

Physics-guided auxiliary multi-task learning for temperature-dependent thermoelectric zT curve prediction

Jie He

jiehe_henu_edu@163.com

Henan University <https://orcid.org/0009-0000-3096-4220>

Research Article

Keywords: Thermoelectric materials, thermoelectric figure of merit, temperature-dependent properties, physics-guided machine learning, multi-task learning, materials informatics

Posted Date: October 21st, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-11439866/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

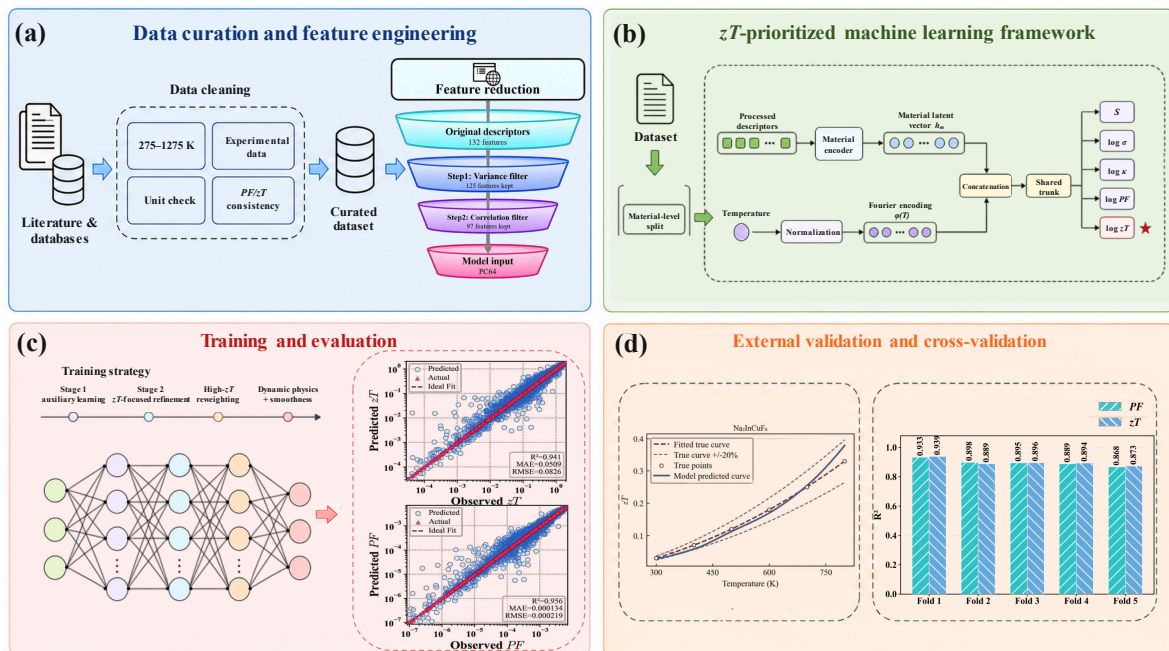
[Read Full License](#)

Additional Declarations: The authors declare no competing interests.

Graphical Abstract

Physics-guided auxiliary multi-task learning for temperature-dependent thermoelectric zT curve prediction

Jie He, Xiaopeng Lian, Kaiyue Liu, Xuyang Ge



Highlights

Physics-guided auxiliary multi-task learning for temperature-dependent thermoelectric zT curve prediction

Jie He, Xiaopeng Lian, Kaiyue Liu, Xuyang Ge

- Predicts temperature-dependent thermoelectric zT curves from composition and temperature.
- Prioritizes zT while learning coupled transport properties.
- Uses progressive physical consistency during multi-task optimization.
- Evaluates transfer across materials with grouped and external validation.

Physics-guided auxiliary multi-task learning for temperature-dependent thermoelectric zT curve prediction

Jie He^a, Xiaopeng Lian^a, Kaiyue Liu^a and Xuyang Ge^{a,*}

^aHenan Key Laboratory of High Efficiency Energy Conversion Science and Technology, Henan International Joint Laboratory of New Energy Materials and Devices, School of Physics and Electronics, Henan University, Kaifeng, 475004, China

ARTICLE INFO

Keywords:

Thermoelectric materials
thermoelectric figure of merit
temperature-dependent properties
physics-guided machine learning
multi-task learning
materials informatics

ABSTRACT

Temperature-dependent thermoelectric performance is central to materials screening because the magnitude and position of the $zT(T)$ maximum determine the useful operating window. We develop a zT -prioritized, physics-guided auxiliary multi-task framework that predicts thermoelectric property curves from composition and temperature. The framework combines Magpie descriptors, Fourier temperature encoding, joint prediction of the Seebeck coefficient S , electrical conductivity σ , thermal conductivity κ , power factor PF , and figure of merit zT , high- zT reweighting, staged optimization, dynamic physical regularization, and curve smoothing. The curated dataset contains 4222 observations from 755 formula-defined materials and 270 elemental systems over 275–1275 K. The ten-seed ensemble gives training-set $R^2 = 0.9563$ for PF and $R^2 = 0.9405$ for zT ; material-grouped out-of-fold evaluation gives $R^2 = 0.8927$ for PF and $R^2 = 0.8765$ for zT , with zT MAE = 0.0764 and RMSE = 0.1166. The model reproduces temperature-dependent responses across the measured ranges of representative materials. External validation on four literature-derived materials and 24 points gives $R^2 = 0.9755$, MAE = 0.0128, and RMSE = 0.0198. The resulting composition-based workflow supports candidate ranking and operating-window selection through coupled property prediction and soft physical supervision.

1. Introduction

Thermoelectric materials directly and reversibly convert thermal and electrical energy, supporting waste-heat recovery, solid-state refrigeration, and thermal management [1–3]. Their performance is quantified by the dimensionless figure of merit $zT = S^2\sigma T/\kappa$, where S is the Seebeck coefficient, σ is the electrical conductivity, κ is the total thermal conductivity, T is the absolute temperature, and $PF = S^2\sigma$ is the power factor. Electrical and thermal transport are strongly coupled [4, 5]: changes that enhance carrier transport can alter S , the electronic contribution to κ , and the optimal temperature range. Their coupled temperature dependence sets the magnitude, position, and width of the $zT(T)$ maximum and the practical operating window [6–8]. Temperature-resolved prediction is therefore required for operating-window selection and composition screening.

Comprehensive experimental characterization of thermoelectric transport curves is laborious because S , σ , and κ must be measured over dense temperature intervals, often with different instruments or sample conditions [9, 10]. First-principles calculations and Boltzmann-transport

simulations provide microscopic insight, while their computational cost and structural requirements limit broad screening across realistic alloying, defects, microstructures, and carrier concentrations [11, 12]. High-throughput calculations combined with machine learning or Bayesian optimization narrow the search for low-thermal-conductivity candidates [13, 14]. Neural-network interatomic potentials extend atomistic thermal-transport calculations and require structure-resolved training data [15]. Literature-derived datasets complement these approaches. Reported curves use nonuniform temperature grids and differ in units, interpolation procedures, digitization quality, synthesis routes, and measurement conditions. Interlaboratory round-robin measurements also quantify uncertainty in the measured transport quantities and the derived figure of merit [16]. We therefore retain source provenance and apply explicit consistency checks when combining literature records. Recent databases have expanded access to thermoelectric data [17, 18], and the sysTEM dataset uses physical relationships among transport properties to curate and validate literature-derived records [19]. These developments motivate the curation checks and physical constraints used in the present study.

*Corresponding author.

✉ m150724231548@163.com (J. He); 15617242080@163.com (X. Lian); liukaiyue0524@163.com (K. Liu); 13017608761@163.com (X. Ge)

ORCID(s):

Machine learning provides a route to nonlinear composition–property relationships and accelerated thermoelectric discovery. Artificial neural networks predict thermoelectric performance from simulated module data [20], while composition-derived descriptors support prediction and screening of experimentally measured properties [21, 22]. Composition-only deep learning learns chemical representations directly from elemental fractions [23], and dopant-aware representations improve the treatment of alloyed systems [24]. Public experimental datasets enable joint analysis of S , σ , κ , PF , and zT over composition and temperature. Attention-based models predict multiple thermoelectric transport quantities across temperature and doping conditions from chemical composition [25]; verified experimental data and multi-task architectures further introduce temperature as an explicit model input [19, 26]. These results establish composition-based learning as a practical surrogate when detailed crystal structures are unavailable.

Temperature-dependent zT prediction requires material-level evaluation because pointwise metrics do not measure peak location or complete curve shape. Random pointwise splitting mixes temperatures from the same material across training and test sets and evaluates interpolation along partially observed curves. We therefore keep all temperatures from a material in the same split and evaluate predictions for materials excluded from training [27]. Peak-location and curve-shape metrics quantify the screening quantities directly.

The five targets have different scales, noise levels, coverage, and relevance to screening. Multi-task learning uses shared information among related targets [28], and task weights distribute optimization across them [29]. In thermoelectrics, PF and zT are derived performance quantities linked to the transport coefficients. The present weighting emphasizes the high- zT region, peak magnitude, and curve shape used for screening, while accounting for the lower frequency of high- zT observations through imbalanced continuous regression [30].

The thermoelectric relations provide physical supervision during optimization. Reconstructing zT from independently predicted S , σ , and κ propagates errors from all three quantities, with the contribution from S amplified by the square term and that from κ entering through division. We introduce the consistency penalty progressively, allowing the auxiliary heads to stabilize before coupling them more strongly to the dedicated zT predictor. Uniformly recalculated PF and zT labels provide a consistent reference for this physical branch [31].

This study develops a zT -prioritized, physics-guided auxiliary multi-task framework that combines composition-derived descriptors with Fourier temperature encoding.

The properties S , σ , κ , and PF provide auxiliary supervision, while the dedicated zT head remains the principal output. High- zT reweighting, two-stage optimization, and progressive physical-consistency regularization allocate learning capacity to the screening target while preserving the thermoelectric relations. We evaluate the framework with material-disjoint testing, curve-level analysis, grouped cross-validation, and independent external validation. The resulting model ranks candidate operating windows and compositions from composition and temperature; structure, process, and uncertainty information provide additional material-specific context. The workflow is summarized in the graphical abstract and follows recent Materials Today Physics studies that connect data curation with reliable thermoelectric machine learning [32, 33].

2. Methodology

2.1. Dataset curation and representation

After quality control, the modeling workbook contained 4222 temperature-dependent observations from 755 formula-defined materials and 270 elemental systems over 275–1275 K. The source inventory in Supplementary Table S1 contains 4433 DOI-attributed observations from 187 DOI sources; an observation supported by more than one article is listed under each DOI, whereas the modeling workbook retains one unique record for each material–temperature coordinate. The feature workbook contained 142 columns: identifiers, temperature and transport quantities, and 132 Magpie composition descriptors. Individual materials contained 1–15 temperature points (median: six). Sparse records were retained for pointwise learning, whereas curve-level analyses used materials with at least four distinct temperatures. Each record included a material identifier, formula, elemental-system label, temperature, measured transport properties, the complete descriptor vector, and source URL. This organization preserved both point-level supervision and the material-level grouping required for curve prediction.

Reported PF and zT values were checked against values recalculated from S , σ , κ , and T . Deviations up to 1% for PF and 5% for zT were accepted; larger discrepancies were verified against the source article, confirmed errors were corrected, and unresolved records were excluded. The modeling table used uniformly recalculated PF and zT . Full curation and consistency-check details are provided in Supplementary Method S1.

Material identity was defined by formula-normalized composition rather than publication or specimen provenance. Each `material_id` mapped to one formula and vice versa. The composition-only model encoded formula-derived information, while publication, preparation route, measurement direction, carrier type, and specimen form

remained available in the literature archive for provenance tracking. Exact duplicate rows were removed, leaving one observation per material–temperature coordinate. No record-level imputation was required.

Chemical formulas were represented by 132 Magpie composition descriptors [34, 35]. All preprocessing was fitted exclusively on the training set. Near-constant descriptors were removed, one feature from each highly correlated pair with $|r| > 0.95$ was discarded, and the remaining features were standardized. Principal component analysis (PCA) then reduced redundancy [36]. As shown in Fig. 1, the feature dimensionality decreased from 132 to 125 after variance filtering, to 97 after correlation filtering, and finally to 96 principal components.

Because many retained features were derived from the same elemental property using different statistical aggregations, Table 1 summarizes the descriptor families represented by the 97 inputs before PCA.

Table 1: Magpie descriptor families retained after correlation filtering (97 inputs before PCA). Statistical operators are omitted for brevity.

Feature	Meaning
Number	Atomic number; reflects elemental identity and nuclear charge.
MendelevNumber	Mendelev chemical-order number; encodes periodic chemical similarity among elements.
MeltingT	Elemental melting temperature; relates to bond strength and thermal stability.
Column	Periodic-table group or column; reflects valence-shell family.
Row	Periodic-table period or row; reflects principal shell and size-related periodic trends.
CovalentRadius	Covalent atomic radius; describes the bonded atomic size scale.
Electronegativity	Elemental electronegativity; measures the tendency to attract electrons in bonding.
NsValence	Number of valence s electrons.
NpValence	Number of valence p electrons.
NdValence	Number of valence d electrons.
NfValence	Number of valence f electrons.
NValence	Total number of valence electrons.
NsUnfilled	Number of unfilled valence s states.
NpUnfilled	Number of unfilled valence p states.
NdUnfilled	Number of unfilled valence d states.
NfUnfilled	Number of unfilled valence f states.
NUnfilled	Total number of unfilled valence states.
GSvolume_pa	Ground-state volume per atom; reflects elemental atomic volume and packing scale.
GSbandgap	Ground-state band gap; reflects elemental electronic gap character.
GSmagmom	Ground-state magnetic moment; reflects magnetic tendency and unpaired-spin character.
SpaceGroupNumber	Element-derived reference space-group statistic used as a compositional descriptor, distinct from compound-level crystal structure.

The positive-valued targets σ , κ , PF , and zT were logarithmically transformed, while the signed Seebeck coefficient was left untransformed. All five targets were standardized using training-set statistics. Exact transformations are given in Supplementary Method S1.

The dataset was grouped by `material_id` before model development so that all temperatures for a given material remained together. Approximately 85% of the observations formed the training set and 15% formed the

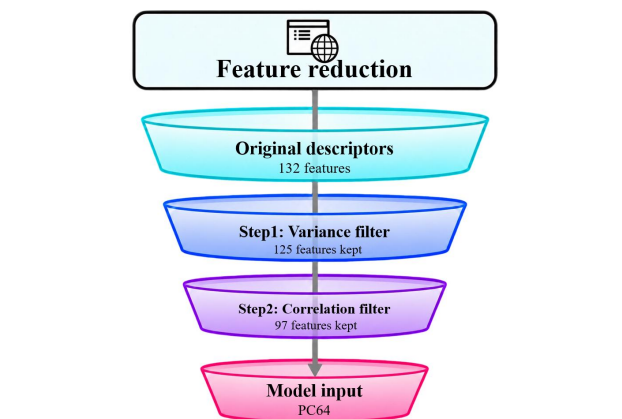


Fig. 1. Feature dimensionality reduction from 132 Magpie descriptors to the 96-dimensional PCA model input.

validation set. Elemental systems could recur across these subsets, while the material grouping prevented temperature records from the same formula from being distributed between training and validation. Primary generalization was assessed independently with ten-fold material-grouped out-of-fold evaluation.

2.2. zT -prioritized temperature-conditioned network

Each input combined a 96-dimensional PCA composition vector with the corresponding temperature. The composition vector was processed by two 256-unit hidden blocks followed by a linear projection to a 128-dimensional latent representation, using layer normalization, SiLU activation, and dropout ($p = 0.10$) [28, 29].

Temperature was normalized to the training range and expanded with four Fourier frequencies [37]:

$$\tilde{T} = \frac{T - T_{\min}}{T_{\max} - T_{\min}}, \quad (1)$$

$$\phi(T) = \left[\tilde{T}, \left\{ \sin(2^k \pi \tilde{T}), \cos(2^k \pi \tilde{T}) \right\}_{k=0}^3 \right]. \quad (2)$$

This produced a 9-dimensional temperature representation capable of expressing nonlinear temperature dependence.

The composition and temperature representations were concatenated and passed through a shared trunk with two 256-unit hidden blocks. Five task-specific heads predicted S , σ , κ , PF , and zT . The dedicated zT head was the principal task, while the remaining properties provided auxiliary supervision. Full layer definitions are given in Supplementary Method S2.

After inverse transformation, the auxiliary predictions were also used to reconstruct the derived properties through the standard thermoelectric relations:

$$\widehat{PF}_{\text{phys}} = \widehat{S}^2 \widehat{\sigma}, \quad \widehat{zT}_{\text{phys}} = \frac{\widehat{PF}_{\text{phys}} T}{\widehat{\kappa}}. \quad (3)$$

The reported prediction, $\widehat{zT}$, came directly from the dedicated task head, while the reconstructed $\widehat{zT}_{\text{phys}}$ provided a soft physical reference during training and a consistency metric.

2.3. zT -prioritized and physics-guided training

All five targets were optimized with Smooth L1 loss [38]. Their relative weights changed between two stages to concentrate the later optimization on zT . For the target order $[S, \sigma, \kappa, PF, zT]$,

$$\mathbf{W}_{\text{task}} = \begin{bmatrix} 0.30 & 0.35 & 0.35 & 0.60 & 6.00 \\ 0.15 & 0.20 & 0.20 & 0.40 & 8.00 \end{bmatrix}, \quad (4)$$

where the rows correspond to stages 1 and 2. Stage 1 promoted shared multi-property learning, whereas stage 2 reduced auxiliary contributions and concentrated optimization on zT . All parameters remained trainable.

High- zT observations were emphasized with a smooth loss weight,

$$w_i = 1 + 2 \cdot \text{sigmoid} \left(\frac{\ln zT_i - \ln zT_{0.70}}{0.35} \right), \quad (5)$$

where $zT_{0.70}$ is the training-set 70th percentile. The weight increases from approximately 1 in the low- zT region to 3 in the high-performance region, following the motivation of imbalanced continuous regression [30].

Task-head and reconstructed PF and zT outputs were coupled by Smooth L1 consistency penalties in logarithmic space. The penalties were introduced with a 30-epoch cosine ramp and weakened in stage 2, allowing the auxiliary heads to stabilize before exerting stronger influence on the principal task. Full definitions and schedules are given in Supplementary Method S3.

A weak curve regularizer was applied to $\widehat{zT}$ on a 300–800 K grid with 25-K spacing by penalizing large second differences,

$$\Delta^2 \widehat{zT}_{m,j} = \widehat{zT}_{m,j+1} - 2\widehat{zT}_{m,j} + \widehat{zT}_{m,j-1}, \quad (6)$$

where m denotes a material and j a temperature-grid position. The term suppresses unsupported local oscillations while allowing flexible curve shapes.

The total objective was

$$\mathcal{L} = \mathcal{L}_{\text{data}} + \lambda_{PF} \mathcal{L}_{PF}^{\text{phys}} + \lambda_{zT} \mathcal{L}_{zT}^{\text{phys}} + \lambda_{\text{sm}} \mathcal{L}_{\text{sm}}, \quad (7)$$

where λ_{PF} and λ_{zT} vary with epoch and training stage. The smoothness coefficient is $\lambda_{\text{sm}} = 0.005$, the curve-pair coefficient is $\lambda_{\text{curve}} = 0.10$, and the physical-regularization schedule is warmed up dynamically over the first 30 epochs. Additional details are provided in Supplementary Methods S3 and S4.

2.4. Training and evaluation

The model was implemented in PyTorch and optimized with AdamW [39]. Stage 1 used up to 150 epochs with an initial learning rate of 1.5×10^{-3} ; stage 2 used up to 60 epochs at 40% of the stage-1 learning rate. The batch size was 256, weight decay was 10^{-4} , and the gradient norm was clipped at 5.0. Early stopping used a patience of 30 epochs, and ReduceLROnPlateau reduced the learning rate by a factor of 0.6 when validation performance plateaued. The same implementation was repeated for ten seeds (42–51), providing an ensemble that averages the learned temperature-conditioned composition response.

The final implementation used ten deterministic runs (seed_001–seed_010) with seeds 42–51. Checkpoint selection and ensemble construction were specified before the out-of-fold evaluation. Positive-valued outputs were aggregated in logarithmic space, and the direct task-head zT output was reported. Full training and ensemble details are given in Supplementary Method S4.

Evaluation was performed at both the observation level and the material-curve level. Pointwise metrics quantify

the error at each measured temperature, whereas curve-level metrics give each eligible material equal weight and therefore prevent densely sampled materials from dominating the summary. The curve analysis uses the measured temperature coordinates, reports the mean curve RMSE, and separately tracks the error in the maximum zT and the temperature at which that maximum occurs. These measures connect numerical accuracy with the screening quantities used to rank compositions and operating windows.

The ten-fold grouped procedure provides an additional material-level estimate by generating out-of-fold predictions for every observation while keeping all temperatures from one material in the same fold. Ensemble predictions are aggregated in logarithmic space for positive-valued properties, and the direct zT head is used for the principal performance metrics. The physics-reconstructed branch is retained as a consistency diagnostic, so pointwise accuracy, curve shape, peak magnitude, and peak location can be interpreted together.

Pointwise performance was evaluated using MAE, RMSE, and R^2 :

$$\text{MAE} = \frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}_i|, \quad (8)$$

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2}, \quad (9)$$

$$R^2 = 1 - \frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{\sum_{i=1}^N (y_i - \bar{y})^2}. \quad (10)$$

All valid material-temperature observations contributed to pointwise evaluation. Curve-level evaluation used materials with at least four distinct experimental temperatures and was restricted to measured temperature coordinates. Mean curve RMSE gives each eligible material equal weight; peak- zT MAE measures maximum-performance error, and peak-temperature MAE measures operating-temperature error. Computational definitions are provided in Supplementary Method S4.

Ten-fold grouped cross-validation used material identifiers as groups, so all temperature records belonging

to one material remained in the same outer fold. For each outer fold, the remaining materials were divided into fold-specific training and internal-validation subsets. Ten independently initialized models were trained per fold, and positive-valued predictions were aggregated in logarithmic space. Performance on the held-out fold was reported for the PF task-head output and the ensemble zT output; the pooled out-of-fold results were used as the generalization estimate. Physics-reconstructed zT values were retained as a consistency metric. The implementation uses 96 PCA components, four Fourier frequencies, dynamic physics scheduling, high- zT weighting, weak smoothing, and ten-seed ensembles.

3. Results and Discussion

3.1. Training-set fit and physical consistency

The final configuration was trained with ten deterministic seeds (42–51). The ensemble achieved $R^2 = 0.9563$, $\text{MAE} = 1.34 \times 10^{-4}$, and $\text{RMSE} = 2.19 \times 10^{-4}$ for PF , together with $R^2 = 0.9405$, $\text{MAE} = 0.0509$, and $\text{RMSE} = 0.0826$ for zT . The direct and physics-reconstructed outputs remained closely aligned, with logarithmic direct-reconstructed MAE values of 0.0399 for PF and 0.0441 for zT . The training-set parity plots are shown in Fig. 2 and Fig. 3; the corresponding direct-reconstructed comparisons are shown in Fig. 4 and Fig. 5.

The ten-seed average stabilizes the learned response across the training domain. High- zT weighting and the two-stage schedule focus capacity on the screening region, while the auxiliary heads retain shared transport information. The resulting representation fits the pointwise and derived quantities together and preserves the relative ordering of materials across the training domain.

The direct-reconstructed comparisons provide a physical consistency check. Agreement with the dedicated PF and zT heads shows that the predicted S , σ , and κ values carry compatible transport information. For each composition and temperature, the model provides a screening-oriented performance estimate together with linked transport quantities for curve analysis and candidate ranking.

The same prediction supplies transport-property estimates, derived performance, and the temperature coordinate of the screening optimum, connecting the pointwise scores with the material-level quantities used below.

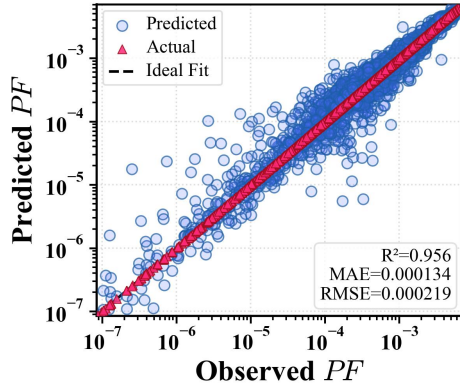


Fig. 2. Training-set prediction parity for the power factor PF from the ten-seed ensemble.

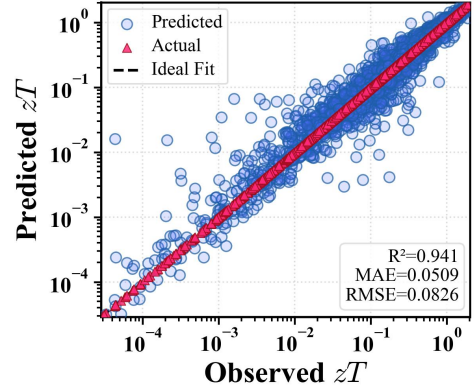


Fig. 3. Training-set prediction parity for the thermoelectric figure of merit zT from the ten-seed ensemble.

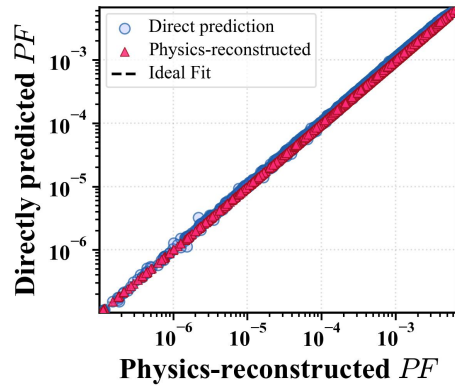


Fig. 4. Physics consistency between direct and reconstructed training-set power factor PF predictions.

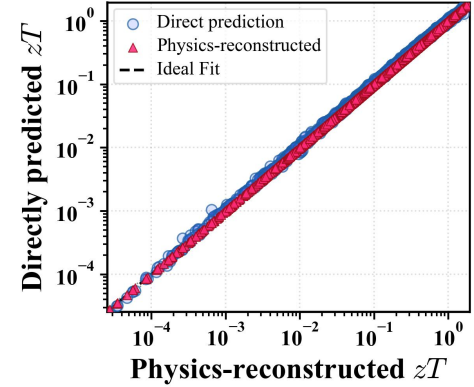


Fig. 5. Physics consistency between direct and reconstructed training-set thermoelectric figure of merit zT predictions.

3.2. Feature attribution in the reduced descriptor space

SHAP-based attribution analysis [40] examines the 96 PCA components for the dedicated $\log PF$ and $\log zT$ heads. Fig. 6 presents the SHAP distributions for PC1–PC64, with PC1–PC32 and PC33–PC64 shown in separate halves of each panel. The two targets exhibit different attribution magnitudes, signs, and feature-value trends across the reduced descriptor space. Mapped original-descriptor rankings and the complete PC1–PC96 loading matrices are provided in the Supplementary Information. Because each principal component combines multiple descriptors, the rankings describe attribution patterns in the reduced representation rather than isolated elemental descriptors.

The attribution maps organize the learned composition response in the reduced descriptor space. The SHAP distributions identify descriptor-space directions

that contribute to the PF and zT outputs. Comparing the two heads separates shared transport information from target-specific information: shared patterns follow the derived relationship between PF and zT , whereas target-specific patterns identify composition directions used by the dedicated performance heads. These maps prioritize composition regions for subsequent screening and follow-up calculations.

The attribution structure supports a direct screening workflow. Components with consistently large absolute SHAP values define the composition directions that most strongly shift the predicted response, while the sign of each contribution distinguishes favorable and unfavorable movement along that direction. Reading the two heads together links the composition signal to both power-factor and figure-of-merit performance and provides a compact basis for ranking candidates before temperature-curve analysis.

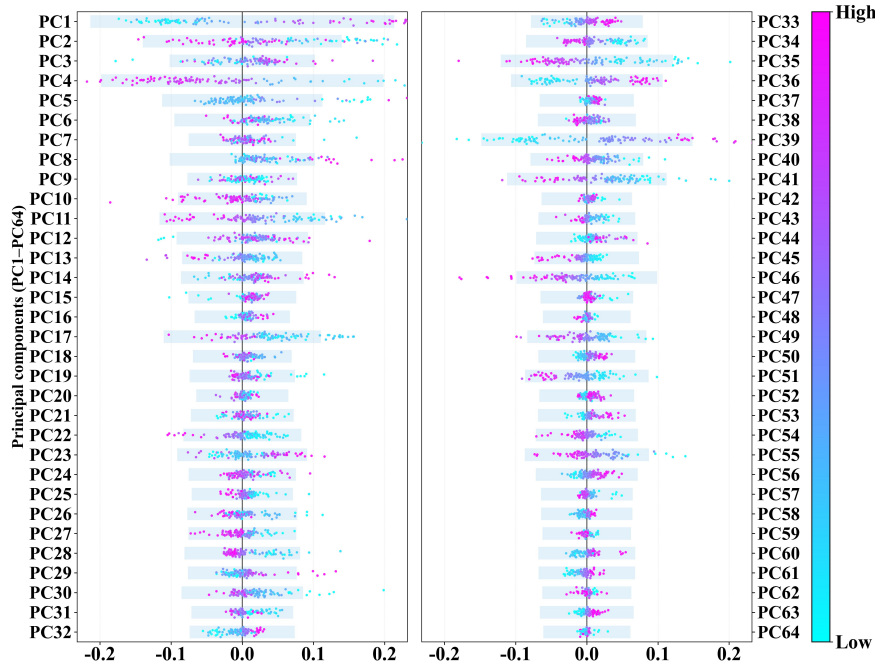
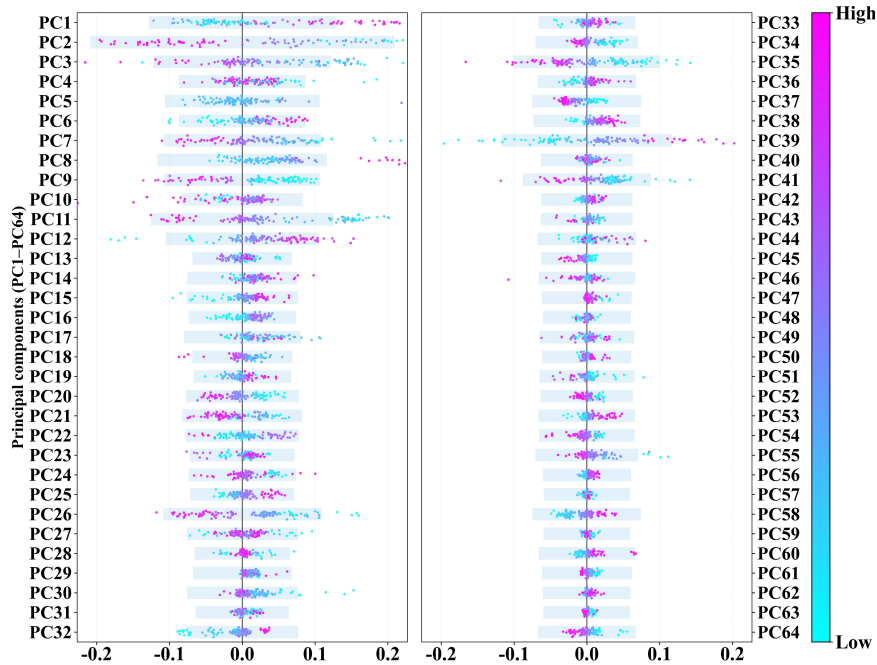
(a) Dedicated log zT head; PC1–PC32 and PC33–PC64 are shown from left to right.

 (b) Dedicated log PF head; PC1–PC32 and PC33–PC64 are shown from left to right.


Fig. 6. SHAP beeswarm summaries for the first 64 principal components. Each point represents one observation; horizontal position gives the SHAP value and color encodes the component value from low (cyan) to high (magenta).

3.3. Ten-fold material-grouped cross-validation

Ten-fold grouped cross-validation kept all temperature records for each material in one outer fold. In each fold, the remaining materials formed training and internal-validation subsets; ten independently initialized models were trained, and positive predictions were aggregated in logarithmic space.

The pooled out-of-fold predictions achieved $R^2 = 0.8927$, $\text{MAE} = 2.00 \times 10^{-4}$, and $\text{RMSE} = 3.54 \times 10^{-4}$ for PF , and $R^2 = 0.8765$, $\text{MAE} = 0.0764$, and $\text{RMSE} = 0.1166$ for zT . Across folds, R^2 ranged from 0.8499–0.9291 for zT and 0.8537–0.9506 for PF ; peak- zT MAE was 0.1141 and peak-temperature MAE was 14.07 K (Fig. 8).

These out-of-fold results quantify transfer across the curated 275–1275 K domain and support operating-window

selection through pointwise accuracy, peak magnitude, and peak temperature.

The paired bars in Fig. 8 show the fold-to-fold behavior directly: PF R^2 values span 0.854–0.951, while zT values span 0.850–0.929. The strongest PF values occur in folds 6 and 9, and the highest zT value occurs in fold 1. The grouped results therefore maintain predictive agreement across distinct material groups.

For each held-out material, the ensemble produces a complete temperature-conditioned response together with auxiliary transport estimates. The grouped evaluation connects pointwise accuracy with material-level ranking, peak magnitude, and peak temperature, providing the quantitative basis for the curve comparisons below.

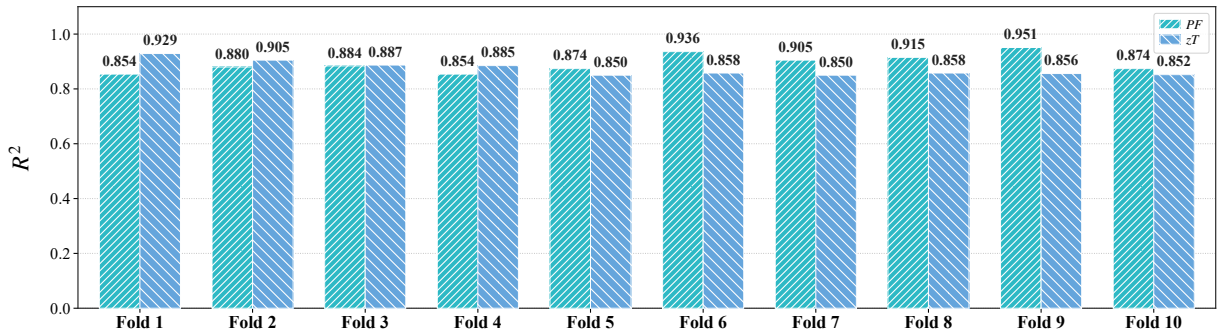


Fig. 8. Held-out R^2 values across ten material-grouped cross-validation folds. PF and zT denote the task-head outputs from ten-seed fold ensembles.

3.4. Temperature-dependent zT curve prediction

The model reproduced the temperature dependence of zT for unseen materials across their measured ranges. Fig. 7 shows eight representative materials, and all 22 held-out material curves from the test partition are provided in the Supplementary Information.

Panel citations Fig. 7(a)–Fig. 7(h) identify the materials. Predictions followed the measured direction of change in all eight panels, with trend MAE values of 0.027–0.136 per 100 K. The curve-level comparison captures trend agreement, peak location, and peak magnitude.

3.5. Literature-derived external validation

After model selection, the final ensemble was evaluated on four literature-linked curve sets: BaSnS_3 (300–800 K), $\text{Na}_2\text{InCuF}_6$ (300–800 K), LaOBiS_2 (300–550 K), and

$\text{MgI}_2\text{Bi}_2\text{S}_3$ (300–550 K) [41–44]. The source records and complete DOI links are listed in the reference section.

The ensemble achieved an overall $R^2 = 0.9755$, $\text{MAE} = 0.0128$, $\text{RMSE} = 0.0198$, and mean absolute relative error of 5.61%. The material-level R^2 values were 0.9787 for BaSnS_3 , 0.9347 for $\text{Na}_2\text{InCuF}_6$, 0.9878 for LaOBiS_2 , and 0.9737 for $\text{MgI}_2\text{Bi}_2\text{S}_3$. Panels Fig. 9(a)–Fig. 9(d) show the four material-specific comparisons, with predicted curves following the observed temperature trends.

Across the four external curves, the ensemble captures both steady increases and changes in curvature over the measured temperature windows. Agreement in magnitude and trend connects the external metrics with the material-level response used for ranking and operating-temperature selection.

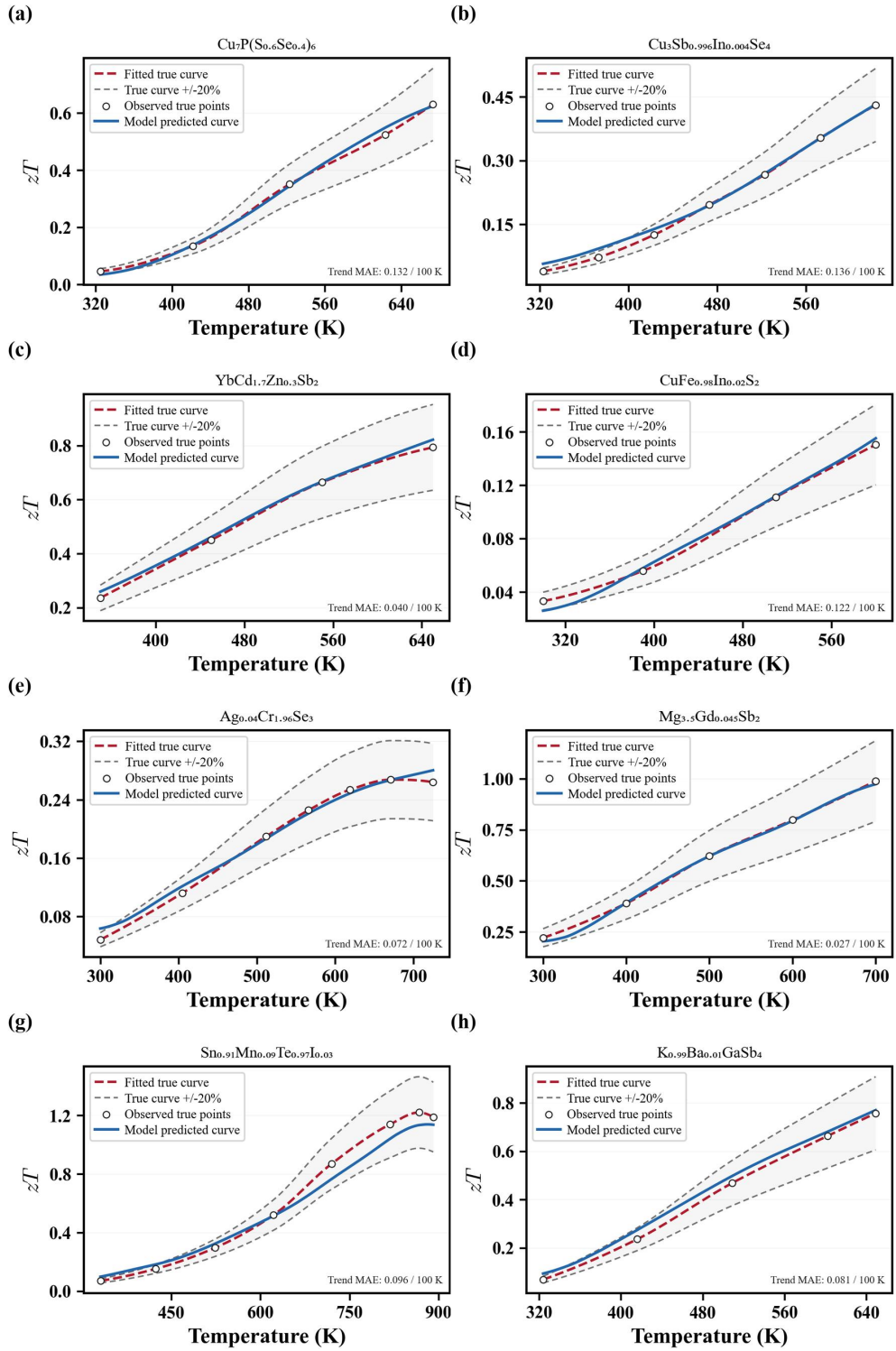


Fig. 7. Representative temperature-dependent zT curves from the material-grouped evaluation. The panels correspond to (a) $\text{Cu}_7\text{P}(\text{S}_{0.6}\text{Se}_{0.4})_6$, (b) $\text{Cu}_3\text{Sb}_{0.996}\text{In}_{0.004}\text{Se}_4$, (c) $\text{YbCd}_{1.7}\text{Zn}_{0.3}\text{Sb}_2$, (d) $\text{CuFe}_{0.98}\text{In}_{0.02}\text{S}_2$, (e) $\text{Ag}_{0.04}\text{Cr}_{1.96}\text{Se}_3$, (f) $\text{Mg}_{3.5}\text{Gd}_{0.045}\text{Sb}_2$, (g) $\text{Sn}_{0.91}\text{Mn}_{0.09}\text{Te}_{0.97}\text{I}_{0.03}$, and (h) $\text{K}_{0.99}\text{Ba}_{0.01}\text{GaSb}_4$. Dashed curves and markers denote the literature-derived observations; solid curves show the model predictions.

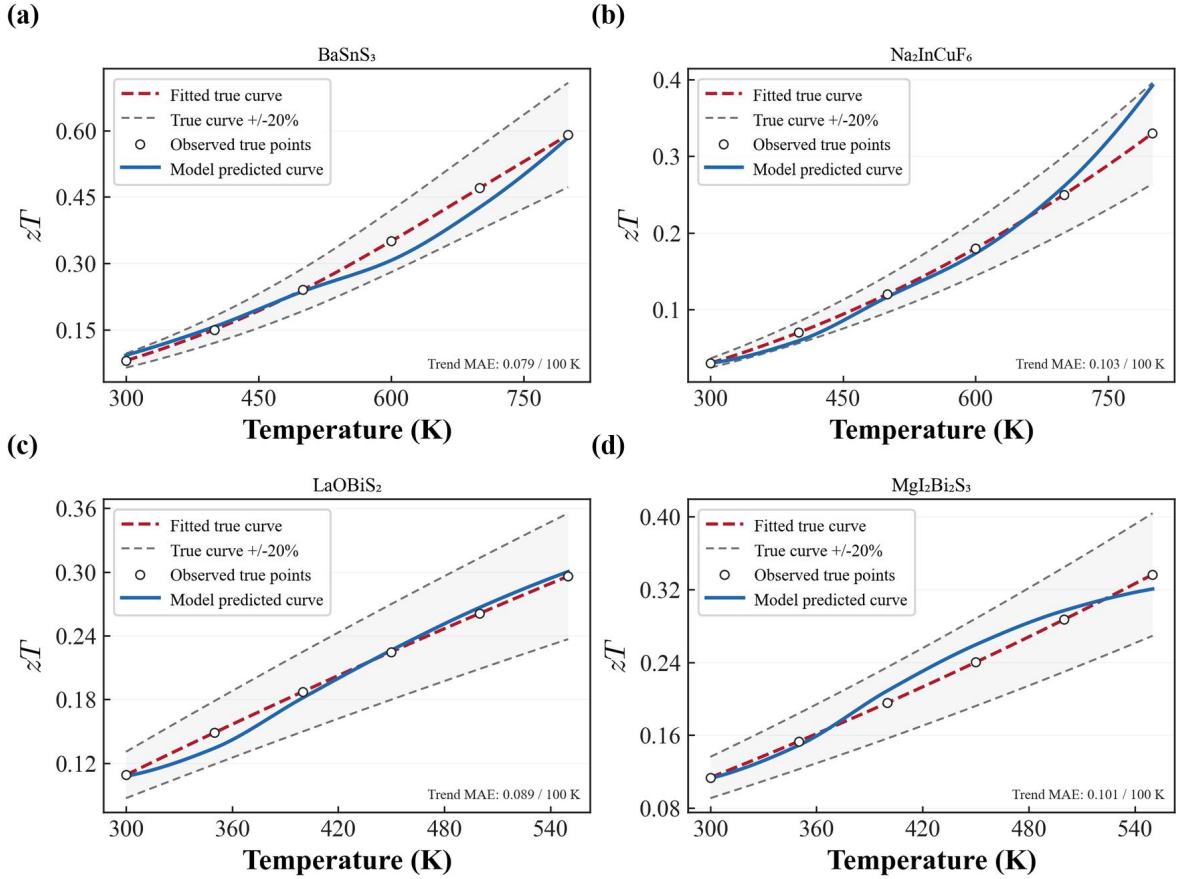


Fig. 9. External validation of ensemble zT predictions for (a) BaSnS_3 , (b) $\text{Na}_2\text{InCuF}_6$, (c) LaOBiS_2 , and (d) $\text{MgI}_2\text{Bi}_2\text{S}_3$. Markers denote the 24 temperature– zT points in the external workbook, dashed lines show fitted curves, and solid lines show model predictions. The source articles are cited in the text and reference list.

4. Conclusion

This study establishes a zT -prioritized, physics-guided auxiliary multi-task framework for predicting temperature-dependent thermoelectric performance from composition and temperature. The framework combines high- zT weighting, two-stage refinement, progressive physical consistency, weak curve regularization, and seed ensembling to describe the coupled transport response.

The final ten-seed ensemble achieved a training-set $R^2 = 0.9405$ for zT and $R^2 = 0.9563$ for PF . Ten-fold material-grouped out-of-fold evaluation yielded $R^2 = 0.8765$, $\text{MAE} = 0.0764$, and $\text{RMSE} = 0.1166$ for zT , together with $R^2 = 0.8927$ for PF . The model reproduced the temperature-dependent $zT(T)$ responses of representative held-out materials. External validation on four literature-derived materials and 24 points yielded $R^2 = 0.9755$, $\text{MAE} = 0.0128$, and $\text{RMSE} = 0.0198$. The composition-based representation supports candidate

ranking and operating-window selection, while structure, carrier concentration, microstructure, synthesis history, and measurement conditions provide additional information for material-specific refinement.

CRedit authorship contribution statement

Conceptualization: X. Lian, J. He. Methodology: X. Ge. Investigation: K. Liu. Writing – original draft: J. He. Writing – review & editing: all authors. Supervision: all authors.

Data availability

The 4222-row modeling workbook and the 24-row external-validation workbook are available from the corresponding author upon reasonable request. The DOI source inventory and DOI-attributed observation counts are provided in Supplementary Table S1; the corrected SI

figures use the 22 held-out curves and the diagnostic files supplied with this manuscript.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

- [1] J. He, T. M. Tritt, Advances in thermoelectric materials research: Looking back and moving forward, *Science* 357 (6358) (2017) eaak9997. <https://doi.org/10.1126/science.aak9997>.
- [2] L. E. Bell, Cooling, heating, generating power, and recovering waste heat with thermoelectric systems, *Science* 321 (5895) (2008) 1457–1461. <https://doi.org/10.1126/science.1158899>.
- [3] D. Beretta, N. Neophytou, J. M. Hodges, M. G. Kanatzidis, D. Narducci, M. Martin-Gonzalez, M. Beekman, B. Balke, G. Cerretti, W. Tremel, A. Zevalkink, A. I. Hofmann, C. Müller, B. Döring, M. Campoy-Quiles, M. Caironi, Thermoelectrics: From history, a window to the future, *Mater. Sci. Eng. R Rep.* 138 (2019) 100501. <https://doi.org/10.1016/j.mser.2018.09.001>.
- [4] G. S. Hegde, A. N. Prabhu, A review on doped/composite bismuth chalcogenide compounds for thermoelectric device applications: Various synthesis techniques and challenges, *J. Electron. Mater.* 51 (2022) 2014–2042. <https://doi.org/10.1007/s11664-022-09513-x>.
- [5] G. J. Snyder, E. S. Toberer, Complex thermoelectric materials, *Nat. Mater.* 7 (2008) 105–114. <https://doi.org/10.1038/nmat2090>.
- [6] S. Puthran, G. S. Hegde, A. N. Prabhu, Review of chalcogenide-based materials for low-, mid-, and high-temperature thermoelectric applications, *J. Electron. Mater.* 53 (2024) 5739–5768. <https://doi.org/10.1007/s11664-024-11310-7>.
- [7] A. Y. Samunin, V. K. Zaitsev, P. P. Konstantinov, M. I. Fedorov, G. N. Isachenko, A. T. Burkov, S. V. Novikov, E. A. Gurieva, Thermoelectric properties of hot-pressed materials based on $\text{Mg}_2\text{Si}_n\text{Sn}_{1-n}$, *J. Electron. Mater.* 42 (2013) 1676–1679. <https://doi.org/10.1007/s11664-012-2372-3>.
- [8] L.-D. Zhao, S.-H. Lo, Y. Zhang, H. Sun, G. Tan, C. Uher, C. Wolverton, V. P. Dravid, M. G. Kanatzidis, Ultralow thermal conductivity and high thermoelectric figure of merit in SnSe crystals, *Nature* 508 (2014) 373–377. <https://doi.org/10.1038/nature13184>.
- [9] B. Jariwala, D. Shah, N. M. Ravindra, Transport property measurements in doped Bi_2Te_3 single crystals obtained via zone melting method, *J. Electron. Mater.* 44 (2015) 1509–1516. <https://doi.org/10.1007/s11664-014-3438-1>.
- [10] S. J. Pandey, G. Joshi, S. Wang, S. Curtarolo, R. M. Gaume, Modeling the thermoelectric properties of Ti_5O_9 magneli phase ceramics, *J. Electron. Mater.* 45 (2016) 5526–5532. <https://doi.org/10.1007/s11664-016-4762-4>.
- [11] S. Curtarolo, G. L. W. Hart, M. Buongiorno Nardelli, N. Mingo, S. Sanvito, O. Levy, The high-throughput highway to computational materials design, *Nat. Mater.* 12 (2013) 191–201. <https://doi.org/10.1038/nmat3568>.
- [12] G. K. H. Madsen, J. Carrete, M. J. Verstraete, BoltzTraP2, a program for interpolating band structures and calculating semi-classical transport coefficients, *Comput. Phys. Commun.* 231 (2018) 140–145. <https://doi.org/10.1016/j.cpc.2018.05.010>.
- [13] J. Carrete, W. Li, N. Mingo, S. Wang, S. Curtarolo, Finding unprecedentedly low-thermal-conductivity half-Heusler semiconductors via high-throughput materials modeling, *Phys. Rev. X* 4 (2014) 011019. <https://doi.org/10.1103/PhysRevX.4.011019>.
- [14] A. Seko, A. Togo, H. Hayashi, K. Tsuda, L. Chaput, I. Tanaka, Prediction of low-thermal-conductivity compounds with first-principles anharmonic lattice-dynamics calculations and bayesian optimization, *Phys. Rev. Lett.* 115 (2015) 205901. <https://doi.org/10.1103/PhysRevLett.115.205901>.
- [15] R. Qu, Y. Lv, Z. Lu, A deep neural network potential to study the thermal conductivity of mnbi_2te_4 and $\text{bi}_2\text{te}_3/\text{mnbi}_2\text{te}_4$ superlattice, *J. Electron. Mater.* 52 (2023) 4475–4483. <https://doi.org/10.1007/s11664-023-10403-z>.
- [16] E. Alleno, D. Bérardan, C. Byl, C. Candolfi, R. Daou, R. Decourt, E. Guilmeau, S. Hébert, J. Hejtmanek, B. Lenoir, P. Masschelein, V. Ohorodnichuk, M. Pollet, S. Populoh, D. Ravot, O. Rouleau, M. Soulier, Invited article: A round robin test of the uncertainty on the measurement of the thermoelectric dimensionless figure of merit of $\text{Co}_{0.97}\text{Ni}_{0.03}\text{Sb}_3$, *Rev. Sci. Instrum.* 86 (1) (2015) 011301. <https://doi.org/10.1063/1.4905250>.
- [17] M. W. Gaultois, T. D. Sparks, C. K. H. Borg, R. Seshadri, W. D. Bonificio, D. R. Clarke, Data-driven review of thermoelectric materials: Performance and resource considerations, *Chem. Mater.* 25 (15) (2013) 2911–2920. <https://doi.org/10.1021/cm400893e>.
- [18] O. Sieriepeklis, J. M. Cole, A thermoelectric materials database auto-generated from the scientific literature using chemdataextractor, *Sci. Data* 9 (2022) 648. <https://doi.org/10.1038/s41597-022-01752-1>.
- [19] L. Z. Tang, L. Purdy, T. Mohanty, L. W. T. Ng, T. D. Sparks, Systematically verified experimental thermoelectric dataset for data-driven approaches, *Integr. Mater. Manuf. Innov.* 15 (1) (2026) 181–193. <https://doi.org/10.1007/s40192-026-00446-5>.
- [20] N. Derebasi, M. Eltez, F. Guldiken, A. Sever, K. Kallis, H. Kilic, Influence of geometrical factors on performance of thermoelectric material using numerical methods, *J. Electron. Mater.* 44 (2015) 2068–2073. <https://doi.org/10.1007/s11664-015-3657-0>.
- [21] Y. Iwasaki, I. Takeuchi, V. Stanev, A. G. Kusne, M. Ishida, A. Kirihara, K. Ihara, R. Sawada, K. Terashima, H. Someya, K.-I. Uchida, E. Saitoh, S. Yoroza, Machine-learning guided discovery of a new thermoelectric material, *Sci. Rep.* 9 (2019) 2751. <https://doi.org/10.1038/s41598-019-39278-z>.
- [22] J. Yan, P. Gorai, B. Ortiz, S. Miller, S. A. Barnett, T. Mason, V. Stevanović, E. S. Toberer, Material descriptors for predicting thermoelectric performance, *Energy Environ. Sci.* 8 (3) (2015) 983–994. <https://doi.org/10.1039/C4EE03157A>.
- [23] D. Jha, L. Ward, A. Paul, W.-k. Liao, A. Choudhary, C. Wolverton, A. Agrawal, ElemNet: Deep learning the chemistry of materials from only elemental composition, *Sci. Rep.* 8 (2018) 17593. <https://doi.org/10.1038/s41598-018-35934-y>.
- [24] G. S. Na, S. Jang, H. Chang, Predicting thermoelectric properties from chemical formula with explicitly identifying dopant effects, *npj Comput. Mater.* 7 (2021) 106. <https://doi.org/10.1038/s41524-021-00564-y>.
- [25] L. M. Antunes, K. T. Butler, R. Grau-Crespo, Predicting thermoelectric transport properties from composition with attention-based deep learning, *Mach. Learn.: Sci. Technol.* 4 (1) (2023) 015037. <https://doi.org/10.1088/2632-2153/acc4a9>.
- [26] L. Z. Tang, T. Mohanty, S. G. Baird, L. W. T. Ng, T. D. Sparks, Multi-task attention for doped thermoelectric properties prediction, *Mach. Learn.: Sci. Technol.* 7 (3) (2026) 035022. <https://doi.org/10.1088/2632-2153/ae6416>.
- [27] K. Li, B. DeCost, K. Choudhary, M. Greenwood, J. Hattrick-Simpers, A critical examination of robustness and generalizability of machine learning prediction of materials properties, *npj Comput. Mater.* 9 (2023) 55. <https://doi.org/10.1038/s41524-023-01012-9>.
- [28] R. Caruana, Multitask learning, *Mach. Learn.* 28 (1997) 41–75. <https://doi.org/10.1023/A:1007379606734>.

- [29] Y. Zhang, Q. Yang, A survey on multi-task learning, *IEEE Trans. Knowl. Data Eng.* 34 (12) (2022) 5586–5609. <https://doi.org/10.1109/TKDE.2021.3070203>.
- [30] P. Branco, L. Torgo, R. P. Ribeiro, A survey of predictive modeling on imbalanced domains, *ACM Comput. Surv.* 49 (2) (2016) 31. <https://doi.org/10.1145/2907070>.
- [31] G. E. Karniadakis, I. G. Kevrekidis, L. Lu, P. Perdikaris, S. Wang, L. Yang, Physics-informed machine learning, *Nat. Rev. Phys.* 3 (2021) 422–440. <https://doi.org/10.1038/s42254-021-00314-5>.
- [32] S. Athar, A. Mecibah, P. Jund, Tackling dataset curation challenges towards reliable machine learning: A case study on thermoelectric materials, *Materials Today Physics* 59 (2025) 101948. <https://doi.org/10.1016/j.mtphys.2025.101948>.
- [33] Z.-L. Wang, T. Funada, T. Onda, Z.-C. Chen, Knowledge extraction and performance improvement of Bi₂Te₃-based thermoelectric materials by machine learning, *Materials Today Physics* 31 (2023) 100971. <https://doi.org/10.1016/j.mtphys.2023.100971>.
- [34] L. Ward, A. Agrawal, A. Choudhary, C. Wolverton, A general-purpose machine learning framework for predicting properties of inorganic materials, *npj Comput. Mater.* 2 (2016) 16028. <https://doi.org/10.1038/npjcompumats.2016.28>.
- [35] L. Ward, A. Dunn, A. Faghaninia, N. E. R. Zimmermann, S. Bajaj, Q. Wang, J. Montoya, J. Chen, K. Bystrom, M. Dylla, K. Chard, M. Asta, K. A. Persson, G. J. Snyder, I. Foster, A. Jain, Matminer: An open source toolkit for materials data mining, *Comput. Mater. Sci.* 152 (2018) 60–69. <https://doi.org/10.1016/j.commatsci.2018.05.018>.
- [36] I. T. Jolliffe, J. Cadima, Principal component analysis: A review and recent developments, *Philos. Trans. R. Soc. A* 374 (2065) (2016) 20150202. <https://doi.org/10.1098/rsta.2015.0202>.
- [37] S. Wang, H. Wang, P. Perdikaris, On the eigenvector bias of fourier feature networks: From regression to solving multi-scale pdes with physics-informed neural networks, *Comput. Methods Appl. Mech. Eng.* 384 (2021) 113938. <https://doi.org/10.1016/j.cma.2021.113938>.
- [38] P. J. Huber, Robust estimation of a location parameter, *Ann. Math. Stat.* 35 (1) (1964) 73–101. <https://doi.org/10.1214/aoms/1177703732>.
- [39] R. Shaiakhmetov, D. Pianini, A. Filaseta, G. D’Angelo, V. Venusti, First-order optimization algorithms: State of the art, classification, and performance: A practitioner’s guide, *Neural Comput. Appl.* 38 (2026) 233. <https://doi.org/10.1007/s00521-026-12014-1>.
- [40] S. M. Lundberg, G. Erion, H. Chen, A. DeGrave, J. M. Prutkin, B. Nair, R. Katz, J. Himmelfarb, N. Bansal, S.-I. Lee, From local explanations to global understanding with explainable ai for trees, *Nat. Mach. Intell.* 2 (1) (2020) 56–67. <https://doi.org/10.1038/s42256-019-0138-9>.
- [41] S. Wang, L. Zhang, Mechanical, optical, and thermoelectric properties of alpha-BaSnS₃ and beta-BaSnS₃, *RSC Adv.* 15 (16) (2025) 12179–12190. <https://doi.org/10.1039/d5ra01106j>.
- [42] H. Saadi, H. Kerrai, E. M. Jalal, M. El Bouziani, First-principles investigation of optical and thermoelectric properties of Na₂InCu(F/Cl)₆ double perovskites for energy harvesting devices, *Vacuum* 236 (2025) 114149. <https://doi.org/10.1016/j.vacuum.2025.114149>.
- [43] O. Shahid, K. Srivastava, J. Prakash, Enhancement of the thermoelectric performance of LaOBiS₂ by Sc and Y-substitution, *Mater. Sci. Eng. B* 324 (2026) 118995. <https://doi.org/10.1016/j.mseb.2025.118995>.
- [44] C. Zhang, X. Jiang, X. Wang, Q. Lu, L. Li, L. Zhou, N. Xu, Manipulating thermoelectric properties of 2d MgI₂/Bi₂S₃ heterostructure based on first-principles calculations, *Physica B: Condensed Matter* 708 (2025) 417210. <https://doi.org/10.1016/j.physb.2025.417210>.