



OPEN

Descriptor selection for predicting interfacial thermal resistance by machine learning methods

Xiaojuan Tian¹✉ & Mingguang Chen²✉

Interfacial thermal resistance (ITR) is a critical property for the performance of nanostructured devices where phonon mean free paths are larger than the characteristic length scales. The affordable, accurate and reliable prediction of ITR is essential for material selection in thermal management. In this work, the state-of-the-art machine learning methods were employed to realize this. Descriptor selection was conducted to build robust models and provide guidelines on determining the most important characteristics for targets. Firstly, decision tree (DT) was adopted to calculate the descriptor importances. And descriptor subsets with topX highest importances were chosen (topX-DT, X = 20, 15, 10, 5) to build models. To verify the transferability of the descriptors picked by decision tree, models based on kernel ridge regression, Gaussian process regression and K-nearest neighbors were also evaluated. Afterwards, univariate selection (UV) was utilized to sort descriptors. Finally, the top5 common descriptors selected by DT and UV were used to build concise models. The performance of these refined models is comparable to models using all descriptors, which indicates the high accuracy and reliability of these selection methods. Our strategy results in concise machine learning models for a fast prediction of ITR for thermal management applications.

Interfacial thermal resistance (ITR) plays an important role for thermal management of ultra-fast electronics and thermoelectric materials^{1–6}. When heat is transferred through an interface, temperature discontinuity exists. The ratio of the temperature discontinuity to the heat flux through the interface is named as ITR. There is a miniaturization trend of electronic devices in recent decades. In terms of nanostructured devices, the ITR become a dominant factor for device performance because the phonon mean free paths are larger than the characteristic length scales under such circumstance⁷. Screening materials with desired ITR is significant for electronics fabrication. For example, materials system with low ITR helps to reduce the energy consumption of electronics, while high ITR materials system is required for excellent thermoelectrics. There are a great number of factors affecting ITR, such as the intrinsic properties of materials and their differences, roughness of surfaces, crystal impurities, binding energy and thickness of films etc^{8,9}. Thus, the accurate prediction of ITR is a high-dimensional problem and difficult to solve with regular mathematic equations.

Traditional models predicting ITR include acoustic mismatch model (AMM) and diffuse mismatch model (DMM)¹⁰. AMM assumes that there is no scattering of photons at the interface, which works well only under ideal conditions at low temperature. The DMM model is built based on the assumption of complete elastic diffusing mismatch, which makes it not suitable for non-elastic circumstances. An improved model named scattering-mediated acoustic mismatch model (SMAMM) incorporates phonon scattering into the original AMM and realizes prediction of ITR in a wide temperature range¹¹. Still, the prediction accuracy of SMAMM model is restricted by the Debye approximation and the reliability of experimental data used to fit the parameters¹². Besides, Foygel's model based on Monte Carlo simulations and percolation theory has been widely adopted to predict the thermal conductivity in carbon nanotube composites^{13–17}. This model simplifies nanotubes as penetrable, rigid and straight cylinders and ignores the waviness and 3D entanglement of carbon nanotubes^{18–20}. Molecular dynamics (MD) simulation is also applied to ITR prediction. Originally, MD simulation is used to analyze the physical movements of atoms and molecules. When it comes to a system of interacting particles, the system properties are predicted by numerically solving empirical or semi-empirical equations defined by classical Newton's law of motion. The interactive forces between different particles are calculated following a potential function (eg. Lennard–Jones potential, tight-binding potential), which is an approximate function at a certain level of accuracy. The results of MD simulation are valid only when the input atomic interactions are

¹Department of Chemical Engineering, China University of Petroleum, Beijing 102249, China. ²Physical Science and Engineering Division, King Abdullah University of Science and Technology (KAUST), Thuwal 23955-6900, Saudi Arabia. ✉email: tian@cup.edu.cn; mingguang.chen@kaust.edu.sa

consistent with the forces in real situations. In some simple cases, this assumption can be fulfilled by carefully select the potential functions. For example, Lennard–Jones potential can be selected for the non-bonded interaction between two particles²¹, while other potentials or methods such as embedded atom model²², environment dependent interatomic potential²³, or tight-binding second moment approximation potentials²⁴ can be adopted for many-body systems. However, it can be extremely hard to mimic forces between real atoms when quantum effects^{25,26}, time²⁷ and size limitations^{28,29} need to be taken into account in biologically important processes. Besides, MD simulation is computationally expensive and time-consuming, which limits its applications as screening tools for specific materials. Lately, machine learning methods have been applied to predict composite thermal conductivity, ITR between graphene and boron nitride, and thermoelectric conversion efficiency^{30–38}. Specifically, Xu group⁸ applied machine learning algorithms as regression tree ensembles of LSBoost, support vector machines, and Gaussian regress processes to build ITR prediction models. Descriptors with a total amount of 35 including property descriptors, compound descriptors, and process descriptors were selected as input. All three models show better prediction accuracy than traditional AMM and DMM, which indicates the prospect of machine learning methods for predicting physical properties. However, it's still very hard for researchers to consider all the 35 descriptors when designing thermal management systems with novel materials. In light of this, we focused on evaluating the 35 descriptors further by machine learning methods and screening minimum but most significant descriptors for ITR prediction.

For data set with modest size, descriptor selection is critical for reaching a robust machine learning model and provide insight on which characteristics are most important for the target^{39,40}. In this work, descriptor selection was firstly conducted according to their importances calculated by decision tree (DT). The importances are the scores assigned to each input feature of a predictive model that indicates its relative contribution to the predicted results. And descriptors with topX (X = 20, 15, 10, 5) highest importances were selected (topX-DT). To verify the transferability of the selected subsets, kernel ridge regression (KRR), Gaussian process regression (GPR) and K-nearest neighbors (KNN) algorithms were used to build models besides DT. R² and root-mean-squared-error (RMSE) of models built from descriptors subsets by all three algorithms were calculated. The metrics for model evaluation were acquired from shuffled and grouped cross-validation. Datasets were randomly split under shuffled cross-validation. Considering identical interface system may exist in both validation set and training set when shuffled, datasets were also grouped by substrate/interlayer/film system to exclude the potential interference on feature importance. It is shown that the performance of all algorithms are stable with descriptor size decreasing to top10-DT. DT has a relatively good performance even when the descriptor size reduces to top5-DT, while the performance of KRR, GPR and KNN is not satisfying. To obtain a more reliable feature subset, univariate selection (UV) was introduced. And the subset selected by UV is named as topX-UV. As a result, there are 15 common descriptors selected by both top20-DT and top20-UV (Top15-DTUV). Meantime, 5 common descriptors exist in both top10-DT and top10-UV (Top5-DTUV). It is confirmed that the model performance is more robust under descriptors selected both by DT and UV than that from descriptors only picked by DT. Besides, descriptors selected by DT and UV has a high overlap with the descriptors used for AMM and DMM and factors testified from previous experimental studies. Thus, the selected descriptors work well for building machine learning models and are valid on the physical point of view. Descriptor selection methods presented in this work are transferrable to predict other materials properties beyond ITR.

Methods

Dataset collection. Original dataset for this study was the experimental data collected from 85 published papers. Xu group organized them and introduced descriptors for predicting ITR by machine learning method^{8,9}. Details of the developed descriptors and collected ITR were explained in the previous work⁴¹. And data could be found in the file named “training dataset for ITR prediction.xlsx” and downloaded directly from <https://doi.org/10.5281/zenodo.3564173>.

Dataset preprocessing. Descriptors were scaled before feeding into models. According to distribution of descriptors, min–max scale and standard scale were applied. Min–max scale is to transform features by scaling each feature to a given range, e.g. between zero and one. For each descriptor, min–max scale is conducted by the following equation, where X.max and X.min are the maximum and minimum value of the descriptor.

$$X_{\text{scaled}} = \frac{X - X_{\text{min}}}{X_{\text{max}} - X_{\text{min}}}$$

The descriptors transformed by min–max scaler include fthick, fmelt, fdensity, sdensity, fAC1x, fAC1y, fAC2x, fAC2y, fIPc, fIPa, smelt, sAC1x, sAC1y, sAC2x, sAC2y, sIPc, and sIPa.

Standard scale is to standardize features by removing the mean and scaling to unit variance. Centering and scaling happen independently on each feature by computing the relevant statistics on the samples in the training set. For each descriptor, standard scale is conducted by the following equation, where μ and s are the mean and standard deviation of the descriptor.

$$X_{\text{scaled}} = \frac{X - \mu}{s}$$

The descriptors transformed by standard scaler are T, fmass, fEb, sEb, and smass.

Algorithms and models. Decision trees (DT) are a non-parametric supervised learning method for classification and regression⁴². It creates a model to predict target by learning with a set of simple if–then–else rules.

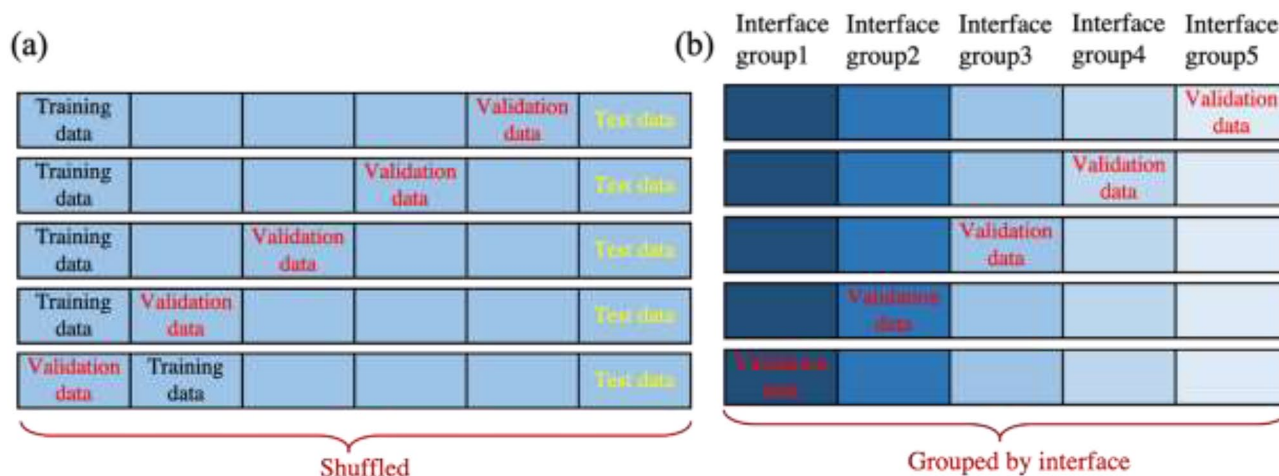


Figure 1. Schemes for (a) shuffled cross-validation and (b) grouped cross-validation.

It is a white box model simple to understand, interpret, and visualize. A representative decision tree algorithm is classification and regression tree (CART), introduced by Leo Breiman⁴³. CART is based on a binary recursive partitioning procedure. The objective of partitioning is to minimize dissimilarity in the terminal nodes for classification and mean-squared-error for regression. The dissimilarity is measured by the loss functions, typically Gini index or cross-entropy for classification trees^{44,45}. For the regression trees applied in our work, each partitioning is made to maximize the reduction in root-mean-squared-error (RMSE)⁴⁶. After all the partitioning has been done, a decision tree is obtained where each branch is a split in a predictor and each end node gives a prediction for the outcome variable. The feature importances in CART could be determined in one shot during training, which is computationally efficient compared with greedy search methods.

KRR is an algorithm combining Ridge regression (linear least squares with l2-norm regularization) with the “kernel trick”^{47,48}. Actually, KRR is a special case of support vector regression. It takes advantage of integral operator kernel functions to map principal components in high-dimensional feature spaces to input space nonlinearly^{49–51}. Radial basis function (RBF) was applied as kernel in our work.

GPR implements Gaussian processes for regression purposes. It can find a probabilistic distribution of new output given the training data and new input data^{52–55}. Both KRR and GPR learn a target function by the “kernel trick”. However, KRR learns a linear function in the space induced by the respective kernel which corresponds to a non-linear function in the original space⁴⁹. While, GPR uses the kernel to define the covariance of a prior distribution over the target functions and uses the observed training data to define a likelihood function. Here, we also applied radial basis function (RBF) as the kernel for the GPR⁵⁶.

Besides descriptor importances from decision tree, feature selection was conducted by selecting the best descriptors based on the UV statistical tests for dimensionality reduction purpose. Here, F-test was adopted to estimate the degree of linear dependency between descriptor and target^{57,58}. Briefly, F-test of equality of variances is a test for the null hypothesis that two normal populations have the same variance⁵⁹. In this situation, F value is the ratio of descriptor variance over target variance. It has an F-distribution if the null hypothesis of equality of variances is true. If F value is either too large or too small, the null hypothesis will be rejected^{60,61}. The built-in function `f_regression` of sklearn library computes the correlation between the descriptor and target, and converts it to an F value automatically. Then the F values are used for descriptor selection.

Algorithm evaluation. Models built by different algorithms and descriptor subsets were evaluated by R^2 and RMSE.

R^2 computes the coefficient of determination. It is calculated by

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - u_i)^2}{\sum_{i=1}^n (y_i - \bar{u})^2}.$$

And RMSE is calculated by

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (y_i - u_i)^2}{n}}$$

where n , y_i , u_i , and \bar{u} are number of data, experimental ITR, predicted ITR, and average experimental ITR values, respectively.

The datasets were handed by shuffled cross-validation and grouped cross-validation for training and model evaluation, as seen in Fig. 1. In terms of the shuffled cross-validation, the original dataset was split into training and cross-validation set (80%) and holdout set (20%) randomly. Models were built by training set and optimal hyperparameters were picked through grid search with fivefold cross-validation. The holdout set was seen and used only once for model evaluation. Thus, the holdout set was named as test data. Under circumstances that

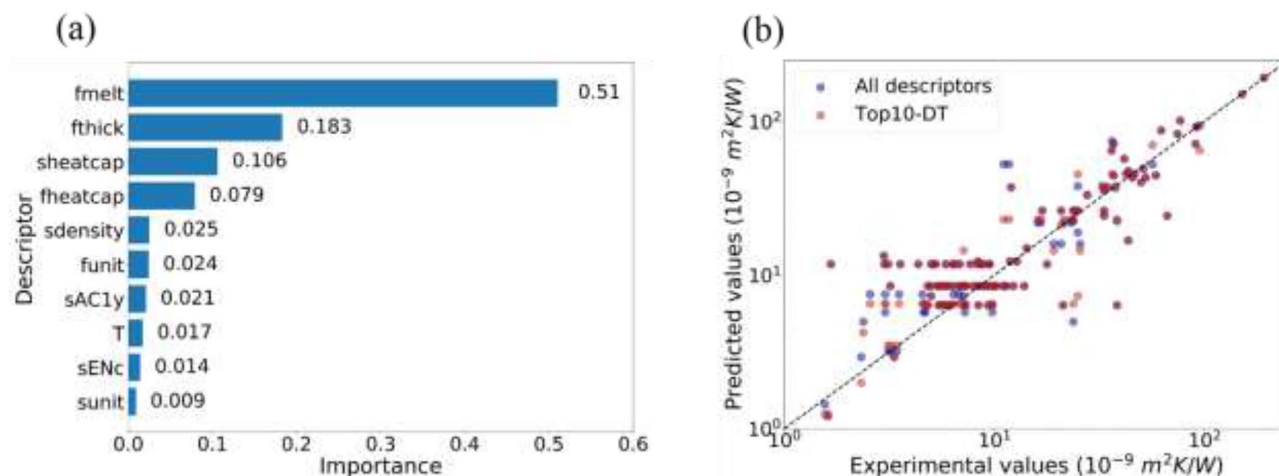


Figure 2. (a) Descriptors with the highest 10 importances (Top10 descriptors). (b) Correlation between the experimental values and values predicted by all descriptors (violet dots) and top10 descriptors (red dots).

the train and test set follow the same probability distribution, the holdout method can provide the most accurate metrics for unseen data, since the metrics obtained from validation set contain bias from hyperparameter optimization⁶². Besides shuffled cross-validation, the dataset was grouped by unique interfaces (film-interlayer-substrate). Every group contain $\sim 20\%$ dataset with some specific interfaces, which is different among these groups. Thus, no identical interface exists in more than one groups. In such case, data among these groups may follow different probability distribution. So the fivefold cross-validation was applied to evaluate model performance. R^2 and RMSE were used as metrics in cross-validation.

Table S1 and S2 in supporting information summarize the grid search space and final hyperparameters picked for various models. Please refer to the github link (<https://github.com/descriptor-selection-ITR/Descriptor-Selection-for-Predicting-Interfacial-Thermal-Resistance-by-Machine-Learning-Methods>) for more details. Fivefold cross-validation was selected because our original dataset was less than 1000 samples. For such small datasets, fivefold cross-validation generally gives better results. Lower fold cross-validation can't train the models well, while higher fold cross-validation allocates few data to test set, making testing results not representative. R^2 and RMSE of predictions from test set were used to evaluate the performance of models.

All analysis were conducted in Scikit-learn package⁶³. The StandardScaler and MinMaxScaler package was used for data preprocessing, DecisionTreeRegressor, KernelRidge, and GaussianProcessRegressor package for the three algorithms mentioned above, and SelectKBest and $f_{\text{regression}}$ for univariate descriptor selection.

Results and discussion

Descriptors selected by decision tree. Dataset was treated firstly by shuffled cross-validation with an ideal assumption that all data follow the same probabilistic distribution, as seen in Fig. 1a. Descriptor selection plays a critical role in building robust and computationally-cheap models. For dataset whose size is not large, descriptor selection is helpful to prevent overfitting and provide insight into which properties are most important for targets. Here, decision tree (DT) was applied to train an ITR prediction model and get the descriptor importances. Figure 2a shows the descriptors with the top10 highest importances (Top10-DT), which occupy a total importance of more than 98%. Among them, the film melting point has a high importance of 51%, and the top4 descriptors possess importance around 88%. Table 1 presents all the descriptors and their corresponding importances. Interestingly, only 20 out of 35 descriptors are selected by decision tree, indicating the existence of uninformative inputs. The descriptors for traditional AMM and DMM include temperature, density, speed of sound (longitudinal and transverse), and unit cell volume. It is worth noting that temperature, density and unit cell volume are all in the Top10-DT. Meanwhile, speed of sound (longitudinal and transverse) has a Pearson correlation coefficient as high as 0.71 with the melting point⁹, while melting point is the most important descriptor according to decision tree. Therefore, useful descriptors confirmed by AMM and DMM are all selected as important descriptors by decision tree successfully. As shown in Fig. 2a, heat capacity and film thickness also act as significant descriptors. The relationship between film thickness and ITR has been observed by experiments and simulations in previous studies^{64,65}. The reason that heat capacity was selected is attributed to the relationship between heat capacity and density. Figure 2b shows the correlation between experimental values and predicted values of test data from DT. It is observed that there are same predicted values for multiple experimental data (horizontal series of data in Fig. 2b). This phenomena occurs since decision tree takes the mean of samples located at the same leaf node as prediction. Thus, the data assigned to the same leaf node has the same predicted value. It is indicated that the DT built from all descriptors and top10-DT have comparable performance.

To verify the transferability of the descriptors selected by DT, kernel ridge regression (KRR), Gaussian process regression (GPR) and K-nearest neighbors (KNN) models were also built under different descriptor subsets. These subsets were named as top x -DT, presenting the descriptors with top x highest importances from DT, as shown in Table 1. Here, top20-DT, top15-DT, top10-DT, and top5-DT were applied as inputs together with all descriptors. Shuffled cross-validation was applied here. R^2 and RMSE of the test data (holdout set) served as

	Descriptor	Importance	Remarks
1	fmelt	5.10×10^{-1}	
2	fthick	1.83×10^{-1}	
3	sheatcap	1.06×10^{-1}	
4	fheatcap	7.94×10^{-2}	
5	sdensity	2.49×10^{-2}	
6	funit	2.41×10^{-2}	
7	sAC1y	2.14×10^{-2}	
8	T	1.66×10^{-2}	
9	sENc	1.37×10^{-2}	
10	sunit	8.93×10^{-3}	
11	interlayer	4.50×10^{-3}	
12	fAC1x	3.07×10^{-3}	interlayer: 1 (exist) or 0 (absent)
13	fEb	3.02×10^{-3}	T: temperature (K)
14	sAC2x	5.73×10^{-4}	fthick: film thickness (nm)
15	fIPa	5.34×10^{-4}	The following columns are labeled for film and substrate by f and s in the front, respectively. (e.g. fheatcap and sheatcap)
16	sIPc	4.48×10^{-4}	heatcap: specific heat capacity (J/gK)
17	fENc	2.33×10^{-4}	melt: melting point (K)
18	sR1	3.47×10^{-5}	density (g/cm ³)
19	smelt	2.92×10^{-5}	unit: Volume per formula unit (10 ⁻²⁹ m ³ /f.u.)
20	fdensity	1.65×10^{-7}	R1: atomic ratio of the first element
21	fmass	0	R2: atomic ratio of the second element
22	sEb	0	AC: AC represents atomic coordinates defined from the periodic table. The group as the x coordinate and the period as the y coordinate as (ACix, ACiy), where i represents the order of the elements of the compound
23	smass	0	ENc: electronegativity for cation
24	fAC1y	0	ENa: electronegativity for anoin
25	fAC2x	0	IPc: ionic potential for cation
26	fAC2y	0	IPa: ionic potential for anion
27	fIPc	0	Eb: binding energy (eV/f.u.)
28	sAC1x	0	mass (u)
29	sAC2y	0	
30	sIPa	0	
31	fr1	0	
32	fr2	0	
33	fENa	0	
34	sR2	0	
35	sENa	0	
Total		1	

Table 1. Descriptor importances from decision tree.

the metrics for model evaluation. It is believed that the performance on holdout set is the most close to that of unseen data, since hyperparameter optimization may result in overfitting to validation set. Commonly, a higher R^2 and lower RMSE indicate a better performance. The R^2 and RMSE for both training set and test set could be found in Table S3 in supporting information. As seen in Fig. 3, DT with all descriptors shows a R^2 of 0.85 and a RMSE of 11, which is comparable to the previous results⁸. Notably, the performance of DT doesn't degrade with the reduction of descriptors. In terms of KRR and GPR, the performance is as good as DT until top10-DT. When the descriptors size decreases further to 5, the performance of KRR and GPR degrades sharply. The performance of KNN model is not as good as the others. Overall, the top10-DT have a total importance of more than 98%, which include the properties used for AMM and DMM. These 10 descriptors have a good transferability from DT to other machine learning models, such as KRR and GPR.

Descriptors selected by univariate testing. To cross validate the descriptors selected by decision tree, univariate selection (UV) was applied. The UV is a totally different algorithms compared with decision tree selection. It filters descriptors based on statistical test. In this work, F-test estimating the degree of linear dependency between descriptor and targets was used. 10 and 20 out of 35 total descriptors were selected by univariate testing, as shown in Table S4 in supporting information. Figure 4 is the Venn diagram showing the amount of common descriptors for top 20 and top 10 descriptors selected by DT and UV. Obviously, there are 15 common descriptors out from 20 picked by both DT and UV. 10 descriptors are never selected. For the top10-DT and

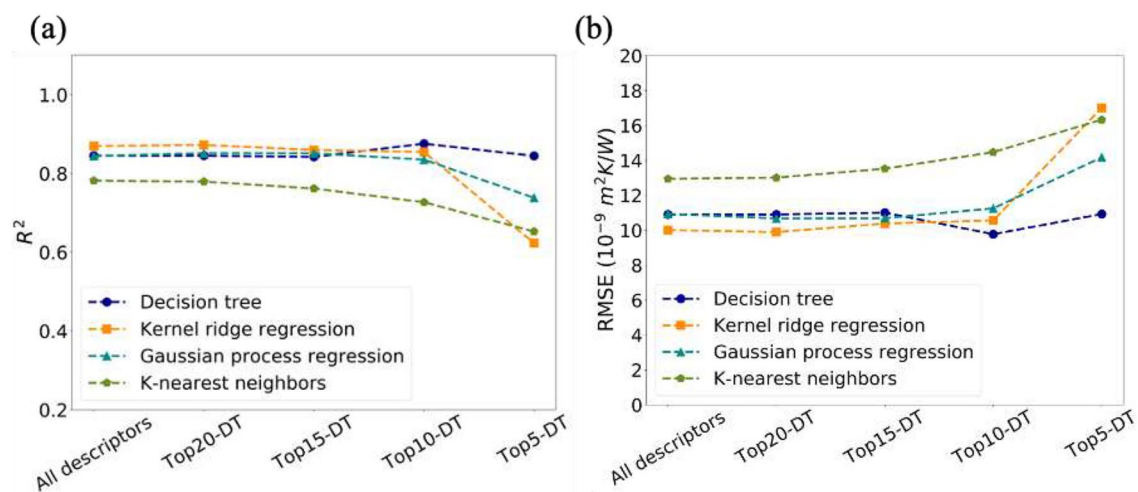


Figure 3. R^2 (a) and RMSE (b) of test data (holdout set) predicted by models built by different descriptors subsets under shuffled cross-validation.

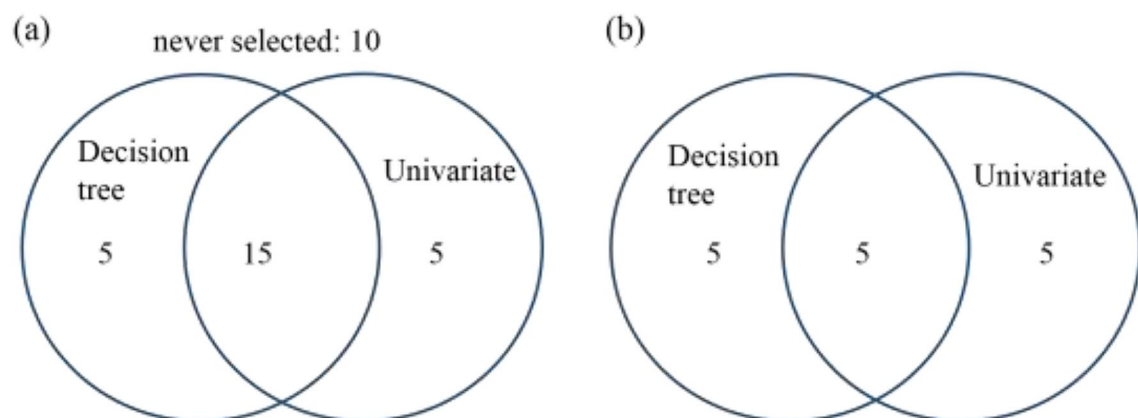


Figure 4. Venn diagram showing the amount of common descriptors picked for both decision tree and univariate testing.

	Descriptors
Top15-DTUV	fmelt, fthick, funit, fheatcap, fdensity, fEb, fENc, fAC1x, T, smelt, sunit, sheatcap, sAC1y, sENc, sIPc
Top5-DTUV	fmelt, fheatcap, funit, sheatcap, sENc

Table 2. Common descriptors selected by decision tree and univariate testing.

top10-UV, there are 5 common. The details of the 15 and 5 common descriptors are shown in Table 2. The 5 common descriptors for top10-DT and top10-UV include the melting point, heat capacity, unit and electronegativity.

Performance of models built by the 15-common descriptors (Top15-DTUV) and 5-common (Top5-DTUV) descriptors were evaluated by R^2 and RMSE under shuffled cross-validation. (Fig. 5) The same as the previous part, performance here is for the test data (holdout set). And train set performance is listed in Table S5 in supporting information. Unlike the descriptors selected by DT only, the descriptor reduction conducted by both DT and UV show a much more stable performance. The R^2 of KRR improves from 0.62 to 0.77 by applying the top5-DTUV instead of top5-DT. And the R^2 of GPR and R^2 of KNN improves from 0.74 to 0.82 and from 0.65 to 0.78 by the same way. At the same time, the RMSE of KRR, GPR and KNN decreases with the utilization of top5-DTUV. Therefore, descriptors selected by a combination of decision tree and univariate testing are more reliable than that selected by only one algorithm.

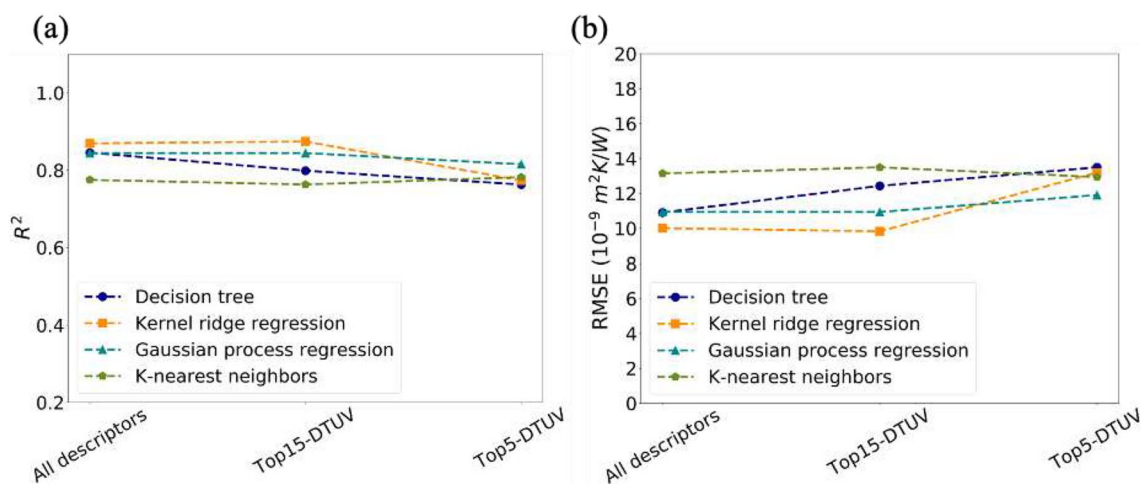


Figure 5. R^2 (a) and RMSE (b) of test data (holdout set) predicted by models built by all descriptors, 15 common descriptors, and 5 common descriptors under shuffled cross-validation.

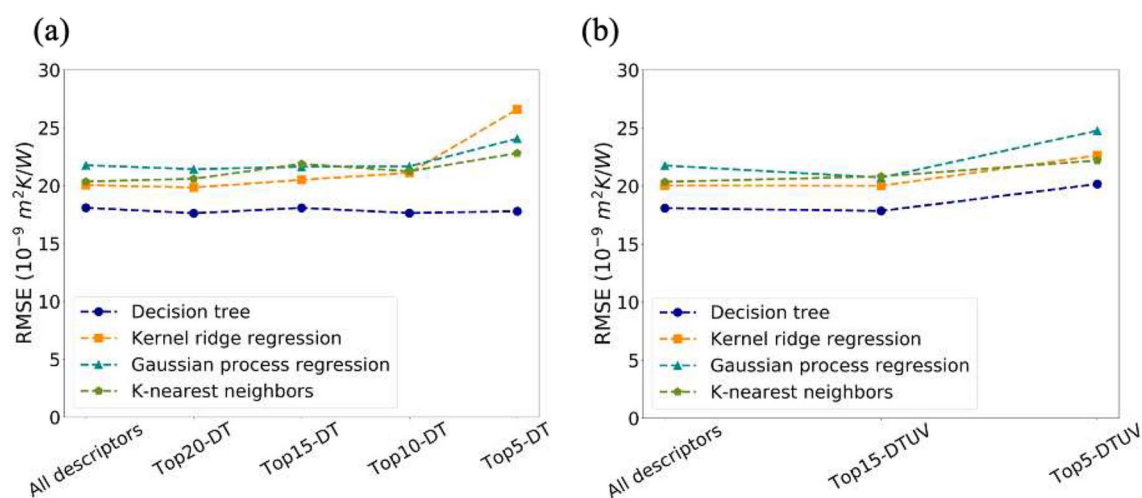


Figure 6. (a) RMSE of validation set by models built by descriptors selected by decision tree under grouped cross-validation. (b) RMSE of validation set by models built by common descriptors selected by decision tree and univariate testing under grouped cross-validation.

Model performance by grouped cross-validation. Beside shuffled cross-validation applied above, grouped cross-validation was investigated. The dataset was split by substrate/interlayer/film systems to guarantee that no identical interface exists in more than one group. Every group includes several distinct interface systems, $\sim 20\%$ of which serves as the validation set. In this case, it is not appropriate to have holdout set since no individual group can represent the others. Figure 6 shows the performance of models built by different descriptors set. The models RMSE for grouped cross-validation is not as good as that of random validation, which is not surprising since the information from many interfaces in validation set is not seen and learnt by machine learning models. Figure 6 shows RMSE values here are around $17 \sim 25$, still lower than that from AMM and DMM models, which are 121 and 91^9 , respectively, confirming the superiority of our models built by grouped cross-validation.

As shown in Fig. 6, DT is best and the most robust among these models. For KNN and KRR, the descriptors selected by both DT and UV show much stable performance than descriptors selected by DT only, which is consistent with the conclusions drawn from shuffled cross-validation. In sum, shuffled cross-validation and grouped cross-validation were both performed in our work. In both systems, the common descriptors from decision tree and univariate testing are more reliable than that selected by only one algorithm. And the selection methods result in concise models with relatively good performance but much lower dimensions.

Although machine learning processes are hard to be understood intuitively, our findings (top5-DTUV) can be explained by classic theories and are well supported by experimental results documented in the literature. For example, melting point directly affect Lennard–Jones interatomic potential by the following equations in molecular simulations^{66,67}:

$$T_m \propto \varepsilon$$

$$4\varepsilon\left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^6\right]$$

where T_m is the melting point, ε is the well depth, σ is the distance at which the intermolecular potential equals to 0, r is the real distance of both particles. Additionally, phonon transport is a dominating mode in thermal transport of nanostructured devices. Capacity (C_V) is a key value when calculating phonon mean free path (l_{ph}) based on kinetic theory⁶⁸:

$$l_{ph} = \frac{3k_L}{C_V v_m}$$

where v_m is the average sound speed, k_L is the lattice thermal conductivities. In other words, the effect of heat capacity on ITR is realized by affecting the phonon transport in nanostructured devices. The effect of electronegativity on ITR varies in different material systems. Thus, the analysis has to be on a case-by-case basis. Overall, electronegativity is always one of the most important factors in both theoretical and experimental exploration of ITR among various material systems^{69–71}.

Conclusions

In conclusion, descriptors selection for ITR prediction was conducted utilizing machine learning methods. Decision tree and univariate testing were applied to determine the important descriptors. Decision tree, kernel ridge regressor, Gaussian process regressor, and K-nearest neighbors were utilized to build models. Dataset was treated by shuffled cross-validation and grouped cross-validation. Performance of different algorithms and descriptors subsets were evaluated by R^2 and RMSE. All models demonstrated relatively good performance when reducing all descriptors to top10-DT, indicating the validity of these selected descriptors. Furthermore, the 5 common descriptors selected both by top10-DT and top10-UV have a higher prediction accuracy than descriptors selected only by DT. These descriptors selected by machine learning methods based on big data collected from real experiments agree with properties affecting ITR heavily from a physical point of view. The characteristic selection methods by machine learning algorithms can not only be used for ITR prediction but also determining important descriptors for other materials properties.

Received: 12 September 2020; Accepted: 28 December 2020

Published online: 12 January 2021

References

- Evans, W. *et al.* Effect of aggregation and interfacial thermal resistance on thermal conductivity of nanocomposites and colloidal nanofluids. *Int. J. Heat Mass Tran.* **51**, 1431–1438 (2008).
- Nan, C. W., Birringer, R., Clarke, D. R. & Gleiter, H. Effective thermal conductivity of particulate composites with interfacial thermal resistance. *J. Appl. Phys.* **81**, 6692–6699 (1997).
- Pei, Q., Zhang, Y., Sha, Z. & Shenoy, V. B. Carbon isotope doping induced interfacial thermal resistance and thermal rectification in graphene. *Appl. Phys. Lett.* **100**, 101901 (2012).
- Wei, Z., Ni, Z., Bi, K., Chen, M. & Chen, Y. Interfacial thermal resistance in multilayer graphene structures. *Phys. Lett. A* **375**, 1195–1199 (2011).
- Yang, H., Bai, G., Thompson, L. J. & Eastman, J. A. Interfacial thermal resistance in nanocrystalline yttria-stabilized zirconia. *Acta Mater.* **50**, 2309–2317 (2002).
- Zhong, H. & Lukes, J. R. Interfacial thermal resistance between carbon nanotubes: Molecular dynamics simulations and analytical thermal modeling. *Phys. Rev. B* **74**, 125403 (2006).
- Hu, L., Desai, T. & Keblinski, P. Determination of interfacial thermal resistance at the nanoscale. *Phys. Rev. B* **83**, 195423 (2011).
- Wu, Y., Fang, L. & Xu, Y. Predicting interfacial thermal resistance by machine learning. *Npj Comput. Mater.* **5**, 56 (2019).
- Zhan, T., Fang, L. & Xu, Y. Prediction of thermal boundary resistance by the machine learning method. *Sci. Rep.* **7**, 1–9 (2017).
- Swartz, E. T. & Pohl, R. O. Thermal boundary resistance. *Rev. Mod. Phys.* **61**, 605 (1989).
- Prasher, R. S. & Phelan, P. E. A scattering-mediated acoustic mismatch model for the prediction of thermal boundary resistance. *J. Heat Transfer* **123**, 105–112 (2001).
- Landry, E. S. & McGaughey, A. J. H. Thermal boundary resistance predictions from molecular dynamics simulations and theoretical calculations. *Phys. Rev. B* **80**, 165304 (2009).
- Foygel, M., Morris, R. D., Anez, D., French, S. & Sobolev, V. L. Theoretical and computational studies of carbon nanotube composites and suspensions: Electrical and thermal conductivity. *Phys. Rev. B* **71**, 104201 (2005).
- Du, F., Fischer, J. E. & Winey, K. I. Effect of nanotube alignment on percolation conductivity in carbon nanotube/polymer composites. *Phys. Rev. B* **72**, 121404 (2005).
- Li, C., Thostenson, E. T. & Chou, T. Dominant role of tunneling resistance in the electrical conductivity of carbon nanotube-based composites. *Appl. Phys. Lett.* **91**, 223114 (2007).
- Haggenmueller, R., Guthy, C., Lukes, J. R., Fischer, J. E. & Winey, K. I. Single wall carbon nanotube/polyethylene nanocomposites: Thermal and electrical conductivity. *Macromolecules* **40**, 2417–2421 (2007).
- Chu, K., Li, W., Jia, C. & Tang, F. Thermal conductivity of composites with hybrid carbon nanotubes and graphene nanoplatelets. *Appl. Phys. Lett.* **101**, 211903 (2012).
- Berhan, L. & Sastry, A. M. Modeling percolation in high-aspect-ratio fiber systems. II. The effect of waviness on the percolation onset. *Phys. Rev. E* **75**, 041121 (2007).
- Schilling, T., Jungblut, S. & Miller, M. A. Depletion-induced percolation in networks of nanorods. *Phys. Rev. Lett.* **98**, 108303 (2007).
- Li, J. *et al.* Correlations between percolation threshold, dispersion state, and aspect ratio of carbon nanotubes. *Adv. Funct. Mater.* **17**, 3207–3215 (2007).
- Vogelsang, R., Hoheisel, C. & Ciccotti, G. Thermal conductivity of the Lennard–Jones liquid by molecular dynamics calculations. *J. Chem. Phys.* **86**, 6371–6375 (1987).
- Daw, M. S., Foiles, S. M. & Baskes, M. I. The embedded-atom method: A review of theory and applications. *Mater. Sci. Rep.* **9**, 251–310 (1993).

23. Justo, J. F., Bazant, M. Z., Kaxiras, E., Bulatov, V. V. & Yip, S. Interatomic potential for silicon defects and disordered phases. *Phys. Rev. B* **58**, 2539 (1998).
24. Cleri, F. & Rosato, V. Tight-binding potentials for transition metals and alloys. *Phys. Rev. B* **48**, 22 (1993).
25. Karplus, M. & Petsko, G. A. Molecular dynamics simulations in biology. *Nature* **347**, 631–639 (1990).
26. Laberge, M. & Yonetani, T. Molecular dynamics simulations of hemoglobin A in different states and bound to DPG: Effector-linked perturbation of tertiary conformations and HbA concerted dynamics. *Biophys. J.* **94**, 2737–2751 (2008).
27. Schaad, O., Zhou, H., Szabo, A., Eaton, W. A. & Henry, E. R. Simulation of the kinetics of ligand binding to a protein by molecular dynamics: Geminate rebinding of nitric oxide to myoglobin. *Proc. Natl. Acad. Sci.* **90**, 9547–9551 (1993).
28. Parrinello, M. & Rahman, A. Polymorphic transitions in single crystals: A new molecular dynamics method. *J. Appl. Phys.* **52**, 7182–7190 (1981).
29. Meller, J. (Nature Publishing Group, 2001).
30. Wei, H., Zhao, S., Rong, Q. & Bao, H. Predicting the effective thermal conductivities of composite materials and porous media by machine learning methods. *Int. J. Heat Mass Transf.* **127**, 908–916 (2018).
31. Yang, H., Zhang, Z., Zhang, J. & Zeng, X. Machine learning and artificial neural network prediction of interfacial thermal resistance between graphene and hexagonal boron nitride. *Nanoscale* **10**, 19092–19099 (2018).
32. Hou, Z., Takagiwa, Y., Shinohara, Y., Xu, Y. & Tsuda, K. Machine-learning-assisted development and theoretical consideration for the Al₂Fe₃Si₃ thermoelectric material. *ACS Appl. Mater. Interfaces* **11**, 11545–11554 (2019).
33. Yan, B., Gao, R., Liu, P., Zhang, P. & Cheng, L. Optimization of thermal conductivity of UO₂-Mo composite with continuous Mo channel based on finite element method and machine learning. *Int. J. Heat Mass Transf.* **159**, 120067 (2020).
34. Chan, H. *et al.* Machine learning a bond order potential model to study thermal transport in WSe₂ nanostructures. *Nanoscale* **11**, 10381–10392 (2019).
35. Wu, Y., Sasaki, M., Goto, M., Fang, L. & Xu, Y. Electrically conductive thermally insulating Bi-Si nanocomposites by interface design for thermal management. *ACS Appl. Nano Mater.* **1**, 3355–3363 (2018).
36. Zhang, Y. & Xu, X. Predicting the thermal conductivity enhancement of nanofluids using computational intelligence. *Phys. Lett. A* **12**, 6500 (2020).
37. Guan, K. *et al.* Estimating thermal conductivities and elastic moduli of porous ceramics using a new microstructural parameter. *J. Eur. Ceram. Soc.* **39**, 647–651 (2019).
38. Hemmati-Sarapardeh, A., Varamesh, A., Amar, M. N., Husein, M. M. & Dong, M. On the evaluation of thermal conductivity of nanofluids using advanced intelligent models. *Int. Commun. Heat Mass Transf.* **118**, 104825 (2020).
39. Ghiringhelli, L. M., Vybiral, J., Levchenko, S. V., Draxl, C. & Scheffler, M. Big data of materials science: Critical role of the descriptor. *Phys. Rev. Lett.* **114**, 105503 (2015).
40. Huang, B. & Von Lilienfeld, O. A. Communication: Understanding molecular representations in machine learning: The role of uniqueness and target similarity. *J. Chem. Phys.* (2016).
41. Wu, Y., Zhan, T., Hou, Z., Fang, L. & Xu, Y. Physical and chemical descriptors for predicting interfacial thermal resistance. *Sci. Data* **7**, 1–9 (2020).
42. Safavian, S. R. & Landgrebe, D. A survey of decision tree classifier methodology. *IEEE T. Syst. Man CY. B.* **21**, 660–674 (1991).
43. Breiman, L., Friedman, J., Stone, C. J. & Olshen, R. A. *Classification and Regression Trees* (CRC Press, Boca Raton, 1984).
44. Berk, R. A. *Statistical Learning from a Regression Perspective* 1–65 (Springer, Berlin, 2008).
45. Lathifah, S. N., Nhita, F., Aditsania, A. & Saepudin, D. in *2019 7th International Conference on Information and Communication Technology (ICoICT)*. 1–5 (IEEE).
46. Xu, M., Watanachaturaporn, P., Varshney, P. K. & Arora, M. K. Decision tree regression for soft classification of remote sensing data. *Remote Sens. Environ.* **97**, 322–336 (2005).
47. Schölkopf, B., Smola, A. & Müller, K. R. Nonlinear component analysis as a kernel eigenvalue problem. *Neural Comput.* **10**, 1299–1319 (1998).
48. Yu, H. & Kim, S. *SVM Tutorial-Classification, Regression and Ranking*. Vol. 1 (2012).
49. Vovk, V. *Empirical Inference* 105–116 (Springer, Berlin, 2013).
50. Zhang, Y., Duch, J. & Wainwright, M. in *Conference on learning theory*. 592–617.
51. An, S., Liu, W. & Venkatesh, S. in *2007 IEEE Conference on Computer Vision and Pattern Recognition*. 1–7 (IEEE).
52. Rasmussen, C. E. *Summer School on Machine Learning* 63–71 (Springer, Berlin, 2020).
53. Seeger, M. Gaussian processes for machine learning. *Int. J. Neural Syst.* **14**, 69–106 (2004).
54. Williams, C. K. I. & Rasmussen, C. E. in *Advances in neural information processing systems*. 514–520.
55. Rasmussen, C. E. *Evaluation of Gaussian Processes and Other Methods for Non-linear Regression* (University of Toronto, Toronto, 1997).
56. Park, J. *et al.* Gaussian process regression (GPR) representation in predictive model markup language (PMML). *Smart Sustain. Manuf. Syst.* **1**, 121 (2017).
57. Golugula, A., Lee, G. & Madabhushi, A. in *2011 Annual International conference of the IEEE engineering in medicine and biology society*. 949–952 (IEEE).
58. Czekaj, T., Wu, W. & Walczak, B. Classification of genomic data: Some aspects of feature selection. *Talanta* **76**, 564–574 (2008).
59. Allingham, D. & Rayner, J. Two-sample testing for equality of variances. (2011).
60. Markowski, C. A. & Markowski, E. P. Conditions for the effectiveness of a preliminary test of variance. *Am. Stat.* **44**, 322–326 (1990).
61. Gunavathi, C. & Premalatha, K. A comparative analysis of swarm intelligence techniques for feature selection in cancer classification. *Sci. World J.* **20**, 14 (2014).
62. Kohavi, R. in *Ijcai*. 1137–1145 (Montreal, Canada).
63. Pedregosa, F. *et al.* Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.* **12**, 2825–2830 (2011).
64. Xu, Y., Kato, R. & Goto, M. Effect of microstructure on Au/sapphire interfacial thermal resistance. *J. Appl. Phys.* **108**, 104317 (2010).
65. Zhan, T., Minamoto, S., Xu, Y., Tanaka, Y. & Kagawa, Y. Thermal boundary resistance at Si/Ge interfaces by molecular dynamics simulation. *AIP Adv.* **5**, 047102 (2015).
66. Xue, L., Keblinski, P., Phillpot, S. R., Choi, S. & Eastman, J. A. Two regimes of thermal resistance at a liquid–solid interface. *J. Chem. Phys.* **118**, 337–339 (2003).
67. Xue, M., Heichal, Y., Chandra, S. & Mostaghimi, J. Modeling the impact of a molten metal droplet on a solid surface using variable interfacial thermal contact resistance. *J. Mater. Sci.* **42**, 9–18 (2007).
68. Xu, Y., Kato, R. & Goto, M. Effect of microstructure on Au/sapphire interfacial thermal resistance. *J. Appl. Phys.* **108**, 104317 (2010).
69. Ma, R., Wan, X., Zhang, T., Yang, N. & Luo, T. Role of molecular polarity in thermal transport of boron nitride-organic molecule composites. *ACS omega* **3**, 12530–12534 (2018).
70. Wang, T., Zhang, C., Snoussi, H. & Zhang, G. Machine learning approaches for thermoelectric materials research. *Adv. Funct. Mater.* **30**, 1906041 (2020).
71. Xue, B. *et al.* From tanghulu-like to cattail-like SiC nanowire architectures: Interfacial design of nanocellulose composites toward high thermal conductivity. *J. Mater. Chem. A* **8**, 14506–14518 (2020).

Acknowledgements

We gratefully acknowledge the financial support from National Natural Science Foundation of China (No. 21808240).

Author contributions

X.T. designed the study, X.T. and M.C. conducted the analysis and wrote the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-020-80795-z>.

Correspondence and requests for materials should be addressed to X.T. or M.C.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021