



# Application of machine learning in predicting the thermal conductivity of single-filler polymer composites

Yinzhou Liu<sup>a,b</sup>, Weidong Zheng<sup>b</sup>, Haoqiang Ai<sup>b</sup>, Hao Zhou<sup>b</sup>, Liyin Feng<sup>b</sup>, Lin Cheng<sup>b,\*</sup>, Ruiqiang Guo<sup>b</sup>, Xiaohan Song<sup>b,\*</sup>

<sup>a</sup> Institute of Advanced Technology, Shandong University, Jinan 250061, China

<sup>b</sup> Shandong Institute of Advanced Technology, Jinan 250100, China



## ARTICLE INFO

### Keywords:

Machine learning  
Thermal conductivity  
Polymer composites

## ABSTRACT

Polymer composites with superior thermal conductivity and low electrical conductivity are pivotal in the cooling of electronic devices. Despite their prevalence, the accurate prediction of the thermal conductivity of these composites remains a challenge. The emergence of machine learning (ML) provides a groundbreaking approach to solving this issue. In this study, we constructed a comprehensive dataset collected from previous experimental papers and successfully predicted the thermal conductivity of single-filler polymer composites using four ML regression algorithms: random forest regression (RFR), gradient boosting decision tree (GBDT), extreme gradient boosting (XGBoost), and Gaussian process regression (GPR). By employing feature engineering to select pertinent features from the original datasets, the accuracy of the four models on the test set was improved, among which GBDT exhibited the highest accuracy with the Pearson correlation coefficient value of 0.981. Factors such as filler volume fraction and matrix thermal conductivity significantly influence the thermal conductivity of composite materials, while the thermal conductivity of the fillers has a relatively minor impact. Additionally, we identified the topological polar surface area (TPSA) as a crucial descriptor for surface modifications, quantifying diverse surface-modifying agents. Due to the incorporation of more descriptors, ML models exhibit higher precision and broader applicability compared to empirical formulas. Our study provides an effective tool for predicting the thermal conductivity of polymer composites with single fillers and underscores the potential of machine learning in accelerating materials design.

## 1. Introduction

The continuing miniaturization and integration of electronic components have led to rapid increases in heat flux density during operation, which can significantly affect their performance and lifespan [1–4]. To cool the electronic components, thermal interface materials (TIMs), commonly consisting of fillers and matrices, have been widely used to dissipate heat in these devices [5–7]. High thermal conductivity fillers like metal and ceramic particles are common [8–10], while matrices, often made from polymer materials like epoxy resin, ensure low electrical conductivity and offer ease of processing [11]. With the advancement of materials science, certain one- and two-dimensional fillers such as carbon nanotubes, graphene, and h-BN, can significantly enhance the thermal conductivity of TIMs due to their exceptionally high thermal conductivity along specific directions, which has aroused widespread interest [12–15].

Evaluating the thermal conductivity of polymer composites is intricate due to the multitude of influencing factors, mainly including the filler type, shape, distribution, and interfacial dynamics [16]. Empirical formulas have been widely used in estimating the thermal conductivity of composite materials for their rapid prediction capabilities [17–19]. However, the primary drawback of empirical formulas lies in their lack of universality and precision. Since they are based on specific datasets and conditions, their accuracy and applicability are limited to only the scenarios closely resembling those from which they were derived. In the pursuit of precision, researchers often opt for directly solving the heat diffusion equation, with commonly employed methods including the finite volume method (FVM) and the finite element method (FEM). Both approaches involve discretizing the solution domain into small elements or volumes and utilizing numerical methods to obtain approximate solutions to the differential equation. While this ensures higher accuracy, it comes at the cost of substantial computational resources and time

\* Corresponding authors.

E-mail addresses: [cheng@sdu.edu.cn](mailto:cheng@sdu.edu.cn) (L. Cheng), [xiaohan.song@iat.cn](mailto:xiaohan.song@iat.cn) (X. Song).

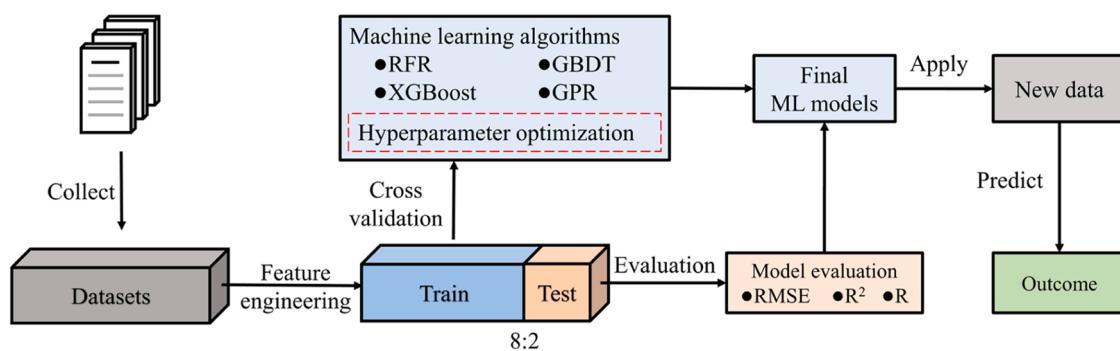


Fig. 1. The workflow diagram of constructing ML models.

requirements [20,21]. In addition, molecular dynamics (MD) simulations and Boltzmann transport equation (BTE) have also been used to obtain the thermal conductivity of composites [22]. MD simulations offer an atomic-scale perspective on material interactions, but their computational scale is generally at the micrometer level, and their accuracy heavily relies on the choice of potential functions [23–25]. The BTE method allows for larger computational systems, reaching millimeter scales; however, this method still involves solving differential equations and requires significant computational resources [26–28]. Therefore, it is imperative to propose a method that can rapidly determine the thermal conductivity of polymer composite materials with high accuracy.

Machine learning (ML) offers a promising alternative to this challenge. By harnessing the thermal conductivity data of selected material combinations, machine learning methods can extrapolate and predict outcomes of a broader set. Previous studies employed various machine learning algorithms to predict thermal conductivity in different materials. For instance, Han et al. [29] applied support vector regression (SVR), Gaussian process regression (GPR), and convolutional neural network (CNN) to predict the thermal conductivity of porous materials. By comparing with empirical formulas, they demonstrated the high accuracy of machine learning algorithms. Alrebdī et al. [30] employed decision tree regression (DTR) and SVR to predict the thermal conductivity of thermoelectric materials. They achieved an  $R^2$  value of 0.97 on the test set by selecting appropriate descriptors and optimizing model hyperparameters. Li et al. [31] compiled a comprehensive database of thermal conductivity for soil from various literature and implemented

six ML algorithms including multivariate linear regression (MLR), GPR, SVR, DTR, random forest regression (RFR), and adaptive boosting methods (AdaBoost), for prediction purposes. They analyzed the importance of different soil parameters and compared the performance of these ML algorithms with traditional empirical models. The study concluded that certain ML algorithms, particularly AdaBoost, demonstrate higher predictive accuracy and showcase the potential of machine learning in efficiently predicting soil properties.

Despite the success of ML in predicting the thermal conductivity of crystalline materials, rare efforts have been devoted to predicting thermal properties of polymer composites using ML methods. This study aims to bridge the observed gap in predicting thermal conductivity for TIMs. Through employing a collected experimental dataset, we trained four regression algorithms: RFR, gradient boosting decision tree (GBDT), extreme gradient boosting (XGBoost), and GPR. By applying feature engineering techniques, the number of descriptors was reduced from 30 to 11 and the performance of all four models on the test set was improved. The four ML models are more accurate than the traditional empirical formulae because they take into account a wider range of factors. With reference to the parameters of the empirical formula, we have analyzed the important descriptors that dominate the thermal conductivity of the composites. Additionally, we identified that the topological polar surface area (TPSA) of surface modifiers significantly influences the bonding strength between the fillers and the matrix in the composite materials. Finally, we utilized the ML models to predict the thermal conductivity of new composites and discussed the impact of surface modifications on thermal conductivity.

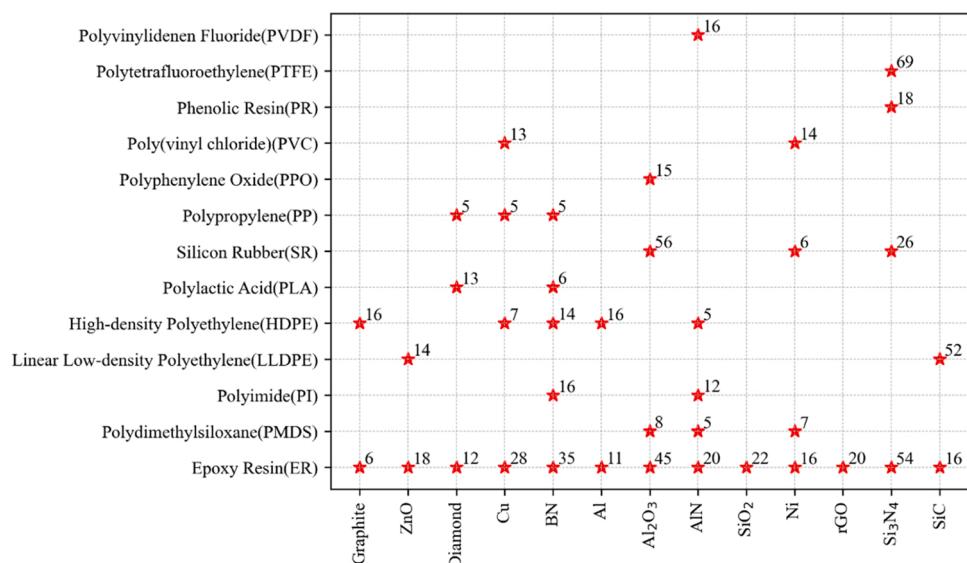


Fig. 2. Combination space of filler and matrix in composites.

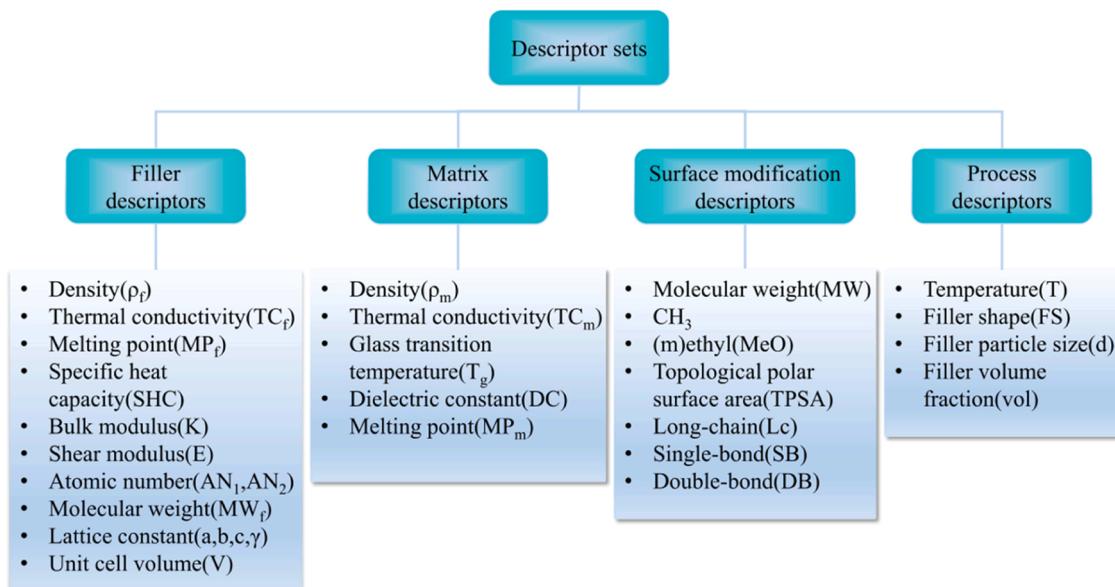


Fig. 3. Schematic diagram of descriptor sets.

## 2. Computational methods

In this paper, the complete workflow of machine learning is illustrated in Fig. 1. The figure depicts various tasks, including the construction of the datasets, training, and evaluation of ML models, and models' prediction on new data.

### 2.1. Datasets

As shown in Fig. 1, we first aggregated a dataset comprising 742 data points sourced from 43 published papers [32–75], which were later fed to train machine learning models. The accumulated dataset comprises 39 polymer materials, consisting of 13 fillers and 13 matrices. The primary fillers include graphite, reduced graphene oxide (rGO), metals, and their oxides. Epoxy resin and high-density polyethylene are the predominant matrices. Fig. 2 illustrates the combined space between the filler and matrix, highlighting the sample count for each unique combination. The sample is defined by matrix, filler, volume fraction, filler

particle size, filler shape, temperature, and surface modifier.

### 2.2. Descriptors

There are many factors that affect the thermal conductivity of composite materials. The primary considerations generally include the thermal conductivity of filler and matrix, the shape, size, volume fraction of filler, and interface thermal resistance, etc [16,76]. In this study, we segmented the 30 descriptors into four descriptor sets: filler descriptors, matrix descriptors, surface modification descriptors, and process descriptors. As shown in Fig. 3, for each descriptor set, we selected descriptors that are highly correlated with the thermal conductivity of composite materials and are easy to obtain. A detailed explanation of each descriptor is presented in Table S1 and the data distribution for each descriptor is shown in Figs. S1–4. Filler descriptors include the material properties of the fillers, such as density ( $\rho_f$ ), thermal conductivity ( $TC_f$ ), melting point ( $MP_f$ ), specific heat capacity (SHC), bulk modulus (K), shear modulus (E), atomic number ( $AN_1, AN_2$ ),

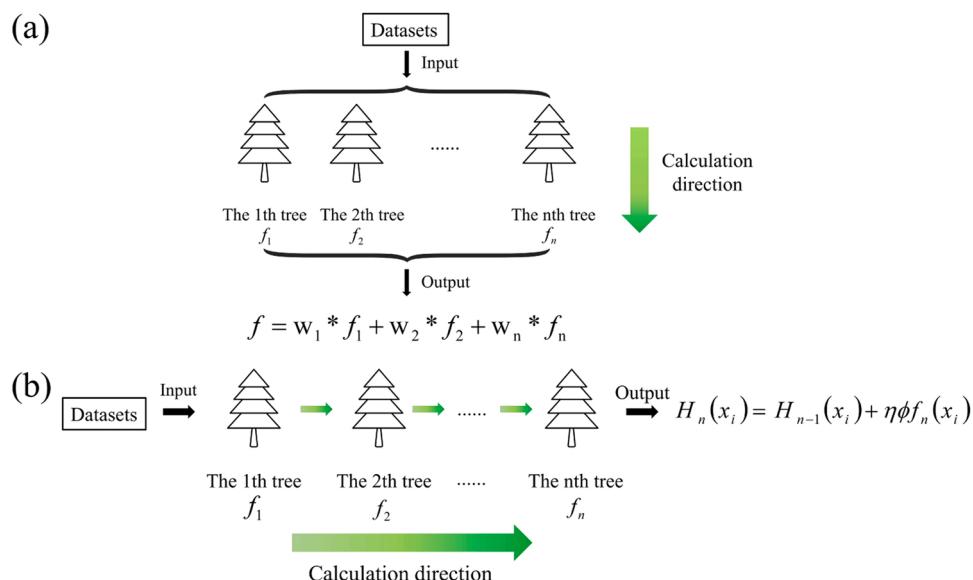


Fig. 4. Integrated learning algorithm frame diagram for (a) RFR, (b) GBDT and XGBoost.

molecular weight ( $MW_f$ ), lattice constant ( $a, b, c, \gamma$ ) and unit cell volume ( $V$ ). Matrix descriptors include the material properties of matrices, such as density ( $\rho_m$ ), thermal conductivity ( $TC_m$ ), glass transition temperature ( $T_g$ ), dielectric constant (DC), and melting point ( $MP_m$ ). Surface modification descriptors contain the material properties of the surface modification molecules (Table S2), which have a great influence on the interfacial thermal resistance of the filler-matrix in the composite [77]. Proper surface treatments can amplify the bond strength between the functional groups on both filler and matrix surfaces [78–81]. The molecular weight (MW),  $CH_3$ , (m)ethyl (MeO), topological polar surface area (TPSA), long-chain (Lc), single-bond (SB), and double-bond (DB) are included in surface modification descriptors. Descriptors other than the material intrinsic properties are included in the process descriptors, mainly the variables controlled during the experiment, such as temperature (T), filler shape (FS), filler particle size (d), and filler volume fraction (vol). Descriptors were collected from published papers and reliable databases, including Materials Project ([www.materialsproject.org](http://www.materialsproject.org)), Material Properties Database ([www.makeitfrom.com](http://www.makeitfrom.com)), and MATWEB Material Property Data ([www.matweb.com](http://www.matweb.com)).

### 2.3. Machine learning algorithm

Traditional machine learning algorithms, such as logistic regression, DTR, and GPR, have been preferred for their conceptual clarity and robust performance. However, when faced with increasingly complex datasets, these time-honored methods are facing difficulties in catering to the intricate requirements of modern industrial applications. In contrast, ensemble learning algorithms have consistently dominated the field due to their outstanding predictive capabilities. This study explored three ensemble learning algorithms: RFR, GBDT, and XGBoost. Their efficacy was compared with that of the traditional machine learning algorithm, GPR, to observe their respective performances. The scikit-learn library was utilized for building the aforementioned machine learning models [82].

The RFR algorithm [83] is a quintessential bagging technique that combines multiple weak estimators, typically decision trees, in a parallel arrangement. The final output is obtained by taking the weighted average of these weak estimators. As shown in Fig. 4(a), due to its parallel architecture, the algorithm enables independent and concurrent execution of multiple weak estimators, thereby providing fast computational speed.

The GBDT and XGBoost algorithms are both classified as Boosting algorithms [84,85]. XGBoost is an evolutionary improvement over GBDT, primarily due to the incorporation of a unique regularization term that effectively controls overfitting and enhances the model's generalization capacity. Fundamentally, these algorithms share the same conceptual framework, and sequentially integrate multiple weak estimators. As shown in Fig. 4(b), each weak estimator receives input from the output data generated by its immediate predecessor, with the n-th estimator being directly influenced by the results of the (n-1)th estimator. In this picture,  $H_n(x_i)$  represents the output result of the previous tree on the sample,  $f_n(x_i)$  represents the output of the n-th tree for sample  $x_n$ ,  $\varphi$  denotes the weights of the n-th tree,  $\eta$  represents the learning rate.

In machine learning algorithms, hyperparameters play a significant role in determining the efficacy and generalization capabilities of models. These hyperparameters are not determined during training and require manual fine-tuning. However, as the number of parameters increases exponentially, tuning becomes time-consuming. Therefore, selecting an appropriate hyperparameter optimization strategy is crucial. Two prominent methods in this field are grid search method and tree-structured Parzen estimator (TPE) optimization [86,87]. The former is an exhaustive search method while the latter falls under Bayesian optimization methods. Given the extensive hyperparameter space for our ensemble algorithms, using grid search would be excessively time-consuming. Henceforth, we opted for TPE algorithm for

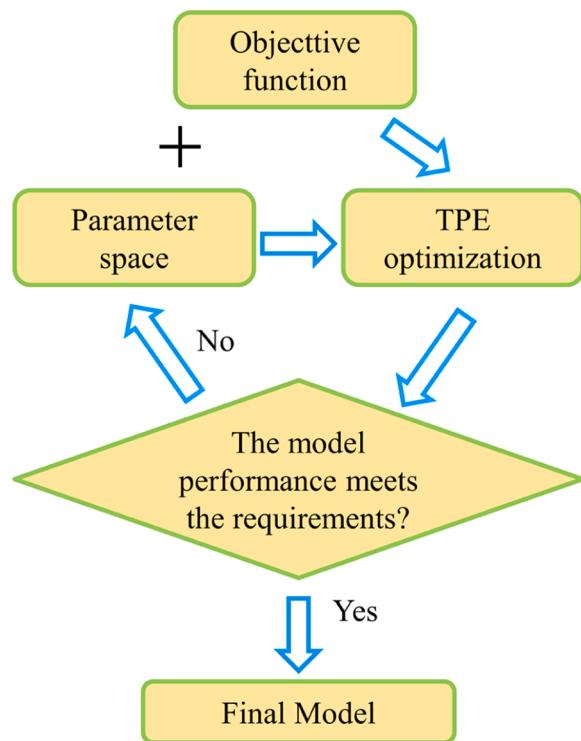


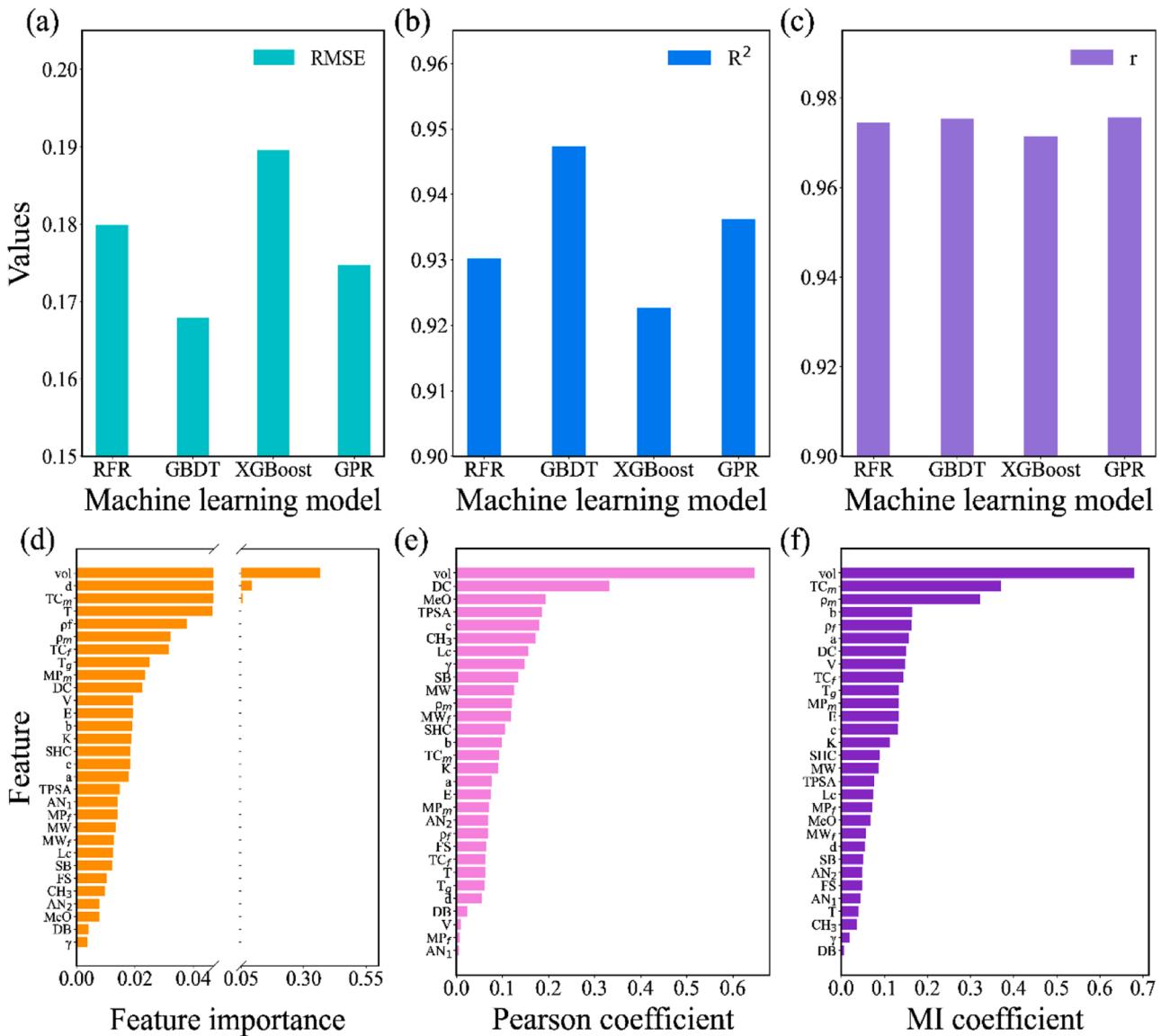
Fig. 5. Flowchart of TPE optimization algorithm.

hyperparameter optimization. Leveraging Bayesian optimization principles, TPE utilizes historical performance data to predictively model the effectiveness of different hyperparameter configurations. This approach not only enables a more efficient exploration of the hyperparameter space but also enhances the optimization process for our machine learning models. The workflow of TPE algorithm is graphically illustrated in Fig. 5. The procedure of this method starts with the definition of the hyperparameter space and the objective function to be optimized. The TPE optimizer is then used to identify the hyperparameter combination that yields the optimal solution for the objective function. If the machine learning model achieves the desired performance under this hyperparameter combination, the model is retained; otherwise, the hyperparameter space need be readjusted and the process repeated. By implementing the TPE approach, we improved our models' performance and significantly reduced the time typically required for hyperparameter tuning.

To strike a balance between assessment accuracy and computational cost, 5-fold cross-validation was adopted. Lower folds may lead to unstable evaluations, while higher folds escalate computational expenses with limited performance gains. And 5-fold cross-validation is a widely used and validated method, which has been widely recognized in practice [31,88].

The evaluation metrics employed for the models included root mean square error (RMSE), coefficient of determination ( $R^2$ ), and Pearson correlation coefficient ( $r$ ), as expressed in Eqs. (1–3). RMSE quantifies the magnitude of error by measuring the square root of the average squared difference between predicted and true values.  $R^2$  assesses model fit, with values ranging from 0 to 1; a value closer to 1 indicates a better fit. And  $r$  measures both strength and direction of linear relationship between variables, ranging from -1 to 1; -1 denotes perfect negative linear correlation, 1 signifies perfect positive linear correlation, while 0 implies no linear correlation.

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



**Fig. 6.** Evaluation metrics of (a) RMSE, (b)  $R^2$ , and (c)  $r$  for RFR, GBDT, XGBoost, and GPR models for the test datasets. (d) Feature importance ranking of GBDT model. (e) Pearson correlation coefficient and (f) mutual information coefficient between descriptors and thermal conductivity of composite materials.

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

$$r = \frac{\sum_{i=1}^n (y_i - \bar{y})(\hat{y}_i - \bar{\hat{y}})}{\sqrt{\sum_{i=1}^n (y_i - \bar{y})^2 \sum_{i=1}^n (\hat{y}_i - \bar{\hat{y}})^2}} \quad (3)$$

where  $n$  is the number of samples,  $y_i$  is the true value of the  $i$ th sample,  $\hat{y}_i$  is the predicted values of the  $i$ th sample,  $\bar{y}$  is the average of the true values of all samples,  $\bar{\hat{y}}$  is the average of the predicted values of all samples.

#### 2.4. Empirical formulas for predicting the thermal conductivity of composite materials

Empirical formulas provide a streamlined approach for estimating the thermal conductivity of polymer composites. Among them, the Maxwell-Eucken model and the Bruggman model are widely used to predict thermal conductivity of composite materials, the corresponding equations are presented below [29]:

Maxwell-Eucken model:

$$k = k_1 \frac{2k_1 + k_2 - 2(k_1 - k_2)v_2}{2k_1 + k_2 + (k_1 - k_2)v_2} \quad (4)$$

Bruggman model :

$$k = \frac{1}{4} \left( (3v_2 - 1)k_2 + [3(1 - v_2) - 1]k_1 + \sqrt{[(3v_2 - 1)k_2 + (3(1 - v_2) - 1)k_1]^2 + 8k_1 k_2} \right) \quad (5)$$

where subscripts 1, and 2 represent the matrix and the filler, respectively;  $k$  represents the thermal conductivity and  $v$  represents the volume fraction.

#### 3. Results and discussion

During the feature selection stage, our goal is to identify and eliminate descriptors that are not only highly redundant but also exhibit low importance in the modeling process [88,89]. We initiated this process by evaluating the correlation among descriptors as well as their connection

**Table 1**

The features obtained after the feature selection.

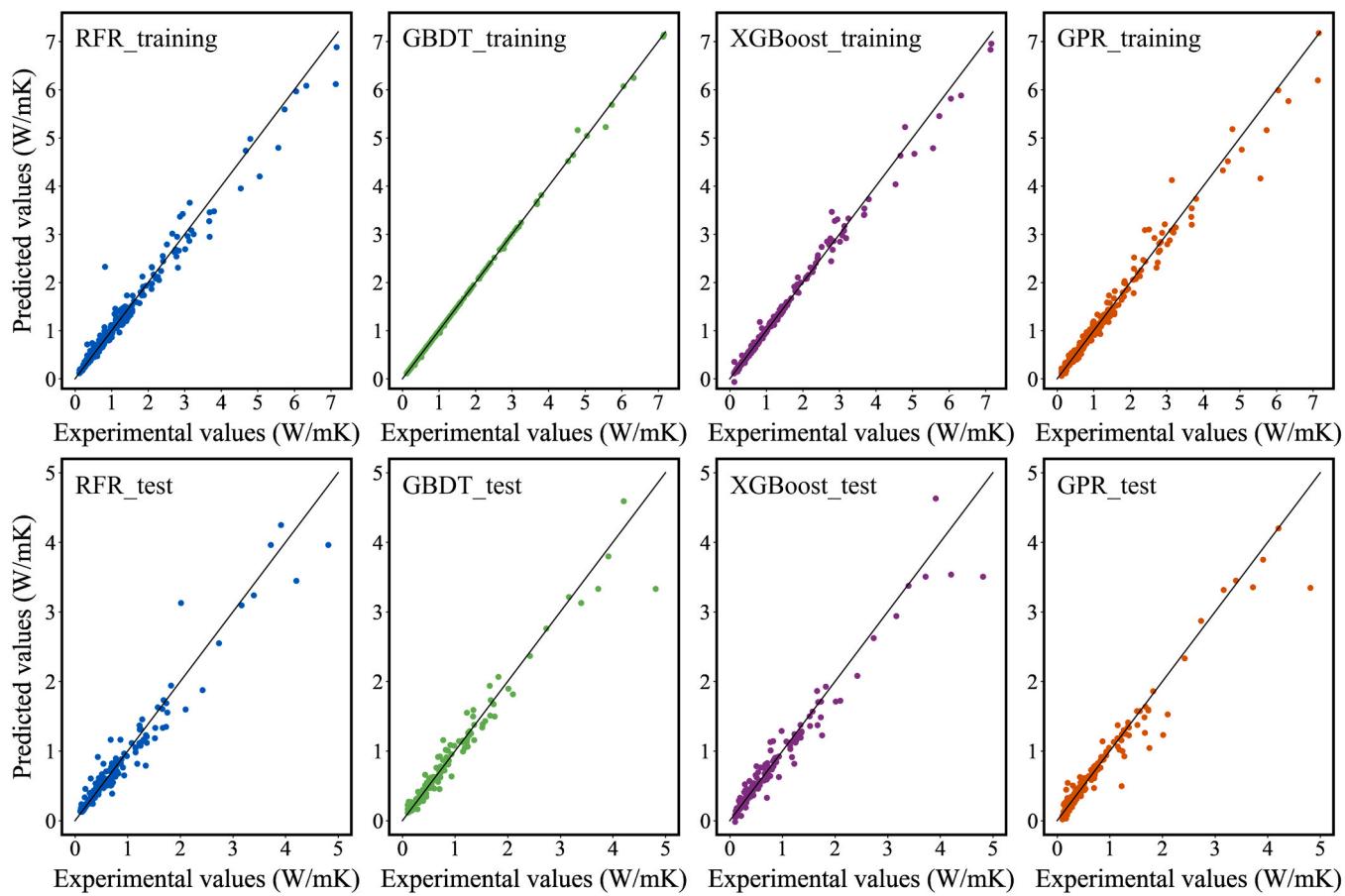
Classification	Descriptors
Process descriptors	T, d, vol.
Filler descriptors	$\rho_f$ , TC <sub>f</sub> , E, V
Matrix descriptors	$\rho_m$ , TC <sub>m</sub> , MP <sub>m</sub>
Surface modification descriptors	TPSA

to the thermal conductivity of the composite material. For this purpose, we first evaluated the performance of the four models on the original datasets using RMSE,  $R^2$ , and  $r$ . As depicted in Fig. 6(a-c), among these models, GBDT model exhibited the highest  $R^2$  and  $r$  values while recording the lowest RMSE on the test set. This indicates that the GBDT model possesses exceptional learning capability among the original datasets. Consequently, we employed the GBDT model for feature selection.

The presence of a strong correlation between two descriptors may suggest redundancy, as they can convey similar information. We utilized the Pearson correlation coefficient to access linear correlations, which is presented in Fig. S5. This metric offers an initial filter by elucidating the linear relationship between descriptors. However, it is crucial that the

Pearson correlation coefficient only captures linear interactions between variables, thus a low Pearson coefficient does not exclude the possibility of nonlinear relationships between variables. Recognizing this limitation, we utilized the mutual information (MI) method to examine potential nonlinear dependencies between variables, as shown in Fig. S6.

However, it is essential to consider that high correlation coefficients in specific subsets of descriptors do not automatically justify the removal of descriptors. Omitting such descriptors without thorough validation within the model could have a negative impact on performance. Therefore, the selection of retained descriptors is determined through multiple analyses including Pearson correlation coefficient, mutual information method, and feature importance based on GBDT model (Fig. 6(d-f)). For instance, it can be seen from Fig. S5 that the lattice constants  $a$  and  $b$  have a strong linear correlation, while the linear correlation coefficient between  $a$  and  $c$  is weak, however, Fig. S6 shows a strong nonlinear correlation between  $a$  and  $c$ . Therefore, based on the ranking of the importance of their features, we decided to remove  $a$  and  $c$  and keep  $b$ . Following this approach, similar analytical methodologies were applied to evaluate other features. After the feature selection process, we reduced the number of descriptors from 30 to 21, the remaining features are listed in Table S3. In addition to the above process, we applied

**Fig. 7.** Performance of the RFR, GBDT, XGBoost, and GPR models on the training set and the test set.**Table 2**

Evaluation results of the RFR, GBDT, XGBoost, and GPR models on training and test sets.

	training_RMSE	training_R <sup>2</sup>	training_r	test_RMSE	test_R <sup>2</sup>	test_r
RFR	0.1346	0.9770	0.9902	0.1804	0.9361	0.9718
GBDT	0.0224	0.9994	0.9997	0.1499	0.9560	0.9811
XGBoost	0.1253	0.9805	0.9913	0.1747	0.9362	0.9756
GPR	0.0798	0.9926	0.9963	0.1667	0.9439	0.9773

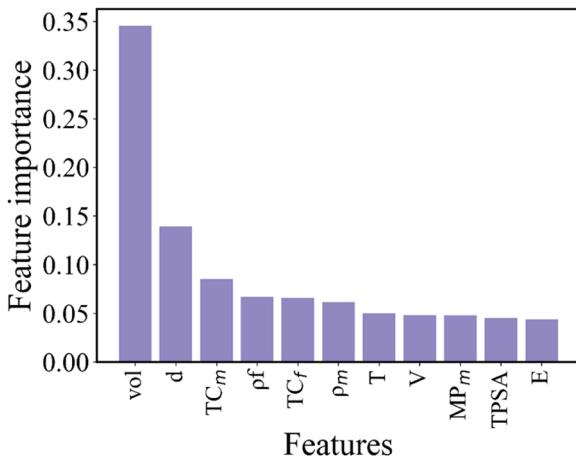


Fig. 8. Feature importance ranking of GBDT model after feature selection.

Recursive Feature Elimination (RFE) to further filter the descriptors that are more important to ML models. The basic idea of this method is to construct the model iteratively (usually based on some kind of machine learning algorithm), eliminating a certain number of the least important features each time, until a predefined number of features is finally reached or the performance of the model no longer improves. After using the RFE method, we further reduced the number of descriptors from 21–11 and the final descriptors are shown in Table 1.

We retrained the RFR, GBDT, XGBoost, and GPR models using descriptors that were retained after feature selection. To mitigate overfitting, we employed 5-fold cross-validation during training. The models' performance on both the training and test sets was assessed using RMSE,  $R^2$ , and r values, as shown in Fig. 7 and Table 2. Among these models, the GBDT model exhibited superior performance with the highest  $R^2$  (0.956) and r (0.981) values along with the lowest RMSE (0.150). This highlights its proficiency in recognizing complex patterns and features within the data while demonstrating exceptional generalization capabilities. Interestingly, despite lacking advantages such as feature amalgamation and model ensembling present in ensemble algorithms, GPR still performed admirably and ranked second only to GBDT models in terms of performance on this dataset. This can likely be attributed to GPR's nature as a probabilistic regression technique [18, 31, 90], which excels at encapsulating uncertainties and data noise to yield precise predictions. Although the accuracy of RFR and XGBoost models is relatively lower compared to other models mentioned above, their respective  $R^2$  ( $>0.93$ ) and r ( $>0.97$ ) values can still accurately describe thermal conductivity of composite materials.

Upon comparing the data from Table 2 with that of Fig. 6(a-c), it is clear that the  $R^2$  and r values have improved for the four models on the test set, and the RMSE values have decreased. This indicates that feature engineering has enhanced the quality of the original datasets and confirms the efficacy of feature engineering.

Identifying important features is crucial for exploring the factors influencing thermal conductivity of composites and enhancing the interpretability of machine learning models. The feature importance ranking of the GBDT model is illustrated in Fig. 8. Evidently, the volume

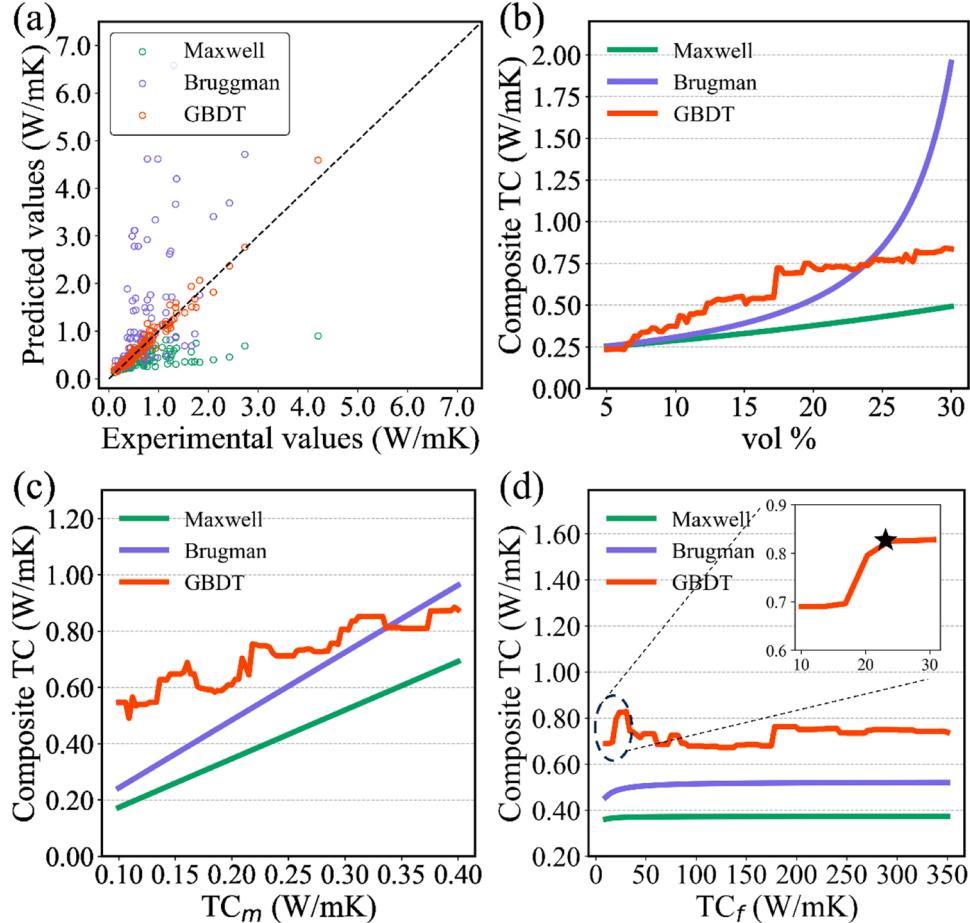


Fig. 9. (a) The correlation between the experimental values and the predicted values of Maxwell-Eucken model, Bruggman model, and GBDT model. The effects of (b) the filler volume fraction and the thermal conductivity of (c) the matrix and (d) filler on thermal conductivity of composites in Maxwell-Eucken model, Bruggman model, and GBDT model.

**Table 3**

The range of values of the filler volume fraction (vol), the thermal conductivity of the matrix ( $TC_m$ ) and filler ( $TC_f$ ) when comparing ML models with empirical formulas.

Descriptors	range	count
vol%	[5–30]	100
$TC_f$ (W/mK)	[10,350]	100
$TC_m$ (W/mK)	[0.1,0.4]	100

fraction of the filler emerges as the most influential descriptor for the thermal conductivity of composite materials, aligning with conventional understanding. The particle size of the filler and thermal conductivity of the matrix closely follow in terms of their characteristic importance, while the thermal conductivity of the filler exhibits relatively lower significance. This implies that solely altering filler type may yield limited impact on the thermal conductivity of composite materials, which contradicts intuition. Notably, topological polar surface area plays a vital role among surface modification molecule descriptors. It is comprehensible since surface-modified molecules establish connections between the matrix and the filler, where topological polar surface area characterizes molecular polarity associated with binding strength to the matrix.

The performance of machine learning models and empirical formulas, such as the Maxwell-Eucken model and Bruggman model, on the test set is illustrated in Fig. 9(a). It should be noted that the applicability of empirical formulas is often limited. For instance, the Maxwell model is only applicable when the volume fraction is less than 40 %, thus we have retained data within this applicable range [18,19]. Furthermore, we observed that predictions from the Bruggman model are inaccurate for fillers with relatively high thermal conductivity. Therefore, data points with filler thermal conductivity exceeding 500 W/mK were excluded. The results clearly demonstrate that the GBDT model exhibits significantly higher accuracy compared to both the Maxwell-Eucken and Bruggman models, thereby highlighting its superiority in machine learning.

Although empirical formulas may not be as accurate as machine learning models, they offer the advantage of being highly interpretable, which can provide a clear understanding of how different factors influence the thermal conductivity of composite materials. Key factors affecting the thermal conductivity of composites, such as the filler volume fraction and the thermal conductivity of the matrix and filler, have been extracted and incorporated into computational expressions of empirical formulas. The influence of these key factors on the thermal conductivity of the composite is plotted in Fig. 9(b-d) for both the machine learning model and empirical formula. The new datasets used here focused on the filler volume fraction (vol), the thermal conductivity of the matrix ( $TC_m$ ) and filler ( $TC_f$ ), with each generating 100 evenly spaced data points in the range shown in Table 3. The remaining descriptors took the average of the original datasets. The results demonstrated that machine learning models predicted similar trends to those obtained from empirical formulas for all three descriptors, highlighting their potential to capture intrinsic physical mechanisms. Additionally, it was observed that an increase in vol and  $TC_m$  positively influenced composite thermal conductivity; however, this relationship varied for  $TC_f$ . As shown in Fig. 9(d),  $TC_f$  has a great positive impact on the thermal conductivity of the composite when the  $TC_f$  is below 20 W/mK. And then, for fillers with thermal conductivity exceeding 20 W/mK, the composite thermal conductivity changes little as the  $TC_f$  increases, which is consistent with our previous analysis of the thermal conductivity descriptor of the filler. And according to Maxwell-Eucken model, when the thermal conductivity of the filler is much greater than that of the matrix, Eq. (4) can be simplified to the following expression:

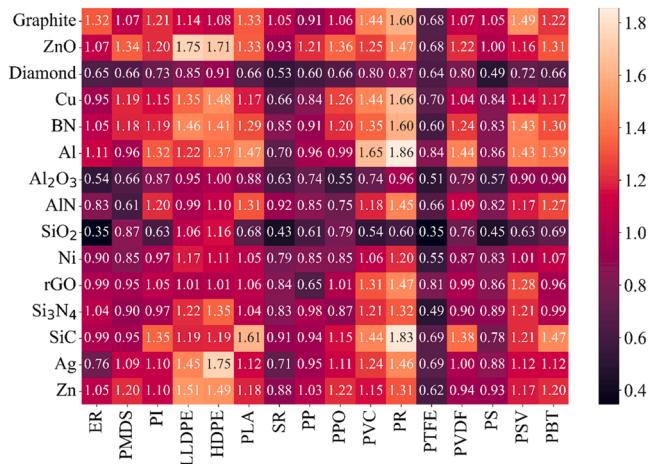


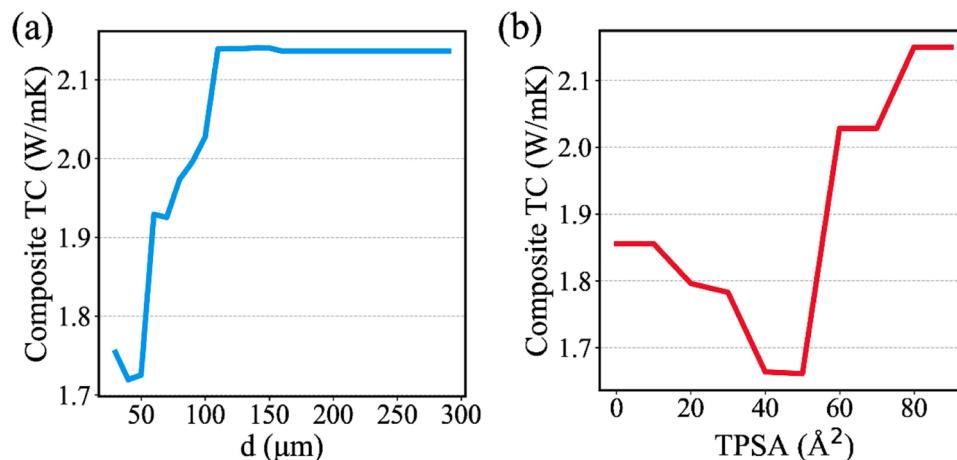
Fig. 10. Heatmap of the predicted thermal conductivity for newly designed composites.

$$k = k_1 \frac{\frac{2 + \frac{k_2}{k_1} - 2 \left(1 - \frac{k_2}{k_1}\right) v_2}{2 + \frac{k_2}{k_1} + \left(1 - \frac{k_2}{k_1}\right) v_2}}{1 - v_2} = k_1 \frac{1 + 2v_2}{1 - v_2} \quad (6)$$

Under these conditions, the thermal conductivity of the composite materials depends solely on  $TC_m$  and vol, so the thermal conductivity of the composites will not continue to increase with the increase of the  $TC_f$ . It is understandable that in cases where vol is not high, a thermal conductivity network may not be formed by the fillers, thereby limiting the contribution of their high thermal conductivity to that of the composites. Meanwhile, this means that there is a threshold for  $TC_f$ , defined as  $TC_f^*$ , which refers to the  $TC_f$  value when the thermal conductivity of the composite no longer changes significantly with the  $TC_f$ . We used the laws mined by the GBDT model to depict the relationship between threshold and  $TC_m$  under different volume fractions as shown in Fig. S7. It is evident that the threshold remains nearly unchanged at around 20 W/mK with variations in both  $TC_m$  and volume fraction. In addition, while the machine learning model exhibits a similar trend to the empirical formula, its specific details are different, and the predictions of the machine learning model show a richer variety. This is attributable to the incorporation of additional descriptors in the machine learning algorithm, which enables a more precise description of the thermal conductivity of composites.

The high accuracy of machine learning models, especially GBDT models, in dealing with multi-dimensional problems gives it great potential in finding composites with target properties. So, we used the GBDT model to predict the new data. We built a prediction dataset comprising 16 matrices and 15 fillers [18,76], and predicted the thermal conductivity of composites at 303 K with filler particle size of 10 μm, volume fraction of 30 %, and without any surface modification. The shape of graphite and BN were set as flake while the remaining fillers were considered granular in shape. The heatmap illustrating the thermal conductivity of composite materials with various combinations of matrices and fillers is presented in Fig. 10. The results illustrate that, under a constant volume fraction, composites made from matrices with higher thermal conductivity tend to exhibit superior thermal conductivity performance. Additionally, although the thermal conductivity of the filler does have a positive impact on the thermal conductivity of composites, this effect is relatively weaker compared to that of the matrix, which aligns with the previously analyzed patterns.

As can be seen in Fig. 8, in addition to the descriptors included in the empirical formulas, the particle size of the filler also has a large effect on the thermal conductivity of the composite. Here we have investigated the effect of particle size on the thermal conductivity of composites in



**Fig. 11.** The effects of (a) the filler diameter and (b) topological polar surface area on thermal conductivity of Al-PR composites.

the range 30–300 μm using the combination of Al and PR as an example. From Fig. 11(a) it can be seen that when the volume fraction of filler is fixed at 30 %, on the whole, the increase in filler particle size positively enhances the thermal conductivity of the composites due to the fact that the larger particle size helps the fillers to contact each other and thus form a thermally conductive network [35,60]. However, this improvement is not unlimited, when the filler particle size is larger than about 110 μm, it no longer has much effect on the thermal conductivity of the composite.

Subsequently, we have investigated the effect of TPSA of the surface modification on the thermal conductivity of the composites, as shown in Fig. 11(b). For the composites formed by Al and PR, the surface modification can increase its thermal conductivity from 1.86 to 2.15 W/mK. In addition, the effect of TPSA on the thermal conductivity of the composites is not a simple monotonic relationship, but a complex non-linear relationship, so the optimal TPSA is different for composites formed by different matrix and filler, which is determined by the matching of surface modification molecules to the filler and the matrix.

#### 4. Conclusion

In this study, we have curated a unique dataset of polymer composites with single fillers and successfully employed four machine learning models (RFR, GBDT, XGBoost, and GPR) to accurately predict the thermal conductivity of these composites. Through feature engineering techniques applied to refine the original datasets, we achieved enhanced performance across all models. Our analysis reveals that the filler volume fraction, filler size, and matrix thermal conductivity exerted significant influence on the thermal conductivity of the composites, while the impact from filler thermal conductivity is relatively minor. Additionally, we find that the topological polar surface area of modified molecules plays a crucial role in determining the connection between the matrix and filler materials.

Compared to empirical formulas commonly used in this field, our machine learning models demonstrated superior accuracy and broader applicability. Furthermore, our new dataset enabled predictions for composites of matrix-filler combinations with high thermal conductivity. Notably, the particle size of the filler and the TPSA of the surface modification reagent also have a large effect on the thermal conductivity of the composites. Overall, our work confirms that machine learning approaches can effectively predict the thermal conductivity of single-filler polymer composites, indicating the great potential for solving complex thermal issues using machine learning.

#### CRediT authorship contribution statement

**Yinzhou Liu:** Writing – original draft, Visualization, Validation, Software, Methodology, Formal analysis. **Liyin Feng:** Software, Investigation. **Hao Zhou:** Data curation. **Haoqiang Ai:** Writing – review & editing. **Xiaohan Song:** Writing – review & editing, Supervision, Project administration, Funding acquisition, Conceptualization. **Weidong Zheng:** Writing – review & editing, Funding acquisition. **Ruiqiang Guo:** Writing – review & editing, Funding acquisition. **Lin Cheng:** Supervision, Resources.

#### Declaration of Competing Interest

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

#### Data availability

Data will be made available on request.

#### Acknowledgements

This study is supported by the Basic Research Project of Natural Science Foundation of Shandong Province (ZR2022QA092). R. G. acknowledges support from the Excellent Young Scientists Fund (Overseas) of Shandong Province (Grant No. 2022HWYQ-091), the Taishan Scholars Program of Shandong Province, and the Natural Science Foundation of Shandong Province (Grant No. ZR2022MA011). W. Z. acknowledges the support from Natural Science Foundation of Shandong Province (Grant No. ZR2023QE223) and the Taishan Scholars Program of Shandong Province.

#### Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.mtcomm.2024.109116.

#### References

- [1] X. Qian, R. Yang, Machine learning for predicting thermal transport properties of solids, Mater. Sci. Eng. R 146 (2021) 100642, <https://doi.org/10.1016/j.mser.2021.100642>.
- [2] D.G. Cahill, P.V. Braun, G. Chen, D.R. Clarke, S. Fan, K.E. Goodson, P. Kebinski, W.P. King, G.D. Mahan, A. Majumdar, H.J. Maris, S.R. Phillpot, E. Pop, L. Shi, Nanoscale thermal transport. II. 2003–2012, Appl. Phys. Rev. 1 (1) (2014) 011305, <https://doi.org/10.1063/1.4832615>.

- [3] J. Chen, X. Xu, J. Zhou, B. Li, Interfacial thermal resistance: past, present, and future, Rev. Mod. Phys. 94 (1) (2022) 025002, <https://doi.org/10.1103/RevModPhys.94.025002>.
- [4] X. Qian, J. Zhou, G. Chen, Phonon-engineered extreme thermal conductivity materials, Nat. Mater. 20 (9) (2021) 1188–1202, <https://doi.org/10.1038/s41563-021-00918-3>.
- [5] A.J. McNamara, Y. Joshi, Z.M. Zhang, Characterization of nanostructured thermal interface materials – a review, Int. J. Therm. Sci. 62 (2012) 2–11, <https://doi.org/10.1016/j.ijthermalsci.2011.10.014>.
- [6] J. Hansson, T.M.J. Nilsson, L. Ye, J. Liu, Novel nanostructured thermal interface materials: a review, Int. Mater. Rev. 63 (1) (2017) 22–45, <https://doi.org/10.1080/09506608.2017.1301014>.
- [7] X. Xu, J. Chen, J. Zhou, B. Li, Thermal conductivity of polymers and their nanocomposites, Adv. Mater. 30 (17) (2018) 1705544, <https://doi.org/10.1002/adma.201705544>.
- [8] D.D.L. Chung, Materials for thermal conduction, Appl. Therm. Eng. 21 (2001) 1593–1605, [https://doi.org/10.1016/S1359-4311\(01\)00042-4](https://doi.org/10.1016/S1359-4311(01)00042-4).
- [9] E.A.A. Jarvis, E.A. Carter, Exploiting covalency to enhance metal-oxide and oxide-oxide adhesion at heterogeneous interfaces, J. Am. Ceram. Soc. 86 (3) (2003) 373–386, <https://doi.org/10.1111/j.1151-2916.2003.tb03309.x>.
- [10] A. Shimamura, Y. Hotta, H. Hyuga, M. Hotta, K. Hirao, Improving the thermal conductivity of epoxy composites using a combustion-synthesized aggregated beta-Si<sub>3</sub>N<sub>4</sub> filler with randomly oriented grains, Sci. Rep. 10 (1) (2020) 14926, <https://doi.org/10.1038/s41598-020-71745-w>.
- [11] M. Harada, N. Hamaura, M. Ochi, Y. Agari, Thermal conductivity of liquid crystalline epoxy/BN filler composites having ordered network structure, Compos., Part B 55 (2013) 306–313, <https://doi.org/10.1016/j.compositesb.2013.06.031>.
- [12] T. Tong, Y. Zhao, L. Delzeit, A. Kashani, M. Meyyappan, A. Majumdar, Dense vertically aligned multiwalled carbon nanotube arrays as thermal interface materials, IEEE Trans. Compon. Packag. Technol. 30 (1) (2007) 92–100, <https://doi.org/10.1109/tcpt.2007.892079>.
- [13] T.C. Fitzgibbons, M. Guthrie, E.S. Xu, V.H. Crespi, S.K. Davidowski, G.D. Cody, N. Alem, J.V. Badding, C. Energy Frontier Research, E. Energy Frontier Research in Extreme, Benzene-derived carbon nanowires, Nat. Mater. 14 (1) (2015) 43–47, <https://doi.org/10.1038/nmat4088>.
- [14] J. Wang, F. Ma, M. Sun, Graphene, hexagonal boron nitride, and their heterostructures: properties and applications, RSC Adv. 7 (27) (2017) 16801–16822, <https://doi.org/10.1039/c7ra00260b>.
- [15] K.M. Shahil, A.A. Balandin, Graphene-multilayer graphene nanocomposites as highly efficient thermal interface materials, Nano Lett. 12 (2) (2012) 861–867, <https://doi.org/10.1021/nl203906r>.
- [16] H. Chen, V.V. Ginzburg, J. Yang, Y. Yang, W. Liu, Y. Huang, L. Du, B. Chen, Thermal conductivity of polymer-based composites: fundamentals and applications, Prog. Polym. Sci. 59 (2016) 41–85, <https://doi.org/10.1016/j.progpolymsci.2016.03.001>.
- [17] S.J. Jafari Nejad, A review on modeling of the thermal conductivity of polymeric nanocomposites, e-Polymers 12 (1) (2012) 025, <https://doi.org/10.1515/epoly.2012.12.1.253>.
- [18] Y. Zhou, S. Wu, Y. Long, P. Zhu, F. Wu, F. Liu, V. Murugadoss, W. Winchester, A. Nautiyal, Z. Wang, Z. Guo, Recent advances in thermal interface materials, ES Mater. Manuf. 7 (2020) 4–24, <https://doi.org/10.30919/esmm5f717>.
- [19] R. Prasher, Thermal interface materials: historical perspective, status, and future directions, Proc. IEEE 94 (8) (2006) 1571–1586, <https://doi.org/10.1109/jproc.2006.879796>.
- [20] C. Demuth, M.A.A. Mendes, S. Ray, D. Trimis, Performance of thermal lattice Boltzmann and finite volume methods for the solution of heat conduction equation in 2D and 3D composite media with inclined and curved interfaces, Int. J. Heat. Mass Transf. 77 (2014) 979–994, <https://doi.org/10.1016/j.ijheatmasstransfer.2014.05.051>.
- [21] Z. Tong, M. Liu, H. Bao, A numerical investigation on the heat conduction in high filler loading particulate composites, Int. J. Heat. Mass Transf. 100 (2016) 355–361, <https://doi.org/10.1016/j.ijheatmasstransfer.2016.04.092>.
- [22] H. Bao, J. Chen, X. Gu, B. Cao, A Review of Simulation Methods in Micro/nanoscale Heat Conduction, ES Energy Environ. 1 (2018) 16–55, <https://doi.org/10.30919/eseec8c149>.
- [23] Q. Anjum, N. Nasir, S.A. Cheema, M. Imran, A.R. Rahman, Z. Tanveer, N. Amin, Y. N. Anjam, Multiscale modeling investigation into the thermal conductivity dynamics of graphene-silver nano-composites: a molecular dynamic study, Dig. J. Nanomat. Biostruct. 17 (2) (2022) 557–568, <https://doi.org/10.15251/djnjb.2022.172.557>.
- [24] Y.A.N. He, Y. Tang, Thermal conductivity of carbon nanotube/natural rubber composite from molecular dynamics simulations, J. Theor. Comput. Chem. 12 (03) (2013) 1350011, <https://doi.org/10.1142/s0219633613500119>.
- [25] Luo, C., Pan, D., Wang, X., 2019. Thermal conductivity of PDMS chains and GNPs/PDMS composites: a molecular dynamics study, in: Proceeding s of the SPIE 10968, Behavior and Mechanics of Multifunctional Materials XIII 1096814, doi:10.1117/1.2.2513800.
- [26] S. Pisipati, J. Geer, B. Sammakia, B.T. Murray, A novel alternate approach for multiscale thermal transport using diffusion in the Boltzmann Transport Equation, Int. J. Heat. Mass Transf. 54 (15–16) (2011) 3406–3419, <https://doi.org/10.1016/j.ijheatmasstransfer.2011.03.046>.
- [27] S. Pisipati, C. Chen, J. Geer, B. Sammakia, B.T. Murray, Multiscale thermal device modeling using diffusion in the Boltzmann Transport Equation, Int. J. Heat. Mass Transf. 64 (2013) 286–303, <https://doi.org/10.1016/j.ijheatmasstransfer.2013.04.011>.
- [28] S.-N. Li, B.-Y. Cao, Fractional Boltzmann transport equation for anomalous heat transport and divergent thermal conductivity, Int. J. Heat. Mass Transf. 137 (2019) 84–89, <https://doi.org/10.1016/j.ijheatmasstransfer.2019.03.120>.
- [29] H. Wei, S. Zhao, Q. Rong, H. Bao, Predicting the effective thermal conductivities of composite materials and porous media by machine learning methods, Int. J. Heat. Mass Transf. 127 (2018) 908–916, <https://doi.org/10.1016/j.ijheatmasstransfer.2018.08.082>.
- [30] T.A. Alrebd, Y.S. Wudil, U.F. Ahmad, F.A. Yakasai, J. Mohammed, F.H. Kallas, Predicting the thermal conductivity of Bi<sub>2</sub>Te<sub>3</sub>-based thermoelectric energy materials: a machine learning approach, Int. J. Therm. Sci. 181 (2022) 107784, <https://doi.org/10.1016/j.ijthermalsci.2022.107784>.
- [31] K.-Q. Li, Y. Liu, Q. Kang, Estimating the thermal conductivity of soils using six machine learning algorithms, Int. Commun. Heat. Mass Transf. 136 (2022) 106139, <https://doi.org/10.1016/j.icheatmasstransfer.2022.106139>.
- [32] M. Ghobadi, F. Rahimzadeh Lotfabad, S.M. Zebarjad, R. Ebrahimi, Quantitative analyzing the effect of h-BN on the thermal conductivity of HDPE-BN composite through multi-objective optimization, Polym. Adv. Technol. 33 (9) (2022) 2966–2974, <https://doi.org/10.1002/pat.5762>.
- [33] C. Zhang, Y. Chen, H. Fang, Z. Tian, F. Meng, H. Lin, The effect of Si<sub>3</sub>N<sub>4</sub> on the thermal and dielectric properties of polytetrafluoroethylene/glass fiber composites, J. Mater. Sci. Mater. Electron. 32 (17) (2021) 21957–21965, <https://doi.org/10.1007/s10854-021-06611-0>.
- [34] S.-W. Xiong, P. Zhang, Q. Zou, Y. Xia, M.-y Jiang, J.-G. Gai, High thermal conductivity and electrical insulation of liquid alloy/ceramic/epoxy composites through the construction of mutually independent structures, Compos. Sci. Technol. 202 (2021) 108596, <https://doi.org/10.1016/j.compscitech.2020.108596>.
- [35] W. Zhang, C. Lu, M. Ge, F. Bu, J. Zhang, Surface modified and gradation-mixed Al<sub>2</sub>O<sub>3</sub> as an effective filler for the polyphenylene oxide (PPO) insulative layer in copper clad laminates, J. Mater. Sci. Mater. Electron. 31 (23) (2020) 21602–21616, <https://doi.org/10.1007/s10854-020-04673-0>.
- [36] W. Shen, W. Wu, C. Liu, Z. Wang, Z. Huang, Achieving a high thermal conductivity for segregated BN/PLA composites via hydrogen bonding regulation through cellulose network, Polym. Adv. Technol. 31 (9) (2020) 1911–1920, <https://doi.org/10.1002/pat.4916>.
- [37] L. Liu, C. Cao, X. Ma, X. Zhang, T. Lv, Thermal conductivity of polyimide/AlN and polyimide/(AlN + BN) composite films prepared by in-situ polymerization, J. Macromol. Sci., Part. A 57 (5) (2020) 398–407, <https://doi.org/10.1080/10601325.2019.1703555>.
- [38] X. He, D. Ou, Y. Ma, S. Wu, Y. Chen, Y. Luo, D. Wu, Enhancing the thermal conductivities of aluminum nitride-polydimethylsiloxane composites via tailoring of thermal losses in filler networks, Polym. Compos. 42 (3) (2020) 1338–1346, <https://doi.org/10.1002/pc.25904>.
- [39] Y. Yuan, Z. Li, L. Cao, B. Tang, S. Zhang, Modification of Si<sub>3</sub>N<sub>4</sub> ceramic powders and fabrication of Si<sub>3</sub>N<sub>4</sub>/PTFE composite substrate with high thermal conductivity, Ceram. Int. 45 (13) (2019) 16569–16576, <https://doi.org/10.1016/j.ceramint.2019.05.194>.
- [40] T. Chen, L. Deng, Thermal conductive and dielectric properties of epoxy resin with bimetal filler of Zn–Cu particle, J. Mater. Sci. Mater. Electron. 30 (10) (2019) 9775–9784, <https://doi.org/10.1007/s10854-019-01314-z>.
- [41] X. Yang, X. Yu, K. Naito, H. Ding, X. Qu, Q. Zhang, Enhanced thermal conductivity of polyimide composites filled with modified h-BN and nanodiamond hybrid filler, J. Nanosci. Nanotechnol. 18 (5) (2018) 3291–3298, <https://doi.org/10.1166/jnn.2018.14630>.
- [42] S.-H. Su, Y. Huang, S. Qu, W. Liu, R. Liu, L. Li, Microdiamond/PLA composites with enhanced thermal conductivity through improving filler/matrix interface compatibility, Diam. Relat. Mater. 81 (2018) 161–167, <https://doi.org/10.1016/j.diamond.2017.12.011>.
- [43] Z. Liu, R. Tu, Q. Liao, H. Hu, J. Yang, Y. He, H. Bian, L. Ma, W. Liu, High thermal conductivity of flake graphite reinforced polyethylene composites fabricated by the powder mixing method and the melt-extruding process, Polymers 10 (7) (2018) 693, <https://doi.org/10.3390/polym10070693>.
- [44] L. Guo, Z. Zhang, R. Kang, Y. Chen, X. Hou, Y. Wu, M. Wang, B. Wang, J. Cui, N. Jiang, C.T. Lin, J. Yu, Enhanced thermal conductivity of epoxy composites filled with tetrapod-shaped ZnO, RSC Adv. 8 (22) (2018) 12337–12343, <https://doi.org/10.1039/c8ra01470a>.
- [45] K. Srinivas, M.S. Bhagyashekhar, Effect of mono Al / Cu and hybrid (Al-Cu) particulates on Thermal conductivity of epoxy composites, Mater. Today Proc. 4 (9) (2017) 9519–9523, <https://doi.org/10.1016/j.matpr.2017.06.216>.
- [46] Kim, Y.-K., 2017. Induced clustering-enabled thermal transport enhancement in polymer composites for efficient thermal interface materials, IEEE, in: proceedings of the Sixteenth IEEE ITHERM Conference 347–351, doi:10.1109/ITHERM.2017.77992492.
- [47] J. Su, X. Liu, M. Charmchi, H. Sun, Experimental and numerical study of anisotropic thermal conductivity of magnetically aligned PDMS/Ni particle composites, Int. J. Heat. Mass Transf. 97 (2016) 645–652, <https://doi.org/10.1016/j.ijheatmasstransfer.2016.02.023>.
- [48] G.B. Olowojoba, S. Eslava, E.S. Gutierrez, A.J. Kinloch, C. Mattevi, V.G. Rocha, A. C. Taylor, In situ thermally reduced graphene oxide/epoxy composites: thermal and mechanical properties, Appl. Nanosci. 6 (7) (2016) 1015–1022, <https://doi.org/10.1007/s13204-016-0518-y>.
- [49] Q. Mu, Q. Li, H. Liu, Enhancing the thermal conductivities of SiO<sub>2</sub>/Epoxy composites by orientation, Polym. Compos. 37 (3) (2016) 818–823, <https://doi.org/10.1002/pc.23239>.
- [50] G. Hou, B. Cheng, F. Ding, M. Yao, P. Hu, F. Yuan, Synthesis of uniform alpha-Si<sub>3</sub>N<sub>4</sub> nanospheres by RF induction thermal plasma and their application in high thermal

- conductive nanocomposites, *ACS Appl. Mater. Interfaces* 7 (4) (2015) 2873–2881, <https://doi.org/10.1021/am5081887>.
- [51] B.Z. Gao, J.Z. Xu, J.J. Peng, F.Y. Kang, H.D. Du, J. Li, S.W. Chiang, C.J. Xu, N. Hu, X.S. Ning, Experimental and theoretical studies of effective thermal conductivity of composites made of silicone rubber and Al<sub>2</sub>O<sub>3</sub> particles, *Thermochim. Acta* 614 (2015) 1–8, <https://doi.org/10.1016/j.tca.2015.06.005>.
- [52] Y.-N. Chan, S.-C. Huang, H.-T. Li, Thermal conductivity of epoxy composites with controlled high loading of ceramic particles, *Jpn. Inst. Electron. Packag., ICEP-IAAC 2015 Proc.* (2015) 258–261, <https://doi.org/10.1109/ICEP-IAAC.2015.7111034>.
- [53] J. Zhang, S. Qi, Preparation and properties of silicon nitride/glass fiber/epoxy composites, *Polym. Compos.* 35 (7) (2014) 1338–1342, <https://doi.org/10.1002/pcc.22785>.
- [54] J. Hou, G. Li, N. Yang, L. Qin, M.E. Grami, Q. Zhang, N. Wang, X. Qu, Preparation and characterization of surface modified boron nitride epoxy composites with enhanced thermal conductivity, *RSC Adv.* 4 (83) (2014) 44282–44290, <https://doi.org/10.1039/c4ra07394k>.
- [55] Y.-X. Fu, Z.-X. He, D.-C. Mo, S.-S. Lu, Thermal conductivity enhancement with different fillers for epoxy resin adhesives, *Appl. Therm. Eng.* 66 (1–2) (2014) 493–498, <https://doi.org/10.1016/j.applthermaleng.2014.02.044>.
- [56] X.M. Dongsheng Xie, Jie Ma, Lin Zhu, Xiangqian Shen, Preparation and thermal conductivity of nickel fiber (powder)/silicone rubber composites induced by magnetic field, *Polym. Polym. Compos.* 22 (2014) 435–457, <https://doi.org/10.1177/096739111402200504>.
- [57] F.Ö. Özmuhiç, D. Balköse, Effects of particle size and electrical resistivity of filler on mechanical, electrical, and thermal properties of linear low density polyethylene-zinc oxide composites, *J. Appl. Polym. Sci.* 130 (4) (2013) 2734–2743, <https://doi.org/10.1002/app.39433>.
- [58] A. Ma, W. Chen, Y. Hou, Mechanical and thermal conductivities of MWCNTs/Si<sub>3</sub>N<sub>4</sub>/epoxy composites, *Polym. -Plast. Technol. Eng.* 52 (15) (2013) 1590–1594, <https://doi.org/10.1080/03602559.2013.828230>.
- [59] S.-Y. Chin, S.-J. Hsiang, Y.-T. Chen, W.-T. Yang, Study on thermal conductive BN/VGCF/polyimide resin composites (IMPACT 2013), *IEEE, Intern. Micro. Packag., Assem. Circuits Technol. Conf.* (2013) 315–318, <https://doi.org/10.1109/IMPACT.2013.6706664>.
- [60] X.H. Wan, L.G. Wang, B. Zuo, E.Z. Wang, G.Z. Zhu, Studies on the thermal conductivity of Al<sub>2</sub>O<sub>3</sub>/epoxy resin composite materials, *Adv. Mater. Res.* 535–537 (2012) 235–238, <https://doi.org/10.4028/www.scientific.net/AMR.535-537.235>.
- [61] R. Voo, M. Mariatti, L.C. Sim, Thermal properties and moisture absorption of nanofillers-filled epoxy composite thin film for electronic application, *Polym. Adv. Technol.* 23 (12) (2012) 1620–1627, <https://doi.org/10.1002/pat.3039>.
- [62] Z. Shi, R. Fu, S. Agathopoulos, X. Gu, W. Zhao, Thermal conductivity and fire resistance of epoxy molding compounds filled with Si<sub>3</sub>N<sub>4</sub> and Al(OH)<sub>3</sub>, *Mater. Des.* 34 (2012) 820–824, <https://doi.org/10.1016/j.matdes.2011.07.012>.
- [63] F. Ren, P.-g Ren, Y.-y Di, D.-m Chen, G.-g Liu, Thermal, mechanical and electrical properties of linear low-density polyethylene composites filled with different dimensional SiC particles, *Polym. Plast. Technol. Eng.* 50 (8) (2011) 791–796, <https://doi.org/10.1080/03602559.2011.551967>.
- [64] M.S. Nurul, M. Mariatti, Effect of thermal conductive fillers on the properties of polypropylene composites, *J. Thermoplast. Compos. Mater.* 26 (5) (2011) 627–639, <https://doi.org/10.1177/0892705711427345>.
- [65] T. Zhou, X. Wang, X. Liu, D. Xiong, Improved thermal conductivity of epoxy composites using a hybrid multi-walled carbon nanotube/micro-SiC filler, *Carbon* 48 (4) (2010) 1171–1176, <https://doi.org/10.1016/j.carbon.2009.11.040>.
- [66] W. Zhou, D. Yu, C. Min, Y. Fu, X. Guo, Thermal, dielectric, and mechanical properties of SiC particles filled linear low-density polyethylene composites, *J. Appl. Polym. Sci.* 112 (3) (2009) 1695–1703, <https://doi.org/10.1002/app.29602>.
- [67] Z. Wenyi, W. Caifeng, A. Qunli, O. Haiyan, Thermal properties of heat conductive silicone rubber filled with hybrid fillers, *J. Compos. Mater.* 42 (2) (2008) 173–187, <https://doi.org/10.1177/0021998307086184>.
- [68] Z. Wen-ying, Q. Shu-hua, W. You-ming, W. Cai-feng, Y. Jiang-long, G. Jian, Thermal conductive composite BN/HDPE plastics, *Polym. Mater. Sci. Eng.* 24 (2008) 83–86, <https://doi.org/10.16865/j.cnki.1000-7555.2008.02.021>.
- [69] E.-S. Lee, S.-M. Lee, D.J. Shanefield, W.R. Cannon, Enhanced thermal conductivity of polymer matrix composite via high solids loading of aluminum nitride in epoxy resin, *J. Am. Ceram. Soc.* 91 (4) (2008) 1169–1174, <https://doi.org/10.1111/j.1551-2916.2008.02247.x>.
- [70] G.-W. Lee, M. Park, J. Kim, J.I. Lee, H.G. Yoon, Enhanced thermal conductivity of polymer composites filled with hybrid filler, *Compos. Part A* 37 (5) (2006) 727–734, <https://doi.org/10.1016/j.compositesa.2005.07.006>.
- [71] Y.P. Mamuny, V.V. Davydenko, P. Pissis, E.V. Lebedev, Electrical and thermal conductivity of polymers filled with metal powders, *Eur. Polym. J.* 38 (2002) 1887–1897, [https://doi.org/10.1016/S0014-3057\(02\)00064-2](https://doi.org/10.1016/S0014-3057(02)00064-2).
- [72] Y. Xu, D.D.L. Chung, C. Mroz, Thermally conducting aluminum nitride polymer-matrix composites, *Compos., Part A* 32 (2001) 1749–1757, [https://doi.org/10.1016/s1359-835x\(01\)00023-9](https://doi.org/10.1016/s1359-835x(01)00023-9).
- [73] I.H. Tavman, Thermal and mechanical properties of copper powder filled poly (ethylene) composites, *Powder Technol.* 91 (1997) 63–67, [https://doi.org/10.1016/s0032-5910\(96\)03247-0](https://doi.org/10.1016/s0032-5910(96)03247-0).
- [74] Y. NAGAI, G.-C. LAI, Thermal conductivity of epoxy resin filled with particulate aluminum nitride powder, *J. Ceram. Soc. Jpn.* 105 (1219) (1997) 197–200, <https://doi.org/10.2109/jcersj.105.197>.
- [75] I.H. Tavman, Thermal and mechanical properties of aluminum powder-filled high-density polyethylene composites, *J. Appl. Polym. Sci.* 62 (12) (1996) 2161–2167, [https://doi.org/10.1002/\(sici\)1097-4628\(19961219\)62:12<2161::aid-app19>3.0.co;2-8](https://doi.org/10.1002/(sici)1097-4628(19961219)62:12<2161::aid-app19>3.0.co;2-8).
- [76] Y. Guo, K. Ruan, X. Shi, X. Yang, J. Gu, Factors affecting thermal conductivities of the polymers and polymer composites: a review, *Compos. Sci. Technol.* 193 (2020) 108134, <https://doi.org/10.1016/j.compscitech.2020.108134>.
- [77] R.J. Warzoha, A.S. Fleischer, Heat flow at nanoparticle interfaces, *Nano Energy* 6 (2014) 137–158, <https://doi.org/10.1016/j.nanoen.2014.03.014>.
- [78] K. Wattanakul, H. Manuspiya, N. Yanumet, Effective surface treatments for enhancing the thermal conductivity of BN-filled epoxy composite, *J. Appl. Polym. Sci.* 119 (6) (2010) 3234–3243, <https://doi.org/10.1002/app.32889>.
- [79] J. Wie, K. Kim, J. Kim, High thermal conductivity composites obtained by novel surface treatment of boron nitride, *Ceram. Int.* 46 (11) (2020) 17614–17620, <https://doi.org/10.1016/j.ceramint.2020.04.063>.
- [80] C. Wei, X. Xu, B. Wei, J. Cheng, P. Chen, Effect of diamond surface treatment on microstructure and thermal conductivity of diamond/W-30Cu composites prepared by microwave sintering, *Diam. Relat. Mater.* 104 (2020) 107760, <https://doi.org/10.1016/j.diamond.2020.107760>.
- [81] C. Pan, K. Kou, Q. Jia, Y. Zhang, G. Wu, T. Ji, Improved thermal conductivity and dielectric properties of hBN/PTFE composites via surface treatment by silane coupling agent, *Compos., Part B* 111 (2017) 83–90, <https://doi.org/10.1016/j.compositesb.2016.11.050>.
- [82] Ge.V. Fabian Pedregosa, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Matthieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, Jake Vanderplas, Alexandre Passos, David Cournapeau, Matthieu Brucher, Matthieu Perrot, Edouard Duchesnay, Scikit-learn: machine Learning in Python, *J. Mach. Learn. Res.* 12 (2011) 2825–2830, <https://doi.org/10.48550/arXiv.1201.0490>.
- [83] Y. Liu, Y. Wang, J. Zhang, New Machine Learning Algorithm: Random Forest, *ICICA (LNCS)*, Springer Berlin Heidelberg, Berlin, Heidelberg, 2012, pp. 246–252.
- [84] J.H. Friedman, Greedy function approximation: a gradient boosting machine, *Ann. Stat.* 29 (5) (2001) 1189–1232, <https://doi.org/10.1214/aos/1013203451>.
- [85] Chen, T., Guestrin, C., 2016. XGBoost: A scalable tree boosting system, 13–17, Proceedings of the Twenty Second ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 785–794, doi:10.1145/2939672.2939785.
- [86] Brindha, S., Gowriswari, S., 2022. Hyperparameters Optimization using Gridsearch Cross Validation Method for machine learning models in Predicting Diabetes Mellitus Risk, IEEE, In: Proceedings of the International Conference on Communication, Computing and Internet of Things (IC3IoT) 1–4, doi:10.1109/IC3IoT53935.2022.9768005.
- [87] M. Liang, B. An, K. Li, L. Du, T. Deng, S. Cao, Y. Du, L. Xu, X. Gao, L. Zhang, J. Li, H. Gao, Improving Genomic Prediction with Machine Learning Incorporating TPE for Hyperparameters Optimization, *Biol. (Basel, Switz.)* 11 (11) (2022) 1647, <https://doi.org/10.3390/biology1111647>.
- [88] X. Tian, M. Chen, Descriptor selection for predicting interfacial thermal resistance by machine learning methods, *Sci. Rep.* 11 (1) (2021) 739, <https://doi.org/10.1038/s41598-020-80795-z>.
- [89] X. Huang, S. Ma, Y. Wu, C. Wan, C.Y. Zhao, H. Wang, S. Ju, High-throughput screening of amorphous polymers with high intrinsic thermal conductivity via automated physical feature engineering, *J. Mater. Chem. A, Mater. Energy Sustain.* 11 (38) (2023) 20539–20548, <https://doi.org/10.1039/d3ta03370h>.
- [90] N. Yadav, N. Chakraborty, A. Tewari, Interval prediction machine learning models for predicting experimental thermal conductivity of high entropy alloys, *Comput. Mater. Sci.* 214 (2022) 111754, <https://doi.org/10.1016/j.commatsci.2022.111754>.

*Supporting Information for*

## **Application of Machine Learning in Predicting the Thermal Conductivity of Single-Filler Polymer Composites**

Yinzhou Liu<sup>a b</sup>, Weidong Zheng<sup>b</sup>, Haoqiang Ai<sup>b</sup>, Hao Zhou<sup>b</sup>, Liyin Feng<sup>b</sup>, Lin Cheng<sup>b\*\*</sup>, Ruiqiang Guo<sup>b</sup>, Xiaohan Song<sup>b\*</sup>

<sup>a</sup> *Institute of Advanced Technology, Shandong University, Jinan 250061, China*

<sup>b</sup> *Shandong Institute of Advanced Technology, Jinan 250100, China*

\* Corresponding author.

\*\* Corresponding author.

E-mail address: [cheng@sdu.edu.cn](mailto:cheng@sdu.edu.cn) (L. Cheng), [xiaohan.song@iat.cn](mailto:xiaohan.song@iat.cn) (X. Song).

**Table S1. Definition of descriptors.**

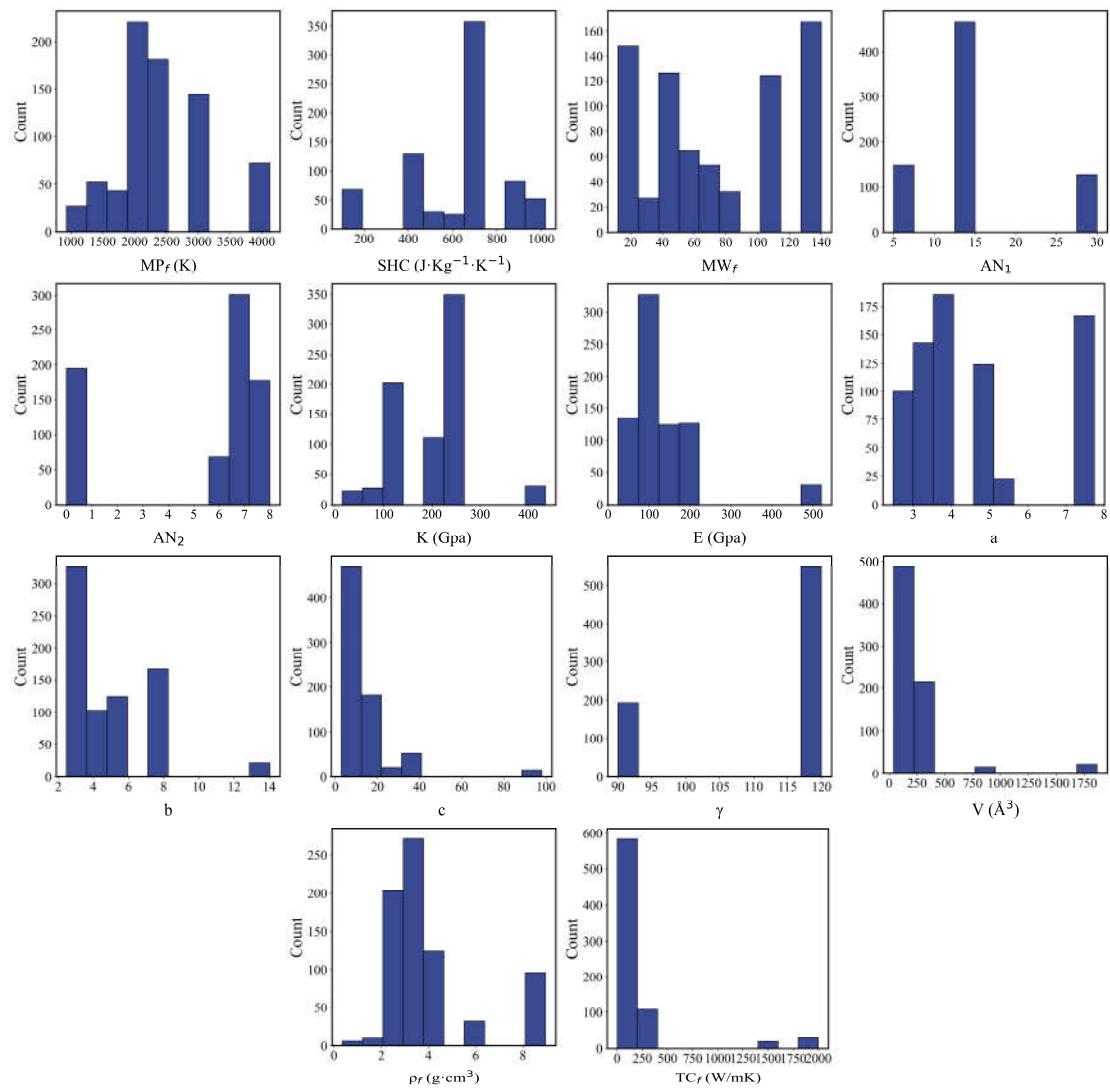
Descriptors		Definition
	$\rho_f$	Density of filler. (g/cm <sup>3</sup> )
	TC <sub>f</sub>	Thermal conductivity of filler. (W/mK)
	MP <sub>f</sub>	Melt point of filler. (K)
	SHC	Specific heat capacity of filler. (J/(Kg • K))
	K	Bulk modulus of filler. (Gpa)
	E	Shear modulus of filler. (Gpa)
	AN <sub>1</sub>	The atomic number of the first atom from left to right in the chemical formula of filler.
Filler	AN <sub>2</sub>	The atomic number of the second atom from left to right in the chemical formula of filler, if the filler consists of only one element, then AN <sub>2</sub> is 0.
	MW <sub>f</sub>	Molecular weight of filler.
	a	Lattice constant, side length of cell. (Å)
	b	Lattice constant, side length of cell. (Å)
	c	Lattice constant, side length of cell. (Å)
	$\gamma$	Lattice constant, the angle between b and c. (°)
	V	Uni cell volume of filler. (Å <sup>3</sup> )
	$\rho_m$	Density of matrix. (g/cm <sup>3</sup> )
	TC <sub>m</sub>	Thermal conductivity of matrix. (W/mK)
Matrix	T <sub>g</sub>	Glass transition temperature of matrix. (K)
	DC	Dielectric constant.
	MP <sub>m</sub>	Melt point of matrix. (K)
	T	Temperature at which the thermal conductivity is measured. (K)
Process	FS	Shape of filler.
	d	Average particle size of filler. (μm)
	vol	Volume fraction of filler. (%)
	MW	Molecular weight of surface modification reagent.
	CH <sub>3</sub>	Methyl number of surface modification reagent.
	MeO	Number of methyl and ethyl groups in surface modification reagent.
Surface modification	TPSA	Topological polar surface area of surface modification reagent. (Å <sup>2</sup> )
	Lc	Number of long-chain in surface modification reagent.
	SB	Number of single-bond in surface modification reagent, except C-H bond.
	DB	Number of double-bond in surface modification reagent, including carbon-carbon double bonds and carbon-oxygen double bond.

**Tabel S2. The molecular formula of surface treatment reagents.**

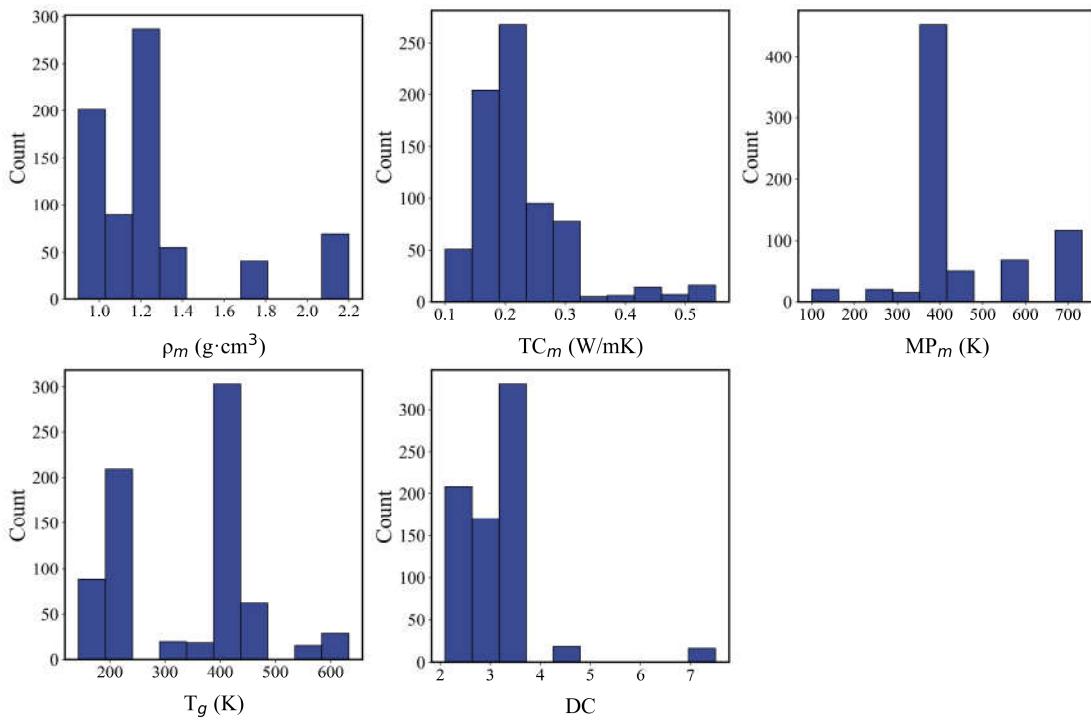
Surface treatment reagents	Molecular formula
F8261	C <sub>14</sub> H <sub>25</sub> F <sub>13</sub> O <sub>3</sub> Si
KH550	C <sub>9</sub> H <sub>23</sub> NO <sub>3</sub> Si
KH560	C <sub>9</sub> H <sub>20</sub> O <sub>5</sub> Si
KH570	C <sub>9</sub> H <sub>18</sub> O <sub>5</sub> Si
S32	C <sub>6</sub> H <sub>14</sub> O <sub>2</sub> Si
Titanate coupling agent	C <sub>12</sub> H <sub>28</sub> O <sub>4</sub> Ti
JH-V171	C <sub>5</sub> H <sub>12</sub> O <sub>3</sub> Si

**Table S3. The features obtained after the selection process.**

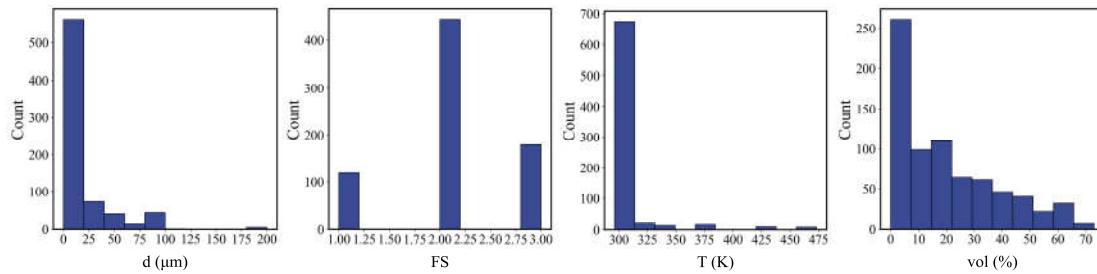
Classification	Descriptors
Process descriptors	T, FS, d, vol.
Filler descriptors	$\rho_f$ , TC <sub>f</sub> , MP <sub>f</sub> , SHC, K, E, AN <sub>2</sub> , MW <sub>f</sub> , c, V
Matrix descriptors	$\rho_m$ , TC <sub>m</sub> , T <sub>g</sub> , DC, MP <sub>m</sub>
Surface modification descriptors	MeO, TPSA



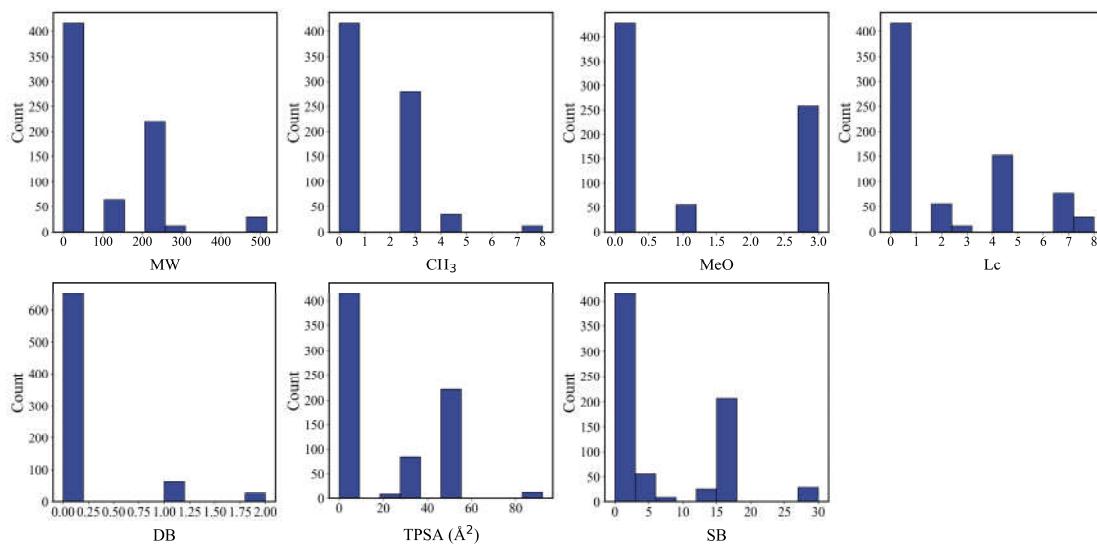
**Fig.S1.** Data distribution of filler descriptors



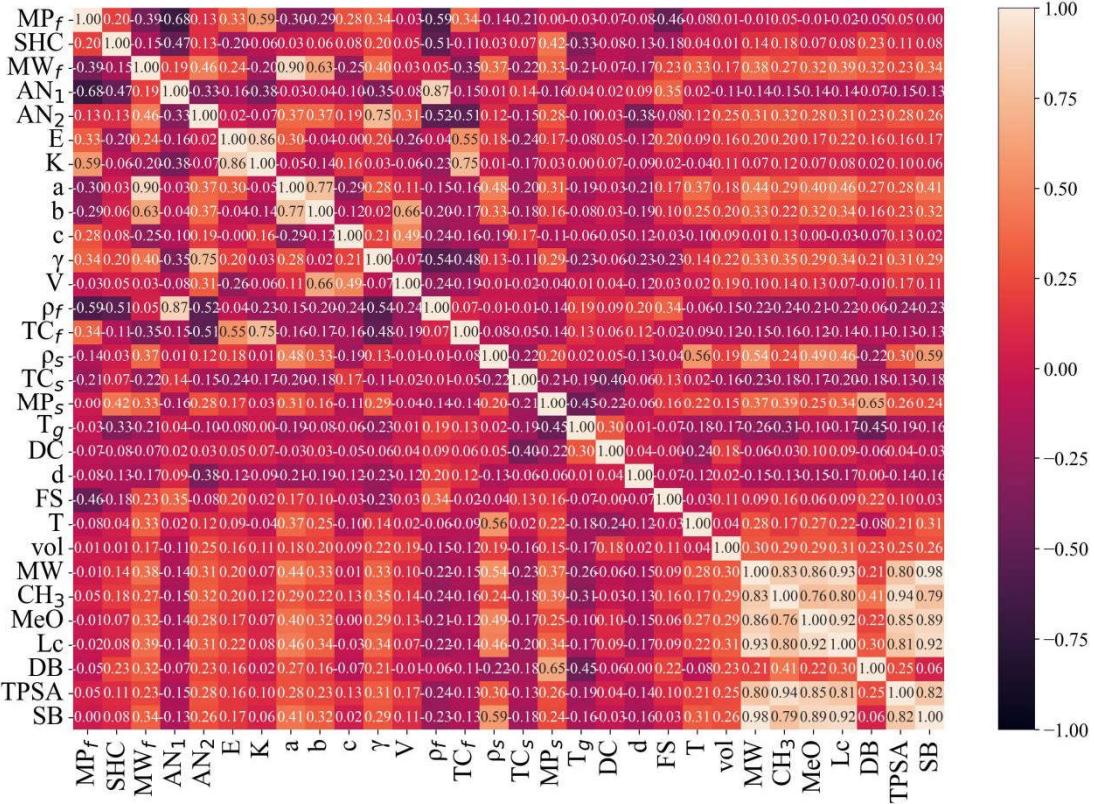
**Fig.S2.** Data distribution of matrix descriptors



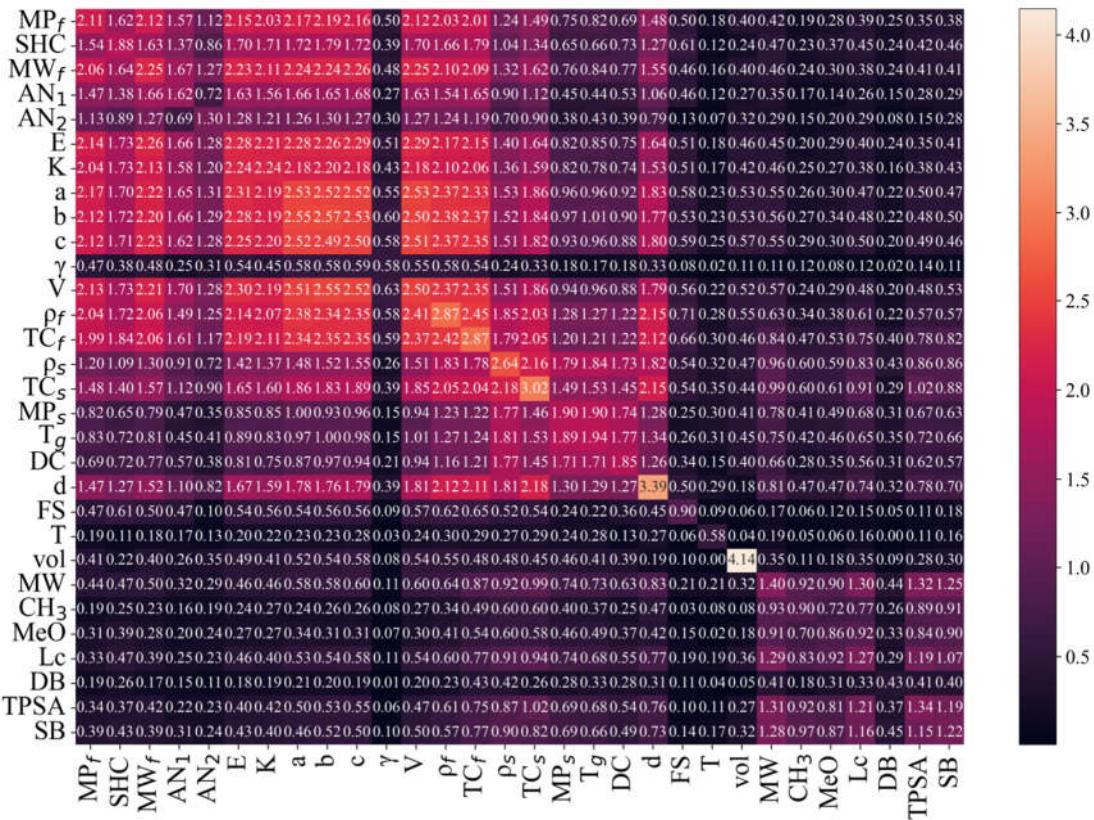
**Fig.S3.** Data distribution of process descriptors



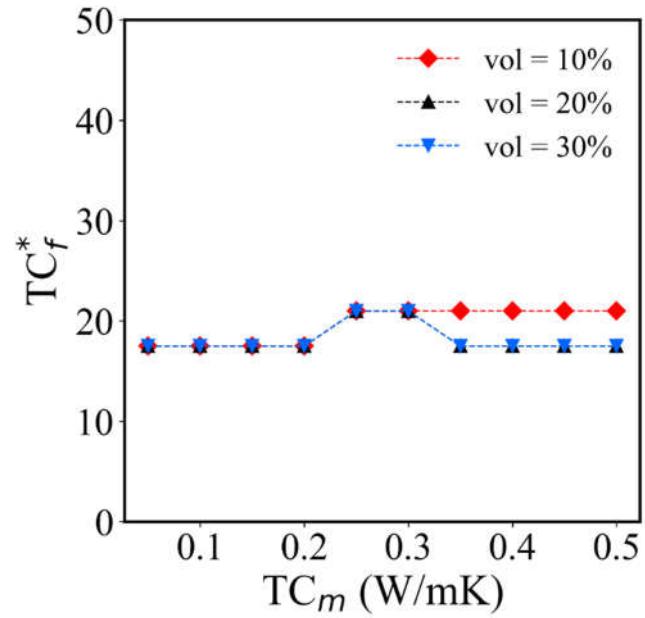
**Fig.S4.** Data distribution of surface modification descriptors



**Fig.S5.** Pearson correlation coefficient heat map between descriptors after feature engineering.



**Fig.S6.** Mutual information values heat map between descriptors after feature engineering.



**Fig. S7.** Effect of  $TC_m$  and vol on thresholds of  $TC_f$ .