Zhang *et al. J Mater Inf* 2024;4:8
DOI: 10.20517/jmi.2024.13

Journal of Materials Informatics

Research Article

Open Access

Check for updates

Thermal stability prediction of copolymerized polyimides via an interpretable transfer learning model

Yu Zhang[1], Yating Fang[1], Ling Li[1], Tongle Xu[1], Fang Peng[2], Xiong Li[1], Guangrui Xu[3], Wei Lv[3], Minjie Li[1], Peng Ding[1,*]

[1]Department of Chemistry, College of Sciences, Shanghai University, Shanghai 200444, China.
[2]School of Materials Science and Engineering, Shanghai University, Shanghai 200444, China.
[3]Shanghai Plastics Research Institute Co., Ltd, Shanghai Huayi Holdings Group, Shanghai 200090, China.

[*]**Correspondence to:** Prof. Peng Ding, Department of Chemistry, College of Sciences, Shanghai University, 99 Shangda Road, Baoshan District, Shanghai 200444, China. E-mail: dingpeng@shu.edu.cn

How to cite this article: Zhang Y, Fang Y, Li L, Xu T, Peng F, Li X, Xu G, Lv W, Li M, Ding P. Thermal stability prediction of copolymerized polyimides via an interpretable transfer learning model. *J Mater Inf* 2024;4:8. https://dx.doi.org/10.20517/jmi.2024.13

Received: 30 Apr 2024 **First Decision:** 28 May 2024 **Revised:** 5 Jun 2024 **Accepted:** 13 Jun 2024 **Published:** 17 Jun 2024

Academic Editor: Xingjun Liu **Copy Editor:** Pei-Yun Wang **Production Editor:** Pei-Yun Wang

Abstract

To address the issues with molecular representation of copolymerized polyimides (PIs) and the mini dataset of PI powders. We constructed an interpretable machine learning (ML) model for PI films using the weighted-additive Morgan Fingerprints with Frequency descriptors and developed an interpretable transfer learning model for PI powders. To enhance Thermal Stability (Temperature at 5% weight loss) of PI films and powders, it is recommended to add conjugated functional groups to diamines, control phenyl ring side chains, and reduce pyridine and hydroxyl groups; select copolyimides (co-PIs); ensure that anhydride is directly connected to the benzene ring in dianhydrides, avoiding aliphatic cycles. It is noteworthy that the close alignment between experimental results and model predictions serves to confirm the model is a reliable prediction tool. It is hoped that this polymer informatics approach will provide further implementation for practical applications of other functional materials.

Keywords: Transfer learning models, copolymerized polyimides, thermal stability

INTRODUCTION

The dream of chemists is to create substances to advance human knowledge and benefit society[1]. Traditionally, chemists have used experiments to achieve these goals. However, this approach is labor-intensive. As we stand on the cliff edge of artificial general intelligence, the collaborative potential between machine learning (ML) and chemistry is vast and promising[2-4].

Accurate and effective property prediction is crucial for polymer design in various applications, including polymer electrolytes[5,6], energy storage[7,8], organic optoelectronics[9,10], and many other applications[11,12]. The demand in these fields for enhanced thermally stable polymers that preserve their useful properties at high temperatures has been significantly increased[13-16]. Polyimides (PIs) are a class of high-performance polymers widely used in the aerospace industries and microelectronics[17-20]. Thermal distortion temperature, good chemical resistance and high mechanical strength are brought about by rigid aromatic heterocyclic structure and strong chain-chain interaction[21-26]. Temperature at 5% weight loss ($T_{5\%}$) refers to the temperature at which a material's mass decreases by 5% during heating, a key indicator of suitability of PIs for high-temperature applications. Therefore, studying the $T_{5\%}$ not only aids in optimizing the formulations of existing materials but may also drive the development of new high-performance PIs. Different dianhydrides and diamines are selected to tailor the properties of PIs. For example, diacid anhydride and m-phenylene diamine yield polyetherimide with excellent toughness and rigidity[27]; pyromellitic dianhydride and 4,4'-oxy-dianiline can be used to produce Kapton™[28] with high thermal resistance and mechanical strength. Experimental studies on the reactions between dianhydrides and diamines have already produced a range of multifunctional PIs that meet diverse requirements. However, it is not feasible for experimenters to exhaust all possible two-component reactions for material design and the discovery of novel PIs[29].

Polymer informatics[30-32] is advancing research on high-performance polymers, particularly in building correlations between the chemical structure and properties of polymers[33-35]. Researchers[36,37] have proposed a theoretical relationship between the chemical structure of polymers and their glass transition temperature (*T*g). These empirical methods assume that the chemical groups within the repeating units of the polymer chain influence *T*g through additive effects with varying weighting factors. For the Quantitative Structure-Property Relationship (QSPR) approach, a large array of molecular descriptors is extracted from the repeating units of polymers, applicable to any chemical structure[38]. For instance, Katritzky *et al.* have extracted over 400 constitutional, geometrical, topological, and quantum chemical descriptors for the repeating units of polymers[38]. Subsequently, multi-step linear regression analysis is employed to train these descriptors, resulting in a strong match between predicted and experimental for 88 homopolymers. However, quantum chemical descriptors require time-consuming density functional theory calculations, and other molecular descriptors can generate many parameters difficult to physically interpret, such as Kier shape indices and topological bond connectivity[36]. Rationally representing polymer maps in a continuous vector space is crucial for applying ML tools in polymer performance prediction. Fingerprints have been proven effective in molecular ML models and have been introduced for polymer-related tasks[39]. Tao *et al.* employed the Morgan fingerprint to represent the simplified molecular-input line-entry system (SMILES) of PI repeating units, counting the occurrence frequency of each substructure[29]. This method engendered a vector containing 121 prominent substructures, effectively reflecting the structural characteristics of the polymers. They established a comprehensive library of PIs based on dianhydrides and diamines/diisocyanates, covering eight million hypothetical PIs. The entire hypothetical dataset was screened, and molecular dynamics validation was conducted. Using graphs as input for molecular representation learning[36] has been proven to be a viable option. Volgin *et al.* developed a graph convolutional neural network to estimate the *T*g of PI with a mean absolute error of approximately 20 K[40]. However, the effect of

graph representation learning on small data sets and the translation of implicit "machine language"[41] into chemical knowledge with clear physicochemical meaning related to target properties still need to be explored. Huang *et al.* proposed a high-throughput screening framework for polymer chains with thermal conductivity (TC) via physical feature engineering[42]. They reduced 320 initial physical descriptors to 20, enabling ML models to achieve an R^2 over 0.80, outperforming traditional graph descriptors. Many efforts have begun to apply Transformer to predict the properties of small organic molecules[43-45]. Due to its complex self-attention mechanism, the model has limited interpretability[46], requires substantial training data, and demands high computational resources. According to Occam's Razor principle[47], ML should employ simple methods as much as possible to solve complex problems.

Currently, Zhang *et al.* proposed a materials genome approach to design and screen silicon-containing acetylene resins with excellent processing properties and heat resistance, but there are a few studies using ML methods to predict $T_{5\%}$[48]. Using only repeating units as research objects may result in missing structural information[49]; synthetic monomers, as substitutes, can solve this problem to a certain extent. Additionally, the issue of copolymerization rates in copolyimides (co-PIs) needs attention, which is often overlooked when using repeating units alone, and no ML studies have yet addressed it. ML studies on PIs often do not specify material types due to limited data, especially for PI powders. The "transfer learning" approach has been extensively tested in research dedicated to predicting the properties of small molecules[50-53]. Wu *et al.* implemented a transfer learning scheme to mitigate the small dataset issue for TC and developed an alternative model based solely on 28 experimental data[54]. Similarly, training on a large dataset of film materials and then fine-tuning for the powder model could be an effective strategy to improve the accuracy of the powder model.

This study aims to explore and develop ML methods for PI powders, enhancing material performance prediction accuracy and efficiency through improved molecular representation techniques and transfer learning models. Additionally, it focuses on enhancing model interpretability to provide actionable insights for materials scientists and engineers. We propose the flowchart of Figure 1: (1) A new method of weighted-additive Morgan Frequency Fingerprints (MFF) for both homopolyimides (homo-PIs) and co-PIs, using the copolymerization rates of dianhydrides and diamines as weights, and the MFF is added to form new descriptors; (2) Training a ML model on PI films and transferring it to the PI powder model to solve the problem of small data sets; (3) Utilizing expert knowledge and the SHapley Additive exPlanations (SHAP) method to interpret the structural characteristics related to performance in both films and powders; (4) The prediction results are highly consistent with the experimental values, and the model has good practical value. This research not only demonstrates the transformative potential of ML in chemical research but also provides new avenues for accelerating scientific discovery and interdisciplinary collaboration.

MATERIALS AND METHODS

Raw molecular dataset and data cleaning

A dataset comprising 270 entries on the T5% $T_{5\%}$ (N_2 atmosphere) of experimentally synthesized PI films, involving various combinations of dianhydrides and diamines, was compiled[25,55-97]. The data are presented in Supplementary Table 1 and illustrated in Supplementary Figures 1 and 2. Experimental procedures are prone to errors, potentially affecting the collected data and making it susceptible to noise. Handling outliers is crucial for ensuring data quality and consistency in model training. Supplementary Figure 3 shows that an outlier is identified when applying the 1.5 * interquartile range (IQR) rule to draw box plots of $T_{5\%}$. In addition, the outcomes of utilizing the method of mean ± 3 standard deviations to pinpoint the outlier are uniform. After removing the outlier, the histogram in Figure 2A reveals that $T_{5\%}$ ranges from 325 to 600 °C, predominantly clustering between 425 and 575 °C.

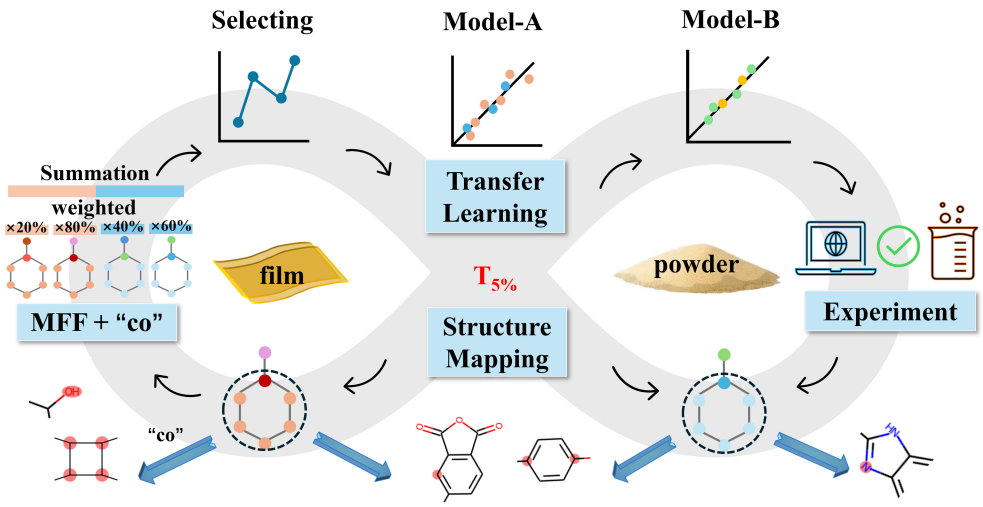

Figure 1. The workflow of our work. MFF: Morgan Frequency Fingerprint.

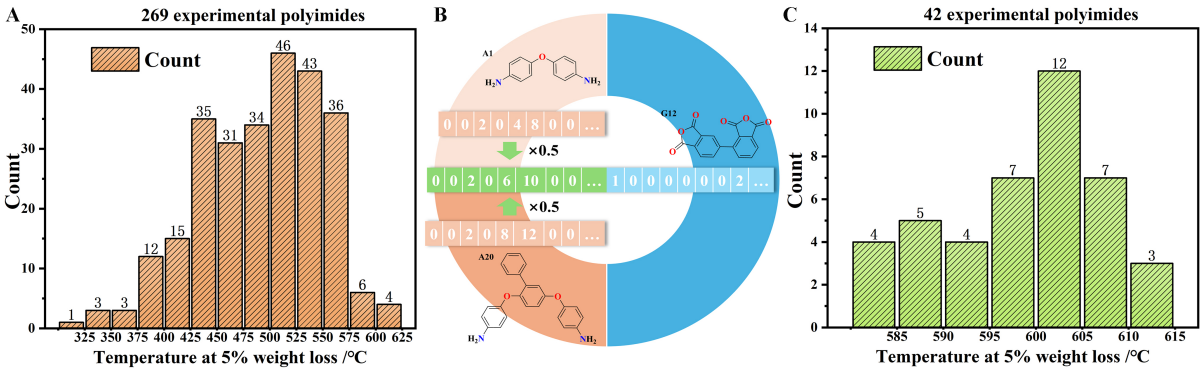

Figure 2. Distribution histogram of the experimental values of $T_{5\%}$ of (A) 269 PI films (outliers deleted) and (C) 42 PI powders. (B) Weighted-additive co-PI MFF molecular representation learning method. PI: Polyimide; MFF: Morgan Frequency Fingerprint.

Descriptor factory

A total of 22 dianhydrides and 67 diamines molecular structures generated 327 and 897 MFFs, respectively. More details on MFFs can be found in Supplementary Material and Supplementary Figure 4. The MFFs[98] were "weighted-additive", as shown in Figure 2B, where the copolymerization rates of diamines were used as weights and then summed to form a new MFF, which was overlaid with the diamines. Similarly, diamines can be integrated with the "weighted-additive" MFF of two dianhydrides. It is important to note that the MFF of the diamines should precede that of the dianhydrides in the combination. Considering the differences between copolymers and homopolymers, an additional "copolymer" descriptor was added. "Copolymer" is one-hot encoded, with co-PIs represented as "1" and homo-PIs as "0". Redundancy may increase the difficulty of data processing. During data cleaning, incomplete features, such as zero values (which do not contain structural characteristics), can be removed. Additionally, 1,225 descriptors were preprocessed, and those with a Pearson correlation coefficient greater than 0.90 were eliminated. The descriptors were subsequently normalized through min-max scaling to confine them within standard ranges[99]. Randomly allocate 85% of the data for the training set and the remaining 15% for the test set.

ML process

These data were input into our ML framework. For the 161 descriptors remaining after preprocessing, nested recursive feature elimination with Ridge, Random Forest, XGBoost, and CatBoost models (RFE-Ridge, RFE-RF, RFE-XGBoost, and RFE-CatBoost) was used to determine the optimal feature subset. The construction of Model-A adheres to the best model for feature selection, which is CatBoost. Individual descriptors from model-A were transferred and combined with model-C individualized descriptors to construct a predictive model for PI powder $T_{5\%}$ (model-D). Model parameters were fine-tuned. One of the critical challenges is opening the black box of ML. The SHAP method is used to nest the above models, providing feedback on descriptor contributions and their positive or negative impacts. The meaning of the filtered descriptors is also discussed. All the above code is executed in Python.

Experimental synthesis and validation

PI powder was prepared using two-step and thermo-imide methods. Some N,N-dimethylacetamide (DMAc) was added to the reactor under N_2 protection. Then, dianhydride and diamine in a certain proportion were added and stirred well. The reaction feeding ratio was shown in Supplementary Table 2 and Table 1. The reaction temperature was controlled, and the viscous precursor polyamide-acid (PAA) was obtained by stirring for 3 h at room temperature. A certain amount of xylene was continued to be added to the reaction system, and the temperature was raised to the reflux temperature and kept for 3 h, and PI powders were obtained using a thermo-imide approach. Materials and test methods are found in the Experimental section of the supporting information. The histogram in Figure 2C shows that $T_{5\%}$ ranges from 580 to 615 °C. Supplementary Figures 5 and 6 are the synthetic routes of PI-a and PI-b, respectively. Experimental section in support information can be viewed.

RESULTS AND DISCUSSION

Feature engineering

The optimal subset is carefully selected to reduce the dimensionality of the data. This subset contains a minimal number of descriptors while maintaining a close relationship to the required material properties[100,101]. The 1,225 descriptors declined to 161 after removing the redundancy with the Pearson correlation coefficient of less than 0.90. Four nesting methods, RFE-Ridge, RFE-RF, RFE-XGBoost, and RFE-CatBoost were used to compare and determine the optimal descriptor sets. The average R^2 and root mean square error (RMSE) of five-fold cross-validation (5-fold-CV) were determined as the metrics to evaluate the model performance. For the RFE-Ridge method, Supplementary Figure 7A indicates no inflection points. This observation suggests that nonlinear algorithms might not be ideal for handling high-dimensional datasets. Considering the selections from the four models, further increasing the number of descriptors to achieve better model results will not be pursued. An effective descriptor must be simpler to obtain than the target property and be as low-dimensional as possible[102]. A subset of 18 selected descriptors was identified as the optimal solution. The other three methods selected optimal feature subsets of 18, 12, and nine descriptors, respectively [Supplementary Table 3, Supplementary Figure 7B and C, Figure 3A]. Guided by the "Occam Shaver" principle[47], the smallest number of descriptors necessary for sufficient accuracy is selected. The RFE-CatBoost model, which offers the best performance with the fewest descriptors, achieves optimal values of 28.84 for RMSE and 0.75 for R^2. Among the subsets selected by RFE-CatBoost [Supplementary Table 3], the descriptors selected twice by the other three were the most. The same preprocessing and feature engineering were applied to the PI powder dataset, reducing 643 descriptors to 11, and ultimately determining an optimal subset of six descriptors based on Supplementary Figure 8A [Supplementary Table 4]. The heat map in Supplementary Figure 9 indicates that the correlation coefficients between descriptors are generally low across both datasets, suggesting a "global" representation.

Table 1. Structures of co-PI powders at different molar ratios, along with the $T_{5\%}$ values predicted by model-B and validated experimentally

No.	A1	ratio	A2	ratio	G1	ratio	G2	ratio	$T_{5\%}^{pred}$	$T_{5\%}^{exp}$	RMSE
a	7	1			7	0.8	12	0.2	612	600	8.9
b	7	0.7	23	0.3	7	1			610	606	

The corresponding structure codes can be found in Supplementary Figures 2 and 3, where the structures are detailed. PI: Polyimide; RMSE: root mean square error.

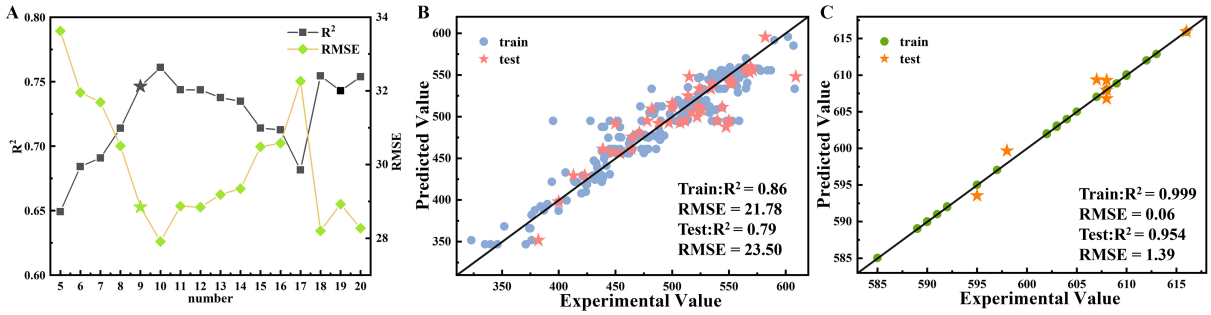

Figure 3. Descriptor sniping and optimal models. (A) The changes in R^2 and RMSE for model-A (5-fold-CV), built using RFE nested CatBoost with different numbers of descriptors, were accurately described. Scatter plots of the best model to predict the $T_{5\%}$ values versus their experimental values for the train and test sets for (B) model-A and (C) model-D. RMSE: Root mean square error; 5-fold-CV: five-fold cross-validation; RFE: recursive feature elimination.

Performance of the four models

The CatBoost algorithm was selected for modeling, with hyperparameters optimized using the grid search method [Supplementary Table 5]. Figure 3B shows the plots of the experimental $T_{5\%}$ values that match the predicted $T_{5\%}$ values. The R^2 and the RMSE of the CatBoost model (model-A) are 0.86 and 21.78, respectively. For the independent test set, the R^2 and the RMSE of Model-A are 0.79 and 23.50. The dataset was randomly split 20 times to ensure the reliability of the ensemble model. As shown in Supplementary Table 6, the average R^2 and RMSE for the training set are 0.86 and 21.60, respectively, while the test set achieved an average R^2 of 0.74 and RMSE of 28.40. The results of the existing model predicting the external test set[103] are shown in Supplementary Table 7. Zhang *et al.* constructed an artificial neural network model based on molecular descriptors to predict the $T_{5\%}$ of silicon-containing aromatic acetylene resins[48]. The R^2 values of their training and test sets were 0.66 and 0.65, respectively. Model-A was described as robust, stable, capable of good generalization and accurate.

In order to address the issue with the small dataset (powders), all descriptors and model parameters from model-A were transferred to model-B. In Supplementary Figure 8B, model-B achieved an R^2 of 0.88 and an RMSE of 2.97 on the training set, and an R^2 of 0.77 and an RMSE of 3.13 on the test set. In contrast to powders, the film formation process undergoes heat treatment, which affects the degree of cross-linking[104] and the *d*-space[105] between polymer chains, thereby influencing its thermal stability and performance characteristics. For the different effects caused by film and powder morphology, descriptors were ranked in model-A [Supplementary Table 8]. These descriptors were then transferred to a separately built powder, model-C [Supplementary Figure 8C], resulting in model-D. The optimal model parameters are shown in Supplementary Table 6. Model-D demonstrated outstanding performance with an R^2 of 0.999 and an RMSE of 0.06 on the training data. When evaluated on the independent test set, it achieved an R^2 of 0.954 and an RMSE of 1.39. These results are shown as a scatter plot in Figure 3C. As shown in Supplementary Table 9, the data set was randomly divided six times, resulting in an average R^2 and RMSE of 0.99 and 0.07,

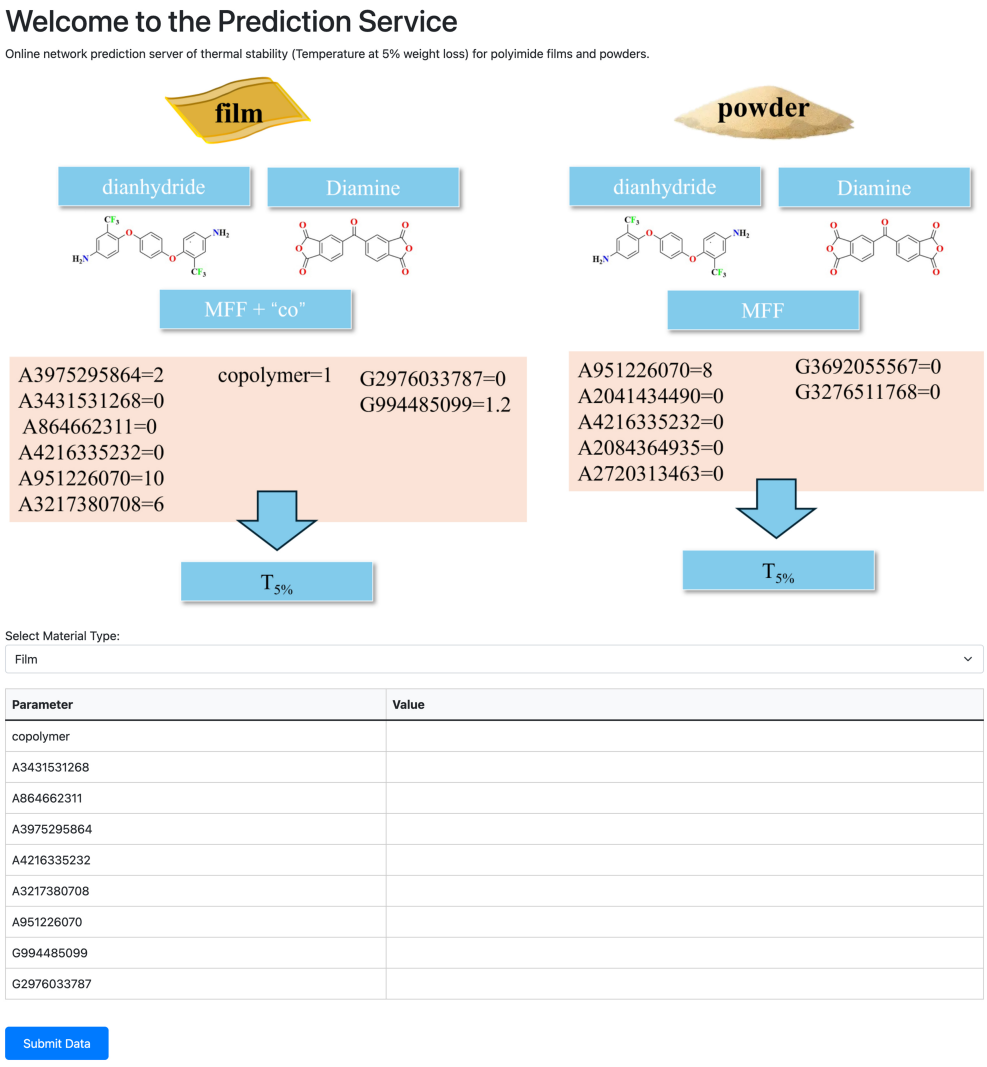

Figure 4. A prediction toolkit for predicting $T_{5\%}$ of PI films and PI powders from descriptor values. PI: Polyimide.

respectively, for the training sets and 0.88 and 2.44, respectively, for the test sets. Due to the insufficient size of the dataset, generalizability was somewhat affected, yet the results remain commendable. Zhang *et al.* constructed an artificial neural network model based on molecular descriptors to predict the $T_{5\%}$, with training and test set R^2 values of 0.66 and 0.65[48]. The transferred model, model-D, is more robust and stable.

To improve the model's user-friendliness, an interactive online prediction tool was developed. As illustrated in Figure 4, researchers can effortlessly leverage this toolkit by entering descriptor values and clicking on "Predict". This straightforward prediction tool allows researchers to quickly predict the $T_{5\%}$ of a candidate PI, whether in film or powder form (https://github.com/TanMengyuan/online_model_predict).

Model interpretation and structure mapping

To further understand the implications of structural descriptors of PI monomers in $T_{5\%}$ and offer guidance for experimental synthesis, SHAP and expert knowledge were leveraged to interpret the meaning of descriptors. Figure 5 illustrates that eigenvalues colored blue correspond to lower model output values,

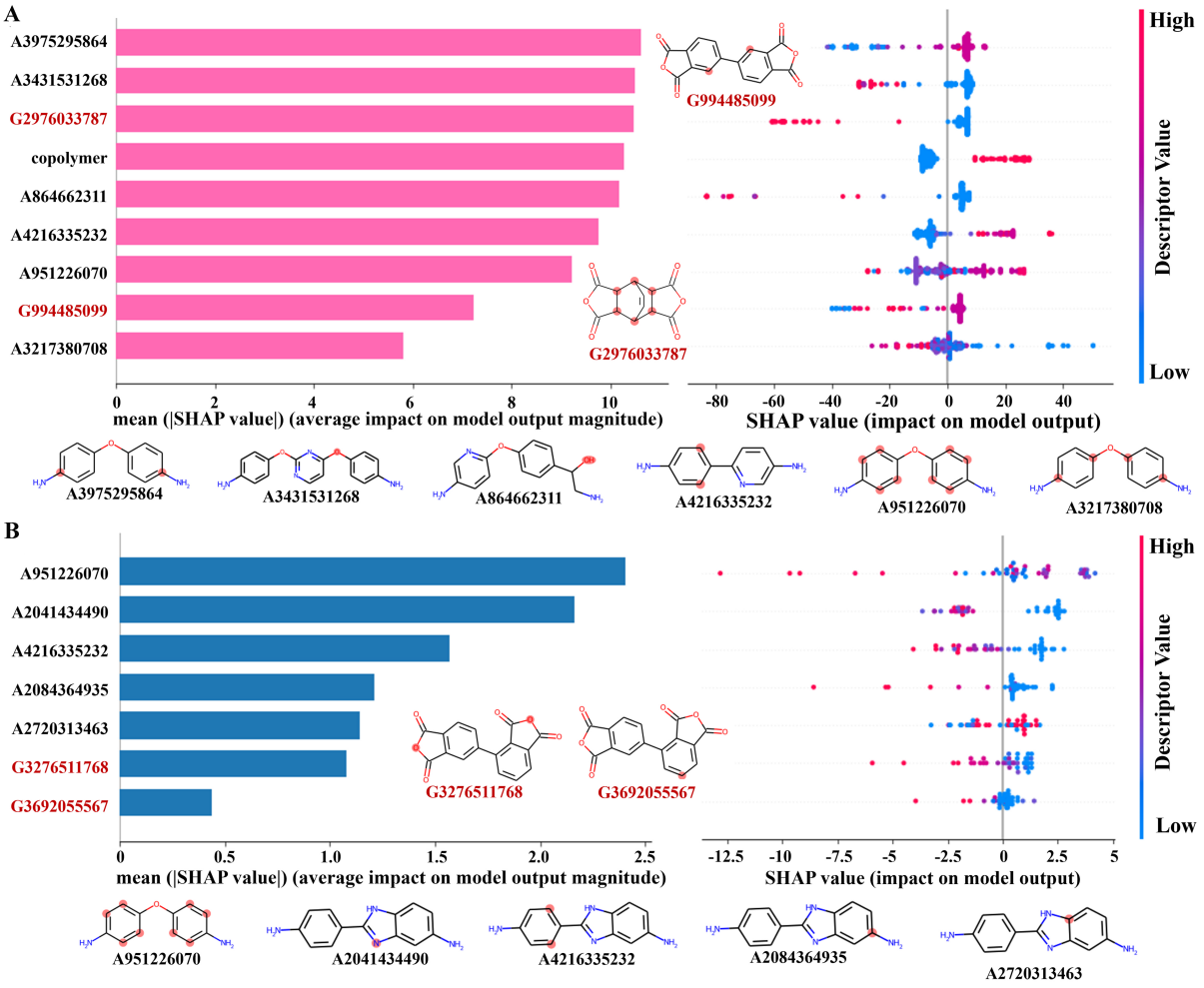

Figure 5. SHAP value of the output of the best descriptors of the SHAP nested CatBoost algorithm for (A) model-A and (B) model-D. On the left, the importance ranking of the best-chosen subset of the SHAP nested CatBoost algorithm. On the right, blue color for negative correlation and red color for positive correlation. SHAP: SHapley Additive exPlanations.

while red ones indicate higher output values. A3975295864, copolymer, and A864662311 have obvious positive effects on $T_{5\%}$, A3431531268, G2976033787, and A4216335232 have a negative correlation with $T_{5\%}$. A951226070 and G994485099 have a positive correlation trend, while A3217380708 has a negative correlation trend in Figure 5A. In Figure 5B, A951226070, A2041434490, A4216335232, A2084364935, A2720313463, and G3692055567 demonstrate a negative effect on $T_{5\%}$ of PI powders, while G3276511768 is shown to positively influence. In it, A951226070 appears in both forms of PI. Figure 5 shows the importance of each filter descriptor for $T_{5\%}$. The larger the SHAP value, the greater the impact of the feature on the model's objectives, warranting more focus on it. The effect of amines on $T_{5\%}$ is more acute than that of anhydrides. The SHAP method facilitated the establishment of a link between $T_{5\%}$ and descriptors, creating a bridge from descriptors to molecular structure, which then directly correlated with $T_{5\%}$ [Figure 1].

A3975295864 represents a benzene ring connected to an amino group. Supplementary Figure 10 shows the diamine molecule with the A3975295864 structure highlighted. Figure 6A demonstrates that A3975295864 exhibits a clear positive correlation, with SHAP values increasing as it increases. This is consistent with practical applications. For diamines, enhancing the π-π interactions between molecules by increasing the

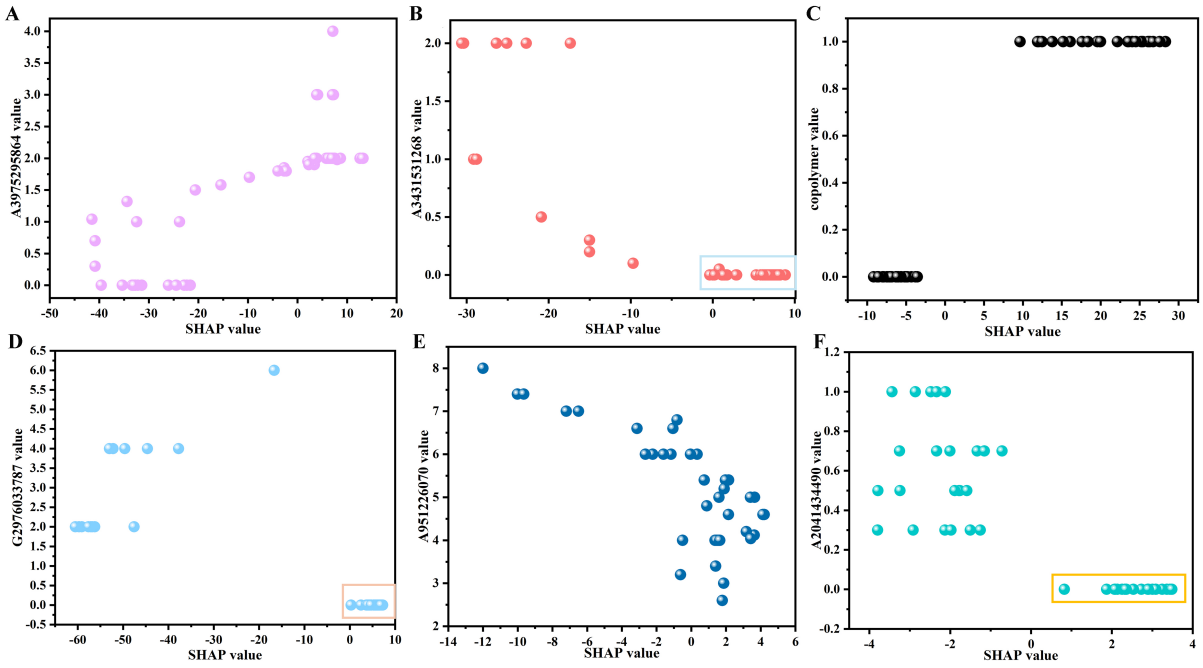

Figure 6. The relationship between the descriptor values and the SHAP values output by the SHAP nested CatBoost algorithm for the four best descriptors of model A (A) A3975295864 (B) A3431531268 (C) copolymer (D) G2976033787 and the two best descriptors of model-B (E) A951226070 (f) A2041434490. SHAP: SHapley Additive exPlanations.

rigid naphthalene rings and reducing the flexible ether bonds in the structure, as seen in 4,4'-(2,6-naphthalenediyl)bis[benzenamine]) (NADA)[58], can lead to a denser stacking of the molecular chains. This orderly and dense arrangement helps to enhance the thermal stability of the polymer, as the more rigid molecular chains reduce internal rotation, thereby strengthening the overall structural stability. A64 and A21 have similar structures, but A64 contains two A3975295864 units while A21 does not. Correspondingly, the $T_{5\%}$ values for PI-190 containing A64 in Supplementary Table 1 are lower than those for the PI containing A21. Similarly, A23 and A24 are compared, and the $T_{5\%}$ values for PI-6 (which also contains two A3975295864 units) are higher than those for the PI featuring A24. Increasing rigid structures, such as benzene and naphthalene rings, can indeed benefit the enhancement of $T_{5\%}$ values of polymers, improving their ability to maintain structural integrity during heating. Therefore, it is advisable to directly attach amino groups to benzene rings in diamines.

A3431531268 represents a carbon-oxygen single bond [Figure 5]. In Supplementary Figure 10, the highlighted sections of the selected dianhydride fragments are all ether bonds. Comparing Supplementary Figure 2, it is observed that only the ether bonds adjacent to pyridine are highlighted, such as in A17 and A41. Therefore, the influence of pyridine requires more attention. The region selected within the blue box in Figure 6B displays that as A3431531268 approaches zero, the SHAP value increases, indicating a more favorable effect on $T_{5\%}$. Introducing pyridine into diamines with biphenyl structures for synthesizing PIs can effectively improve their solubility[79]. However, incorporating these structures might compromise the material's thermal stability[106], as the increased flexibility and larger steric hindrance could reduce the tight packing of the molecules, thereby affecting the material's performance at high temperatures. The A26, A41 and A43 fragments can reflect the regularity. A41 and A43 both feature two ether bonds linked to pyridine, with an A3431531268 value of 2, whereas A26, which lacks these, has a value of 0. This is corroborated in Supplementary Table 1 entries PI-66, 141, and 215, where the anhydrides are the same, and PIs containing

A26 exhibit higher $T_{5\%}$ values than those containing A41 and A43. To reduce the A3431531268 value, diamines should minimize the occurrence of ether bonds linked to pyridine as much as possible.

The influence of copolymer on $T_{5\%}$ is positively correlated [Figures 5 and 6C]. The "1" for copolymer represents co-PI, and "0" denotes homo-PI. The SHAP values for co-PIs are positive, while those for homo-PIs are negative. Researchers predominantly use co-PIs to improve various material properties[75,80,93,94], yet their effectiveness varies depending on the specific property. Co-PIs may be favored in future work involving $T_{5\%}$.

In Figures 5 and 6D, when the G2976033787 value is 0, the SHAP values are positive, and the $T_{5\%}$ is higher. G2976033787 indicates the number of branches in aliphatic rings, which are highlighted in Supplementary Figure 11. Although a higher number of branches tends to increase the SHAP values, in this case, all the SHAP values are negative, suggesting that the best performance is observed when there are no aliphatic rings. Compared to aliphatic compounds, aromatic compounds show superior effectiveness in enhancing the thermal stability of PIs[64,74,83]. G9 has two more side chains than G4, as seen in Supplementary Table 1 entries PI-59 and 58, with 59 having a higher $T_{5\%}$ value. Comparing G1 and G14, G1 is linked to an aromatic benzene ring, while G14 is linked to an aliphatic four-membered ring [Supplementary Figure 3]. The $T_{5\%}$ values for PIs with the G14 structure, such as entries PI-155 and 156 in Supplementary Table 1, are significantly lower than those containing the G1 structure, such as entries PI-20, 21, and 22. It is aromatic rings that should be used in dianhydrides rather than aliphatic rings, for better $T_{5\%}$. Supporting information in Supplementary Figure 12 discusses the remaining five descriptors, their relationship with structure, and their impact on performance.

A951226070 is a descriptor that establishes a certain correlation between films and powders, representing the number of unsubstituted hydrogens (vacancies) on the benzene ring [Supplementary Figure 13]. Figure 6E shows that A951226070 values between 4-5 result in positive SHAP values, which benefit $T_{5\%}$. Excessive substitution of hydrogen atoms on the benzene ring increases side chains and disrupts molecular chain regularity. This spatial hindrance reduces the packing density of the molecular chains, thereby decreasing the thermal stability of the PI[58]. In Supplementary Table 2, with the same dianhydride G1, as the amount of diamine A1 (A951226070 = 8) increases, the $T_{5\%}$ values for PI-2 to PI-4 gradually decrease. Therefore, it is necessary to control the frequency of side chains on the benzene ring in the diamine structure. Likewise, the impact of A951226070 on the $T_{5\%}$ of PI films can be seen in Supplementary Figure 12.

Once the structure A2041434490 [Supplementary Figure 13] appears, it leads to a decrease in $T_{5\%}$, as demonstrated in Figure 6F. This is attributed to the nitrogen atom in the main chain. The localized lone pair of electrons on the sp^2 orbit of the nitrogen atom[74] reduces the thermal stability of PI when nitrogen is added to its main chain. This point is confirmed in Supplementary Table 2, where PI-9 to PI-11, all using the same dianhydride G1, show a gradual decline in $T_{5\%}$ as the concentration of diamine A71 (A2041434490 = 1) rises. Consequently, nitrogen atoms should not be present in the main chain of diamines. The remaining four descriptors are discussed in the supporting information Supplementary Figure 14.

Experimental verification results

Based on the known structures, diamines A70 and A71, which exhibit more negative structural effects, are excluded, while any dianhydride with minimal impact on performance is considered suitable. Different ratios of selected dianhydrides and diamines are chosen to form PI powders for experimental validation.

The thermogravimetric and Fourier transform infrared (FTIR) spectra of the prepared PI-a and PI-b powders were shown in Supplementary Figures 15 and 16. The statistical data of 5%, 10% thermogravimetric temperature and 1,000 °C residual amount of PI powder was shown in Supplementary Table 10. The PI powder can maintain good thermal stability before 500 °C. The thermal decomposition temperatures of PI-a at 5% and 10% in nitrogen are 600 and 620 °C, respectively. Those of PI-b at 5% and 10% in nitrogen are 606 and 623 °C, respectively, showing good thermal stability. Table 1 lists different ratios of dianhydride and diamine codes, along with model predictions and experimental values. The experimental and predicted RMSE values for PI-a and PI-b are 8.9. A low RMSE is a positive indicator, suggesting that the predictions made by model-B are accurate.

CONCLUSION

We have implemented an effective strategy using a novel MFF approach to address the co-PI molecular expression method, transferring from the PI film model to the PI powder model. This allows for the prediction of the $T_{5\%}$ for PI powders and reveals the relationship between structure and performance across different material forms. The enhanced MFF, which weights according to molar ratios and aggregates for co-PIs, has constructed a robust PI-film model (model-A). Descriptors and the model were transferred to develop a superior PI-powder model (model-D). Model-A exhibits excellent robustness and generalizability, achieving an R^2 of 0.86 for the training set and 0.79 for the test set. Model-D is more outstanding, with R^2 values of 0.999 for the training set and 0.954 for the test set. The link between structure and performance is bridged by descriptors, highlighting the structural characteristics associated with each descriptor. Overall, to improve the $T_{5\%}$ of PI film materials: (1) Add more conjugated functional groups to diamines but control the number of side chains on benzene rings, reducing pyridine and hydroxyl groups; (2) Co-PI is a good choice; (3) In dianhydrides, there are no aliphatic rings and anhydrides are directly connected to benzene rings. In parallel, for enhancing the $T_{5\%}$ of PI powder materials: (1) Control the number of side chains on benzene rings in diamines, reducing the presence of nitrogen atoms in the main chain; (2) Ensure anhydrides are directly connected to benzene rings. More importantly, the results predicted by model-B align with the experimental outcomes (RMSE value of 8.9), demonstrating the reliability of the model as a prediction tool. Structure, descriptors, and performance are interconnected. Additionally, it is encouraging that the online prediction platform (https://github.com/TanMengyuan/online_model_predict) assists researchers in the field with rapid predictions.

It is acknowledged that attempting to postulate unknown chemical structures and property relationships through models presents challenges. This endeavor requires innovative data and new experimental results to train the models, enabling them to capture new scientific laws and insights. Additionally, the influence of experimental parameters is a critical factor that cannot be ignored. Continued efforts will be made to conduct in-depth studies exploring the adaptability of different experimental conditions within the model, aiming to provide valuable guidance for designing a wider range of functional materials.

DECLARATIONS

Authors' contributions

Conceived the idea and designed the project: Zhang Y, Ding P, Xu T
Performed data analysis and interpretation: Zhang Y, Li X, Xu T
Supervised the project: Ding P, Li M, Xu G, Lv W
Drafted the manuscript: Zhang Y, Fang Y, Li L, Xu T, Peng F
Revised and finalized the manuscript: Zhang Y, Ding P, Li M
All authors read and approved the final manuscript.

Availability of data and materials

Supplementary Materials are available from the *Journal of Materials Informatics* or the authors. The dataset for polyimide films and the predicted dataset are provided with this article.

Financial support and sponsorship

This work was supported by the National Key Research and Development Program of China (2022YFB3707800) and the Key Program of Science and Technology of Yunnan Province (No. 202302AB080022).

Conflicts of interest

Ding P is the guest editor of the Special Issue. Xu G and Lv W are affiliated with Shanghai Plastics Research Institute Co., Ltd, while the other authors have declared that they have no conflicts of interest.

Ethical approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

REFERENCES

1. Yaghi OM, O'Keeffe M, Ockwig NW, Chae HK, Eddaoudi M, Kim J. Reticular synthesis and the design of new materials. *Nature* 2003;423:705-14. DOI PubMed
2. Bubeck S, Chandrasekaran V, Eldan R, et al. Sparks of artificial general intelligence: early experiments with GPT-4. ArXiv. [Preprint] Apr 13, 2023. [accessed on 2024 Jun 14]. Available from: https://doi.org/10.48550/arXiv.2303.12712.
3. Li Y, Wu Y, Han Y, et al. Local environment interaction-based machine learning framework for predicting molecular adsorption energy. *J Mater Inf* 2024;4:4. DOI
4. Yang Y, Xu B, Zong H. Physics infused machine learning force fields for 2D materials monolayers. *J Mater Inf* 2023;3:23. DOI
5. Wang Y, Xie T, France-lanord A, et al. Toward designing highly conductive polymer electrolytes by machine learning assisted coarse-grained molecular dynamics. *Chem Mater* 2020;32:4144-51. DOI
6. Xie T, France-Lanord A, Wang Y, et al. Accelerating amorphous polymer electrolyte screening by learning to reduce errors in molecular dynamics simulated properties. *Nat Commun* 2022;13:3415. DOI PubMed PMC
7. Luo H, Chen S, Liu L, et al. Core–shell nanostructure design in polymer nanocomposite capacitors for energy storage applications. *ACS Sustain Chem Eng* 2019;7:3145-53. DOI
8. Hu H, Zhang F, Luo S, Chang W, Yue J, Wang C. Recent advances in rational design of polymer nanocomposite dielectrics for energy storage. *Nano Energy* 2020;74:104844. DOI
9. St John PC, Phillips C, Kemper TW, et al. Message-passing neural networks for high-throughput polymer screening. *J Chem Phys* 2019;150:234111. DOI PubMed
10. Munshi J, Chen W, Chien T, Balasubramanian G. Transfer learned designer polymers for organic solar cells. *J Chem Inf Model* 2021;61:134-42. DOI PubMed
11. Bai Y, Wilbraham L, Slater BJ, Zwijnenburg MA, Sprick RS, Cooper AI. Accelerated discovery of organic polymer photocatalysts for hydrogen evolution from water through the integration of experiment and theory. *J Am Chem Soc* 2019;141:9063-71. DOI PubMed PMC
12. Liang J, Xu S, Hu L, Zhao Y, Zhu X. Machine-learning-assisted low dielectric constant polymer discovery. *Mater Chem Front* 2021;5:3823-9. DOI
13. Rajendran S, Palani G, Kanakaraj A, et al. Metal and polymer based composites manufactured using additive manufacturing - a brief review. *Polymers* 2023;15:2564. DOI PubMed PMC
14. Oladele IO, Omotosho TF, Adediran AA, Beatriz Morales-cepeda A. Polymer-based composites: an indispensable material for present and future applications. *Int J Polym Sci* 2020;2020:1-12. DOI
15. Wang L, Yang C, Wang X, et al. Advances in polymers and composite dielectrics for thermal transport and high-temperature applications. *Compos Part A Appl S* 2023;164:107320. DOI
16. Jayalath S, Herath M, Epaarachchi J, Trifoni E, Gdoutos EE, Fang L. Durability and long-term behaviour of shape memory polymers

and composites for the space industry - a review of current status and future perspectives. *Polym Degrad Stabil* 2023;211:110297. DOI

17. Gouzman I, Grossman E, Verker R, Atar N, Bolker A, Eliaz N. Advances in polyimide-based materials for space applications. *Adv Mater* 2019;31:e1807738. DOI PubMed
18. Ding M. Isomeric polyimides. *Prog Polym Sci* 2007;32:623-68. DOI
19. Liaw D, Wang K, Huang Y, Lee K, Lai J, Ha C. Advanced polyimide materials: syntheses, physical properties and applications. *Prog Polym Sci* 2012;37:907-74. DOI
20. Li Y, Sun G, Zhou Y, Liu G, Wang J, Han S. Progress in low dielectric polyimide film - a review. *Prog Org Coat* 2022;172:107103. DOI
21. Song N, Yao H, Ma T, et al. Decreasing the dielectric constant and water uptake by introducing hydrophobic cross-linked networks into co-polyimide films. *Appl Surf Sci* 2019;480:990-7. DOI
22. Song N, Shi K, Yu H, et al. Decreasing the dielectric constant and water uptake of co-polyimide films by introducing hydrophobic cross-linked networks. *Eur Polym J* 2018;101:105-12. DOI
23. Liu TQ, Zheng F, Ma X, et al. High heat-resistant polyimide films containing quinoxaline moiety for flexible substrate applications. *Polymer* 2020;209:122963. DOI
24. Liu B, Zhou Y, Dong L, Lu Q, Xu X. Enhanced thermal conductivity in copolymerized polyimide. *iScience* 2022;25:105451. DOI PubMed PMC
25. Lian R, Lei X, Xiao Y, et al. Synthesis and properties of colorless copolyimides derived from 4,4'-diaminodiphenyl ether-based diamines with different substituents. *Polym Chem* 2021;12:4803-11. DOI
26. Jiao L, Du Z, Dai X, Wang H, Yao H, Qiu X. Multifunctional polyimide films with superheat-resistance, low coefficient of thermal expansion and fluorescence performance. *Polymer* 2022;247:124792. DOI
27. Hicyilmaz A, Celik Bedeloglu A. Applications of polyimide coatings: a review. *SN Appl Sci* 2021;3:363. DOI
28. Takekoshi T. Polyimides. In: Kirk-othmer encyclopedia of chemical technology. 2000. DOI
29. Tao L, He J, Munyaneza NE, et al. Discovery of multi-functional polyimides through high-throughput screening using explainable machine learning. *Chem Eng J* 2023;465:142949. DOI
30. Batra R, Song L, Ramprasad R. Emerging materials intelligence ecosystems propelled by machine learning. *Nat Rev Mater* 2021;6:655-78. DOI
31. Chen L, Pilania G, Batra R, et al. Polymer informatics: current status and critical next steps. *Mat Sci Eng R* 2021;144:100595. DOI
32. Kuenneth C, Ramprasad R. polyBERT: a chemical language model to enable fully machine-driven ultrafast polymer informatics. *Nat Commun* 2023;14:4099. DOI PubMed PMC
33. Chen G, Shen Z, Iyer A, et al. Machine-learning-assisted de novo design of organic molecules and polymers: opportunities and challenges. *Polymers* 2020;12:163. DOI PubMed PMC
34. Choi J, Yu S, Yang S, Cho M. The glass transition and thermoelastic behavior of epoxy-based nanocomposites: a molecular dynamics study. *Polymer* 2011;52:5197-203. DOI
35. Huang X, Ju S. Tutorial: AI-assisted exploration and active design of polymers with high intrinsic thermal conductivity. *J Appl Phys* 2024;135:171101. DOI
36. Tao L, Chen G, Li Y. Machine learning discovery of high-temperature polymers. *Patterns* 2021;2:100225. DOI PubMed PMC
37. Uddin MJ, Fan J. Interpretable machine learning framework to predict the glass transition temperature of polymers. *Polymers* 2024;16:1049. DOI PubMed PMC
38. Katritzky AR, Kuanar M, Slavov S, et al. Quantitative correlation of physical and chemical properties with chemical structure: utility for prediction. *Chem Rev* 2010;110:5714-89. DOI PubMed
39. Mannodi-kanakkithodi A, Chandrasekaran A, Kim C, et al. Scoping the polymer genome: a roadmap for rational polymer dielectrics design and beyond. *Mater Today* 2018;21:785-96. DOI
40. Volgin IV, Batyr PA, Matseevich AV, et al. Machine learning with enormous "synthetic" data sets: predicting glass transition temperature of polyimides using graph convolutional neural networks. *ACS Omega* 2022;7:43678-91. DOI PubMed PMC
41. Zhang S, He X, Xia X, et al. Machine-learning-enabled framework in engineering plastics discovery: a case study of designing polyimides with desired glass-transition temperature. *ACS Appl Mater Interfaces* 2023;15:37893-902. DOI PubMed
42. Huang X, Ma S, Zhao CY, Wang H, Ju S. Exploring high thermal conductivity polymers via interpretable machine learning with physical descriptors. *npj Comput Mater* 2023;9:191. DOI
43. Honda S, Shi S, Ueda HR. SMILES Transformer: Pre-trained molecular fingerprint for low data drug discovery. ArXiv. [Preprint] Nov 12, 2019. [accessed on 2024 Jun 14]. Available from: https://arxiv.org/abs/1911.04738.
44. Ying C, Cai T, Luo S, et al. Do transformers really perform bad for graph representation? ArXiv. [Preprint] Nov 24, 2021. [accessed on 2024 Jun 14]. Available from: https://arxiv.org/abs/2106.05234.
45. Irwin R, Dimitriadis S, He J, Bjerrum EJ. Chemformer: a pre-trained transformer for computational chemistry. *Mach Learn Sci Technol* 2022;3:015022. DOI
46. Barredo Arrieta A, Díaz-rodríguez N, Del Ser J, et al. Explainable artificial intelligence (XAI): concepts, taxonomies, opportunities and challenges toward responsible AI. *Inform Fusion* 2020;58:82-115. DOI
47. Wen C, Wang C, Zhang Y, et al. Modeling solid solution strengthening in high entropy alloys using machine learning. *Acta Mater* 2021;212:116917. DOI

48. Zhang S, Du S, Wang L, et al. Design of silicon-containing arylacetylene resins aided by machine learning enhanced materials genome approach. *Chem Eng J* 2022;448:137643. DOI
49. Xu C, Wang Y, Barati Farimani A. TransPolymer: a transformer-based language model for polymer property predictions. *npj Comput Mater* 2023;9:64. DOI
50. Vermeire FH, Green WH. Transfer learning for solvation free energies: from quantum chemistry to experiments. *Chem Eng J* 2021;418:129307. DOI
51. Panapitiya G, Girard M, Hollas A, et al. Evaluation of deep learning architectures for aqueous solubility prediction. *ACS Omega* 2022;7:15695-710. DOI PubMed PMC
52. Zhang D, Xia S, Zhang Y. Accurate prediction of aqueous free solvation energies using 3D atomic feature-based graph neural network with transfer learning. *J Chem Inf Model* 2022;62:1840-8. DOI PubMed PMC
53. Chen C, Ye W, Zuo Y, Zheng C, Ong SP. Graph networks as a universal machine learning framework for molecules and crystals. *Chem Mater* 2019;31:3564-72. DOI
54. Wu S, Kondo Y, Kakimoto M, et al. Machine-learning-assisted discovery of polymers with high thermal conductivity using a molecular design algorithm. *npj Comput Mater* 2019;5:66. DOI
55. Qiao W. Synthesis and properties of polyimides derived from m-xylylenediamine monomer. *Fine Chem* 2022;39:1141-7. (in Chinese) Available from: http://www.finechemicals.com.cn/jxhg/article/abstract/202111181175?st=search. [Last accessed on 14 Jun 2024]
56. Yang Z, Kang C, Guo H, Gao L. Synthesis and properties of polyimide films for flexible OLED displays. *Acta Polym Sin* 2021;52:1308-15. DOI
57. Wang T, Su YM, Chen F, Li WM. Synthesis and characterization of polyimides based on twisted non-coplanar backbone containing indolocarbazole. *Chin J Struc Chem* 2021;40:1611-20. DOI
58. Luo JR, Liu YD, Liu H, et al. Synthesis and characterization of polyimides with naphthalene ring structure introduced in the main chain. *Materials* 2022;15:8014. DOI PubMed PMC
59. Li Q, Zhang S, Liao G, Yi C, Xu Z. Novel fluorinated hyperbranched polyimides with excellent thermal stability, UV-shielding property, organosolubility, and low dielectric constants. *High Perform Polym* 2018;30:872-86. DOI
60. Liu Z, Shen Y, Li X, Shi Q, Liu B, Matsumoto T. Synthesis and properties of the novel polyimides containing cyano and biphenyl moieties. *High Perform Polym* 2018;30:1183-92. DOI
61. Yu B, Jiang C, Wang C. Synthesis and characterization of highly transparent fluorinated copolymer polyimide. *Fine Chem* 2019;36. (in Chinese) Available from: http://www.finechemicals.com.cn/jxhg/article/abstract/201904030272?st=search. [Last accessed on 14 Jun 2024]
62. Liu Y, Zhou L, Sheng S, Song C, Hou H, Song C. Synthesis and properties of super-high temperature diketone anhydride polyimides. *Chin J Appl Chem* 2019;36:658-63. DOI
63. Cui X, Yang J. Synthesis and characterization of high performance thermoplastic polyimide used the anti-riot bomb. In: International Conference on Electronical, Mechanical and Materials Engineering (ICE2ME); Wuhan, China; 2019. pp. 207-10. DOI
64. Abdulhamid MA, Ma X, Ghanem BS, Pinnau I. Synthesis and characterization of organo-soluble polyimides derived from alicyclic dianhydrides and a dihydroxyl-functionalized spirobisindane diamine. *ACS Appl Polym Mater* 2019;1:63-9. DOI
65. Zhang H, Wang W, Chen G, Zhang A, Fang X. Melt-processable semicrystalline polyimides based on 1,4-bis(3,4-dicarboxyphenoxy)benzene dianhydride (HQDPA): synthesis, crystallization, and melting behavior. *Polymers* 2017;9:420. DOI PubMed PMC
66. Liu Y, Huang J, Tan J, et al. Synthesis and characterization of intrinsic high-barrier polyimide derived from a novel diamine monomer containing rigid planar moiety. *J Polym Sci Part A Polym Chem* 2017;55:2373-82. DOI
67. Wang C, Zhao X, Tian D, Wang D, Chen C, Zhou H. Synthesis and characterization of novel polyimides derived from 4,4'-bis(5-amino-2-pyridinoxy)benzophenone: effect of pyridine and ketone units in the main. *Des Monomers Polym* 2017;20:97-105. DOI PubMed PMC
68. Wu F, Zhou X, Yu X. Synthesis and characterization of novel star-branched polyimides derived from 2,2-bis[4-(2,4-diaminophenoxy)phenyl]hexafluoropropane. *RSC Adv* 2017;7:35786-94. DOI
69. Lei Y, Shu Y, Peng J, Tang Y, Huo J. Synthesis and properties of low coefficient of thermal expansion copolyimides derived from biphenyltetracarboxylic dianhydride with p-phenylenediamine and 4,4'-oxydialinine. *e Polymers* 2016;16:295-302. DOI
70. Huang X, Pei X, Wang L, Mei M, Liu C, Wei C. Design and synthesis of organosoluble and transparent polyimides containing bulky substituents and noncoplanar structures. *J Appl Polym Sci* 2016;133:app.43266. DOI
71. Li B, Yan Z, Zhang T, et al. Synthesis and properties of novel colorless and thermostable polyimides containing cross-linkable bulky tetrafluorostyrol pendant group and organosoluble triphenylmethane backbone structure. *J Polym Sci* 2020;58:2355-65. DOI
72. Yuan C, Sun Z, Wang Y. Synthesis and characterization of a novel organo-soluble polyimide containing hydroxyl and bis-tert-butyl substituted triphenylpyridine units. *J Polym Res* 2020;27:220. DOI
73. Mirsamiei A. Synthesis and properties of polyimides derived from bis-(Aminophenoxy) containing naphthalene, [Phenyl] propane and [Methyl] cyclohexane segment and 4, 4'-carbonyldiphthalic anhydride. *J Macromol Sci A* 2018;55:519-25. DOI
74. Wang Z, Guo L, Han S, Qi H, Cheng Y, Liu F. Polyimides from an asymmetric hydroxyl-containing aliphatic-aromatic diamine synthesized via henry reaction. *J Polym Sci Part A Polym Chem* 2017;55:3413-23. DOI
75. Zhao H, Chen G, Zhou Y, Li X, Fang X. Synthesis and characterization of organosoluble and transparent polyimides derived from *trans*-1,2-bis(3,4-dicarboxyphenoxy)cyclohexane dianhydride. *J Appl Polym Sci* 2015;132:app.42317. DOI

76. Liu C, Mei M, Pei X, Huang X, Wei C. Aromatic polyimides with tertbutyl-substituted and pendent naphthalene units: synthesis and soluble, transparent properties. *Chin J Polym Sci* 2015;33:1074-85. DOI
77. Li Q, Xiong H, Pang L, et al. Synthesis and characterization of thermally stable, hydrophobic hyperbranched polyimides derived from a novel triamine. *High Perform Polym* 2015;27:426-38. DOI
78. Chen Y, Zhang Q. Synthesis, characterization and properties of aromatic copolyimides containing Bi-benzimidazole moiety. *J Polym Res* 2015;22:78. DOI
79. Guan Y, Wang C, Wang D, et al. High transparent polyimides containing pyridine and biphenyl units: synthesis, thermal, mechanical, crystal and optical properties. *Polymer* 2015;62:1-10. DOI
80. Wang Y, Wang N, Yu Z, Li G, Zhang X. Novel dye-containing copolyimides: synthesis, characterization and effect of chain entanglements on developed electrospun nanofiber morphologies. *J Polym Res* 2015;22:65. DOI
81. Tapaswi PK, Choi M, Nagappan S, Ha C. Synthesis and characterization of highly transparent and hydrophobic fluorinated polyimides derived from perfluorodecylthio substituted diamine monomers. *J Polym Sci Part A Polym Chem* 2015;53:479-88. DOI
82. Aguilar-lugo C, Santiago-garcía JL, Loría-bastarrachea MI, Guzmán-lucero D, Alexandrova L, Aguilar-vega M. Synthesis, characterization, and structure-property relationships of aromatic polyimides containing 4,4'-diaminotriphenylmethane. *J Polym Res* 2016;23:49. DOI
83. Kumar A, Tateyama S, Yasaki K, et al. Ultrahigh performance bio-based polyimides from 4,4'-diaminostilbene. *Polymer* 2016;83:182-9. DOI
84. Liu S, Zhang Y, Wang X, Tan H, Song N, Guan S. Synthesis and properties of hyperbranched polyimides derived from tetra-amine and long-chain aromatic dianhydrides. *RSC Adv* 2015;5:107793-803. DOI
85. Yang G, Zhang R, Huang H, Liu L, Wang L, Chen Y. Synthesis of novel biobased polyimides derived from isomannide with good optical transparency, solubility and thermal stability. *RSC Adv* 2015;5:67574-82. DOI
86. Zhou Y, Chen G, Zhao H, Song L, Fang X. Synthesis and properties of transparent polyimides derived from trans-1,4-bis(2,3-dicarboxyphenoxy)cyclohexane dianhydride. *RSC Adv* 2015;5:53926-34. DOI
87. Wang Y. Synthesis and characterization of novel polyimides derived from 2,4-bis(4-aminophenoxy)pyrimidine. *High Perform Polym* 2014;26:978-85. DOI
88. Tapaswi PK, Choi M, Jung YS, Cho HJ, Seo DJ, Ha C. Synthesis and characterization of fully aliphatic polyimides from an aliphatic dianhydride with piperazine spacer for enhanced solubility, transparency, and low dielectric constant. *J Polym Sci Part A Polym Chem* 2014;52:2316-28. DOI
89. Li Y, Liu C, Jiao L, Song G, Zhao X, Dang G. Synthesis of new autophotosensitive semiaromatic hyperbranched polyimides with excellent mechanical properties and low birefringences. *High Perform Polym* 2014;26:569-77. DOI
90. Guan Y, Wang D, Song G, et al. Synthesis and characterization of novel polyimides derived from 3,6-bis(4-aminophenoxy)pyridazine. *High Perform Polym* 2014;26:455-62. DOI
91. Cao X, Song Y, Matsumoto T, Liu B. Synthesis and properties of cyano group-containing polyimides with high peel strength. *High Perform Polym* 2016;28:953-61. DOI
92. Chen Y, Zhang Q. Synthesis and properties of polyimides derived from diamine monomer containing bi-benzimidazole unit. *J Polym Res* 2014;21:424. DOI
93. Yao H, Zhang Y, Liu Y, et al. Synthesis and properties of cross-linkable high molecular weight fluorinated copolyimides. *J Polym Sci Part A Polym Chem* 2014;52:349-59. DOI
94. Chen Y, Zhang Q, Sun W, Lei X, Yao P. Synthesis and gas permeation properties of hyperbranched polyimides membranes from a novel (A_2+B_2B'+B_2)-type method. *J Membr Sci* 2014;450:138-46. DOI
95. Guan Y, Wang D, Wang Z, et al. Synthesis and characterization of novel polyimides from 4,4'-bis(5-amino-2-pyridinoxy)diphenyl ether, 4,4'-bis(5-amino-2-pyridinoxy)diphenyl thioether and 4,4'-bis(5-amino-2-pyridinoxy)diphenyl sulfone. *RSC Adv* 2014;4:50163-70. DOI
96. Liu C, Pei X, Mei M, Chou G, Huang X, Wei C. Synthesis and characterization of organosoluble, transparent, and hydrophobic fluorinated polyimides derived from 3,3'-diisopropyl-4,4'-diaminodiphenyl-4''-trifluoromethyltoluene. *High Perform Polym* 2016;28:1114-23. DOI
97. Ding M. Polyimides: chemistry, structure-property relationships and materials. 2nd editon. 2012. (in Chinese) Available from: https://www.ecsponline.com/yz/B405754A7A8064FF096F194B3B878E2BD000.pdf. [Last accessed on 14 Jun 2024]
98. Huang X, Zhao C, Wang H, Ju S. AI-assisted inverse design of sequence-ordered high intrinsic thermal conductivity polymers. *Mater Today Phys* 2024;44:101438. DOI
99. Persson P, Bergström R, Lunell S. Quantum chemical study of photoinjection processes in dye-sensitized TiO_2 nanoparticles. *J Phys Chem B* 2000;104:10348-51. DOI
100. Nakajima T, Sawada K. Discovery of Pb-free perovskite solar cells via high-throughput simulation on the K computer. *J Phys Chem Lett* 2017;8:4826-31. DOI PubMed
101. Zhang Y, Zheng Z, Chen F, Li M, Ding P, Lu W. Accelerating the discovery of N-annulated perylene organic sensitizers via an interpretable machine learning model. *J Mol Struc* 2024;1296:136855. DOI
102. Butler KT, Davies DW, Cartwright H, Isayev O, Walsh A. Machine learning for molecular and materials science. *Nature* 2018;559:547-55. DOI PubMed
103. Chen Y, Liu Y, Min Y. Synthesis and properties comparison of low dielectric silicon containing polyimides. *Materials* 2022;15:2755.

DOI

104. Zhang W, Hu Z, Lu Y, et al. Molecular dynamics simulation on the heat transfer in the cross-linked poly(dimethylsiloxane). *J Phys Chem B* 2023;127:10243-51. DOI PubMed

105. Le NL, Wang Y, Chung T. Synthesis, cross-linking modifications of 6FDA-NDA/DABA polyimide membranes for ethanol dehydration via pervaporation. *J Membr Sci* 2012;415-6:109-21. DOI

106. Guan Y, Wang D, Song G, et al. Novel soluble polyimides derived from 2,2'-bis[4-(5-amino-2-pyridinoxy)phenyl]hexafluoropropane: preparation, characterization, and optical, dielectric properties. *Polymer* 2014;55:3634-41. DOI