Journal of
Materials Chemistry A

ROYAL SOCIETY
OF CHEMISTRY

PAPER

View Article Online
View Journal

Check for updates

Cite this: DOI: 10.1039/d5ta08841k

Deep learning framework for accurate prediction and high-throughput search of the thermoelectric figure of merit in skutterudites

Victor Posligua, *[a] Karina Landivar, [a] Elena R. Remesal, [a] Gerda Rogl, [b] Peter F. Rogl, [b] Javier Fdez Sanz, [a] Jesús Prado-Gonjal, [c] Antonio M. Márquez [a] and Jose J. Plata *[a]

Received 30th October 2025
Accepted 17th April 2026

DOI: 10.1039/d5ta08841k

rsc.li/materials-a

The integration of artificial intelligence and machine learning is rapidly transforming the landscape of materials discovery, facilitating unprecedented acceleration in the exploration of vast chemical spaces and the prediction of material properties. However, the adoption of these advanced techniques has not been uniform across all subfields of materials science; thermoelectrics, for instance, has experienced a relatively slower penetration. This lag can be attributed to several inherent challenges, including the physical complexity of thermoelectric phenomena, the scarcity and reliability issues of available data, and limitations concerning the applicability of general models. To address these challenges, in this work, we have developed a machine learning model, based on neural networks, specifically for the accurate prediction of the thermoelectric figure of merit (zT) in skutterudites. This model demonstrates high accuracy, with prediction errors approaching the range of experimental uncertainties reported for zT measurements. Furthermore, it offers the crucial capability of extracting design rules grounded in the underlying physics and chemistry of these materials, providing valuable insights for optimization. Most importantly, our model is applicable for the high-throughput screening of extensive chemical spaces, facilitating the efficient discovery of novel and high-performance thermoelectric materials.

Introduction

The advent of artificial intelligence and machine learning, ML, in the 21st century has been a global transformative breakthrough, comparable to the transition from vacuum tubes to integrated circuits or the development of the internet in the 20th century. Science and technology are now experiencing a paradigm shift, with ML techniques playing an increasingly vital role in fields such as proteomics[1] or drug discovery.[2] Material science is no exception. Traditional methods of material discovery, which often rely on trial-and-error approaches, are slow and inefficient, typically requiring decades to translate a new material from the laboratory to practical application. As an alternative, ML is revolutionizing materials science by enabling the rapid exploration of vast chemical spaces and the prediction of material properties with unprecedented accuracy. However, its adoption has varied across subfields, with some areas seeing faster integration than others. For example, battery technology,[3] catalysis,[4] and metallurgy[5] have seen a robust increase in the use of ML, while others such as thermoelectrics lag behind. The late and relatively slow adoption of ML in thermoelectrics[6,7] can be attributed to a number of factors, which can be summarized as: (i) physical complexity, (ii) data scarcity and reliability, and (iii) applicability.

Thermoelectric efficiency is typically measured by the thermoelectric figure of merit, zT. This magnitude directly depends on electronic (Seebeck coefficient and electrical conductivity) and thermal (thermal conductivity) properties. The dual nature of zT, encompassing both electronic and thermal transport phenomena, presents a key challenge: the strong interdependence between these properties, which makes enhancing zT particularly difficult.[8] Additionally, transport properties are strongly sensitive to synthesis and processing conditions. Small changes in variables such as chemical composition,[9] grain size,[10] or carrier concentration[11] can significantly alter some of these properties, even by orders of magnitude. This is why some of the earliest applications of ML in thermoelectrics have focused on the specific prediction of thermal conductivity,[12,13] Seebeck coefficient,[14,15] or electrical conductivity.[16]

The lack of sufficient and reliable databases for thermoelectric materials is another impediment to the widespread adoption of ML in this field. The sensitivity of transport properties to a large set of variables makes it extremely difficult to

---

[a] Departamento de Química Física, Facultad de Química, Universidad de Sevilla, Seville, E-41012, Spain. E-mail: jplata@us.es

[b] Institute of Materials Chemistry, University of Vienna, Waehringerstraße 42, Wien, A-1090, Austria

[c] Departamento de Química Inorgánica, Universidad Complutense de Madrid, Madrid, Spain

establish a complete enough dataset that not only includes thermoelectric properties but also a sufficiently detailed characterization of the samples. During the last years, the scientific community is addressing this issue through: (i) automatic data mining and curation of existing experimental reports,[17–19] and (ii) theoretical predictions, which complement experimental efforts by generating large datasets of material properties through simulations.[20] However, both experimental and theoretical databases have to be used with caution. On the one hand, experimental measurement of transport properties is associated with moderate uncertainties. It is well documented that measurement discrepancies for zT, in different round-robin studies can vary between ±11.5% and ±17.0% from the averages, depending on the temperature range and the material.[21,22] Some pioneering works have attempted to address this issue through the in-house synthesis of large experimental datasets, with properties measured under the same conditions.[23] This approach has been successfully applied to medium-sized chemical spaces, such as subsets of Ag–Cu-based chalcopyrites, but remains feasible only for a few laboratories and can be very time- and resource-intensive for larger chemical spaces. On the other hand, theoretical prediction of transport properties is severely hampered by the trade-off between accuracy and computational cost. Over the past decade, it has been demonstrated that first-principles-based predictions of transport properties can be highly accurate,[24–27] however this accuracy is closely linked to a high computational cost, which impedes the generation of very large datasets. The main challenge here is the development of high-throughput frameworks that combine accuracy, low computational cost, and automation.[28]

The obstacles mentioned above have led to the creation of only a few ML models for predicting zT. While some models focus on specific materials and explore minor modifications,[29–31] others aim to be more broadly applicable.[32–34] However, most of these models remain mostly structure-agnostic, which highlights the necessity for a more focused strategy to address two indirect obstacles: synthesizability and dopability.[35] By confining the chemical space to a well-defined structure with easily verifiable synthesizability, while still being large enough to allow for the discovery of improved candidates, more applicable ML models can be developed. This has been done before with Heusler alloys,[36,37] which present a large compositional space, a well defined structure and its stability can be initially assessed with simple valence electron counting;[38] however, beyond synthesizability, finding high zT also requires the possibility of doping control. Tuning the right carrier concentration is critical for optimizing zT. Skutterudites are a good example of a material family that fits perfectly well with the characteristics mentioned above. Not only are their structures governed by fundamental electron-counting rules,[39] but the use of multiple fillers can: (i) tremendously expand their chemical space and (ii) be used to tailor their carrier concentration. In this work, we have created an ML-based model to predict the thermoelectric figure of merit of skutterudites and search for new high-performance thermoelectric materials.

Methodology

This study employes an integrated computational framework combining ML techniques and domain-specific descriptors to predict thermoelectric properties of skutterudite-based materials. The workflow begins with the generation of descriptors derived from chemical composition followed by normalization to ensure consistent data scaling. These descriptors are then processed through advanced model training approaches to achieve robust and accurate predictions. LightGBM[40] was used for pre-training, while deep learning models using TensorFlow/Keras[41,42] were applied for predictions. The framework is designed to ensure both predictive accuracy and reproducibility offering valuable insights into the underlying physicochemical principles.

Data collection and storage

The dataset used in this study was curated from experimental results and literature sources, focusing on skutterudite-based thermoelectric materials. This initial stage is crucial, as data quality is fundamental to success of any ML study.

The database construction began with a collection of experimental data on skutterudites synthesized between 1996 and 2022.[43] Data from this collection were further augmented through manual digitalization of relevant literature within that collection, as well as utilizing the Starry Data repository.[18] The data are manually curated to ensure the use of actual rather than nominal compositions and to exclude samples with significant segregation or secondary phases, which are typically poorly characterized. This enhances the reliability of the dataset for machine learning model development. In total, the database includes approximately 4000 skutterudites with diverse compositions, carrier concentrations at 300 K (n_{300}) and temperature (T) profiles, along with labels like thermoelectric properties (*e.g.*, zT). For this study, we considered skutterudite composition that could incorporate up to five distinct atoms in the filler position, three cations and three anions, enabling a wide range compositional variability (Fig. S1 in SI). The data were stored and managed using MongoDB,[44] a NoSQL database, to ensure scalability and efficient querying.

Feature engineering

Feature engineering in this study was designed to emphasize the compositional properties of skutterudite-based thermoelectric materials, capturing the key chemical and physical attributes that influence their performance. By focusing on descriptors tied directly to the elemental composition, the aim is to build a model that reflects the fundamental relationships between composition and thermoelectric properties.

Atomic descriptors were extracted using RDKit,[45] Mendeleev[46] and Pymatgen[47] libraries, specifically chosen to highlight essential compositional attributes. These descriptors include ionization potential (ip), electron affinity (ea), electronegativity (elecneg), valence electrons (val_e) and atomic mass (mass) of the filler-ratlers (f), anions (a), and cations (c), providing a detailed representation of the chemical and compositional attributes of the materials. Properties such as ionization potential, electron affinity, and electronegativity offer insights into the electronic structure and bonding characteristics that are critical for optimizing carrier transport in thermoelectric materials. Meanwhile, the mass of constituent atoms plays an important role in tuning phonon scattering, thereby modifying lattice thermal conductivity.

In addition to these primary descriptors, derived features were calculated to enhance the representation of compositional variability and its impact on thermoelectric properties. Weighted averages (aver) and standard deviations (dev) of descriptors such as ionization potential, electron affinity and electronegativity were computed using atomic fractions. These measures capture both the central tendency and variability in the compositional properties. Ratios, such as anion-to-cation fractions, were derived to reflect charge carrier balance, while the total number of valence electrons in the system provided additional insight into its electronic structure and potential contribution to thermoelectric properties.

The final feature set comprised 37 descriptors, including intrinsic properties like T and n_{300}, along with the derived features explained above (a detailed list of all descriptor is provided in Table S1 in the SI). To assess potential redundancy and interdependencies among the features, we computed the Pearson correlation matrix across all descriptors and the target variable zT (Fig. S2). This analysis revealed some moderate correlations between chemically related descriptors, *e.g.*, valence electrons and electronegativity. However, no features were removed to preserve the interpretability of the model and allow the network to capture potentially nonlinear interactions between descriptors. For instance, although features such as electronegativity, ionization potential, and electron affinity are correlated, their importance may differ depending on whether the thermoelectric performance of p-type or n-type semiconductors is being evaluated, and they have been simultaneously used in previous studies.[34]

Data processing, splitting and normalization

Starting from the curated skutterudite dataset and the 37 compositional descriptors defined above, we first apply a residual analysis using LightGBM to identify and remove outliers from the development dataset. The resulting filtered development pool is then split into training (80%) and validation (20%) subsets, while a compositionally distinct external test set is kept fully separate. Feature normalization is performed using a StandardScaler function (from SciKit-Learn package[48]) fitted only on the training subset and the same transformation is applied to the validation and external test sets before training and evaluation.

Statistical outlier detection and treatment. We used LightGBM regressor as a residual diagnostic tool to identify outliers in the development dataset. This method is particularly suitable for this task because it can handle noisy experimental data, capture complex feature interactions and provide an efficient statistical screening of potentially inconsistent entries. The residuals between the LightGBM predictions and the experimental zT values were analysed using the 3σ method,[49,50] a statistical criterion based on the 68-95-99.7 rule of a normal distribution. The degree of deviation was quantified using an E value, calculated as:

$$E = \frac{x_i - \bar{x}}{x_{\text{std}}} \tag{1}$$

where x_i is the residual for the i-th zT value, $\bar{x}$ is the mean residual and x_{std} is the standard deviation of the residuals.

Residuals with E values within the range [−3, 3] were considered acceptable and retained in the filtered development set, as they fall within the expected range of a normal distribution. Residuals outside this interval were classified as outliers and removed from further model development (Fig. 1a). This outlier treatment was performed before splitting the filtered development dataset into training and validation subsets. For the present dataset, the LightGBM-based residual analysis revealed approximately 4005 acceptable data points and 75

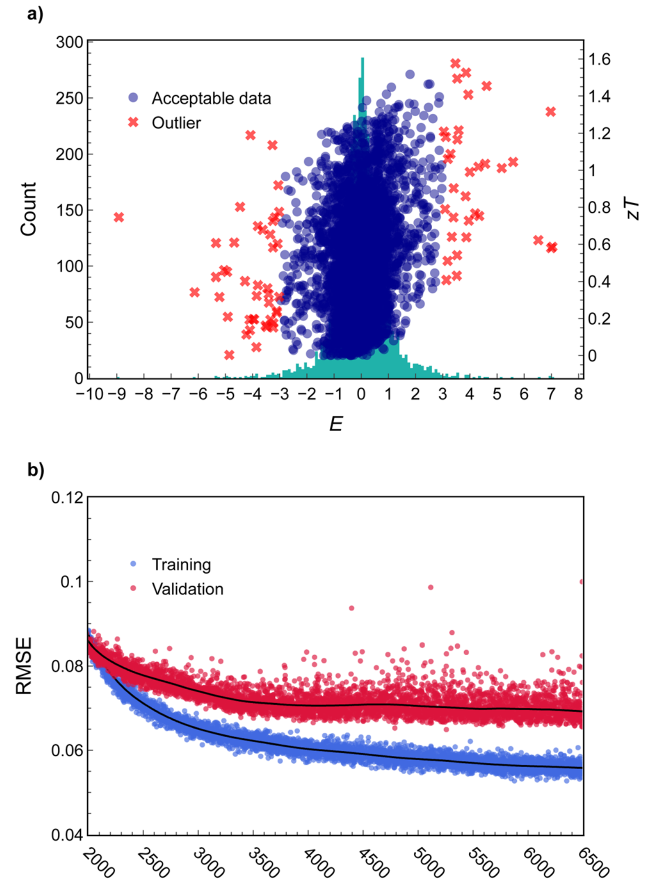

Fig. 1 (a) Residual-based outlier detection using the E-value metric. Histogram shows the distribution of all E values, while scatter points represent corresponding zT values. Acceptable data (blue dots) fall within the normal distribution bounds ($|E| \leq 3$) and outliers (red crosses) exceed this threshold. (b) Root mean squared error (RMSE) as a function of training epochs for both training (blue) and validation (red) sets. Locally weighted scatterplot smoothing (LOWESS) trendlines highlight convergence behavior and show that validation RMSE plateaus without overfitting.

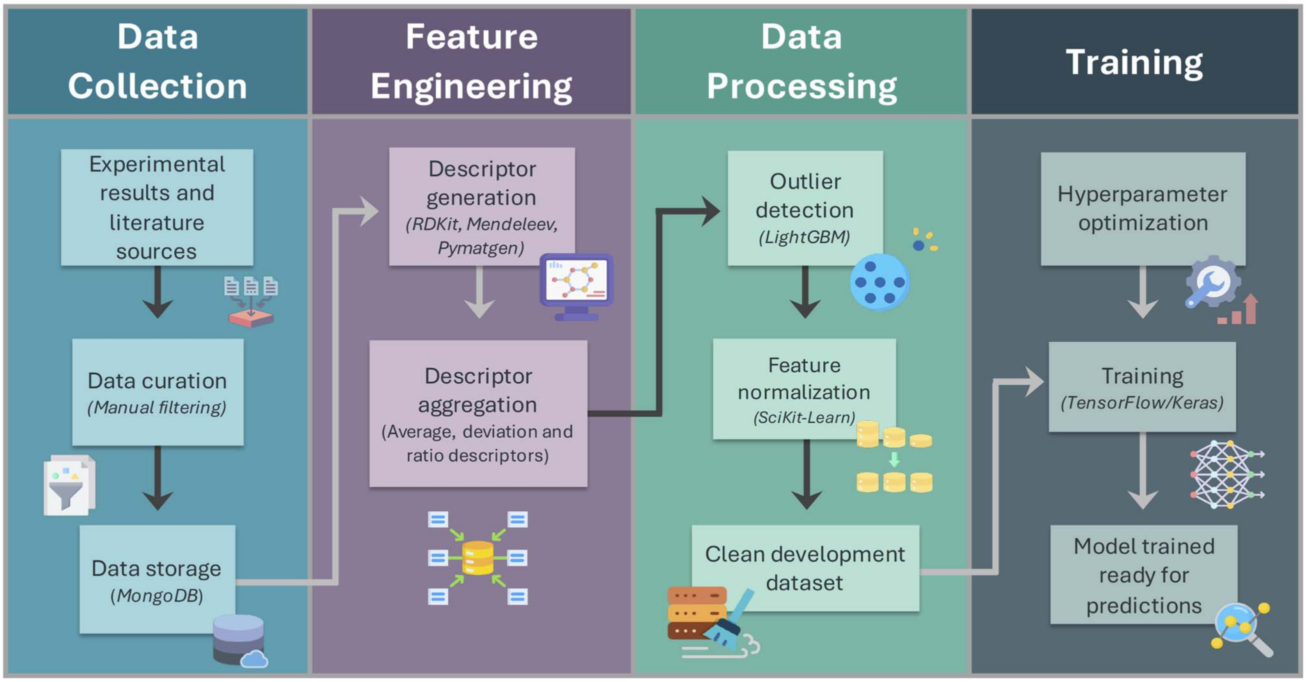

Fig. 2 Workflow of the framework for prediction of zT. The pipeline includes data collection from experiments and literature, descriptor generation, feature engineering, outlier detection, and model training using deep learning.

outliers. Further inspection of the excluded points suggested potential causes such as repeated counting, incorrect typification or digitalization, and measurement inaccuracies. Removing these entries reduced inconsistencies in the development dataset and improved the overall reliability of the subsequent model training.

Since the outlier detection step relies on a supervised LightGBM regressor trained on the development dataset, it may in principle bias the retained data toward compositions that are more consistent with the overall trends learned by the model. In the present study, however, this effect is limited because only a small fraction of entries is removed and a separate robustness check showed that the identified outliers changed only marginally when the LightGBM residual analysis was repeated with and without prior descriptor standardization. In addition, the excluded points are discarded entirely rather than being relabeled or reused. Furthermore, the independent external test set is not subjected to the LightGBM filtering, so the external benchmark remains unaffected by this outlier selection step.

Dataset splitting and feature normalization. After outlier removal, the filtered development dataset was randomly divided into training (80%) and validation (20%) sets using a fixed random seed. The training set was used to fit the neural network parameters, while the validation set was used to monitor convergence and guide the selection of training epoch range. In parallel, the external test set was defined independently and remained fully separated from the development dataset throughout the whole workflow.

Feature normalization was performed using the StandardScaler function fitted only on the training set. The same scaling parameters were then applied to transform the validation and external test sets. This procedure ensures that the normalization step is learned exclusively from the training data and avoids potential scaling leakage into the validation or external benchmarks.

In this way, the workflow (see Fig. 2) consists of: (i) generation of the descriptor matrix from the curated skutterudite dataset, (ii) residual analysis using LightGBM to identify and remove outliers from the development dataset, (iii) splitting of the filtered development pool into training and validation sets, (iv) feature normalization using the StandardScaler function fitted on the training set and applying the same transformation to the validation and external test sets and (v) training and evaluation of the model.

Model architecture and training procedure

The model was developed using TensorFlow/Keras to predict zT values of skutterudite-based materials.

Hyperparameter optimization was carried out using a custom script, allowing efficient exploration of the parameter space while maintaining computational feasibility. The optimized hyperparameters included:

- Number of hidden layers: tested from 1 to 20 to determine the optimal model depth for capturing complex relationships.
- Activation function: evaluated tan$h$, ReLU, eLU and LeakyReLU with varying slopes. LeakyReLU with a slope of 0.05 was selected for its ability to handle negative values while maintaining gradient flow.
- L2 regularization: strengths ranging from 10^{-5} to 0.75 were tested to balance weight penalization and model complexity. A value of 0.1 was chosen as it mitigated overfitting by

discouraging overly complex models, which was particularly beneficial given the moderate dataset size.

- Dropout rate: explored between 0.05 and 0.75, with 0.05 identified as optimal to balance neuron co-adaptation and model complexity.
- Learning rate: values from 10^{-6} to 5×10^{-2} were examined to explore both fine-tuning and rapid convergence scenarios, with 6×10^{-6} providing stable and efficient convergence.
- Batch size: sizes ranging from 2 to 128 were considered, with 32 balancing computational efficiency and gradient stability.

The final model consisted of 10 hidden layers with 37 neurons per layer and was trained using the AdamW optimizer.[51] Model performance was evaluated using the root mean squared error (RMSE) and the coefficient of determination R^2. RMSE was used as the primary metric because it directly quantifies the prediction error on the same scale as the target variable, while R^2 was used as a complementary measure of how well the variance in the experimental data was captured by the model. The number of training epochs was selected from the convergence behaviour of the training and validation RMSE shown in Fig. 1b. Based on this analysis, 6500 epochs were used for the final reported model. Validation metrics were therefore used only as an internal estimate during model development, whereas generalization was assessed separately using independent test and benchmark sets. To assess the sensitivity of the internal validation metrics to the particular 80/20 split, this procedure was repeated using five different random seeds (7, 13, 21, 42 and 84). Across these runs, the validation RMSE was 0.079 ± 0.012 and the validation R^2 was 0.931 ± 0.022, indicating that the model performance is reasonably stable with respect to the specific train/validation partition. To ensure reproducibility, deterministic settings were used during training, including fixed random seed initialization and TensorFlow operation determinism, so that repeated runs under identical conditions produced consistent results.

DFT calculations

All structures were fully relaxed (atoms and lattice) using first-principles density functional theory calculations performed with the VASP[52,53] program and projector-augmented wave potentials.[54] The total energies were computed utilizing the meta-GGA exchange–correlation functional r^{2}SCAN[55] and the core electrons described by the potentials suggested by Calderon *et al.*[56] A high kinetic energy cutoff of 500 eV and a dense *k*-point mesh corresponding to at least 2000 *k*-points per reciprocal atom (*i.e.*, the total number of *k*-points multiplied by the number of atoms in the unit cell) were used to sample the Brillouin zone. The wave function was considered converged when the energy difference between two consecutive electronic steps was less than 10^{-9} eV. To obtain the optimized conventional unit cell geometry, both the atomic positions and the lattice vectors were fully relaxed until the maximum force component acting on any atom was less than 10^{-6} eV Å^{-1}, employing a supplementary support grid to mitigate the noise in the computed forces. To compare the electronic structure of pristine and doped skutterudites, band structure unfolding[57] was performed using the easyunfold code.[58] Force constants and phonon density of states were obtained combining hiPhive[59] and phonopy[60] packages using the hiPhive wrapper.[27,28]

Results and discussion

Predictive performance and benchmarking

Training and validation. After hyperparameter optimization, the final model achieved a training RMSE of 0.056 and R^2 score of 0.966, while the validation set yielded an RMSE of 0.068 and R^2 of 0.947, indicating strong generalization and predictive robustness. Notably, these RMSE values are well below the standard deviations of the corresponding zT distributions ($\sigma = 0.304$ for training and $\sigma = 0.297$ for validation), demonstrating that the prediction errors are small relative to the inherent variability in the data, indicative of a well-trained model.

While these metrics confirm the overall performance of the model, a closer inspection of the training and validation predictions (Fig. 3a and b) reveals a consistent underestimation of high zT values, particularly above 1.2. This deviation likely stems from the absence of descriptors capturing performance enhancements introduced during post-synthesis processing, *e.g.*, high-pressure torsion,[61] spark plasma sintering,[62] cold sintering process,[63] spinodal decomposition[64] or secondary phase segregation,[65] that are known to significantly boost zT. These factors are often inconsistently reported or not systematically quantified in the literature and were thus excluded from the final descriptor set used for model development. As a result, the model learns only from compositional and structural descriptors, limiting its ability to capture extreme zT enhancements arising from non-compositional factors. However, this conservative bias may prove advantageous for screening applications by minimizing false positives and prioritizing compositions that perform well under standard synthesis conditions.

To contextualize these results, we benchmarked our deep learning model against a LightGBM regressor trained on the same curated training and validation sets (excluding outliers). The LightGBM model yielded lower training RMSE (0.041) and higher R^2 (0.982), but its validation metrics (RMSE = 0.122, R^2 = 0.848) showed a greater gap between training and validation performance. This suggests that the model may be overfitting, effectively memorizing patterns rather than learning generalizable trends. In contrast, the deep learning model, though slightly less accurate on the training set, exhibited better validation performance, indicating its enhanced ability to capture nonlinear and complex interactions among the compositional descriptors.

Test set and comparison with other ML-based models. We also compared the performance of our model with recent ML efforts focused on thermoelectric materials. For example, Lee *et al.*[29] developed a model for SnSe-based systems, achieving a validation mean absolute error (MAE) of 0.102 and R^2 of 0.756. Similarly, Li *et al.*[50] applied LightGBM to high-entropy GeTe systems, reporting a validation RMSE of 0.090 and R^2 of 0.954

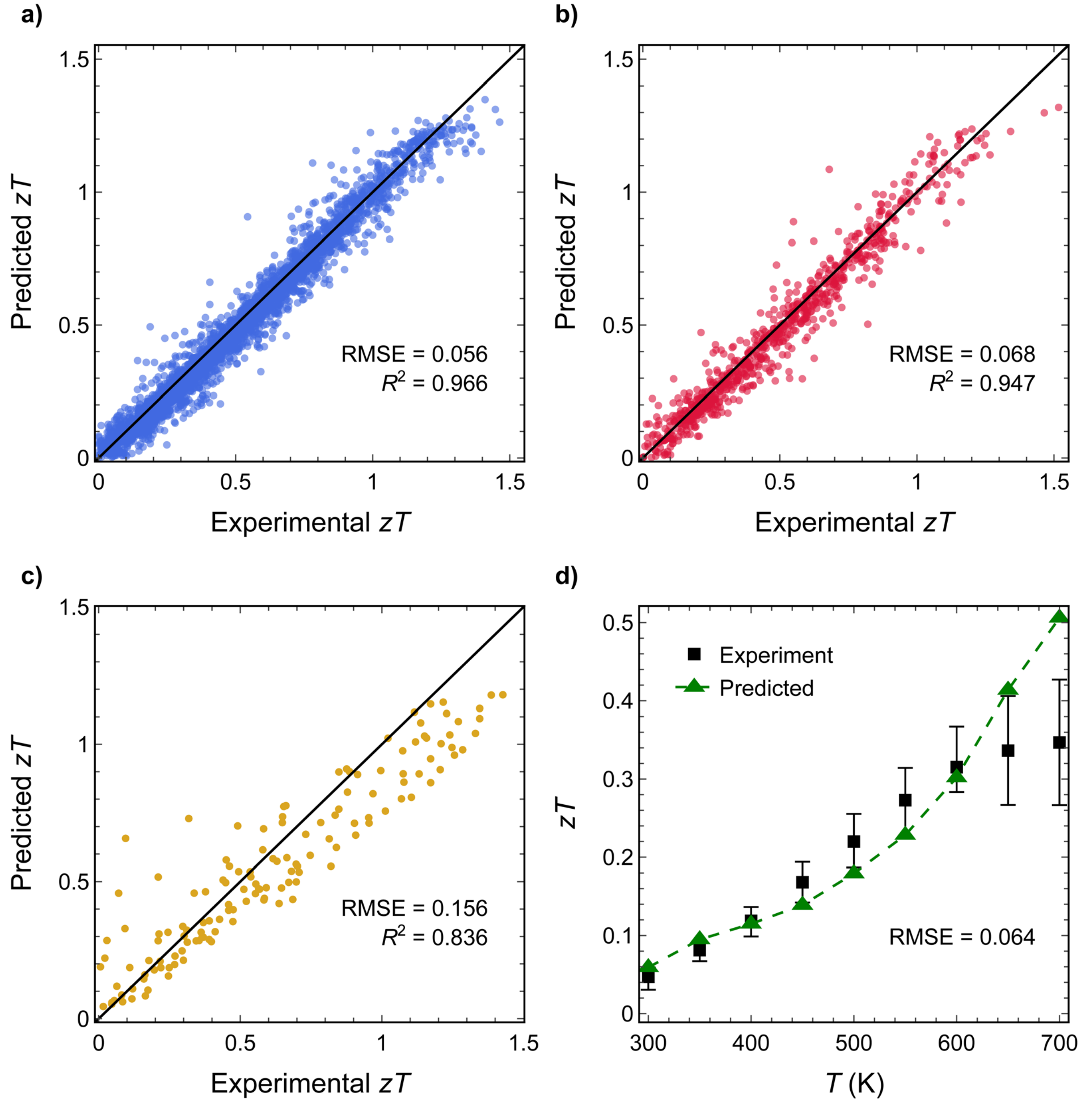

Fig. 3 Model performance on (a) training (blue), (b) validation (red), (c) external test set (orange) and (d) comparison of round robin experimental data (black squares with associated uncertainty bars) with predicted *zT* values (green triangles) for $Co_{0.97}Ni_{0.03}Sb_3$.[22]

using a standard 20% random validation split. In contrast, our model yielded both a lower RMSE and a higher R^2 on a compositionally diverse validation set, underscoring its improved predictive accuracy. While general models evaluate generalization capacity using benchmarks like MRL,[35] ESTM[66] or the dataset reported by Chernyavsky *et al.*,[67] an external set containing only skutterudite-based materials (see Data availability) has been built and carefully selected to exclude any overlap with the development dataset, *i.e.*, no composition appearing in the external test set is present in the training or validation sets at any temperature or carrier concentration. This test set was specifically designed to challenge the model, featuring multifilled systems, Co-substitutions and Sb-vacancies.[68–71] On this benchmark, our model achieved an RMSE of 0.156 and R^2 of 0.836 (Fig. 3c), indicating good predictive performance for chemically complex and experimentally relevant materials. This level of accuracy is particularly meaningful in the context of high-*zT* candidate discovery, where moderate prediction errors (∼0.1) remain acceptable relative to the target–property scale. As such, the model provides a useful tool for identifying high-performance candidates and guiding experimental efforts within the skutterudite family of thermoelectric materials.

Model accuracy *vs.* experimental uncertainty. While the model demonstrates strong predictive accuracy on both the validation and external test sets, it is important to acknowledge the inherent variability in experimentally reported *zT* values. In practice, the same skutterudite-based sample can exhibit considerable variation in zT measurements due to differences in methods and characterization protocols across research groups. For instance, different approximations can be considered when computing heat capacity, C_p, to obtain lattice thermal conductivity, κ_l. To investigate this variability and assess the reliability of the model under realistic experimental conditions, a comparative analysis has been performed using a round robin dataset reported by Alleno *et al.*[22] This dataset consists of zT measurements for the same $Co_{0.97}Ni_{0.03}Sb_3$ sample by multiple laboratories using different measurements systems. It thus provides a unique benchmark to quantify experimental uncertainty and serves as a rigorous reference to evaluate the accuracy of the model predictions. To evaluate how the model aligns with this experimentally observed variability, we compared the predicted zT values for $Co_{0.97}Ni_{0.03}Sb_3$ at different temperatures against the round robin results (Fig. 3d). The predicted values are in good agreement with the measurements; indeed, in the mid-temperature range, the predictions fall within the experimental uncertainty range of zT. The largest deviation occurs at 700 K, but this deviation is no greater than the RMSE computed for the test set. This deviation can be attributed to two main factors: (i) the presence of phase segregation ($CoSb_2$ and cubic Sb) in the reported samples[22] and (ii) the lack of experimental data to confirm the nominal stoichiometry and carrier concentration. Based on other works used as references for the synthesis of the samples,[72] the carrier concentration for that stoichiometry is around 3×10^{-20} cm^{-3}. However, the lattice parameter obtained in the samples synthesized for the round robin are not consistent with previous reports. Indeed, they would indicate a lower Ni concentration and thus a lower carrier concentration.[72] This mismatch in lattice parameters and inferred carrier concentrations likely underlies the tendency of the model to overestimate zT, especially at higher temperatures, where carrier optimization plays a critical role in n-type skutterudite performance. In addition to the round-robin tests, other reports indicate substantial variability in zT measurements for similar Ni-doped $CoSb_3$ samples, even with small variations in Ni content, synthesis method, and microstructure. For instance, Zhang *et al.* reported a zT of 0.57 for a $Co_{0.975}Ni_{0.025}Sb_3$ sample at 700 K.[73] Katsuyama *et al.* and Wang *et al.* independently reported zT values of 0.48 and 0.5, respectively, for $Co_{1-x}Ni_xSb_3$ with $x = 0.06$ and $x = 0.075$ at 700 K using vacuum quartz tubes for the synthesis,[74,75] which are very close to the predicted values. He *et al.* also reported zT values ranging between 0.35 and 0.59 at 700 K for $Co_{3.9}Ni_{0.1}Sb_3$ samples, obtained by subsequent ball milling, hot pressing, and annealing, depending on the presence of pores.[76] These discrepancies highlight the significant influence of synthesis conditions and microstructural variations on thermoelectric performance, underscoring the challenges in developing accurate predictive models for zT.

Model analysis

To further interpret the model predictions, a SHapley Additive exPlanations (SHAP) analysis[77] was conducted on both the validation and external test sets. Including both sets ensures the analysis reflects not only the internal consistency of the model but also the generalizability to compositionally distinct and unseen systems. SHAP values quantify the contribution of each feature to the predicted output, enabling identification of the most influential descriptors driving the predicted zT.

For an individual sample, a positive SHAP means that the corresponding descriptor pushes the predicted zT to a higher value, whereas a negative SHAP value means that it lowers the prediction. The magnitude of the SHAP value indicates how strongly that descriptor influences the model output. In the SHAP summary plots (Fig. 4 and 6), each point represents one sample and is colored according to the descriptor value, from low (blue) to high (red). This allows the trends to be interpreted directly: if high descriptor values are mainly associated with positive SHAP values, increasing that descriptor tends to increased the predicted zT; conversely, if high descriptor values are mainly associated with negative SHAP values, increasing that descriptor tends to reduce the predicted zT. SHAP therefore provides a clear way to connect physically motivated descriptors with the model predictions, while still describing learned associations rather than strict causality.

As shown in Fig. 4, the top-ranked features span thermal, electronic and compositional domains. Temperature, T, and semiconductor type, p_n, emerge as key drivers, consistent with their known influence on charge carrier behavior and transport mechanisms. The SHAP color gradient for T indicates that higher temperatures (in red) are systematically associated with

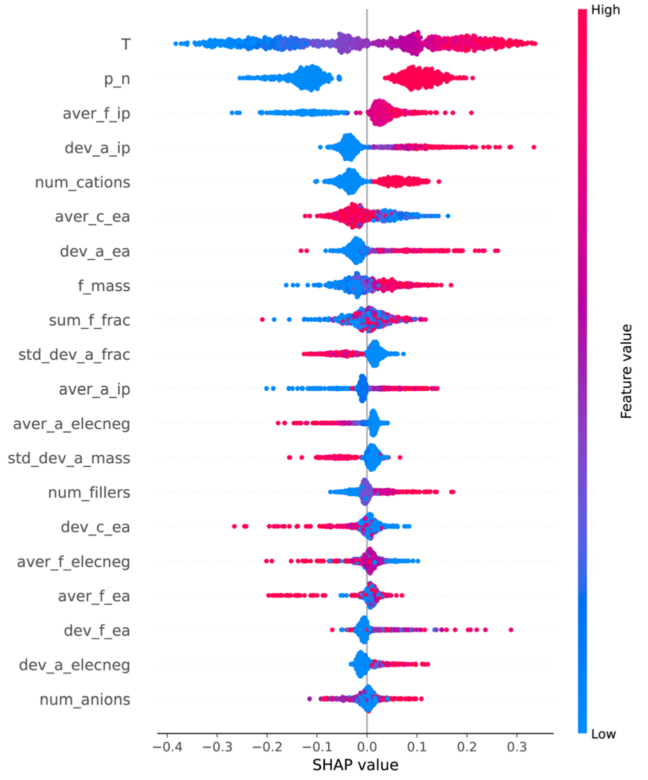

Fig. 4 SHAP summary plot for validation and external test set, highlighting the most impactful features and how their values influence predicted zT.

increased SHAP values, meaning that the model has learned to predict higher zT at elevated temperatures.[78] This aligns with previous reports classifying skutterudites as mid- to high-temperature thermoelectric materials, which perform optimally in the 700–800 K range. The semiconductor type feature, p_n, is binary (0 for p-type, 1 for n-type) and the SHAP distribution shows that n-type samples (in red) consistently yield positive SHAP values. This suggests the model has captured the tendency of n-type skutterudites to achieve higher zT, consistent with their favorable conduction band filling and ability to exploit secondary pockets.[79] This trend is further explored in a focused analysis of double-filled skutterudites presented later.

Some features linked to anions exhibits a broad SHAP distribution, specially those with high values contributing positively to zT. Notably, the standard deviation of the ionization potential and the standard deviation of the electron affinity are ranked among most important features. This suggests that local fluctuations in anionic site potential, likely tied to electronegativity variation, modulate both the electronic structure and scattering processes. Thus, chemical inhomogeneity in anion environments can affect carrier mobility and lattice thermal conductivity in filled skutterudites.

Substituting Sb with atoms of different ionization potentials and electron affinities modifies the typical four-member anion rings in the skutterudite framework, potentially impacting both the electronic structure and phonon transport. The antibonding states of these rings are the responsible of the secondary pocket that increases the power factor in skutterudites.[80] A distortion of these bonds can modify the energy separation between bonding and antibonding states, thereby changing the energy of the secondary pocket.

As proof of concept, the electronic structure of $CoSb_3$ doped with Ge and Se was analyzed to understand the potential influence of doping on zT. First, a 2 × 2 × 2 supercell was modeled, including Ge and Se atoms at opposite corners of one anionic four-membered ring. The unfolded band structure and Ge and Se projections (Fig. 5a) were then calculated for direct comparison with pristine $CoSb_3$ (Fig. S3). In addition to a small reduction of the band gap, both the main and secondary pockets of the conduction band were not well-defined,

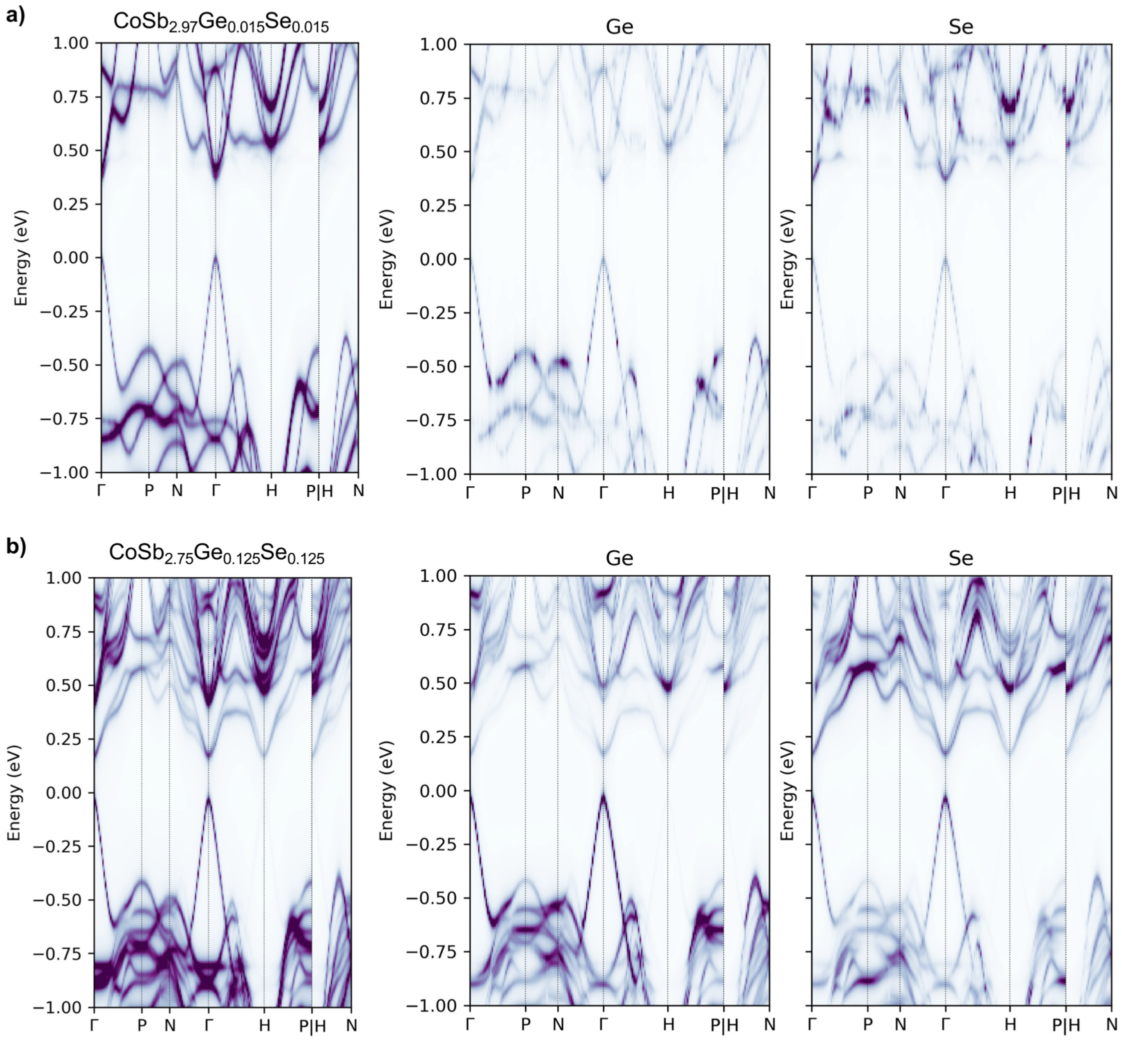

Fig. 5 Unfolded band structures and Ge and Se projections for (a) $CoSb_{2.97}Ge_{0.015}Se_{0.015}$ and (b) $CoSb_{2.75}Ge_{0.125}Se_{0.125}$.

suggesting that doping with Ge and Se breaks the degeneracy of these states. If Ge and Se states are projected and their intensity is magnified, the splitting of states becomes clearer. Additionally, Se states at H are almost at the same energy as the main pocket. To confirm this trend, the same doping was performed in a smaller cell, increasing the Ge and Se content (Fig. 5b). The effects mentioned previously are reinforced here: the band gap is drastically reduced, and the splitting of the states of both pockets is enhanced. A third pocket, almost at the same energy as the main pocket centered at Γ, is observed at H with a strong contribution from Ge and Se. Based on these results, it can be highlighted that appropriate doping of the anionic sites can not only modify the energy of the secondary pocket but also lead to a band convergence phenomenon that should enhance the power factor. Very recently, Wang *et al.* reported what they term "electrical compensation-bonding modulation" method, in which Sb_4 rings are reconfigured through doping with Te, Ge, and even Se.[81] This strategy not only modulates the bond hierarchy, improving transport properties and opening the door to the use of electronegative fillers, but also achieves zT values close to 2. Our model reproduces these results with high accuracy up to zT around 1.2, beyond which it slightly underestimates the values (Fig. S4). These discrepancies can be related to the absence of a significant amount of data in the training set with zT values above 1.2; however, they demonstrate that the model can predict with good accuracy new strategies based on previously undescribed physico-chemical phenomena.

Overall, this SHAP analysis not only validates the importance of expected thermoelectric descriptors such as T and semiconductor type but also reveals nuanced trends in compositional and electronic variability. These observations can not only complement previous experimental discoveries—providing a better interpretation of the roles of descriptors (*e.g.*, ionization potential, filler mass, and anion electronegativity)—but also anticipate and discover new trends when combined with high-throughput screening searches led by the ML model.

High-throughput screening

As mentioned in the introduction, one of the strongest advantages of building an ML model around skutterudites is their large chemical space and applicability. By considering electron counting rules, we can perform an exhaustive enumeration of chemically valid skutterudite candidates for rapid screening using our model. From an experimental standpoint, fillers offer the greatest opportunity for optimization without compromising synthesizability. In particular, we focused on the general formula $R1_xR2_yCo_4Sb_{12}$ (where x, $y = 0.1$, 0.2 and $x + y = 0.3$), reflecting experimentally viable filler concentrations. As potential fillers, elements that have been previously used as rattlers in skutterudite systems were selected. The final list contains over seven thousand compounds, including p- and n-type materials with varying carrier concentrations at 700 K.

Among the top 100 compounds (Table 1), experimental reports exist for some compositions. For instance, Salvador *et al.* reported Ca,Yb double-filled skutterudites.[82] Although they did not synthesize a sample with the precise stoichiometry $Yb_{0.1}Ca_{0.2}Co_4Sb_{12}$, they characterized a close composition by electron probe microanalysis as $Yb_{0.12}Ca_{0.15}Co_4Sb_{12.09}$, with a zT value of 0.75 at 700 K.[82] Using this experimental stoichiometry, the model slightly overestimates the predicted zT, possibly due

Table 1 List of double-filled n-type skutterudites with highest zT using ML model

Composition	zT	Composition	zT	Composition	zT	Composition	zT
$Gd_{0.1}Y_{0.2}Co_4Sb_{12}$	1.16	$Sr_{0.1}Tb_{0.2}Co_4Sb_{12}$	1.06	$Sm_{0.1}Ca_{0.2}Co_4Sb_{12}$	1.04	$Ba_{0.1}Y_{0.2}Co_4Sb_{12}$	1.01
$Sr_{0.1}Li_{0.2}Co_4Sb_{12}$	1.15	$Mg_{0.1}Y_{0.2}Co_4Sb_{12}$	1.06	$Ba_{0.1}Ca_{0.2}Co_4Sb_{12}$	1.04	$Eu_{0.1}Gd_{0.2}Co_4Sb_{12}$	1.00
$Y_{0.1}Dy_{0.2}Co_4Sb_{12}$	1.15	$Ca_{0.1}Al_{0.2}Co_4Sb_{12}$	1.06	$Sr_{0.1}Ce_{0.2}Co_4Sb_{12}$	1.04	$Tl_{0.1}Eu_{0.2}Co_4Sb_{12}$	1.00
$Eu_{0.1}Y_{0.2}Co_4Sb_{12}$	1.14	$Tb_{0.1}Li_{0.2}Co_4Sb_{12}$	1.06	$Ba_{0.1}K_{0.2}Co_4Sb_{12}$	1.04	$Tl_{0.1}Tb_{0.2}Co_4Sb_{12}$	1.00
$Gd_{0.1}Yb_{0.2}Co_4Sb_{12}$	1.13	$Sr_{0.1}Sm_{0.2}Co_4Sb_{12}$	1.06	$Ta_{0.1}Sr_{0.2}Co_4Sb_{12}$	1.04	$Sn_{0.1}Yb_{0.2}Co_4Sb_{12}$	1.00
$Li_{0.1}Ca_{0.2}Co_4Sb_{12}$	1.12	$Sm_{0.1}Yb_{0.2}Co_4Sb_{12}$	1.06	$Pr_{0.1}Y_{0.2}Co_4Sb_{12}$	1.04	$Mg_{0.1}Yb_{0.2}Co_4Sb_{12}$	0.99
$Sm_{0.1}Y_{0.2}Co_4Sb_{12}$	1.12	$Eu_{0.1}Dy_{0.2}Co_4Sb_{12}$	1.06	$Y_{0.1}Li_{0.2}Co_4Sb_{12}$	1.03	$Tb_{0.1}Gd_{0.2}Co_4Sb_{12}$	0.99
$Ce_{0.1}Yb_{0.2}Co_4Sb_{12}$	1.11	$Eu_{0.1}Ce_{0.2}Co_4Sb_{12}$	1.06	$Al_{0.1}Tl_{0.2}Co_4Sb_{12}$	1.03	$Yb_{0.1}Ca_{0.2}Co_4Sb_{12}$	0.98
$Tb_{0.1}Y_{0.2}Co_4Sb_{12}$	1.11	$Sn_{0.1}Gd_{0.2}Co_4Sb_{12}$	1.06	$Ca_{0.1}Tb_{0.2}Co_4Sb_{12}$	1.03	$Al_{0.1}Eu_{0.2}Co_4Sb_{12}$	0.98
$Ce_{0.1}Y_{0.2}Co_4Sb_{12}$	1.10	$Eu_{0.1}Yb_{0.2}Co_4Sb_{12}$	1.06	$Nd_{0.1}Dy_{0.2}Co_4Sb_{12}$	1.03	$Al_{0.1}Sm_{0.2}Co_4Sb_{12}$	0.97
$Y_{0.1}Yb_{0.2}Co_4Sb_{12}$	1.10	$Pr_{0.1}Ce_{0.2}Co_4Sb_{12}$	1.06	$Tl_{0.1}Gd_{0.2}Co_4Sb_{12}$	1.03	$Tl_{0.1}Pr_{0.2}Co_4Sb_{12}$	0.96
$Tb_{0.1}Yb_{0.2}Co_4Sb_{12}$	1.10	$Gd_{0.1}Ce_{0.2}Co_4Sb_{12}$	1.05	$Yb_{0.1}Li_{0.2}Co_4Sb_{12}$	1.03	$Al_{0.1}Nd_{0.2}Co_4Sb_{12}$	0.96
$Sn_{0.1}Y_{0.2}Co_4Sb_{12}$	1.09	$Mg_{0.1}Tb_{0.2}Co_4Sb_{12}$	1.05	$Mg_{0.1}Dy_{0.2}Co_4Sb_{12}$	1.02	$Al_{0.1}Pr_{0.2}Co_4Sb_{12}$	0.96
$Tb_{0.1}Ce_{0.2}Co_4Sb_{12}$	1.08	$Pr_{0.1}Ca_{0.2}Co_4Sb_{12}$	1.05	$Sr_{0.1}Nd_{0.2}Co_4Sb_{12}$	1.02	$K_{0.1}Yb_{0.2}Co_4Sb_{12}$	0.96
$Dy_{0.1}Yb_{0.2}Co_4Sb_{12}$	1.08	$Sr_{0.1}Gd_{0.2}Co_4Sb_{12}$	1.05	$Pr_{0.1}Dy_{0.2}Co_4Sb_{12}$	1.02	$Al_{0.1}Yb_{0.2}Co_4Sb_{12}$	0.94
$Ce_{0.1}Dy_{0.2}Co_4Sb_{12}$	1.08	$Sr_{0.1}Eu_{0.2}Co_4Sb_{12}$	1.05	$Ta_{0.1}Ca_{0.2}Co_4Sb_{12}$	1.02	$Mg_{0.1}Sn_{0.2}Co_4Sb_{12}$	0.91
$Sr_{0.1}Yb_{0.2}Co_4Sb_{12}$	1.08	$Nd_{0.1}Y_{0.2}Co_4Sb_{12}$	1.05	$Ca_{0.1}Dy_{0.2}Co_4Sb_{12}$	1.02	$Yb_{0.1}Ta_{0.2}Co_4Sb_{12}$	0.90
$Gd_{0.1}Dy_{0.2}Co_4Sb_{12}$	1.07	$Sn_{0.1}Tb_{0.2}Co_4Sb_{12}$	1.04	$Mg_{0.1}Gd_{0.2}Co_4Sb_{12}$	1.02	$Tl_{0.1}Ba_{0.2}Co_4Sb_{12}$	0.90
$Ba_{0.1}Li_{0.2}Co_4Sb_{12}$	1.07	$Sr_{0.1}Dy_{0.2}Co_4Sb_{12}$	1.04	$Ca_{0.1}Eu_{0.2}Co_4Sb_{12}$	1.02	$Sr_{0.1}Ba_{0.2}Co_4Sb_{12}$	0.90
$Tl_{0.1}Yb_{0.2}Co_4Sb_{12}$	1.07	$Nd_{0.1}Ca_{0.2}Co_4Sb_{12}$	1.04	$Pr_{0.1}Sr_{0.2}Co_4Sb_{12}$	1.01	$Ce_{0.1}Sn_{0.2}Co_4Sb_{12}$	0.90
$Ce_{0.1}Ca_{0.2}Co_4Sb_{12}$	1.07	$Sm_{0.1}Dy_{0.2}Co_4Sb_{12}$	1.04	$K_{0.1}Sr_{0.2}Co_4Sb_{12}$	1.01	$Pr_{0.1}Mg_{0.2}Co_4Sb_{12}$	0.90
$Y_{0.1}Ca_{0.2}Co_4Sb_{12}$	1.07	$Nd_{0.1}Yb_{0.2}Co_4Sb_{12}$	1.04	$Ce_{0.1}Li_{0.2}Co_4Sb_{12}$	1.01	$Ba_{0.1}Al_{0.2}Co_4Sb_{12}$	0.90
$Sm_{0.1}Ce_{0.2}Co_4Sb_{12}$	1.07	$Pr_{0.1}Yb_{0.2}Co_4Sb_{12}$	1.04	$Li_{0.1}Dy_{0.2}Co_4Sb_{12}$	1.01	$Gd_{0.1}Pr_{0.2}Co_4Sb_{12}$	0.90
$Sn_{0.1}Dy_{0.2}Co_4Sb_{12}$	1.07	$Tb_{0.1}Dy_{0.2}Co_4Sb_{12}$	1.04	$Sn_{0.1}Ta_{0.2}Co_4Sb_{12}$	1.01	$Mg_{0.1}Ta_{0.2}Co_4Sb_{12}$	0.89
$Nd_{0.1}Ce_{0.2}Co_4Sb_{12}$	1.06	$Sn_{0.1}Ca_{0.2}Co_4Sb_{12}$	1.04	$Sn_{0.1}Li_{0.2}Co_4Sb_{12}$	1.01	$Sr_{0.1}Tl_{0.2}Co_4Sb_{12}$	0.89

to secondary phases such as Yb_2O_3 and CaO reported in the samples (Fig. S5). As expected, better agreement is found when experimental samples do not contain secondary phases. For instance, the low Yb content in $Y_{0.05}Yb_x$–$CoSb_3$ minimizes the appearance of nanoprecipitates or phase segregation,[83] obtaining better agreement between our predictions—adapted to the actual compositions—and the experimental values (Fig. S5). Other experimental reports supporting the double-filler skutterudite screening include Ca,Ba–$CoSb_3$,[84] Ca,Ce–$CoSb_3$,[85] Ba,Sr–$CoSb_3$,[86] and Yb,Ce–$CoSb_3$.[87] Most importantly, beyond its predictive accuracy, the model demonstrates that it can be used to efficiently explore large chemical spaces to identify high-*zT* candidates.

Previous SHAP analysis was performed using a large data set in which all features were considered and modified, which can make it more difficult to extract design principles based on filler optimization. That is why a SHAP analysis has been performed on the double-filled skutterudite list (Fig. 6). The analysis, performed separately for p- and n-type samples to reveal potential differences in performance optimization, used a background sample size of 1000 to ensure a balance between computational efficiency and attribution stability. While in the previous SHAP analysis, the main features were well-known variables such as *T*, semiconductor type, or the amount of fillers, here all these variables have been kept fixed, so the analysis focused on filler features. First, feature importance varies depending on the type of skutterudite. For p-type skutterudites, the variables with the greatest impact on *zT* are the standard deviation of the fillers' masses, the standard deviation of the fillers' ionization potentials, and their average electronegativity. The ionization potential of the fillers influences the position of the filler states in the valence band, thereby modifying the band topology and carrier concentration. Moreover, the presence of rattlers with different masses should reduce the lattice thermal conductivity. Meanwhile, for n-type skutterudites, the three most important features—in addition to the standard deviation of the fillers' masses and ionization potentials—are the standard deviations of the fillers' electron affinities. Filler-related features such as electron affinity and electronegativity define the position of the filler states in the conduction band. Additionally, large differences in electron affinity are linked to fillers with different sizes and masses, which can produce more efficient phonon scattering. In both types of semiconductors, the standard deviation of the fillers' masses ranks among the top three features. Following this analysis, we investigated the influence of fillers on thermal conductivity by exploring the phonon density of states of filled skutterudites containing some of the most frequent candidate fillers from Table 1. In each case, the projected phonon density of the rattling atoms was calculated to evaluate the potential for dual-frequency resonant phonon scattering (Fig. 7 and S6).[88] On one hand, heavy atoms such as Yb or Dy present their phonon density of states at very low frequencies, increasing scattering rates of acoustic modes and reducing their group velocities.[89] These rattlers are key to reduce the large contribution of the acoustic modes to the thermal conductivity of these materials, as can be observed in the cumulative thermal conductivity of $CoSb_3$ (Fig. 7). On the other hand, lighter atoms such as Mg and Ca present their vibrational pDOS in the 2–6 THz range enhancing scattering phenomena of low energy optic modes. The contribution of low energy optical modes is not as important as acoustic modes but represents around 15–20% of the total κ_l. Top candidates of the list combine both, heavy and light atoms, thereby promoting broadband phonon scattering essential for significant thermal

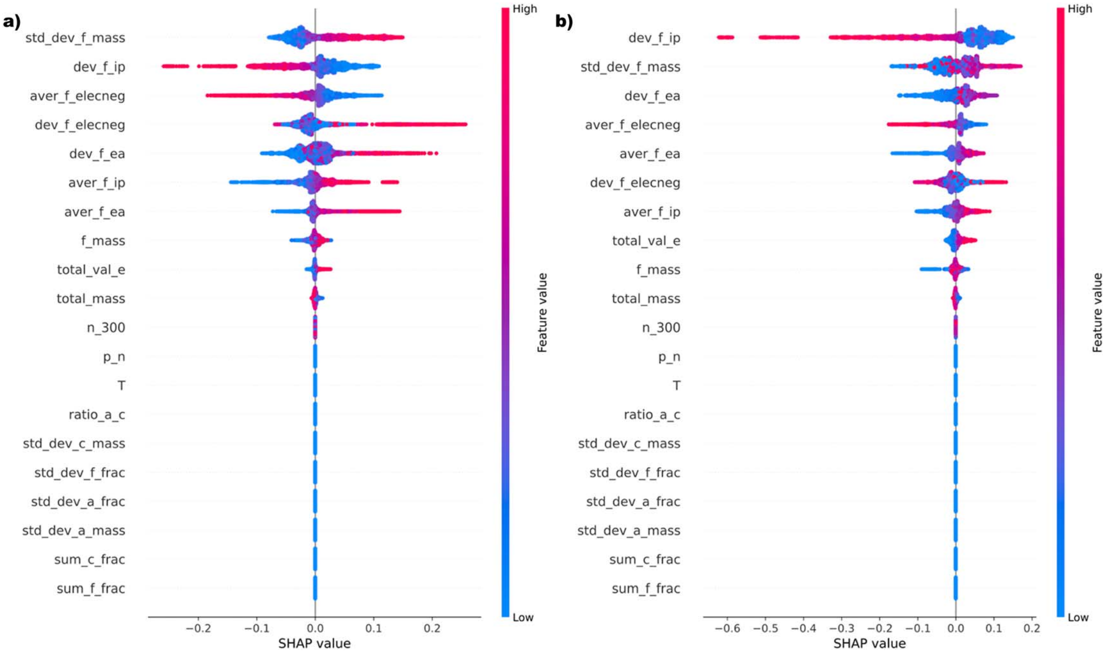

Fig. 6 SHAP summary plots for (a) p-type and (b) n-type double-filled skutterudites.

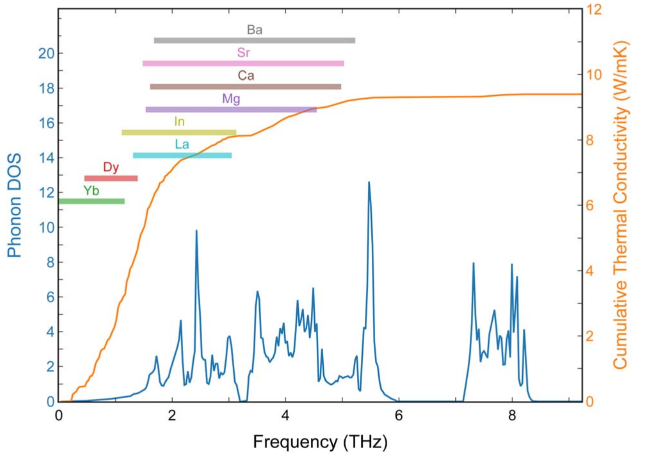

Fig. 7 Phonon density of states (blue) and cumulative lattice thermal conductivity at 300 K (orange) for $CoSb_3$. The color bars indicate the frequency ranges of the rattler-projected phonon density of states. These rattler-projected phonon densities are also shown in Fig. S5.

conductivity reduction. This phenomenon can also be observed in some of the synthesized n-type samples with the highest reported *zT*, where 3 or 4 rattlers (In, Sr, Ba, Yb) are combined.[69,90] Our DFT calculations corroborate that most of these rattlers present resonant frequencies in different areas of the spectra (Fig. 7 and S6), maximizing the number of scattering processes, while the ML model is in good agreement with experimental results (Fig. S7). These results highlight how the model provides critical insights into the design principles for optimizing the thermoelectric performance of skutterudites through strategic filler selection.

Conclusions

This work presents a robust methodological framework designed to address critical challenges in applying artificial intelligence and machine learning to thermoelectric materials. Specifically, this approach tackles physical complexity, data scarcity and reliability, and the broad applicability of ML models for the prediction of *zT* in skutterudites.

To address data scarcity, a combination of digital repositories and manual digitalization of works comprising diverse compositions and carrier concentrations led to a dataset with more than 4000 entries. Data curation constitutes the cornerstone of the methodology regarding reliability. Each experimental entry was analyzed, with particular attention paid to accurately representing actual chemical compositions, rather than nominal ones, and carrier concentrations. To further ensure data integrity and prevent noise from propagating into the model, feature normalization and statistical outlier detection techniques were implemented. We leveraged LightGBM for residual analysis and applied the 3σ method to identify and exclude outliers, thereby minimizing inconsistencies and improving the overall reliability of the development dataset. The inherent physical complexity of thermoelectric phenomena, stemming from the strong interdependence between electronic and thermal transport properties, is addressed by using meaningful features. These features are directly related to thermal and electrical transport properties, including elemental properties and derived features such as weighted averages and standard deviations of ionization potential, electron affinity, and electronegativity. Furthermore, the application of SHAP analysis is critical for verifying that the model not only performs predictions but also implicitly acquires and reflects the fundamental physical laws governing thermoelectricity, such as the influence of temperature and semiconductor type on *zT*. The applicability of the model is ensured by two key pillars. First, the selection of skutterudites, with stability determined by simple, well-defined rules and carrier concentration levels tunable through the nature and concentration of their rattlers, ensures the model's capability to discover new materials. Second, the exclusive use of simple compositional features and elemental properties ensures that the process of exploring new candidates is computationally fast and resource-efficient.

To the best of our knowledge, our manually curated dataset represents the largest collection of skutterudites for thermoelectric applications, which underpins the strong potential of our ML model. We have demonstrated the model's high accuracy, achieving an RMSE of 0.068 and an R^2 of 0.947 during validation. These metrics are comparable to, or even surpass, those obtained by other ML models focused on different families of thermoelectric materials. Most importantly, the errors associated with the model align well with the inherent experimental uncertainties typically observed in thermoelectric measurements, which can vary significantly across different laboratories and characterization protocols. In part, the model's accuracy stems from the data curation, which involved: (i) minimizing the presence of poorly characterized samples or those reported with secondary phases, (ii) including large variability in temperature, carrier concentration, and composition across the entire periodic table, and (iii) selecting samples with low, medium, and high *zT* values to ensure the model performs well over the widest possible range.

Beyond its accuracy, it has been shown that the model has captured a deep understanding of the physical and chemical phenomena controlling thermoelectric performance. For instance, the model identifies how anion substitutions, such as doping with Ge and Se, can drastically modify the topology of the conduction band, leading to band convergence phenomena that enhance the power factor and consequently *zT*. Similarly, regarding rattlers, the model discerns the potential benefits of combining rattlers with differing resonant frequencies to enhance phonon scattering processes and efficiently reduce thermal conductivity. Finally, this ML model proves to be a powerful tool for exploring vast chemical spaces and efficiently identifying new thermoelectric materials with high *zT*, while requiring affordable computational resources. We have demonstrated that the model can identify compositions that have recently been experimentally reported as promising. To further contribute to the scientific community, we are currently developing an online API, which will facilitate researchers to use our model for identifying promising candidates for their experimental synthesis.

Author contributions

J. J. P. and A. M. M. conceived and initiated the research project. V. P., K. L., G. R., P. F. R., E. R. R. and J. J. P. collected the data. V. P. and K. L. trained and optimized the model. J. J. P. and A. M. M. and J. F. S. performed all calculations. All authors contribute to the analysis presented in the main text. V. P. and J. J. P. wrote the first draft. All authors discussed the results and contributed to the final paper.

Conflicts of interest

There are no conflicts to declare.

Data availability

Data sets are available at the ZENODO repository (https://doi.org/10.5281/zenodo.19217112).

Supplementary information (SI) is available. See DOI: https://doi.org/10.1039/d5ta08841k.

Acknowledgements

This work was funded by grant PID2022-138063OB-I00 funded by MICIU/AEI/10.13039/501100011033 and by FEDER, UE, and by grants TED2021-130874B-I00 and TED2021-129569A-I00 funded by MICIU/AEI/10.13039/501100011033 and by the "European Union NextGenerationEU/PRTR. We thankfully acknowledge the computer resources at Lusitania (Cenits-COMPUTAEX), Red Española de Supercomputación, RES (QHS-2024-1-0022 and QHS-2024-2-0020) and Albaicín (Centro de Servicios de Informática y Redes de Comunicaciones – CSIRC, Universidad de Granada).

Notes and references

1 J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Žídek, A. Potapenko, A. Bridgland, C. Meyer, S. A. A. Kohl, A. J. Ballard, A. Cowie, B. Romera-Paredes, S. Nikolov, R. Jain, J. Adler, T. Back, S. Petersen, D. Reiman, E. Clancy, M. Zielinski, M. Steinegger, M. Pacholska, T. Berghammer, S. Bodenstein, D. Silver, O. Vinyals, A. W. Senior, K. Kavukcuoglu, P. Kohli and D. Hassabis, *Nature*, 2021, **596**, 583–589.

2 J. Vamathevan, D. Clark, P. Czodrowski, I. Dunham, E. Ferran, G. Lee, B. Li, A. Madabhushi, P. Shah, M. Spitzer and S. Zhao, *Nat. Rev. Drug Discovery*, 2019, **18**, 463–477.

3 K. A. Severson, P. M. Attia, N. Jin, N. Perkins, B. Jiang, Z. Yang, M. H. Chen, M. Aykol, P. K. Herring, D. Fraggedakis, M. Z. Bazant, S. J. Harris, W. C. Chueh and R. D. Braatz, *Nat. Energy*, 2019, **4**, 383–391.

4 J. A. Keith, V. Vassilev-Galindo, B. Cheng, S. Chmiela, M. Gastegger, K.-R. Müller and A. Tkatchenko, *Chem. Rev.*, 2021, **121**, 9816–9872.

5 T. DebRoy, T. Mukherjee, H. Wei, J. Elmer and J. Milewski, *Nat. Rev. Mater.*, 2021, **6**, 48–68.

6 L. M. Antunes, Vikram, J. J. Plata, A. V. Powell, K. T. Butler and R. Grau-Crespo, in *Machine Learning in Materials Informatics: Methods and Applications*, ACS Publications, 2022, pp. 1–32.

7 N. K. Barua, S. Lee, A. O. Oliynyk and H. Kleinke, *JPhys Energy*, 2025, **7**, 021001.

8 G. J. Snyder and E. S. Toberer, *Nat. Mater.*, 2008, **7**, 105–114.

9 K. Park, K. Ahn, J. Cha, S. Lee, S. I. Chae, S.-P. Cho, S. Ryee, J. Im, J. Lee, S.-D. Park, M. J. Han, I. Chung and T. Hyeon, *J. Am. Chem. Soc.*, 2016, **138**, 14458–14468.

10 S. I. Kim, K. H. Lee, H. A. Mun, H. S. Kim, S. W. Hwang, J. W. Roh, D. J. Yang, W. H. Shin, X. S. Li, Y. H. Lee, G. J. Snyder and S. W. Kim, *Science*, 2015, **348**, 109–114.

11 P. Jood, J. P. Male, S. Anand, Y. Matsushita, Y. Takagiwa, M. G. Kanatzidis, G. J. Snyder and M. Ohta, *J. Am. Chem. Soc.*, 2020, **142**, 15464–15475.

12 Y. Luo, M. Li, H. Yuan, H. Liu and Y. Fang, *npj Comput. Mater.*, 2023, **9**, 4.

13 N. K. Barua, E. Hall, Y. Cheng, A. O. Oliynyk and H. Kleinke, *Chem. Mater.*, 2024, **36**, 7089–7100.

14 A. Furmanchuk, J. E. Saal, J. W. Doak, G. B. Olson, A. Choudhary and A. Agrawal, *J. Comput. Chem.*, 2018, **39**, 191–202.

15 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 and S. Yorozu, *Sci. Rep.*, 2019, **9**, 2751.

16 G. Han, Y. Sun, Y. Feng, G. Lin and N. Lu, *Adv. Electron. Mater.*, 2023, **9**, 2300042.

17 M. W. Gaultois, A. O. Oliynyk, A. Mar, T. D. Sparks, G. J. Mulholland and B. Meredig, *APL Mater.*, 2016, **4**, 053213.

18 Y. Katsura, M. Kumagai, T. Kodani, M. Kaneshige, Y. Ando, S. Gunji, Y. Imai, H. Ouchi, K. Tobita, K. Kimura and K. Tsuda, *Sci. Technol. Adv. Mater.*, 2019, **20**, 511–520.

19 O. Sierepeklis and J. M. Cole, *Sci. Data*, 2022, **9**, 648.

20 F. Ricci, W. Chen, U. Aydemir, G. J. Snyder, G.-M. Rignanese, A. Jain and G. Hautier, *Sci. Data*, 2017, **4**, 1–13.

21 H. Wang, S. Bai, L. Chen, A. Cuenat, G. Joshi, H. Kleinke, J. König, H. W. Lee, J. Martin, M.-W. Oh, W. D. Porter, Z. Ren, J. Salvador, J. Sharp, P. Taylor, A. J. Thompson and Y. C. Tseng, *J. Electron. Mater.*, 2015, **44**, 4482–4491.

22 E. Alleno, D. Bérardan, C. Byl, C. Candolfi, R. Daou, R. Decourt, E. Guilmeau, S. Hébert, J. Hejtmánek, B. Lenoir, P. Masschelein, V. Ohorodnichuk, M. Pollet, S. Populoh, D. Ravot, O. Rouleau and M. Soulier, *Rev. Sci. Instrum.*, 2015, **86**, 011301.

23 Y. Zhong, X. Hu, D. Sarker, X. Su, Q. Xia, L. Xu, C. Yang, X. Tang, S. V. Levchenko, Z. Han and J. Cui, *J. Mater. Chem. A*, 2023, **11**, 18651–18659.

24 G. Pizzi, D. Volja, B. Kozinsky, M. Fornari and N. Marzari, *Comput. Phys. Commun.*, 2014, **185**, 422–429.

25 A. Ganose, J. Park, A. Faghaninia, R. Woods-Robinson, K. Persson and A. Jain, *Nat. Commun.*, 2021, **12**, 2222.

26 W. Li, J. Carrete, N. A. Katcho and N. Mingo, *Comput. Phys. Commun.*, 2014, **185**, 1747–1758.

27 J. J. Plata, V. Posligua, A. M. Márquez, J. F. Sanz and R. Grau-Crespo, *Chem. Mater.*, 2022, **34**, 2833–2841.

28 J. Santana-Andreo, A. M. Márquez, J. J. Plata, E. J. Blancas, J.-L. González-Sánchez, J. F. Sanz and P. Nath, *ACS Appl. Mater. Interfaces*, 2024, **16**, 4606–4617.

29 Y.-L. Lee, H. Lee, T. Kim, S. Byun, Y. K. Lee, S. Jang, I. Chung, H. Chang and J. Im, *J. Am. Chem. Soc.*, 2022, **144**, 13748–13763.

30 Z. Hu, W. Wu, Q. Wang and X. Shao, *J. Phys. Chem. C*, 2022, **126**, 12735–12741.

31 W. Li and M. Liu, *ACS Appl. Electron. Mater.*, 2023, **5**, 4523–4533.

32 Y. Li, J. Zhang, K. Zhang, M. Zhao, K. Hu and X. Lin, *ACS Appl. Electron. Mater.*, 2022, **14**, 55517–55527.

33 J. Qu, Y. R. Xie, K. M. Ciesielski, C. E. Porter, E. S. Toberer and E. Ertekin, *npj Comput. Mater.*, 2024, **10**, 58.

34 N. K. Barua, S. Lee, A. O. Oliynyk and H. Kleinke, *ACS Appl. Electron. Mater.*, 2024, **17**, 1662–1673.

35 G. S. Na, S. Jang and H. Chang, *npj Comput. Mater.*, 2021, **7**, 106.

36 X. Jia, Y. Deng, X. Bao, H. Yao, S. Li, Z. Li, C. Chen, X. Wang, J. Mao, F. Cao, J. Sui, J. Wu, C. Wang, Q. Zhang and X. Liu, *npj Comput. Mater.*, 2022, **8**, 34.

37 A. Tukmakova and P. Graziosi, *ACS Appl. Energy Mater.*, 2024, **7**, 10496–10508.

38 S. Anand, K. Xia, V. I. Hegde, U. Aydemir, V. Kocevski, T. Zhu, C. Wolverton and G. J. Snyder, *Energy Environ. Sci.*, 2018, **11**, 1480–1488.

39 H. Luo, J. W. Krizan, L. Muechler, N. Haldolaarachchige, T. Klimczuk, W. Xie, M. K. Fuccillo, C. Felser and R. J. Cava, *Nat. Commun.*, 2015, **6**, 6489.

40 G. Ke, Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, Q. Ye and T.-Y. Liu, *Adv. Neural. Inf. Process Syst.*, 2017, **30**.

41 M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, Y. Jia, R. Jozefowicz, L. Kaiser, M. Kudlur, J. Levenberg, D. Mané, R. Monga, S. Moore, D. Murray, C. Olah, M. Schuster, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. Tucker, V. Vanhoucke, V. Vasudevan, F. Viégas, O. Vinyals, P. Warden, M. Wattenberg, M. Wicke, Y. Yu and X. Zheng, *TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems*, 2015, software available from https://tensorflow.org.

42 F. Chollet, *et al.*, *Keras*, 2015, https://keras.io.

43 G. Rogl and P. Rogl, *Crystals*, 2022, **12**, 1843.

44 MongoDB Inc., *MongoDB, Version 6.0*, https://www.mongodb.com, 2023.

45 G. Landrum, *RDKit: Open-Source Cheminformatics*, 2023, release 2023.03.1, https://www.rdkit.org.

46 *Mendeleev – A Python Resource for Properties of Chemical Elements, Ions and Isotopes, Ver. 1.0.0*, 2014, https://github.com/lmmentel/mendeleev.

47 S. P. Ong, W. D. Richards, A. Jain, G. Hautier, M. Kocher, S. Cholia, D. Gunter, V. L. Chevrier, K. A. Persson and G. Ceder, *Comput. Mater. Sci.*, 2013, **68**, 314–319.

48 F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot and É. Duchesnay, *J. Mach. Learn. Res.*, 2011, **12**, 2825–2830.

49 R. Lehmann, *J. Surv. Eng.*, 2013, **139**, 157.

50 Y. Li, J. Zhang, K. Zhang, M. Zhao, K. Hu and X. Lin, *ACS Appl. Mater. Interfaces*, 2022, **14**, 55517–55527.

51 I. Loshchilov and F. Hutter, *International Conference on Learning Representations*, 2019.

52 G. Kresse and J. Hafner, *Phys. Rev. B:Condens. Matter Mater. Phys.*, 1993, **47**, 558–561.

53 G. Kresse and J. Furthmüller, *Phys. Rev. B:Condens. Matter Mater. Phys.*, 1996, **54**, 11169–11186.

54 P. E. Blöchl, *Phys. Rev. B:Condens. Matter Mater. Phys.*, 1994, **50**, 17953–17979.

55 J. W. Furness, A. D. Kaplan, J. Ning, J. P. Perdew and J. Sun, *J. Phys. Chem. Lett.*, 2020, **11**, 8208–8215.

56 C. E. Calderon, J. J. Plata, C. Toher, C. Oses, O. Levy, M. Fornari, A. Natan, M. J. Mehl, G. L. W. Hart, M. B. Nardelli and S. Curtarolo, *Comput. Mater. Sci.*, 2015, **108**(part A), 233–238.

57 V. Popescu and A. Zunger, *Phys. Rev. Lett.*, 2010, **104**, 236403.

58 B. Zhu, S. R. Kavanagh and D. Scanlon, *J. Open Source Softw.*, 2024, **9**, 5974.

59 F. Eriksson, E. Fransson and P. Erhart, *Adv. Theory Simul.*, 2019, **2**, 1800184.

60 A. Togo, *J. Phys. Soc. Jpn.*, 2023, **92**, 012001.

61 G. Rogl, D. Setman, E. Schafler, J. Horky, M. Kerber, M. Zehetbauer, M. Falmbigl, P. Rogl, E. Royanian and E. Bauer, *Acta Mater.*, 2012, **60**, 2146–2157.

62 D. Zhao, H. Geng and X. Teng, *J. Alloys Compd.*, 2012, **517**, 198–203.

63 A. Serrano, O. Caballero-Calero, C. Granados-Miralles, G. Gorni, C. Manzano, M. Rull-Bravo, A. Moure, M. Martín-González and J. Fernández, *J. Alloys Compd.*, 2023, **931**, 167534.

64 X. Meng, W. Cai, Z. Liu, J. Li, H. Geng and J. Sui, *Am. Mineral.*, 2015, **98**, 405–415.

65 S. Zhang, S. Xu, H. Gao, Q. Lu, T. Lin, P. He and H. Geng, *J. Alloys Compd.*, 2020, **814**, 152272.

66 G. S. Na and H. Chang, *npj Comput. Mater.*, 2022, **8**, 214.

67 D. Chernyavsky, J. van den Brink, G.-H. Park, K. Nielsch and A. Thomas, *Adv. Theory Simul.*, 2022, **5**, 2200351.

68 G. Rogl, A. Grytsiv, E. Royanian, P. Heinrich, E. Bauer, P. Rogl, M. Zehetbauer, S. Puchegger, M. Reinecker and W. Schranz, *Acta Mater.*, 2013, **61**, 4066–4079.

69 G. Rogl, A. Grytsiv, K. Yubuta, S. Puchegger, E. Bauer, C. Raju, R. Mallik and P. Rogl, *Am. Mineral.*, 2015, **95**, 201–211.

70 D.-K. Shin and I.-H. Kim, *J. Electron. Mater.*, 2016, **45**, 1234–1239.

71 J. Prado-Gonjal, M. Phillips, P. Vaqueiro, G. Min and A. V. Powell, *ACS Appl. Energy Mater.*, 2018, **1**, 6609–6618.

72 H. Anno, K. Matsubara, Y. Notohara, T. Sakakibara and H. Tashiro, *J. Appl. Phys.*, 1999, **86**, 3780–3786.

73 X. Zhang, Q. Lu, J. Zhang, Q. Wei, D. Liu and Y. Liu, *J. Alloys Compd.*, 2008, **457**, 368–371.

74 S. Katsuyama, M. Watanabe, M. Kuroki, T. Maehata and M. Ito, *J. Appl. Phys.*, 2003, **93**, 2758–2764.

75 B. Wang, D. Fang, W. Yi, S. Zhao, J. Li, J. Li, Y. Zhao and H. Jin, *Ceram. Int.*, 2021, **47**, 17753–17759.

76 Q. He, S. Hu, X. Tang, Y. Lan, J. Yang, X. Wang, Z. Ren, Q. Hao and G. Chen, *Appl. Phys. Lett.*, 2008, **93**, 042108.

77 S. M. Lundberg and S.-I. Lee, *Adv. Neural Inf. Process. Syst.*, 2017, **30**, 4768–4777.

78 G. Schierning, R. Chavez, R. Schmechel, B. Balke, G. Rogl and P. Rogl, *Transl. Mater. Res.*, 2015, **2**, 025001.

79 Y. Tang, Z. M. Gibbs, L. A. Agapito, G. Li, H.-S. Kim, M. B. Nardelli, S. Curtarolo and G. J. Snyder, *Nat. Mater.*, 2015, **14**, 1223–1228.

80 R. Hanus, X. Guo, Y. Tang, G. Li, G. J. Snyder and W. G. Zeier, *Chem. Mater.*, 2017, **29**, 1156–1164.

81 Y. Wang, J. Wang, X. Xu, Y. Wang, B. Jia, S. Li, K. Nielsch and J. He, *Adv. Energy Mater.*, 2026, **16**, e05077.

82 J. Salvador, J. Yang, H. Wang and X. Shi, *J. Appl. Phys.*, 2010, **107**, 043705.

83 D. Qin, W. Shi, X. Wang, C. Zou, C. Shang, X. Cui, H. Kang, Y. Lu and J. Sui, *Inorg. Chem. Front.*, 2024, **11**, 1724–1732.

84 G. Rogl, A. Grytsiv, E. Bauer, P. Rogl and M. Zehetbauer, *Intermetallics*, 2010, **18**, 394–398.

85 A. Khan, Z. Wang, M. A. Sheikh, D. J. Whitehead and L. Li, *J. Phys. D: Appl. Phys.*, 2010, **43**, 305302.

86 S. Bai, X. Huang, L. Chen, W. Zhang, X. Zhao and Y. Zhou, *Appl. Phys. A*, 2010, **100**, 1109–1114.

87 S. Ballikaya, N. Uzar, S. Yildirim, J. R. Salvador and C. Uher, *J. Solid State Chem.*, 2012, **193**, 31–35.

88 J. Yang, W. Zhang, S. Bai, Z. Mei and L. Chen, *Appl. Phys. Lett.*, 2007, **90**, 192111.

89 W. Li and N. Mingo, *Phys. Rev. B:Condens. Matter Mater. Phys.*, 2014, **89**, 184304.

90 X. Shi, J. Yang, J. R. Salvador, M. Chi, J. Y. Cho, H. Wang, S. Bai, J. Yang, W. Zhang and L. Chen, *J. Am. Chem. Soc.*, 2011, **133**, 7837–7846.