; Sahni, N.; Bordas, S. P. A. Machine learning approaches to identify and design low thermal conductivity oxides for thermoelectric applications. *Data Centric Eng.* **2020**, *1*, e8. DOI
94. Al-Fahdi, M.; Lin, C.; Shen, C.; Zhang, H.; Hu, M. Rapid prediction of phonon density of states by crystal attention graph neural network and high-throughput screening of candidate substrates for wide bandgap electronic cooling. *Mater. Today Phys.* **2025**, *50*, 101632. DOI
95. Dong, L.; Li, W.; Bu, X. H. Predicting thermal transport properties in phononic crystals via machine learning. *Appl. Phys. Lett.* **2024**, *124*, 162201. DOI

96. Guo, Z.; Roy Chowdhury, P.; Han, Z.; et al. Fast and accurate machine learning prediction of phonon scattering rates and lattice thermal conductivity. *npj Comput. Mater.* **2023**, *9*, 95. DOI
97. Zhou, Z.; Cao, G.; Liu, J.; Liu, H. High-throughput prediction of the carrier relaxation time via data-driven descriptor. *npj Comput. Mater.* **2020**, *6*, 149. DOI
98. Al-Fahdi, M.; Rurali, R.; Hu, J.; Wolverton, C.; Hu, M. Accelerating Discovery of extreme lattice thermal conductivity by crystal attention graph neural network (CATGNN) using chemical bonding intuitive descriptors. *arXiv* **2024**; arXiv:2410.16066. <http://dx.doi.org/10.48550/arXiv.2410.16066>. (accessed 3 Jun 2025)
99. You, H. J.; Chiang, Y. T.; Bansil, A.; Lin, H. Effects of four-phonon scattering and wave-like phonon tunneling effects on thermoelectric properties of  $\text{Mg}_2\text{GeSe}_4$  using machine learning. *arXiv* **2024**; arXiv:2411.10605. <https://doi.org/10.48550/arXiv.2411.10605>. (accessed 3 Jun 2025)
100. Yang, Y.; Lin, Y.; Dai, S.; et al. HH130: a standardized database of machine learning interatomic potentials, datasets, and its applications in the thermal transport of half-Heusler thermoelectrics. *Digit. Discov.* **2024**, *3*, 2201–10. DOI
101. Li, Y.; Zhang, J.; Zhang, K.; Zhao, M.; Hu, K.; Lin, X. Large data set-driven machine learning models for accurate prediction of the thermoelectric figure of merit. *ACS Appl. Mater. Interfaces* **2022**, *14*, 55517–27. DOI
102. Xu, Y.; Liu, X.; Wang, J. Prediction of thermoelectric-figure-of-merit based on autoencoder and light gradient boosting machine. *J. Appl. Phys.* **2024**, *135*, 074901. DOI
103. Wang, Y.; Zhong, C.; Zhang, J.; Yao, H.; Chen, J.; Lin, X. High-Performance stacking ensemble learning for thermoelectric figure-of-merit prediction. *Mater. Design* **2025**, *249*, 113552. DOI
104. Madavali, B.; Nagarjuna, C.; Dewangan, S. K.; Ahn, B.; Hong, S. J. Predicting the thermoelectric figure of merit in p-type BiSbTe-based alloys using artificial neural network modeling. *Mater. Today Commun.* **2024**, *40*, 109396. DOI
105. Jia, X.; Deng, Y.; Bao, X.; et al. Unsupervised machine learning for discovery of promising half-Heusler thermoelectric materials. *npj Comput. Mater.* **2022**, *8*, 34. DOI
106. Vaitesswar, U. S.; Bash, D.; Huang, T.; et al. Machine learning based feature engineering for thermoelectric materials by design. *Digit. Discov.* **2024**, *3*, 210–20. DOI
107. Xu, Y.; Jiang, L.; Qi, X. Machine learning in thermoelectric materials identification: feature selection and analysis. *Comput. Mater. Sci.* **2021**, *197*, 110625. DOI
108. Fan, T.; Oganov, A. R. Combining machine learning models with first-principles high-throughput calculation to accelerate the search of promising thermoelectric materials. *J. Mater. Chem. C* **2025**, *13*, 1439–48. DOI
109. Chen, C.; Ong, S. P. A universal graph deep learning interatomic potential for the periodic table. *Nat. Comput. Sci.* **2022**, *2*, 718–28. DOI
110. Long, Y.; Zhong, C.; Ma, X.; et al. Inverse design of high-performance thermoelectric materials via a generative model combined with experimental verification. *ACS Appl. Mater. Interfaces* **2025**, *17*, 19856–67. DOI
111. Zhong, C.; Zhang, J.; Wang, Y.; et al. High-performance diffusion model for inverse design of high  $T_c$  superconductors with effective doping and accurate stoichiometry. *InfoMat* **2024**, *6*, e12519. DOI
112. Zhong, C.; Zhang, J.; Lu, X.; et al. Deep generative model for inverse design of high-temperature superconductor compositions with predicted  $T_c > 77$  K. *ACS Appl. Mater. Interfaces* **2023**, *15*, 30029–38. DOI