

Cite this: *J. Mater. Chem. C*, 2023,
11, 11643

Experimentally validated machine learning predictions of ultralow thermal conductivity for SnSe materials†

N. K. Barua,[‡] A. Golabek,[‡] A. O. Oliynyk^b and H. Kleinke^{‡*}

Machine-learning (ML) models are used to predict optimal thermoelectric properties for efficient thermoelectric devices. Often, ML models utilize available databases or published sources that might be inconsistent. Herein, we report a boosting ML model – eXtreme gradient boosting (XGBoost) – built from our own lab-generated data with weighted element-to-chemical property features, which predicts the ultralow ($<1 \text{ W m}^{-1} \text{ K}^{-1}$) total thermal conductivity (κ) for p- and n-type doped bulk SnSe materials prior to the synthesis. The metrics of the model included a coefficient of determination (R^2) of 0.94, a root-mean-square error (RMSE) of $\kappa = 0.05 \text{ W m}^{-1} \text{ K}^{-1}$ and a mean absolute error (MAE) of $0.04 \text{ W m}^{-1} \text{ K}^{-1}$ on the validation set using the fivefold cross validation method. The model was able to accurately predict the thermal conductivity values it was trained for, *i.e.*, the Na–Ag–Sn–Se series. The κ values for $\text{Na}_{0.033}\text{Ag}_{0.015-0.016}\text{Sn}_{0.963-0.961}\text{Se}$ were predicted to be $0.54 \text{ W m}^{-1} \text{ K}^{-1}$ on average and experimentally found to be $0.55 \text{ W m}^{-1} \text{ K}^{-1}$. The model also successfully discriminated at low temperatures within the series, with $\text{Na}_{0.033}\text{Ag}_{0.015}\text{Sn}_{0.961}\text{Se}$ predicted to have $\kappa = 0.85 \text{ W m}^{-1} \text{ K}^{-1}$ and measured to have $\kappa = 0.80 (\pm 0.04) \text{ W m}^{-1} \text{ K}^{-1}$, and similarly, $\text{Na}_{0.033}\text{Ag}_{0.016}\text{Sn}_{0.963}\text{Se}$ with a predicted $\kappa = 1.06 \text{ W m}^{-1} \text{ K}^{-1}$ and a measured $\kappa = 0.98 (\pm 0.05) \text{ W m}^{-1} \text{ K}^{-1}$. We pushed the model to the limits to predict the κ values of Cl-doped SnSe, although the training set did not include any κ values with Cl. The predicted and measured values and trends were in good agreement with the RMSE and MAE values achieved by XGBoost's model for this new experimental test dataset and on average agreed with the experimentally determined κ values to be within 9%.

Received 24th April 2023,
Accepted 1st August 2023

DOI: 10.1039/d3tc01450a

rsc.li/materials-c

1. Introduction

Machine learning (ML) is a data-driven method and a branch within artificial intelligence, which utilizes algorithms that can predict, classify, and guide towards the desired outcomes based on the training data. It has found applications in autonomous vehicle technologies,¹ structural biology,² images and voice recognition,^{3,4} and materials science.^{5,6} In materials science, researchers are developing ML methods to analyse data and target novel materials with specific properties. In particular, recent successes include batteries, photovoltaics, and thermoelectric applications.^{7–9} This is due to the advantages the ML

methodology offers in saving time, cost, and resources compared to the usual Edisonian experiments or calculations.^{10,11} Although rare, ML has demonstrated the ability to support experimental validations on predictions outside the expected conditions. For example, Balachandran *et al.*¹² used a two-step ML approach to discover high ferroelectric Curie temperature perovskites supported by experimental validation. Min *et al.*¹³ built an ML model and then validated its predictions experimentally for the optimization of Ni-rich cathode battery materials. Zhuo *et al.*¹⁴ experimentally discovered a phosphor for application in LED devices using ML studies.

In recent decades, the applications of ML in materials science have garnered significant interest to help in improving the efficiency of thermoelectric devices. Optimizing thermoelectric properties such as electrical conductivity (σ), Seebeck coefficient (α), power factor ($\alpha^2\sigma$), and thermal conductivity (κ) to enhance thermoelectric figure-of-merit (zT) through the choice of the best materials in the right stoichiometric proportion forms the crux of the research.^{15–20} In an ideal scenario, high values of σ , α , $\alpha^2\sigma$ and low κ values, especially ultralow κ values below $1.0 \text{ W m}^{-1} \text{ K}^{-1}$ lead to efficient thermoelectric

^a Department of Chemistry, University of Waterloo, 200 University Ave W, Waterloo, ON, Canada. E-mail: kleinke@uwaterloo.ca

^b Department of Chemistry, Hunter College, City University of New York, New York 10065, USA

† Electronic supplementary information (ESI) available: ML workflow map; occurrence of elements in the training set; feature explanation, performance metrics data, scatterplots of test, residual plots, and experimental data. See DOI: <https://doi.org/10.1039/d3tc01450a>

‡ Authors made equal contributions.

devices, governed by $zT = T\alpha^2\sigma\kappa^{-1}$. However, the complicated inter-dependency between α , σ , $\alpha^2\sigma$ and κ presents a challenge to achieve a higher zT value beyond unity, the traditional benchmark for state-of-the-art thermoelectric materials. Few studies have used ML to answer the problems of parallel optimization of thermoelectric properties for higher performance in known materials and search for unknown potential thermoelectric materials by exploring their chemical space of elements, and their respective elemental properties.^{21–24} These reported studies use ML for data obtained from literature or online databases using RESTful API (representational state transfer application programming interface) services by presenting potential thermoelectric properties or candidates using the ML model accuracy. Although presenting an enhanced model accuracy by displaying ML metrics is a prerequisite in published studies, utilizing the model for new experiments is at least equally important. For instance, several studies have demonstrated the prediction of the total κ or the lattice thermal conductivity (κ_l) using a dataset built from the literature. However, most models developed have not been used to validate the results with new experimental data (Table 1).^{25–29} Moreover, these models do not implement the presence of elemental proportions as feature vectors – an elemental vector matrix within a dataset that contains the amount of constituent elements derived from their respective chemical composition – in the final training set. The significance of the vector matrix is vital when using dopants in the training dataset, as high performing thermoelectric materials almost always include various dopants. In a recent report, the high significance of this matrix was demonstrated by the prediction of properties by just using the composition.²¹ However, the vector matrix also has a limitation with the compositional chemical space depending on the elemental prevalence in the training dataset, making it harder to predict properties for compositions containing elements outside the training set.²¹ Besides the length of the training dataset, its diversification with multi-doped materials gives an ML model more confidence in recognizing the underlying complex relationship between the doping on the composition and the target property. So far, only a few studies have demonstrated the benefit of machine learning with the support of experimental evidence for thermoelectric property predictions. For instance, a very recent report by Lee *et al.*³⁰ used the gradient boosted regression tree (GBRT) model to

predict zT with experimentally measured data using elemental properties and electronic structure features followed by utilization of the model's accuracy with the experimentally generated test set. Iwasaki *et al.*³¹ used classical models and a neural network (NN) model³² – a subset of ML that mimics the behaviour of biological neurons in the human brain, to predict the Seebeck coefficient α based on atomic weight, quantum numbers, and lattice mismatch features. Nonetheless, the electronic structure and lattice mismatch features used in this study also require knowledge of the crystal structure, not just the composition with the requirement of an elemental vector matrix. Hou *et al.*³³ used a Gaussian process regression (GPR) model to optimize the power factor of $\text{Al}_2\text{Fe}_3\text{Si}_3$.

Herein, we introduce our ML model developed using consistent experimental data of SnSe-based materials generated in our laboratory. The focus on SnSe instead of other highly efficient thermoelectric materials stemmed from the non-toxic elements and their ultralow thermal conductivity. For example, single crystals of p-type SnSe³⁴ and Br-doped n-type SnSe³⁵ achieved zT peak values of 2.6 at 923 K and 2.8 and 773 K, respectively. Especially once it was realized that contamination of Sn was the main reason for the less stellar performance of polycrystalline samples,³⁶ such samples attained outstanding zT values in excess of 2 too after different purification processes.^{37–39}

We concentrated on thermal conductivity, as this is the best property of this material compared to other state-of-the-art thermoelectric materials, and strongly depends on the impurity oxygen concentration in contrast to the Seebeck coefficient. Moreover, thermal conductivity is a combination of both phonon and electron scattering, while electrical conductivity only depends on the latter. The motivation behind the selection of XGBoost⁴⁰ arose from its built-in regularization to prevent over-fitting, availability of a wide variety of hyperparameters to improve its performance, ability to handle missing values in a dataset and tree-based algorithm that can identify complex trends with faster and good predictions. Moreover, XGBoost also aids in quantifying the significance of features through feature importance to achieve high prediction accuracy. Additionally, we implemented specific features such as elemental vector matrix sensitivity for small doping quantities down to 0.001 atoms per formula unit in our training dataset to describe the elements in differently doped SnSe materials. Ultimately, we experimentally validated the model's prediction of the total thermal conductivity of several hitherto not yet prepared SnSe materials with their synthesis and property measurements.

Table 1 Comparison of the dataset size, model metrics and prediction property of various studies

Author ^a	Model	Dataset size	Dataset source	Test set R^2	Predicted property	Experimental validation
Chen <i>et al.</i> ²⁵	RF	100	Literature	0.93	$\log(\kappa_l)$	No
Bhattacharjee <i>et al.</i> ²⁶	GPR	110	Literature	0.97	$\log(\kappa_l)$	No
Loftis <i>et al.</i> ²⁷	SR ^b	347	Literature	0.92 ^b	κ_l	No
Jaafreh <i>et al.</i> ²⁸	RF	2145	Computational	0.96	κ_l	No
Wang <i>et al.</i> ²⁹	XGBoost	5486	AFLOW database	0.90	κ	No
Lee <i>et al.</i> ³⁰	GBRT	263	Generated in lab, computational	0.74	κ	Yes
Iwasaki <i>et al.</i> ³¹	NN	112	Generated in lab	—	α	Yes
Hou <i>et al.</i> ³³	GPR	—	Generated in lab, literature	0.99	PF ^c	Yes
This work	XGBoost	776	Generated in lab	0.84	κ	Yes

^a Reference number. ^b Symbolic regression. ^c Power factor.

2. End-to-end ML workflow

2.1 Data collection and feature generation

In general, an ML workflow begins with data processing as shown in Fig. S1 (ESI[†]). The data construction for the experimental dataset (EXPD) was initiated by gathering the experimentally generated 799 data from 75 samples with different p- and n-type dopants of SnSe materials synthesized in our laboratory (Table S1, ESI[†]).⁴¹ EXPD includes two groups of features:

(1) Experimental data features (EDF) were composed with the input of synthesis procedures arranged column-wise in the form of features: type of dopant, methods of quenching, melting and reduction processes, ball milling, hot press and annealing conditions, density, and κ values with their related measurement temperatures. These differently doped SnSe materials had their κ values calculated from the thermal diffusivity (D), measured by the TA Instruments LASERFLASH DLF-1 system under an ultra-pure argon atmosphere, the specific heat (C_p) calculated using the Dulong–Petit method, and the density (ρ) determined with the Archimedes method *via* $\kappa = DC_p\rho$. The experimental error of the thermal conductivity measurement was estimated to be 5% based on prior experience.^{42–44} The number of data points collected for D depended on the sample stability with relation to the appearance of bubbles on the surface after the high-temperature diffusivity measurement, surface oxidation, and decomposition of material in pellet form into powder; typically we measured from room temperature to 926 K or less.⁴¹ The selected elements Li, Na, Cu, Ag, Au, Ge, Bi, S, Cl, and Br for this study as dopants for SnSe fell in the category of low environmental impact and medium or high abundance on earth.⁴⁵ The elements Pb, As, Sb, As, and Te were not used in this study because of their toxic nature. The prevalence of the elements in the compositions used in the EXPD is illustrated in Fig. 1. The tabular strategy (see details in Table S2, ESI[†]) gave us the insight to implement the choice of elements for p- and n-type, undoped, doped, or double-doped materials in EXPD. We did not introduce any crystal structure parameters as features in our study as we dealt exclusively with SnSe variants in the two well-established modifications.

(2) Elemental stoichiometry features (ESFs) contain elemental vector matrices as features representing the amounts of elements that project their presence in the chemical composition of doped SnSe materials. The values of the element feature under ESFs

were first derived from their corresponding stoichiometric quantities in the theoretical material composition. The corrected values obtained after weighing the elements replaced these derived values to preserve the accuracy of the data. If a composition did not contain the elements present in the ESF, then the value of these elements under the ESFs was set to zero. Besides ESF and EDF, the construction of weighted element-to-chemical property features was carried out using a separate dataset – the Oliyuk elemental descriptors (OELDs).⁴⁶ OELDs contain detailed information on the chemical and physical properties of elements in the periodic table. These weighted element-to-chemical property features are composed of two groups:

(3) Weighted elemental stoichiometry features (WESFs) were generated using two steps with ESFs and OELDs by NumPy,⁴⁷ a Python library. The first step involved the separation of the composition in EXPD into its constituent elements and stoichiometric proportions. The second step mapped the separated proportions to their related elemental properties – chemical and physical, in OELDs (Table S3, ESI[†]).

(4) Weighted elemental descriptive statistics features (WEDSF) are generated by applying statistics (average, maximum, minimum, sum, difference, and variance) to WESF using NumPy mathematical calculations (Table S4, ESI[†]).⁴⁷

Overall, the EXPD and OELD datasets in Excel sheet format were loaded as two separate Pandas data frames (Python library) using the Jupyter Notebook development environment leveraged by Anaconda open-source distribution for data-preprocessing. The two data frames using NumPy generated WESF and WEDSF data frames. Finally, EXPD combined with WESF and WEDSF led to the creation of the raw training dataset (Table S3, ESI[†]).

2.2 Dataset curation and feature split

Data curation was carried out on the training dataset to ensure that the values are not unrealistic, missing, duplicated, or without a proper numerical format. The data curation process modified the raw training dataset into 776 data and 164 features. The features contained 163 independent variables (as x -vector data) and one dependent variable (as y -vector data) – lowest total thermal conductivity.

2.3 Feature selection

Feature selection is a salient step of the ML workflow for supervised learning (Fig. S1, ESI[†]).⁴⁸ It improves the model performance by suggesting the removal of non-essential features. These non-essential features take additional computational time and create a negative impact on the model with information that is either irrelevant, noisy, or redundant to the dependent variable. Similarly, for our training dataset, not all features can be used by an ML model. Screening of features for relevancy in model selection and evaluation studies is required. Moreover, a general rule of thumb in ML states that the number of features used by an ML model should be approximately 10% of the volume of the training set. Thus, to select important features, we chose two embedded methods for feature selection – L1-norm-regularized linear Regression (LASSO)⁴⁹ and XGBoost feature importance (FeaImpXG).⁴⁰

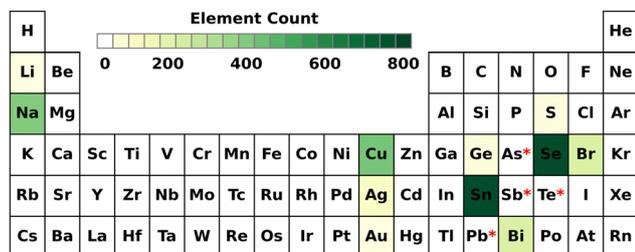


Fig. 1 Heat map representation of the periodic table for the element prevalence in the training data set. * Elements not considered due to their toxic nature.

LASSO is useful when the number of features exceeds the number of training data. The hyperparameter alpha in LASSO is associated with the regularization strength.⁴⁹ The higher the value of alpha, the more aggressive the penalization of the features. Moreover, this regularization prevents the risk of overfitting by reducing the coefficients of the least-predictive variables to zero. Thus, the features were weighted by coefficients in a linear combination for correlation with the dependent variable, κ . The higher the value, positive or negative, the greater the importance of the feature towards the dependent variable. LASSO penalized the less important features – potentially to zero – to reduce their impact on model performance. While the features with a co-efficient of zero were removed, all features with positive and negative correlation coefficients were considered (Fig. 2). Consequently, the features with positive and negative correlation coefficients were prioritized over zero coefficient features.

In FeaimpXG, the model computes the significance of the features through an importance score. The score is based on the features' participation in making key decisions with boosted decision trees. FeaimpXG ranks the features by their relative importance in predicting an outcome variable (Fig. 3 and Fig. S2, ESI[†]). For FeaimpXG, the same hyperparameters were used for both feature selection and model evaluation. This is to maintain the influence of features on the dependent variable. The number of boosting trees, max depth, sub-sample and reg_lambda were set to 100, 3, 0.5 and 3 respectively (Table S5, ESI[†]). The remaining parameters were set to default values.⁵⁰

Even though not all the elements present in EXPD were identified as significant features by the two feature selection methods, we opted to include all the elements that are present in the ESF matrix. This is because even a minor modification in stoichiometry can have a substantial impact on target properties. Additionally, ESF is also useful for differentiating compositions that have the same weighted features or target property values. The two feature selection methods trimmed the training dataset

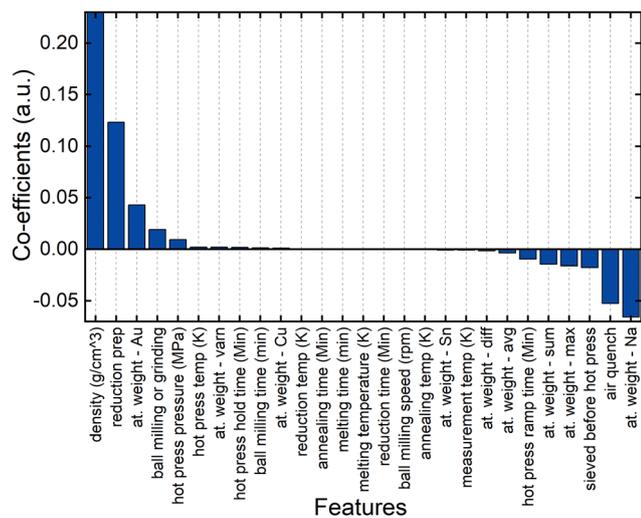


Fig. 2 Important features based on co-efficient score, as suggested by the LASSO feature selection method, along with their correlation with κ .

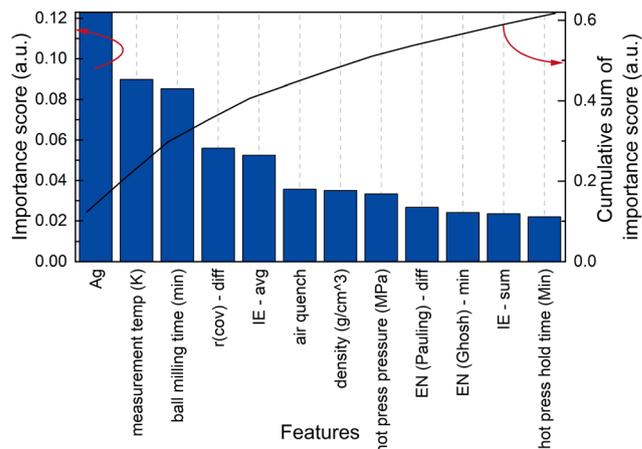


Fig. 3 Blue bars: features ranking based on importance score as suggested by XGBoost. Black curve: cumulative sum of the features. The expanded version of the plot is illustrated in Fig. S2 (ESI[†]).

down to 64 features – 63 independent variables and one dependent variable. Furthermore, the FeaimpXG also provided insights for the hyperparameter schema selection in the XGBoost model during the model selection.

2.4 Model selection and evaluation

In the model selection, we determined the optimal model and its hyperparameters before using it in the test set. We employed four ML algorithms for the k -fold cross-validation – XGboost,⁴⁰ random forest (RF),⁵¹ support vector regressor (SVR),⁵² and k -nearest neighbour (k -NN).⁵³ In XGBoost, it combines the predictions of multiple weak learners to create a more accurate model. In RF, an ensemble of decision trees provides the prediction of the target property through the average of the predictions of all the decision trees. In SVR, a kernel-based method, the extent of error acceptable in the model is defined to find the optimal hyperplane to fit the data for predictions. The SVR algorithm requires a dataset without missing values or noise, whereas ensemble methods such as XGBoost and RF work well with missing values. In the k -NN algorithm, the values are predicted by identifying k nearest neighbors (actual data points) and their corresponding target values. The value of k , an integer, is determined by a Euclidean distance metric that measures the similarity between the new predicted value and the k nearest neighbors. In this method, the algorithm is sensitive to irrelevant features and noisy data, as it assumes equal importance to all features. However, it is easy to implement because only a few parameters are required. To find the optimal model among the four ML algorithms, an evaluation method is essential.

There are a few methods to evaluate the models. However, the k -fold cross-validation method is one of the most robust and preferred methods for evaluating the prediction accuracy of an ML model.⁵⁴ The k -fold library randomly splits the whole training data set into k training and validation sets using a fixed random seed. After the split, the k -fold method trains multiple train-validation sets (Fig. S3, ESI[†]). In our study, the value of k was set to five due to the dataset length. The fivefold cross-validation method picked a unique set of values from the validation sets while

being trained and validated five times. This prevented the validation of data in the training set more than once. Out of the whole dataset, we used only 80% of the training set and 20% of the validation set, using 63 features that were obtained from the feature selection methods.⁵⁵ The hyperparameters used in the algorithms were chosen by following the model documentation page instruction and by the hyperparameter optimization method. These decided hyperparameters consisted of values conservative towards overfitting (Table S6, ESI†). The models from the *k*-fold cross-validation method were evaluated using three metrics: coefficient of determination (R^2), root mean squared error (RMSE) and mean absolute error (MAE) (Table S7, ESI†).⁵⁶

The four ML algorithms demonstrated stable trends of R^2 , RMSE, and MAE across all the folds of the validation sets. This suggests that the models are not overfitting. Thus, the fivefold approach mitigates overfitting, and provides insight into model performance and the limitation of one-sided validation results, namely insufficient training of data.

2.5 Model validation

XGBoost was found to be the optimal model for the test data analysis through the fivefold cross-validation process. Although XGBoost outperformed RF, SVR and *k*-NN in validation sets, the underperformed models were still considered for test data. The test dataset contained seven dopants with multiple measurement temperatures, bringing the count of the dataset to 57 samples, as illustrated in Table 2. The tuning of the hyperparameters was not performed beyond the model selection process to avoid p-hacking, which is an unethical human operation where the evaluation of the model is performed on held-out test data to pick ideal model hyperparameters to achieve the desired statistical results. The models' hyperparameters remained fixed during the model validation. Furthermore, the test set was kept separately and used after each model had the optimal hyperparameter tuning achieved in the model selection step.⁵⁷ The performance metrics of all models used for the test set are captured in Table S8, ESI†.

3. Results and discussion

Based on the knowledge of properties that influence κ , we chose electronegativity, ionization energy, total valence electrons, atomic weight, and covalent radius as elemental properties.

Table 2 Materials used for experimental validation, along with the predicted and measured κ values around 300 K

ID	SnSe material	Predicted κ ($\text{W m}^{-1} \text{K}^{-1}$)	Measured ^c κ ($\text{W m}^{-1} \text{K}^{-1}$)
C1	$\text{Na}_{0.032}\text{Au}_{0.015}\text{Sn}_{0.963}\text{Se}^a$	0.99	0.98 ± 0.05
C2	$\text{Na}_{0.033}\text{Ag}_{0.015}\text{Sn}_{0.961}\text{Se}^a$	0.80	0.85 ± 0.04
C3	$\text{Na}_{0.033}\text{Ag}_{0.016}\text{Sn}_{0.963}\text{Se}^a$	0.87	0.94 ± 0.04
C4	$\text{Na}_{0.034}\text{Au}_{0.015}\text{Sn}_{0.959}\text{Se}^a$	0.93	0.96 ± 0.05
C5	$\text{SnSe}_{0.940}\text{Cl}_{0.060}^b$	0.93	1.01 ± 0.05
C6	$\text{Sn}_{0.999}\text{Se}_{0.940}\text{Br}_{0.060}^b$	0.89	0.78 ± 0.04
C7	$\text{Sn}_{1.002}\text{Se}_{0.900}\text{Br}_{0.100}^b$	0.97	0.91 ± 0.04

^a p-Type dopant. ^b n-Type dopant. ^c Measured κ values with an estimated experimental error of 5%.

These values of the elemental properties were referred from OELDs. The elemental properties chosen were limited to six because having more properties will lead to more features and could be detrimental to model performance or could result in overfitting.

Among the various electronegativity scales available, Pauling and Ghosh were chosen. We avoided the inclusion of electronegativity scales derived from elemental properties already present as features in our dataset. For instance, the Mulliken scale is calculated from ionization energies, Allred–Rochow from covalent radii, and Allen from valence electrons. Moreover, the features of electronegativity, valence electrons, covalent bonding, and ionization energy influence the phonon transport which then impacts κ . The atomic weight of the elements or the average molar mass of the composition has a strong impact as well. After the consolidation of the features in the benchmark dataset, feature selection methods were applied.

Applying the two feature selection methods to the benchmark training dataset allowed us to pick important features relevant to this study. In the LASSO method, experimental features such as density, reduction temperature, an atomic weight of Na and Au, maximum and a sum of atomic weight, and sieved before hot press (Fig. 2) were selected. The FeaimpXG method along with the cumulative sum method applied to importance scores of features ranked Ag, ball milling time, measurement temperature, the difference of maximum and minimum covalent radius and average ionization potential as significant features (Fig. 3). Prior to the application of the feature selection methods, based on the scientific intuition, we expected that atomic weight difference, density, cation/anion atomic-% ratio, and calculated heat capacity based on Dulong–Petit would support the importance of mass in influencing κ . To our surprise, except for the calculated heat capacity the rest of the expected features were not considered as important features by the feature selection methods to predict κ . This demonstrates how we can learn from ML besides the domain knowledge of chemistry to understand the factors affecting the prediction of target properties. Thus, the combined results of the two feature selection methods narrowed our selection of features to air and water quench, ball milling speed and time, number of reductions, annealing time and temperature, melting temperature and time, hot press temperature, pressure and ramp time, cation/anion atomic-% ratio, p-type/n-type, weighted minimum, average, difference, variance and sum of Pauling electronegativity, weighted average, difference and minimum of Ghosh electronegativity, weighted sum, minimum and average of ionization energy, weighted minimum, average, difference and variance of molar mass, weighted sum and minimum of total valence electrons and weighted sum, average, minimum, difference and variance of covalent radius (Table S3, ESI†). The selected features depend on the experimental parameters and elemental properties that affect κ values.

With the selected features in place, we moved on to the model selection method. For this method, XGBoost, RF, *k*-NN, and SVR were the chosen candidates, as each of them functions with a unique algorithmic complexity to perform predictions.^{40,51–53}

We did not choose linear regression due to poor metrics (*e.g.*, $R^2 = 0.32$ for the test set, Fig. S4, ESI†) during the preliminary stage of the ML workflow development and did not choose neural networks due to the quantity of the dataset being below the order of thousands. We noticed that XGBoost yielded the best results within the training set (Fig. 4 and Fig. S5, ESI†).

Most of the data were found along the ideal fit line, with a couple of the data points slightly deviating. The most underestimated values occurred at the lowest end of the measured values for the $\text{Cu}_{0.022}\text{Sn}_{0.980}\text{Se}$ and $\text{Na}_{0.012}\text{Cu}_{0.008}\text{Ge}_{0.025}\text{Sn}_{0.994}\text{S}_{0.05}\text{Se}_{0.95}$ samples. The underestimation (highest positive relative error) was from 16 to 20%; however, considering that these data were some of the lowest thermal conductivity values, we found the discrepancies acceptable. The most overestimated sample was $\text{Cu}_{0.022}\text{Sn}_{0.980}\text{Se}$, measured at the highest temperature of 926 K. The error of 70% for this sample with the predicted $\kappa = 0.19 \text{ W m}^{-1} \text{ K}^{-1}$ vs. $0.11 \text{ W m}^{-1} \text{ K}^{-1}$ is at least in part understandable based on the ultralow absolute values, as the difference was only $0.08 \text{ W m}^{-1} \text{ K}^{-1}$. Furthermore, the highlighted $\text{Cu}_{0.011}\text{Sn}_{0.9991}\text{Se}$ at 521 K had a 38% overestimation error, while $\text{Na}_{0.012}\text{Cu}_{0.008}\text{Sn}_{0.979}\text{Se}$ and $\text{Cu}_{0.062}\text{Sn}_{0.941}\text{Se}$ were underestimated with a 32% and a 20% error at 680 K and 875 K, respectively. These deviations could be attributed to the presence of Cu in the composition, which was underrepresented in our dataset.

Thereafter, we moved to fivefold cross-validation. After this, the models were tested for signs of over-fitting. A major point to note was that there were studies that report only the average metrics using *k*-fold cross-validation. However, it is essential to show and verify how each fold independently performed. In our *k*-fold study not only did we report the average metrics of R^2 ,

MAE, and RMSE of each model in the fivefold cross-validation, but also included the metrics for each fold. Thus, our *k*-fold had five training and validation sets with randomized data for each model. It also gave us confidence that our model mitigated over-fitting as all the five sets chosen gave consistent output in the metrics (Table S7, ESI†).

All datapoints from the fivefold validation sets were grouped and plotted into a parity plot. The fivefold study showed that the XGBoost performed higher than the rest of the models. We believe that the over- or underestimated predictions of some doped SnSe materials data points with the linear fit in Fig. 5 are likely due to factors beyond our control such as the stability of the samples during the measurements.⁴¹ For instance, the most overestimated sample was $\text{Na}_{0.010}\text{Ag}_{0.011}\text{Sn}_{0.974}\text{Se}$ at 324 K and 774 K with a 24% and 28% error, respectively, followed by $\text{Cu}_{0.022}\text{Sn}_{0.980}\text{Se}$, $\text{Cu}_{0.08}\text{Sn}_{0.919}\text{Se}$ and $\text{Sn}_{1.002}\text{Se}$ at room temperatures of 775 K, and 873 K and 825 K at relative errors between 23% and 30%. The next group of underestimated samples belonged again to the Cu-doped SnSe materials such as $\text{Cu}_{0.005}\text{Sn}_{0.998}\text{Se}$, $\text{Cu}_{0.022}\text{Sn}_{0.980}\text{Se}$, $\text{Cu}_{0.011}\text{Sn}_{0.991}\text{Se}$, and $\text{Na}_{0.012}\text{Cu}_{0.008}\text{Sn}_{0.979}\text{Se}$. Interestingly, we observed that $\text{Cu}_{0.022}\text{Sn}_{0.980}\text{Se}$ and $\text{Cu}_{0.011}\text{Sn}_{0.991}\text{Se}$ formed a bubble after the measurement explaining the over- and underestimation by the model. We have additionally presented the residual plots in Fig. S6 (ESI†) to identify these problematic samples.

The selected models were evaluated using the metrics R^2 , RMSE, and MAE. In the validation set, the models – in the order of XGBoost, RF, *k*-NN, and SVR – demonstrated fivefold cross-validation average R^2 values of 0.94, 0.84, 0.85, and 0.79, RMSE

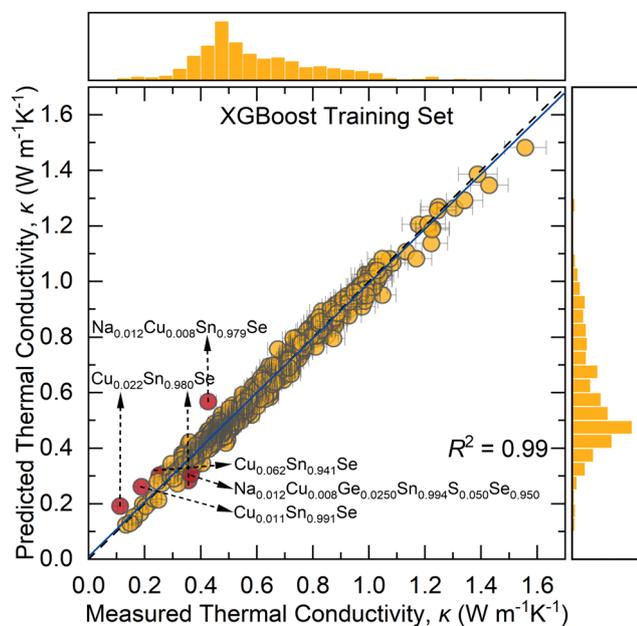


Fig. 4 XGBoost model predicted vs. measured κ values of the training set with R^2 score. Histogram plot on the top: distribution of the datapoints for measured κ . Histogram plot on the right: population of the datapoints for predicted κ . Solid blue line: linear fit; dashed black line: ideal fit. Experimental error bars: estimated 5% error.

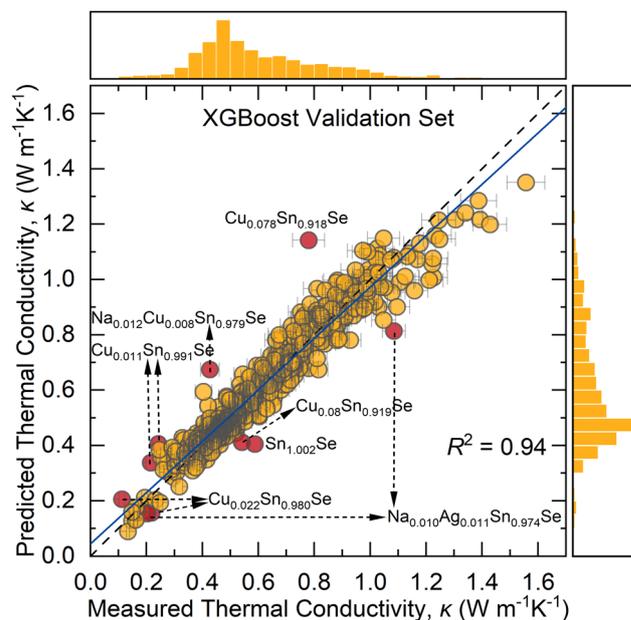


Fig. 5 XGBoost model predicted vs. measured κ values of the combined five validation sets with R^2 score, using a fivefold cross validation method. Histogram plot on the top: distribution of the datapoints for measured κ . Histogram plot on the right: population of the datapoints for predicted κ . Solid blue line: linear fit; dashed black line: ideal fit. Experimental error bars: estimated 5% error.

values of 0.05, 0.09, 0.09, and $0.10 \text{ W m}^{-1} \text{ K}^{-1}$ and MAE values of 0.04, 0.06, 0.07, and $0.07 \text{ W m}^{-1} \text{ K}^{-1}$, respectively (Table S7, ESI†).⁵⁶ The overall metrics indicated XGBoost to be the best model to be used in the experimental test set (Fig. 5 and Fig. S7, ESI†). Furthermore, the consistent metrics for XGBoost across all the folds negated the possibility of one-sided biased model performance.

3.1 Model prediction on the new test set with experimental validation

In the experimental model validation, we tested the model performance by comparing the κ prediction results of the test data with the experimental κ data. To test our model with new experimental results, we prepared seven new p- and n-type SnSe materials illustrated in Table 2 (C1–C7) and determined κ at about eight different temperatures. For the n-type, we used Br as before, plus Cl to challenge the model to predict κ data without having first encountered κ values from that dopant, as Cl was not included in the training set. The Br-doped composition of C6 was used with different synthesis conditions than in the training set, to investigate whether the model can predict κ for such a change too.

After the synthesis of C1–C7 was carried out and their respective κ data experimentally determined, the latter were compared to the predicted values from the test set of all the models. As expected from the model selection, the performance metrics (Table S8, ESI†) showed that XGBoost performed significantly better than the other models. Moreover, Fig. 6 illustrated that the XGBoost model not only performed well on the validation data but also showed a similar dominant performance

when applied to the new experimental test data (Fig. S8, ESI†). In Fig. 6, the overestimation of data points mainly belonged to Br-doped SnSe (C6) at 372 K and 472 K, which was due to the difference in synthetic route in the test set compared to the same composition present in the training dataset. We also found that the κ value of (Na, Ag)-double-doped SnSe (C2) at medium temperature was overestimated with a 27% percent error. The data points with high κ values fit better in the blue model line of fit. However, the samples with κ values of $0.4\text{--}0.9 \text{ W m}^{-1} \text{ K}^{-1}$ fit better with the dashed line (ideal fit). From Fig. 7, 8 and Fig. S9, S10 (ESI†), it can be observed how the models understood the trend of κ changes with temperature. However, with Cl-doped SnSe (C5), the model slightly departs with the trend of κ with temperature as observed in other compositions in the test set because the model has not seen any Cl-containing samples in the training dataset. Even then, the model made accurate predictions for C5 throughout the temperature series – with exceptions of overestimation at 872 K and an underestimation at 921 K (Fig. 6). The latter is likely caused by potential instabilities beyond the phase transition from the orthorhombic *Pnma* space group to the orthorhombic *Cmcm* space group occurring around 810 K.⁵⁸

Fig. 7 illustrates the superiority of XGBoost compared to RF, SVR and *k*-NN using the example $\text{Na}_{0.033}\text{Ag}_{0.016}\text{Sn}_{0.963}\text{Se}$ (C3). XGBoost made outstanding predictions of κ values even beyond 800 K, after the phase transition of SnSe,⁵⁸ including the downward trend. The factors that affect the phase transition such as pressure, temperature, and chemical composition were already present as features (Table S5, ESI†). Using this observation, we can suggest that ML can be used in the areas where we lack scientific methods of understanding the properties in the phase transition region. Studying the properties with small

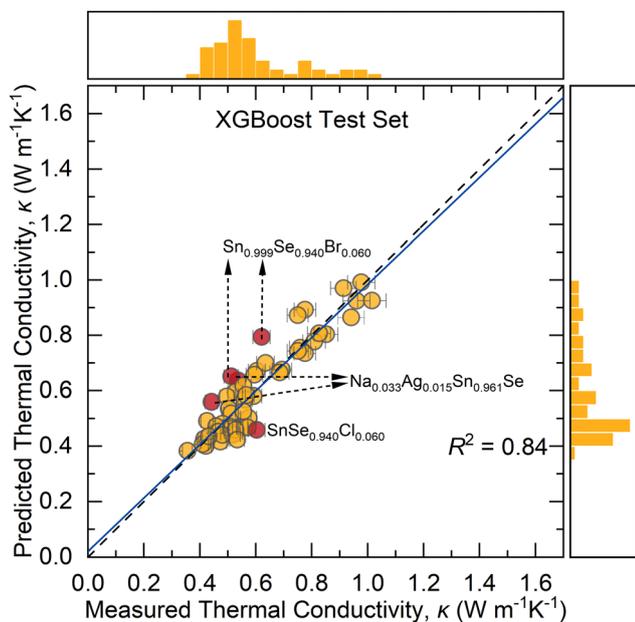


Fig. 6 XGBoost model predicted vs. measured κ values of the experimental test set with R^2 score. Histogram plot on the top: distribution of the datapoints for measured κ . Histogram plot on the right: population of the datapoints for predicted κ . Solid blue line: linear fit; dashed black line: ideal fit. Experimental error bars: estimated 5% error.

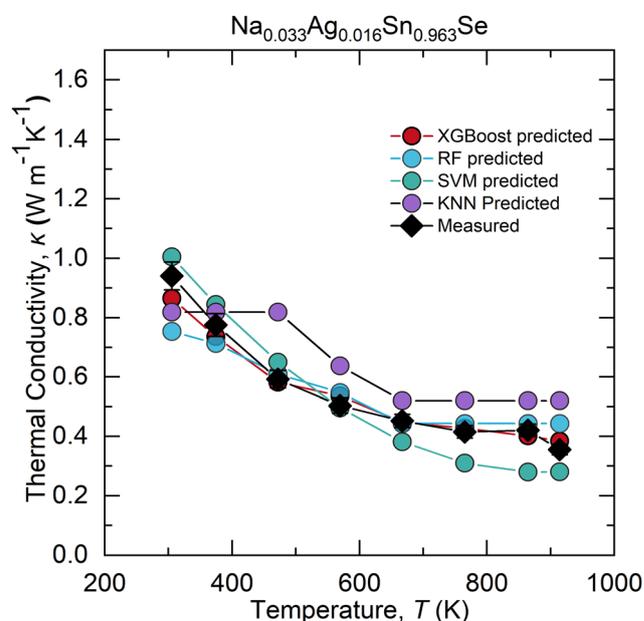


Fig. 7 Temperature-dependent comparison between predicted and experimentally measured κ of $\text{Na}_{0.033}\text{Ag}_{0.016}\text{Sn}_{0.963}\text{Se}$ (C3) using XGBoost, RF, SVR, and *k*-NN models. Experimental error bars: estimated 5% error.

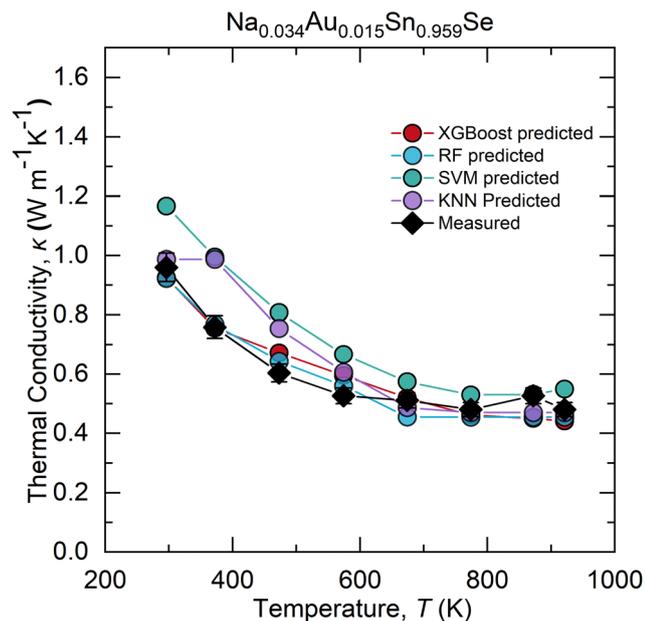


Fig. 8 Temperature-dependent comparison between predicted and experimentally measured κ of $\text{Na}_{0.034}\text{Au}_{0.015}\text{Sn}_{0.959}\text{Se}$ (C4) using XGBoost, RF, SVM, and k -NN models. Experimental error bars: estimated 5% error.

doping amounts at the transition temperature with the use of density functional theory is expensive and nearly impossible.¹¹ Moreover, molecular dynamics might not be ideal to study properties around the transition temperature.⁵⁹ Thus, this is an area where ML methods have the potential to shine over these traditional computational methods.

Similarly, the predicted and measured κ values of $\text{Na}_{0.034}\text{Au}_{0.015}\text{Sn}_{0.959}\text{Se}$ (C4) are compared in Fig. 8, again demonstrating the outstanding predictions by XGBoost, including the region above 800 K. Moreover, XGBoost correctly predicted higher values for Na, Au doping than for Na, Ag doping, something that may be counterintuitive based on the general trend of heavier elements causing lower thermal conductivity.

3.2 Comparison of ML prediction with experimental studies on $\text{Sn}_{0.999}\text{Se}_{0.94}\text{Br}_{0.06}$ and $\text{SnSe}_{0.94}\text{Cl}_{0.06}$

The n-doped SnSe materials C5 and C6 with the same amount of 6% of two different group 17 elements were chosen for a comparative study of measured and predicted κ values. Without ML, it would be difficult to predict whether the Cl-doped C5 will exhibit higher or lower thermal conductivity because of (at least) two opposing trends: the lower mass of Cl compared to Br might cause higher κ , while in turn its larger mass difference to the Se atom it replaced caused larger mass fluctuation and thus potentially lower κ . The different electronegativity and hence character of the Sn–Cl bonds as well as different sizes further complicated matters. As these parameters were all part of the data given to the XGBoost model, we anticipated it could correctly predict changes without having encountered κ data for any Cl-doped case. Indeed, as shown in Fig. 9, XGBoost made the correct prediction of higher κ values for the Cl-doped

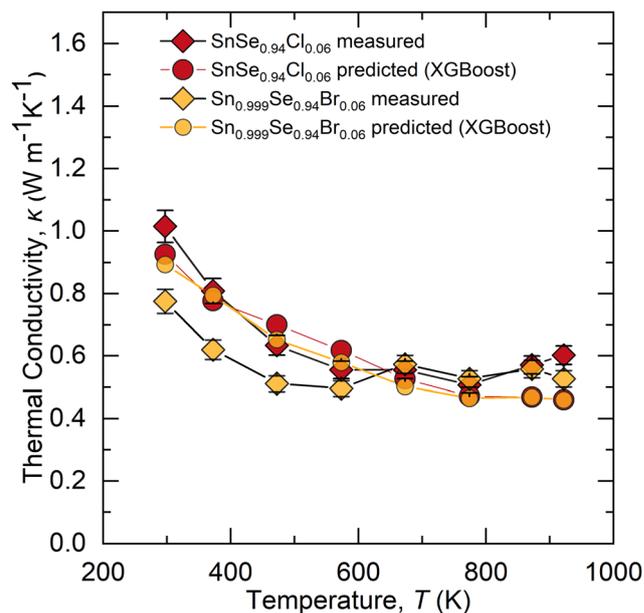


Fig. 9 Temperature-dependent comparison of $\text{SnSe}_{0.94}\text{Cl}_{0.06}$ (C5) and $\text{Sn}_{0.999}\text{Se}_{0.94}\text{Br}_{0.06}$ (C6) for measured and predicted κ . Experimental error bars: estimated 5% error.

C5. This confirms that XGBoost correctly utilized the differences in the elemental features of Cl and Br in its predictions (Fig. S10, ESI[†]).

3.3 Final validation by enlarging the size of the test data

To alleviate potential concerns that the test size was too small⁶⁰ with 57 data, amounting to 7% of the sum of the training and validation data (621 + 155 = 776), we randomly picked eleven samples from the training and validation data and moved them to the test set. This brought the new test size up to 154 data (T2), *i.e.* 23% of the sum of the training and validation data (543 + 136 = 679).

This procedure did not change XGBoost's performance in the training set with still $R^2 = 0.99$ and caused a small increase in R^2 of the test set from 0.84 to 0.86 (Fig. S11, ESI[†]).

4. Conclusions

In conclusion, we developed an ML model trained to predict ultralow thermal conductivity from our dataset of doped SnSe materials all synthesised in our lab. The feature selection methods aided in choosing the right features that enhanced the performance of the ML models. The model performance metrics through fivefold cross-validation guided us to select XGBoost as the superior model for the experimental validation study. Not only did XGBoost predict the data overall correctly, but it was also able to handle the hitherto unencountered use of Cl as a dopant, including predicting the difference between Cl- and Br-doping.

Furthermore, we presented here two important observations: first, we noticed the feature selection methods provided a precise selection of significant features from a wide variety of

features to explain the prediction of κ through chemistry. Second, the model even made accurate predictions beyond the phase transition temperature. Thus, our study showed how we can learn from ML to expand on the domain knowledge, besides presenting the model performance.

In addition, we believe that our work provides a promising strategy and has a broader impact on developing ML models using quality experimental data with the introduction of the elemental vector matrix features to assist in the prediction of optimal properties of thermoelectric materials.

Ultimately, ML models using more comprehensive data bases are envisioned to streamline the exploratory work in the thermoelectric community, leading to the discovery of advanced materials. Of course, one will have to pay attention to not include erroneous data, which in turn could create false predictions.

Author contributions

As per CRediT (Contributor Roles Taxonomy) – N. K. Barua: conceptualization, visualization, software (ML development), data analysis, and writing (original draft, review, and editing); A. Golabek: data curation (synthesis, generation of training and test data), conceptualization, dataset preparation, and data analysis; A. O. Oliynyk: data analysis, visualization, manuscript and code-review; H. Kleinke: funding acquisition, supervision, manuscript (review and editing).

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

Financial support from the Natural Sciences and Engineering Research Council of Canada in the form of a Discovery Grant (RGPIN-2020-04145) is appreciated.

References

- 1 Y. Almalioglu, M. Turan, N. Trigoni and A. Markham, *Nat. Mach. Intell.*, 2022, **4**, 749–760.
- 2 J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Židek, 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.
- 3 Y. Yu, J. Li, S. A. Solomon, J. Min, J. Tu, W. Guo, C. Xu, Y. Song and W. Gao, *Sci. Robot.*, 2022, **7**, eabn0495.
- 4 Z. Zhou, K. Chen, X. Li, S. Zhang, Y. Wu, Y. Zhou, K. Meng, C. Sun, Q. He, W. Fan, E. Fan, Z. Lin, X. Tan, W. Deng, J. Yang and J. Chen, *Nat. Electron.*, 2020, **3**, 571–578.
- 5 C. Gao, X. Min, M. Fang, T. Tao, X. Zheng, Y. Liu, X. Wu and Z. Huang, *Adv. Funct. Mater.*, 2022, **32**, 2108044.
- 6 K. Choudhary, B. DeCost, C. Chen, A. Jain, F. Tavazza, R. Cohn, C. W. Park, A. Choudhary, A. Agrawal, S. J. L. Billinge, E. Holm, S. P. Ong and C. Wolverton, *npj Comput. Mater.*, 2022, **8**, 59.
- 7 C. Lv, X. Zhou, L. Zhong, C. Yan, M. Srinivasan, Z. W. Seh, C. Liu, H. Pan, S. Li, Y. Wen and Q. Yan, *Adv. Mater.*, 2022, **34**, 2101474.
- 8 M. Srivastava, J. M. Howard, T. Gong, M. Rebello Sousa Dias and M. S. Leite, *J. Phys. Chem. Lett.*, 2021, **12**, 7866–7877.
- 9 T. Wang, C. Zhang, H. Snoussi and G. Zhang, *Adv. Funct. Mater.*, 2020, **30**, 1906041.
- 10 K. Pels, P. Dickson, H. An and T. Kodadek, *ACS Comb. Sci.*, 2018, **20**, 61–69.
- 11 J. Peng, D. Schwalbe-Koda, K. Akkiraju, T. Xie, L. Giordano, Y. Yu, C. J. Eom, J. R. Lunger, D. J. Zheng, R. R. Rao, S. Muy, J. C. Grossman, K. Reuter, R. Gómez-Bombarelli and Y. Shao-Horn, *Nat. Rev. Mater.*, 2022, **7**, 991–1009.
- 12 P. V. Balachandran, B. Kowalski, A. Sehirlioglu and T. Lookman, *Nat. Commun.*, 2018, **9**, 1668.
- 13 K. Min, B. Choi, K. Park and E. Cho, *Sci. Rep.*, 2018, **8**, 15778.
- 14 Y. Zhuo, A. Mansouri Tehrani, A. O. Oliynyk, A. C. Duke and J. Brgoch, *Nat. Commun.*, 2018, **9**, 4377.
- 15 J. He and T. M. Tritt, *Science*, 2017, **357**, eaak9997.
- 16 L. R. Macario, Y. Shi, P. Jafarzadeh, T. Zou, J. B. Kycia and H. Kleinke, *ACS Appl. Mater. Interfaces*, 2019, **11**, 45629–45635.
- 17 P. Jafarzadeh, M. Oudah, A. Assoud, N. Farahi, E. Müller and H. Kleinke, *J. Mater. Chem. C*, 2018, **6**, 13043–13048.
- 18 Q. Guo, A. Assoud and H. Kleinke, *Adv. Energy Mater.*, 2014, **4**, 1400348.
- 19 G. J. Snyder and E. S. Toberer, *Nat. Mater.*, 2010, **7**, 101–110.
- 20 S. Sun, M. Li, X. L. Shi and Z. G. Chen, *Adv. Energy Mater.*, 2023, **13**, 2203692.
- 21 A. Y. T. Wang, S. K. Kauwe, R. J. Murdock and T. D. Sparks, *npj Comput. Mater.*, 2021, **7**, 1–10.
- 22 R. E. A. Goodall and A. A. Lee, *Nat. Commun.*, 2020, **11**, 6280.
- 23 L. Ward, A. Agrawal, A. Choudhary and C. Wolverton, *npj Comput. Mater.*, 2016, **2**, 1–7.
- 24 K. T. Schütt, H. E. Saucedo, P. J. Kindermans, A. Tkatchenko and K. R. Müller, *J. Chem. Phys.*, 2018, **148**, 241722.
- 25 L. Chen, H. Tran, R. Batra, C. Kim and R. Ramprasad, *Comput. Mater. Sci.*, 2019, **170**, 109155.
- 26 D. Bhattacharjee, K. Kundavu, D. Saraswat, P. R. Raghuvanshi and A. Bhattacharya, *ACS Appl. Energy Mater.*, 2022, **5**, 8913–8922.
- 27 C. Loftis, K. Yuan, Y. Zhao, M. Hu and J. Hu, *J. Phys. Chem. A*, 2021, **125**, 435–450.
- 28 R. Jaafreh, Y. S. Kang and K. Hamad, *ACS Appl. Mater. Interfaces*, 2021, **13**, 57204–57213.
- 29 X. Wang, S. Zeng, Z. Wang and J. Ni, *J. Phys. Chem. C*, 2020, **124**, 8488–8495.
- 30 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.
- 31 Y. Iwasaki, I. Takeuchi, V. Stanev, A. G. Kusne, M. Ishida, A. Kirihara, K. Ihara, R. Sawada, K. Terashima, H. Someya, K. ichi Uchida, E. Saitoh and S. Yoroazu, *Sci. Rep.*, 2019, **9**, 2751.

- 32 A. C. Mater and M. L. Coote, *J. Chem. Inf. Model.*, 2019, **59**, 2545–2559.
- 33 Z. Hou, Y. Takagiwa, Y. Shinohara, Y. Xu and K. Tsuda, *ACS Appl. Mater. Interfaces*, 2019, **11**, 11545–11554.
- 34 L.-D. Zhao, S.-H. Lo, Y. Zhang, H. Sun, G. Tan, C. Uher, C. Wolverton, V. P. Dravid and M. G. Kanatzidis, *Nature*, 2014, **508**, 373–377.
- 35 C. Chang, M. Wu, D. He, Y. Pei, C. F. Wu, X. Wu, H. Yu, F. Zhu, K. Wang, Y. Chen, L. Huang, J. F. Li, J. He and L. D. Zhao, *Science*, 2018, **360**, 778–783.
- 36 Y. K. Lee, Z. Luo, S. P. Cho, M. G. Kanatzidis and I. Chung, *Joule*, 2019, **3**, 719–731.
- 37 W. Wei, C. Chang, T. Yang, J. Liu, H. Tang, J. Zhang, Y. Li, F. Xu, Z. Zhang, J. F. Li and G. Tang, *J. Am. Chem. Soc.*, 2018, **140**, 499–505.
- 38 S. Chandra and K. Biswas, *J. Am. Chem. Soc.*, 2019, **141**, 6141–6145.
- 39 C. Zhou, Y. K. Lee, Y. Yu, S. Byun, Z. Z. Luo, H. Lee, B. Ge, Y. L. Lee, X. Chen, J. Y. Lee, O. Cojocaru-Mirédin, H. Chang, J. Im, S. P. Cho, M. Wuttig, V. P. Dravid, M. G. Kanatzidis and I. Chung, *Nat. Mater.*, 2021, **20**, 1378–1384.
- 40 T. Chen and C. Guestrin, *Proc. ACM SIGKDD Int. Conf. Knowl. Discov. Data Min.*, 2016, pp. 785–794.
- 41 A. J. Golabek, MS Thesis, Univ. Waterloo, Canada, 2023.
- 42 H. Wang, W. D. Porter, H. Böttner, J. König, L. Chen, S. Bai, T. M. Tritt, A. Mayolett, J. Senawiratne, C. Smith, F. Harris, P. Gilbert, J. W. Sharp, J. Lo, H. Kleinke and L. I. Kiss, *J. Electron. Mater.*, 2013, **42**, 1073–1084.
- 43 D. C. Ramirez, L. R. Macario, X. Cheng, M. Cino, D. Walsh, Y. C. Tseng and H. Kleinke, *ACS Appl. Energy Mater.*, 2020, **3**, 2130–2136.
- 44 C. Sturm, L. R. Macario, T. Mori and H. Kleinke, *Dalton Trans.*, 2021, **50**, 6561–6567.
- 45 B. Elzey, N. Whitehead, V. Norman, C. M. Babyak, J. T. Morningstar, D. Pollard and S. O. Fakayode, *Int. J. Environ. Anal. Chem.*, 2017, **97**, 573–587.
- 46 S. Lee, C. Chen, G. Garcia and A. Oliynyk, *ChemRxiv*, 2023, DOI: [10.26434/chemrxiv-2023-0nlzl](https://doi.org/10.26434/chemrxiv-2023-0nlzl).
- 47 C. R. Harris, K. J. Millman, S. J. van der Walt, R. Gommers, P. Virtanen, D. Cournapeau, E. Wieser, J. Taylor, S. Berg, N. J. Smith, R. Kern, M. Picus, S. Hoyer, M. H. van Kerkwijk, M. Brett, A. Haldane, J. F. del Río, M. Wiebe, P. Peterson, P. Gérard-Marchant, K. Sheppard, T. Reddy, W. Weckesser, H. Abbasi, C. Gohlke and T. E. Oliphant, *Nature*, 2020, **585**, 357–362.
- 48 S. S. Young, F. Yuan and M. Zhu, *Mol. Inf.*, 2012, **31**, 707–710.
- 49 R. Shrinkage, *J. R. Stat. Soc. Ser. B*, 1996, **58**, 267–288.
- 50 H. Zheng, J. Yuan and L. Chen, *Energies*, 2017, **10**, 1168.
- 51 L. Breiman, *Mach. Learn.*, 2001, **45**, 5–32.
- 52 B. E. Boser, I. M. Guyon and V. N. Vapnik, *Proc. Fifth Annu. ACM Work. Comput. Learn. Theory*, 1992, pp. 144–152.
- 53 T. M. Cover and P. E. Hart, *IEEE Trans. Inf. Theory*, 1967, **13**, 21–27.
- 54 T. T. Wong, *Pattern Recognit.*, 2015, **48**, 2839–2846.
- 55 Z. Xiong, Y. Cui, Z. Liu, Y. Zhao, M. Hu and J. Hu, *Comput. Mater. Sci.*, 2020, **171**, 109203.
- 56 D. Chicco, M. J. Warrens and G. Jurman, *PeerJ Comput. Sci.*, 2021, **7**, e623.
- 57 A. Y. T. Wang, R. J. Murdock, S. K. Kauwe, A. O. Oliynyk, A. Gurlo, J. Brgoch, K. A. Persson and T. D. Sparks, *Chem. Mater.*, 2020, **32**, 4954–4965.
- 58 M. Sist, J. Zhang and B. B. Iversen, *Acta Crystallogr., Sect. B: Struct. Sci., Cryst. Eng. Mater.*, 2016, **72**, 310–316.
- 59 N. J. English and C. J. Waldron, *Phys. Chem. Chem. Phys.*, 2015, **17**, 12407–12440.
- 60 Y. Xu and R. Goodacre, *J. Anal. Test.*, 2018, **2**, 249–262.