Hu *et al. J. Mater. Inf.* **2025**, *5*, 1
**DOI:** 10.20517/jmi.2024.59

**Journal of Materials Informatics**

**Research Article**

**Open Access**

Check for updates

# Discovering polyimides and their composites with targeted mechanical properties through explainable machine learning

**Weilong Hu[1,2], Enzhe Jing[1,2], Haoke Qiu[1,2,*], Zhao-Yan Sun[1,2,*]**

[1]State Key Laboratory of Polymer Physics and Chemistry & Key Laboratory of Polymer Science and Technology, Changchun Institute of Applied Chemistry, Chinese Academy of Sciences, Changchun 130022, Jilin, China.
[2]School of Applied Chemistry and Engineering, University of Science and Technology of China, Hefei 230026, Anhui, China.

[*]**Correspondence to:** Haoke Qiu, Prof. Zhao-Yan Sun, State Key Laboratory of Polymer Physics and Chemistry & Key Laboratory of Polymer Science and Technology, Changchun Institute of Applied Chemistry, Chinese Academy of Sciences, Renmin Street 5625, Changchun 130022, Jilin, China. E-mail: hkqiu@ciac.ac.cn; zysun@ciac.ac.cn

**How to cite this article:** Hu, W.; Jing, E.; Qiu, H.; Sun, Z. Y. Discovering polyimides and their composites with targeted mechanical properties through explainable machine learning. *J. Mater. Inf.* **2025**, *5*, 1. https://dx.doi.org/10.20517/jmi.2024.59

**Received:** 14 Oct 2024 **First Decision:** 13 Nov 2024 **Revised:** 6 Dec 2024 **Accepted:** 19 Dec 2024 **Published:** 4 Jan 2025

**Academic Editors:** Peng Ding, Yong Xu **Copy Editor:** Pei-Yun Wang **Production Editor:** Pei-Yun Wang

## Abstract

Polyimides (PIs) are widely used in industries for their exceptional mechanical properties and thermal resilience. Despite their benefits, the traditional development process for PIs is time-consuming, often lagging behind the increasing demand for materials with tailored properties. In this study, we introduce a machine learning-based approach to predict and optimize the mechanical properties of PI materials and their composites. We developed six predictive models to assess PI structures under various conditions, aiming to enhance our understanding of PI mechanical behavior and facilitate the discovery of high-performance PI structures. By analyzing the substructures within top-performing PIs, we identified key structural motifs that contribute to improved tensile strength, modulus, and elongation at break. Furthermore, we examined the influence of fillers on PI composites, revealing that rigid fillers such as $SiO_2$ and graphene oxide (GO) significantly improve mechanical properties, with GO showing versatile enhancement across multiple mechanical properties. We then screened 800,000 virtual PI structures by using our predictive models, identifying several candidates with targeted mechanical properties. These findings provide a basis for the future experimental validation of optimal PI structures and fillers, offering an efficient pathway to accelerate the design of PI materials with targeted mechanical properties. Our study can also be extended to other materials research, serving as a valuable paradigm for the design of polymers and their composites.

**Keywords:** Explainable machine learning, polyimides, mechanical property

## INTRODUCTION

Polyimides (PIs) are a versatile class of high-performance polymers that find applications across diverse fields due to their remarkable mechanical strength, thermal stability, and resistance to chemicals. In aerospace, PIs are widely used as matrix resins and adhesives in composite materials, providing the strength and resilience required for demanding operational environments[1-5]. In the electronics industry, PI films serve as flexible substrates in flexible circuit boards and displays due to their robustness and durability under stress[6-9]. The thermal stability of PIs also makes it indispensable in battery technologies, where high mechanical resistance is essential for maintaining safety during multiple recharging cycles[10-13]. These applications address the significance of mechanical properties, such as tensile strength, modulus, and elongation at break, which are fundamental to the long-term functionality of PI materials.

The development of PI composites has broadened the application spectrum of PIs even further. By incorporating various fillers, such as graphene oxide (GO), carbon nanotubes, or glass fibers, the mechanical properties of PIs have been greatly enhanced, allowing for tailored performance in specific use cases[14-16]. For instance, flexural strength can be improved by optimizing filler content, and tensile strength and modulus are enhanced through the addition of carbon-based fillers. These efforts have resulted in a diverse array of PI materials, each with unique combinations of properties tailored to particular demands. However, the diversity and complexity of filler systems pose considerable challenges for reliably establishing structure-property relationships, especially when mechanical performance is involved. This complexity highlights the need for systematic approaches to predict and optimize the mechanical properties of PIs and their composites.

Despite substantial progress through empirical methods, the synthesis and optimization of PI materials with different chemical structures and fillers are often time-consuming due to the extensive range of possible chemical modifications and various filler types. The traditional trial-and-error approach to material design is labor-intensive, resource-heavy, and slow to adapt to the evolving demands of modern applications[17-24]. As the demand for high-performance materials increases, there is a pressing need for more efficient discovery and development.

The integration of artificial intelligence (AI) and machine learning (ML) offers promising solutions to these challenges by enabling data-driven material discovery through the establishment of quantitative structure-property relationships (QSPRs) for targeted properties. Recent advances in ML have led to extensive research on PI properties, mainly focusing on thermal and dielectric characteristics rather than mechanical properties[25-30]. For example, Zhang *et al.* created a QSPR model by using ML to predict glass transition temperature ($T_g$), a thermal property critical to many PI applications[31]. In addition, Uddin *et al.* employed molecular descriptors and fingerprinting techniques to predict $T_g$[32]. Our previous work also utilized graph neural networks to predict properties containing both thermal and dielectric characteristics[33]. Dong *et al.* further developed predictive models for PI composites that exhibit high thermal conductivity and low dielectric constants[34].

However, studies addressing mechanical properties using ML remain limited. Mechanical performance, such as tensile strength and modulus, is critical for the applicability of PIs in engineering applications, and understanding these properties becomes even more essential when fillers are incorporated into the polymer matrix. Most ML-based studies have focused on the structural information of the polymer itself, neglecting the influence of fillers[33,35-39]. This simplification fails to capture the practical complexity of PI composites, where fillers play a key role in determining mechanical performance. As a result, a significant gap exists

between ML predictions and experimental practices, especially when addressing the intricate interplay between PI structures and filler systems.

In response to these gaps, this study develops a ML-based framework specifically aimed at predicting the mechanical properties of PI polymers and composites with various fillers. We constructed two datasets: one for PI polymers and one for PI composites with diverse fillers. These datasets enable ML predictions for tensile strength, modulus, and other key mechanical properties for both PI polymers and their composites. Our framework, which screens over 800,000 virtual PI structures, identifies promising candidates with target mechanical performance. We also compare the contributions of PI substructures and filler types on the mechanical properties, and propose several candidates with targeted mechanical properties. Our work offers a scalable approach that can be applied to other polymer systems[40], providing a valuable tool for screening polymer materials and polymer nanocomposites.

## MATERIALS AND METHODS

The workflow of our work is shown in Figure 1. We first collected data on PI materials from various literature sources, focusing on four mechanical properties: elastic modulus ($E$), elongation at break ($\varepsilon$), tensile strength ($\sigma$), and tensile modulus ($E_t$). Based on this, we created two datasets: one for PI polymers (PI-H) and the other for PI composites with fillers (PI-F). This allowed us to conduct a more targeted analysis, utilizing multiple ML models to predict mechanical properties and selecting the model with the best predictive performance. To evaluate the importance of features, we employed SHapley Additive exPlanations (SHAP) analysis, which identified the key factors influencing each mechanical property. We analyzed PI structures with superior mechanical properties in both datasets, investigating the contributions of substructures and filler types on the mechanical properties. We then applied the ML model to a database of 800,000 virtual PI structures, identifying six structures with high values of mechanical properties. Finally, we compared the mechanical properties of these six PIs with those of PI with fillers.

### Data collection and processing

We manually extracted 1,172 data points from literature sources, observing that a significant portion of the PIs contained fillers, prompting us to include detailed filler information. Consequently, we compiled two datasets: the PI-H dataset, which includes data on PI structures with one or more mechanical properties ($E$, $\varepsilon$, $\sigma$, and $E_t$), and the PI-F dataset, which contains filler types, filler loadings, and processing conditions such as temperature and duration. Comprehensive details on the data sources, distribution, and pairwise comparisons for these properties can be found in Supplementary Tables 1 and 2. As the data points of $E$ in both datasets are limited (e.g., 15 data points in PI-H and 66 data points in PI-F), we will not build models for these properties as they are insufficient for reliable training. In this work, we only focus on $\varepsilon$, $\sigma$, and $E_t$. To illustrate the diversity of PI structures in our dataset, we have visualized their chemical space using the t-SNE method [Supplementary Figure 1]. The results demonstrate a diverse range of molecular structures in both PI-H and PI-F datasets, indicating the reliability of our datasets for ML. Moreover, the PI-F dataset covers a broader chemical space due to its larger size. Supplementary Figures 2 and 3 show the original and normalized data distributions for all six datasets, revealing a nearly normal distribution after normalization. This suggests that the datasets are well-suited for predictive modeling. In addition, we observed changes in mechanical property ranges with the addition of fillers, such as increased $E_t$ and $\sigma$ and decreased $\varepsilon$, consistent with expected trends.

To further understand the relationships among mechanical properties, we visualized the pairwise distributions for both PI datasets in Supplementary Figure 4. It is shown that there is a lack of significant correlations between each pair of mechanical properties, suggesting complex, non-linear interactions

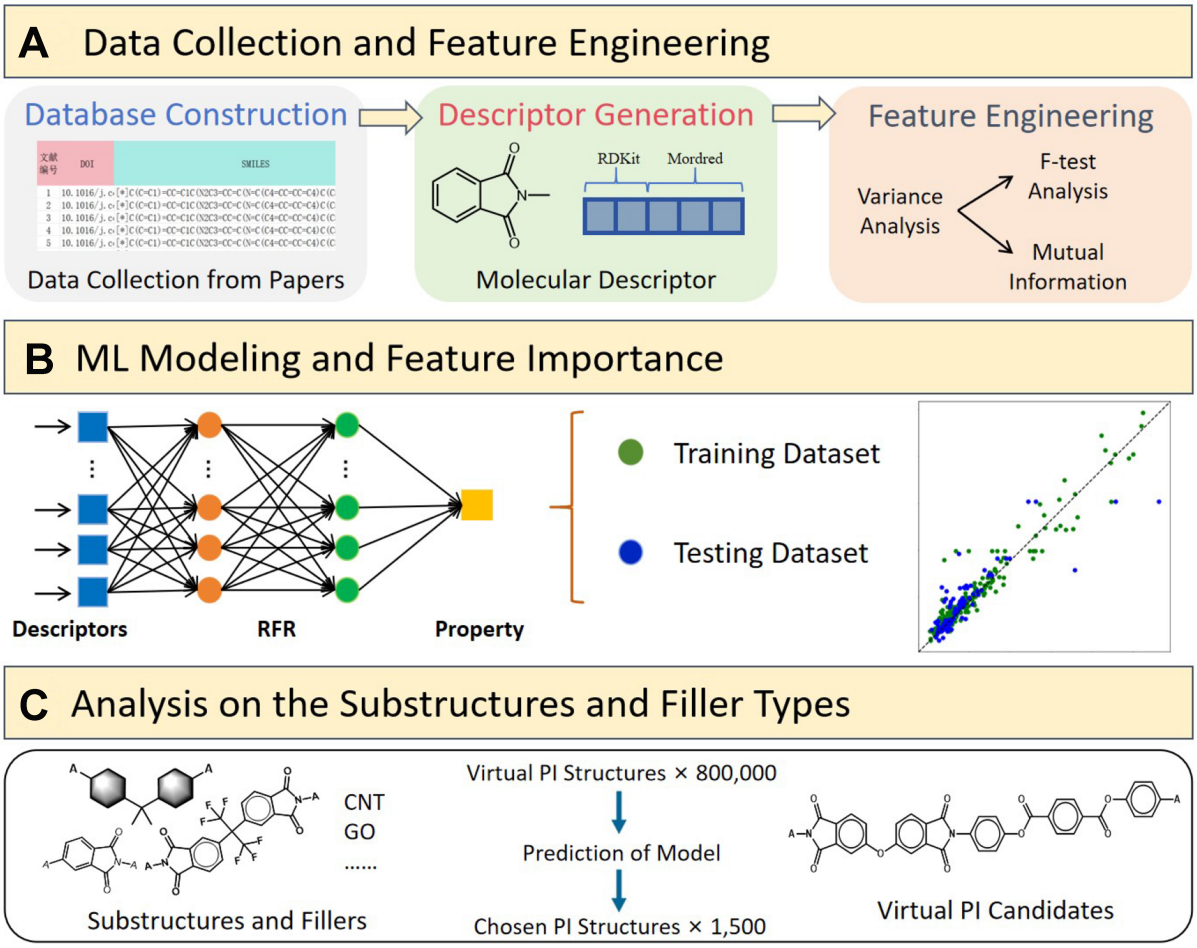


**Figure 1.** Workflow used in this work. (A) Data collection and feature engineering; (B) ML modeling and feature importance; (C) analysis on the substructures and filler types. ML: Machine learning.

among them. This underscores the challenge of designing PI materials with targeted mechanical profiles, as modifying one property may not directly predict or correlate with changes in others.

### Feature engineering and model building

We evaluated 11 common ML models to predict mechanical properties, including random forest regression (RFR), linear regression (LR), support vector regression (SVR), ridge regression (RR), Gaussian process regression (GPR), AdaBoost regression (ABR), gradient boosting regression (GBR), bagging regression (BR), extra trees regression (ETR), decision tree regression (DTR), and multi-layer perceptron (MLP). To identify the top-performing models, we compared a total of 66 models using $R^2$ as performance metrics. RFR, ETR, GBR, and BR consistently performed better, so we selected these four for further analysis. Model prediction results and detailed comparisons are available in Supplementary Table 3.

For the feature selection in the PI-H dataset, we first converted the repeating units of PI into SMILES based on the IMG2SMI method developed by Campos *et al.*[41], an approach that has demonstrated strong predictive performance in recent literature[42-45]. These SMILES were then translated into 1,813 descriptors using RDKit and Mordred packages. In the PI-F dataset, we included additional descriptors for filler type, processing temperature, and processing time. Given a large number of features, we applied three screening

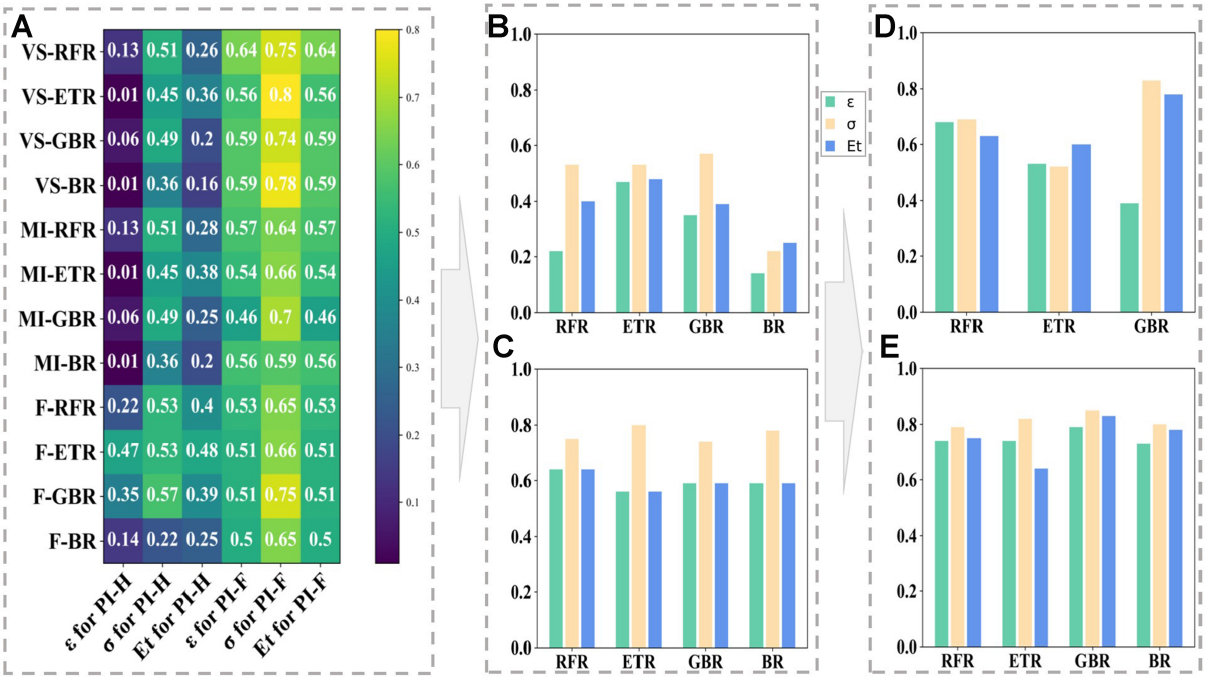


**Figure 2.** Model performance comparisons across different optimization stages. (A) Heatmap showing prediction accuracy for various models under multiple descriptor filtering methods; (B) Performance metrics for the PI-H dataset after descriptor selection using the F-test and initial HPO with GS; (C) Performance metrics for the PI-F dataset following descriptor selection with VS and GS-based HPO; (D) Enhanced model performance for the PI-H dataset after extensive HPO using Optuna; (E) Improved model performance for the PI-F dataset with Optuna optimization. HPO: Hyperparameter optimization; GS: grid search; VS: variance screening.

methods: variance screening (VS)[46], F-test[47], and mutual information (MI)[48], to improve prediction accuracy by filtering out irrelevant features. The best screening method for each dataset was determined based on performance results. For PI-H, the F-test yielded optimal results by retaining descriptors with an F-value greater than 0.05, reflecting a strong linear correlation with target outcomes. For PI-F, VS was the most effective, likely due to the diverse filler types in the dataset, which MI and the F-test may overlook, thereby affecting accuracy. These screening results are depicted in Figure 2.

We applied hyperparameter optimization (HPO) using grid search (GS) and Optuna[49]. GS provided an initial model refinement, while Optuna, with its Bayesian optimization approach, offered precise and reliable results. This combined approach helped us select the six best-performing models. Detailed information on the GS process is available in Supplementary Table 4.

For the PI-H dataset [Figure 2B], BR consistently underperformed compared to RFR, ETR, and GBR, leading us to focus on these three models for further optimization with Optuna. For the PI-F dataset [Figure 2C], all four models (RFR, ETR, GBR, and BR) demonstrated similar performance, so we included them in the Optuna optimization phase. Final optimized results are presented as bar charts in Figure 2D and E, with additional data provided in Supplementary Table 5.

To ensure model reliability, we randomly divided the dataset into training and testing sets (8:2 ratio) and repeated experiments with varying random seeds to create different splits. The models consistently achieved similar predictive performance, as summarized in Supplementary Table 6.

## RESULTS AND DISCUSSION

### Model training results

We evaluated the performance of our ML models on both training and testing sets for the PI-H and PI-F datasets, with the results illustrated in Figure 3. Blue and red symbols represent predicted values for the training and testing sets, respectively. To assess model performance, we calculated coefficient of determination ($R^2$), mean absolute error (MAE), mean squared error (MSE), root mean square error (RMSE) and mean absolute percentage error (MAPE) on the test set. The specific results for these metrics are provided in Supplementary Table 7. In addition, Supplementary Table 8 provides details of the six final models, including training time, prediction speed, and hyperparameters. These data provide additional dimensions for evaluating model performance.

In the PI-H dataset, the RFR model yielded the highest predictive accuracy for $\varepsilon$, achieving an $R^2$ value of 0.68. For $\sigma$ and $E_t$ the GBR model performed best, with $R^2$ values of 0.83 and 0.78 respectively. This pattern indicates that highly correlated mechanical properties ($\sigma$ and $E_t$) within the same dataset are often well predicted by the same model. Meanwhile, for the PI-F datasetthe GBR model consistently outperformed other models across all mechanical properties, attaining $R^2$ values of 0.79 for $\varepsilon$, 0.85 for $\sigma$ and 0.83 for $E_t$. This high accuracy emphasizes the importance of incorporating filler type and processing conditions when predicting the mechanical properties of PI composites.

### Physical insights from ML

Following model construction, we conducted SHAP analysis to evaluate feature importance and uncover the primary factors affecting the mechanical properties of PIs in both datasets. Figures 4 and 5 present the top ten most influential features for each property across the PI-H and PI-F datasets.

In the PI-H dataset, $\varepsilon$ is predominantly influenced by the molecular distance edge between tertiary carbon atoms (MDEC-33), suggesting that interatomic distances within the molecular structure significantly influence the flexibility of materials. Other key features include MDEN-33 and TIC-4, representing distance edges between tertiary nitrogen atoms and higher-order molecular neighborhood information content, respectively. This highlights the importance of both local and extended molecular connectivity in determining the elasticity of PI materials.

For $\sigma$ the most critical features identified are GATS5d and AATSC2d, which relate to the valence electrons and Moreau-Broto autocorrelation of atomic sigma electron contributions. This result, long with similar descriptors such as ATSC6d and AATS6, underscores the significance of electronic characteristics and structural information on the tensile performance of PIs. These features indicate that tensile strength in PI materials is not only structure-dependent but also influenced by electronic distribution, suggesting a coupling between the mechanical strength of a polymer and its atomic-level electronic interactions.

In terms of tensile modulus $E_t$ similar electronic and structural descriptors also play a major role, especially for those associated with Moreau-Broto autocorrelation. In addition, the ratio of rotatable bonds (RotRatio) also contributes to the modulus. The prevalence of features such as AATSC2d and AATSC6Z suggests that the tensile modulus depends on both electronic properties and atomic connectivity. The correlations among mechanical properties, evidenced by the recurrence of shared features such as AATSC2d, indicate that certain intrinsic structural characteristics of PI materials simultaneously influence similar mechanical properties such as tensile strength and tensile modulus.

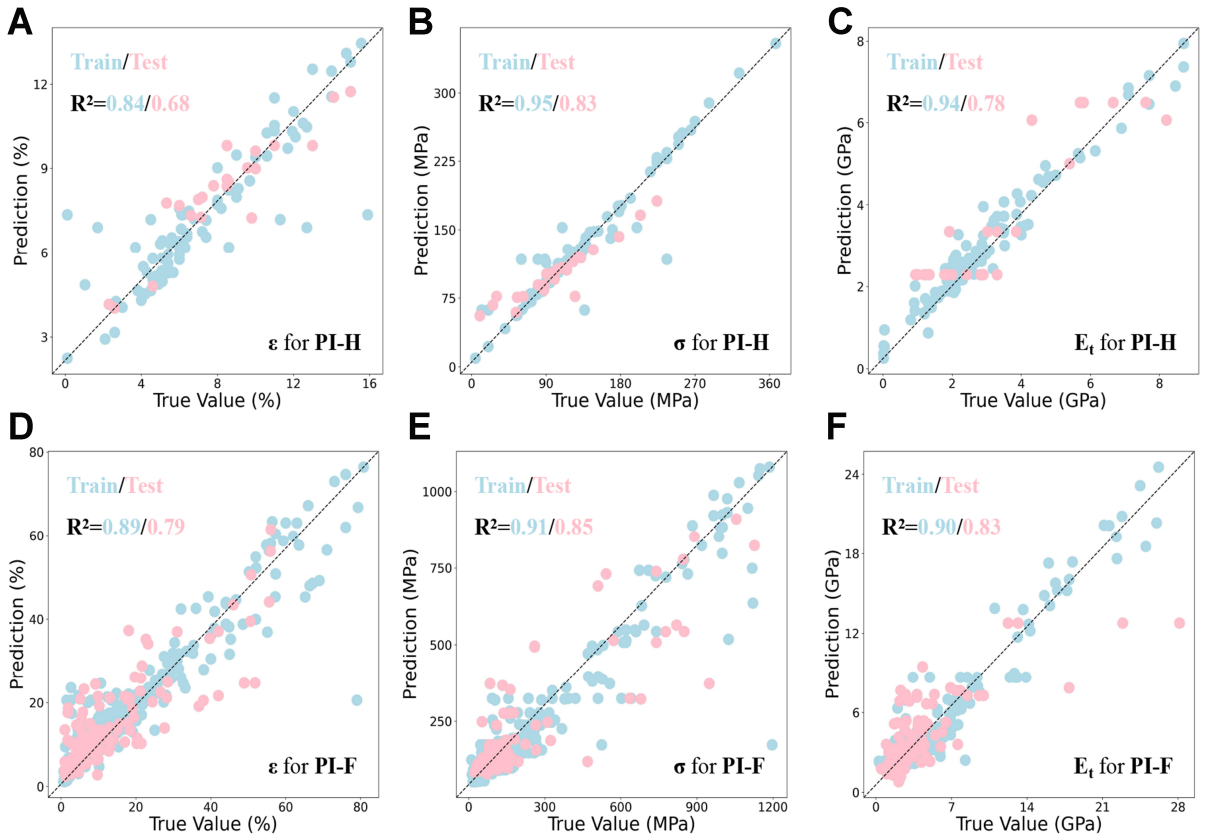


**Figure 3.** Predictive performance of the best model for each mechanical property across the PI-H and PI-F datasets: (A) $\varepsilon$ for the PI-H dataset using RFR; (B) $\sigma$ for the PI-H dataset using the GBR; (C) $E_t$ for the PI-H dataset using GBR; (D) $\varepsilon$ for the PI-F dataset using GBR; (E) $\sigma$ for the PI-F dataset using GBR; (F) $E_t$ for the PI-F dataset using GBR. RFR: Random forest regression; GBR: gradient boosting regression.

The PI-F dataset, which includes filler types and processing conditions, revealed additional insights. For elongation at break ($\varepsilon$), descriptors such as AATSC6dv emphasize the role of electronic interactions weighted by valence electrons. The presence of fillers appears to influence the elasticity of polymers by introducing variability in electronic and molecular structure. In analyzing tensile strength ($\sigma$) and tensile modulus ($E_t$) within the PI-F dataset, many descriptors were related to fillers, with adjacency matrix-based features such as VR1_A playing a crucial role. The impact of filler types on these properties suggests that the interaction between the polymer matrix and filler particles directly affects the composite's mechanical performance, likely due to variations in bonding strength and filler distribution within the matrix.

Note that filler size, dispersion, and composite microstructure significantly affect the mechanical properties of polymer composites[50,51]. While such data were not consistently available in our sources, we included two important processing parameters, mixing time (t_p) and mixing temperature (T_p), that influence filler dispersion in the PI matrix. These parameters were collected as descriptors and shown to have strong feature importance for tensile modulus and tensile strength, as illustrated in Figure 5.

The identified descriptors, such as GATS8dv and AATSC6dv, primarily capture the contributions of atomic $\sigma$-electrons and Moreau-Broto autocorrelation, which are indicative of the number of $\sigma$-bonds in PI structures. A higher proportion of $\sigma$-bonds, coupled with fewer $\pi$-bonds, reduces molecular conjugation and rigidity, resulting in more flexible structures. This flexibility positively correlates with elongation at break, as

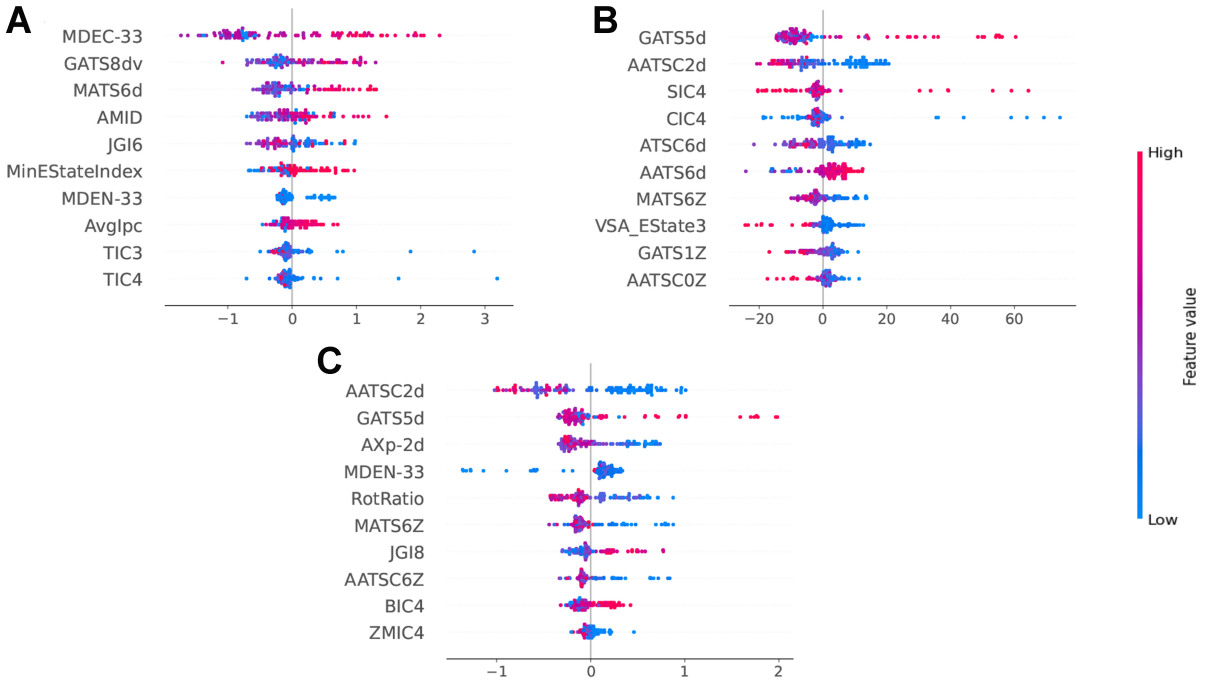


**Figure 4.** Feature importance analysis of the best model in the PI-H dataset showing (A) SHAP values for $\varepsilon$; (B) SHAP values for $\sigma$; (C) SHAP values for $E_t$. SHAP: SHapley Additive exPlanations.

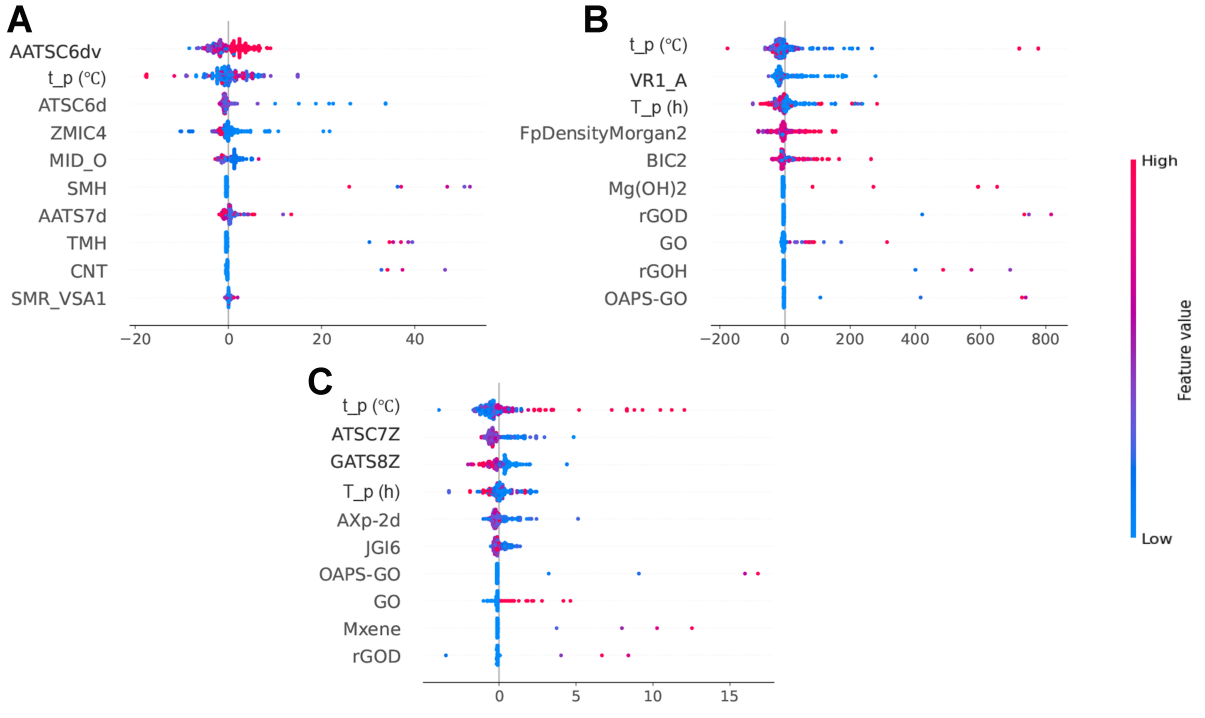


**Figure 5.** Feature importance analysis of the best model in the PI-F dataset showing (A) SHAP values for $\varepsilon$; (B) SHAP values for $\sigma$; and (C) SHAP values for $E_t$. SHAP: SHapley Additive exPlanations.

confirmed by the positive relationship of these descriptors with $\varepsilon$ in Figures 4A and 5B. These findings highlight how ML-derived descriptors provide meaningful insights into the molecular factors influencing PI mechanical properties. Across both datasets, $\varepsilon$, $\sigma$, and $E_t$ exhibit interrelationships, as demonstrated by those shared descriptors. This analysis suggests that while the intrinsic properties of PI determine certain mechanical behaviors, composite structures rely on additional factors such as filler types and processing conditions, leading to complex and property-specific influences.

### Influence of substructures on mechanical properties of PI

To explore the role of substructures in determining PI mechanical properties, we analyzed the top one-third of high-performing PI structures from each dataset. We identified their most common substructures and focused on the ten most frequent ones, calculating the proportion of these substructures among high-performance PIs. The results are shown in Figure 6, illustrating the structural features contributing to the mechanical properties of PIs.

To clearly compare the contribution of substructures on different mechanical properties, we highlighted some substructures with potential functional significance (marked in red in Figure 6). Specifically, Figure 6A and B reveals that substructures I and II appear among the top ten most influential substructures only for the PI-H dataset, while substructure III is uniquely influential in the PI-F dataset. This suggests a distinction in the structural characteristics associated with mechanical properties between PI materials and PI composites with fillers. For example, structures I and II may contribute to increased elongation at break due to their intrinsic flexibility, while substructure III, containing trifluoromethyl groups, may enhance fracture elongation through hydrogen bonding interactions with fillers or by modifying interfacial compatibility within the composite.

The similarities between Figure 6C and D and 6E and F imply that the substructures influencing $\sigma$ and $E_t$ share commonalities, possibly due to a high correlation between these properties. For instance, substructures IV appear in both Figure 6C and E, while substructure VI is exclusive to Figure 6D and F. This suggests that distinct substructural motifs are associated with $\sigma$ and $E_t$ for PI materials and composites. The hydroxyl groups in substructure VI may improve interfacial compatibility with fillers, thereby enhancing these mechanical properties in PI composites.

### Impact of filler types on mechanical properties of PI-F

In PI composites, fillers are essential in enhancing their mechanical properties. We then analyzed the types of fillers associated with the top one-third of high-performing PI composites in terms of mechanical properties, and we identified three commonly used fillers: $SiO_2$, GO, and $TiO_2$. Then, we calculated the average enhancement of mechanical properties due to filler incorporation, expressed as the ratio of each property in PI-F composites to those in PI-H polymers. These enhancement ratios for $\varepsilon$, $\sigma$ and $E_t$ are shown in Figure 7.

The type and content of fillers significantly influence the performance of PI composites. As shown in Figure 7A, $\varepsilon$ is primarily influenced by the intrinsic structure of PI itself, and the filler addition sometimes decreases $\varepsilon$ (most ratios are less than 1.0), likely due to changes in polymer-polymer interactions. In contrast, the presence of fillers substantially improves $\sigma$ and $E_t$. For example, $SiO_2$ significantly increases both $\sigma$ and $E_t$, with enhancement ratios reaching as high as 17.5 and 16.5, respectively. GO fillers also enhance $\sigma$, with ratios ranging from approximately 2.5 to 9.0, and $E_t$ from 3.0 to 15.0. $TiO_2$ provides a moderate improvement, with enhancement ratios of up to 5.0 for $\sigma$ and 13.0 for $E_t$. These results indicate that the incorporation of fillers such as $SiO_2$, GO, and $TiO_2$ generally increases tensile strength and modulus while decreasing elongation at break. Interestingly, certain GO-filled composites demonstrate a higher

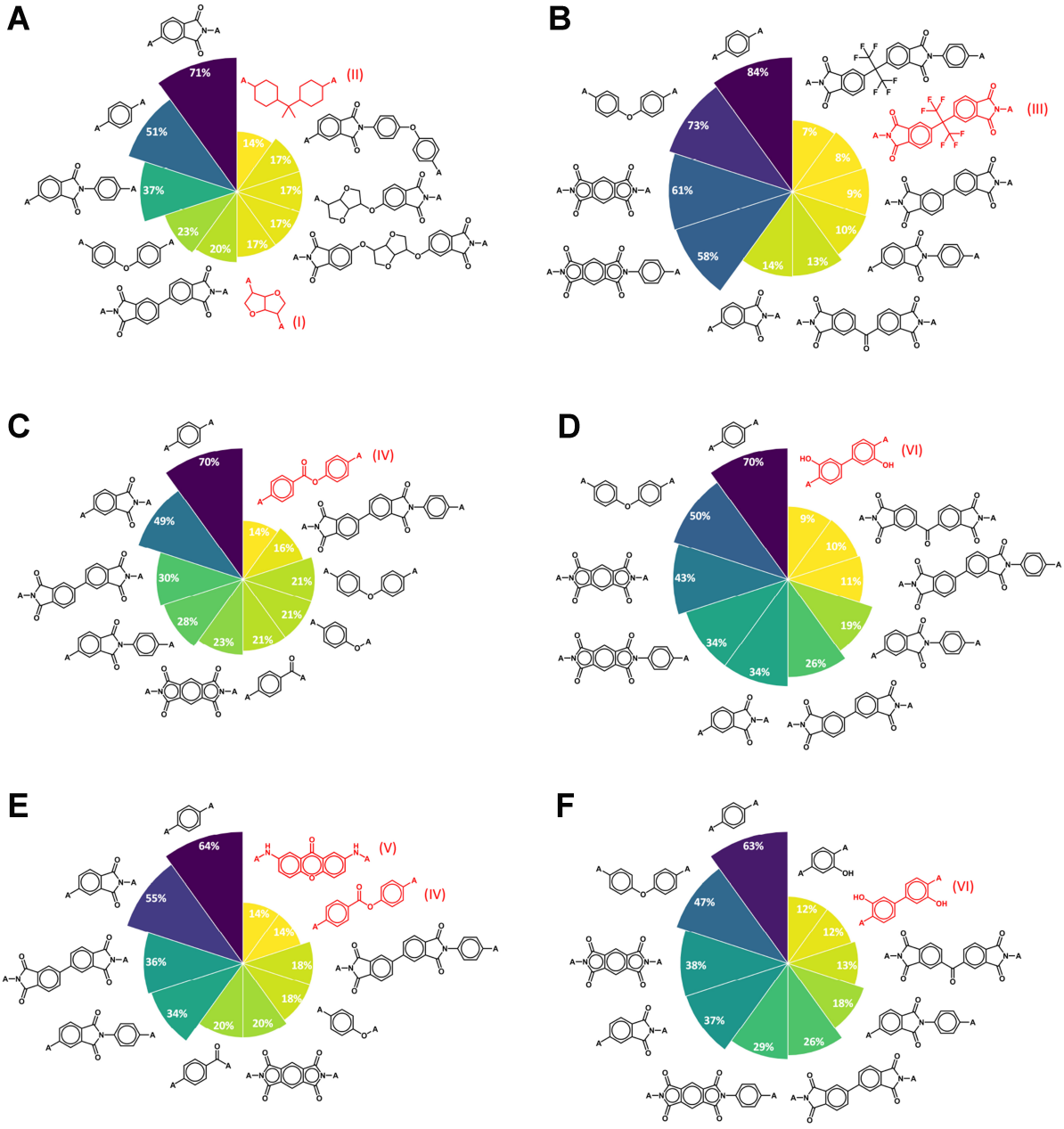


**Figure 6.** Comparison of PI substructures in the PI-H and PI-F datasets based on mechanical properties. (A), (C), and (E) show the substructures present in PI-H dataset exhibiting high performance for $\varepsilon$, $\sigma$, and $E_t$, respectively. (B), (D), and (F) show the substructures found in PI-F dataset exhibiting high performance for $\varepsilon$, $\sigma$, and $E_t$, respectively. PI: Polyimide.

elongation at break compared to pure PI when the GO content is low (ratio > 1). Conversely, higher GO content significantly enhances tensile strength and modulus. This behavior likely arises from the $\pi$-$\pi$ stacking interactions between GO and the benzyl groups in PI molecules, which enhance stiffness at higher filler concentrations while maintaining flexibility at lower concentrations.

### Discovery of high-performance PIs via model predictions

To provide experimental direction for synthesizing high-performance PI materials, we created a virtual dataset of 800,000 PI structures, randomly sampled from an initial dataset of 8,000,000 structures. Using a

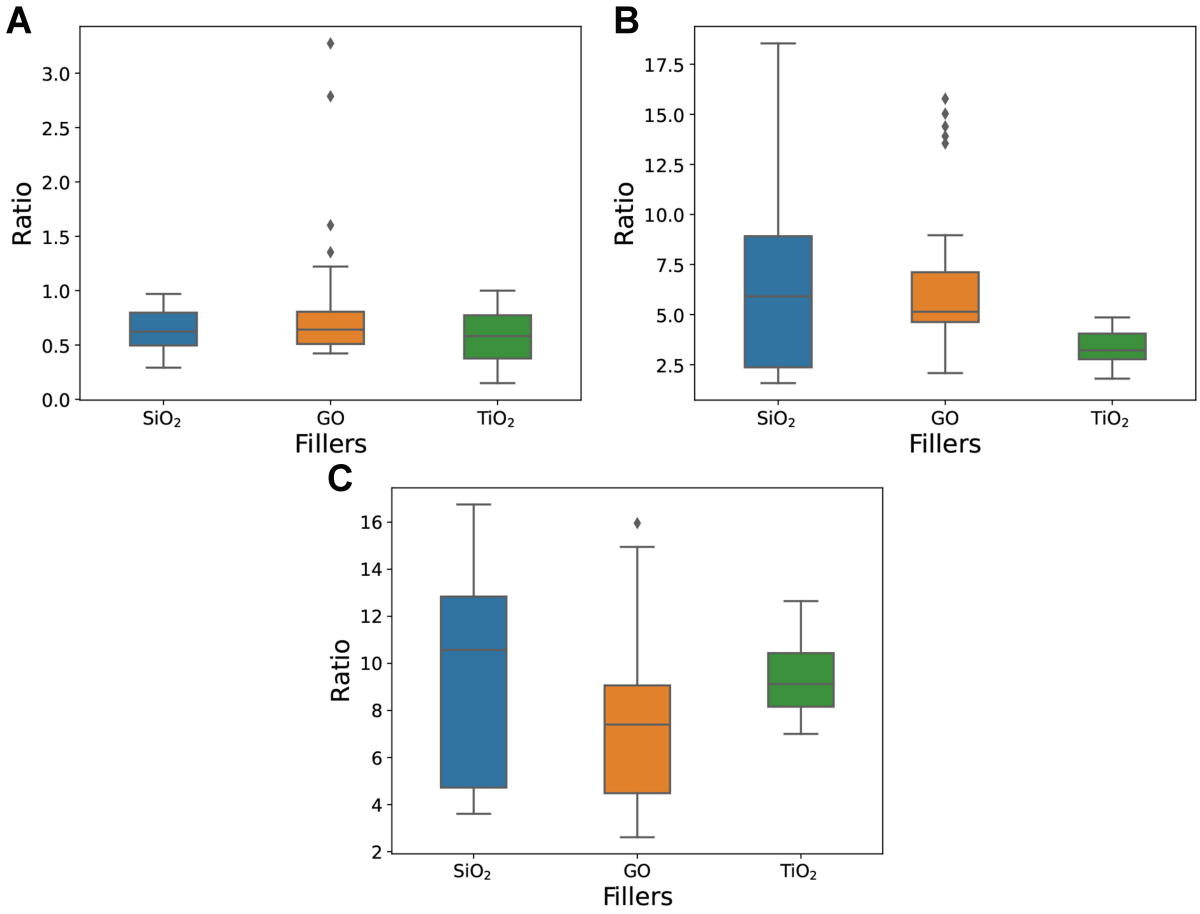


**Figure 7.** Box plots of the PI-F dataset illustrating the impact of filler types on (A) $\varepsilon$, (B) $\sigma$, and (C) $E_t$. Ratios of each property in PI-F composites to those in PI-H polymers are used to represent the influence of fillers on each mechanical property. PI: Polyimide.

model trained on the PI-H dataset, we predicted mechanical properties for these virtual structures, with full results available at https://github.com/Weilong-Hu/PI. Based on the prediction results of the model and our experience, we screened some PIs with single excellent mechanical properties and some with excellent combined mechanical properties.

Figure 8A presents 10,000 randomly selected data points along with the six representative high-performance PIs. Figure 8B provides structural depictions of these six top-performing PIs. Structures #1, #2, and #3 correspond to the high values for $\varepsilon$, $\sigma$, and $E_t$, respectively, and structures #4, #5, and #6 show balanced mechanical properties across multiple metrics. For selecting PIs with high performance in a single property, we ranked all 800,000 structures based on their predicted values for $\varepsilon$, $\sigma$, or $E_t$ and identified the top ten structures for each property. From these, we randomly selected a representative structure to illustrate in Figure 8B. For the selection of PIs with balanced mechanical performance, we normalized all predicted properties using the maximum values of $\varepsilon$, $\sigma$, and $E_t$ in the PI-H dataset. A comprehensive performance score was then calculated by multiplying these normalized values, enabling the identification of structures with well-balanced mechanical properties. We further evaluated the potential benefits of adding fillers to these structures, utilizing a model trained on the PI-F dataset to predict mechanical property enhancements with various fillers.

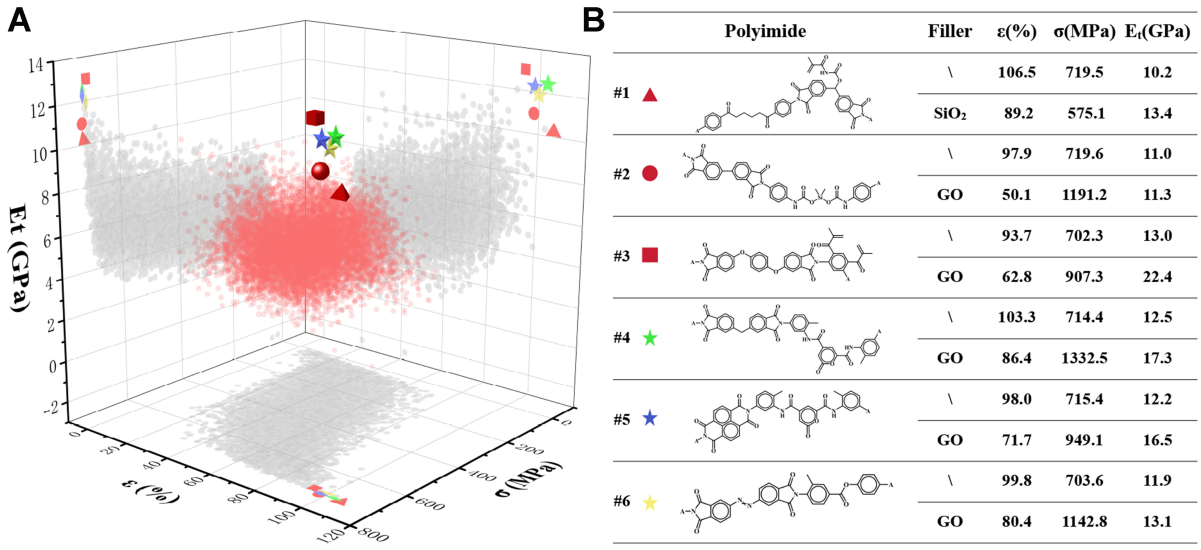


**Figure 8.** Prediction and evaluation of 800,000 virtual PI structures to identify high-performance candidates with enhanced mechanical properties for PI materials. Six PI structures were selected, in which three based on their outstanding individual mechanical properties and three chosen for their overall balanced performance. (A) 3D scatter plots showing 10,000 random PI data points (red color) and the six high-performance PI structures, each marked with distinct symbols: triangles for the structure with high $\varepsilon$, circles for high $\sigma$, squares for high $E_t$, and stars for structures with balanced mechanical properties. Each high-performance structure is projected onto three 2D planes, shown in color, for better visualization and comparison with the general distribution (gray shadow) of random PI data; (B) Structures of six PI candidates and the corresponding mechanical properties for both PI materials and PI composites. PI: Polyimide.

The data in Figure 8B indicate that while filler addition generally reduces $\varepsilon$, certain fillers can minimize this impact. For structure #1, we selected fillers that minimize the performance reduction. For structures #2 and #3, fillers were chosen to maximize $\sigma$ and $E_t$, respectively. Structures #4, #5, and #6 were optimized with fillers that enhance the balanced mechanical properties. Interestingly, GO was consistently selected for structures #4, #5, and #6, implying its effectiveness in significantly enhancing PI composites. This result also aligns with the findings on filler performance in the PI-F dataset.

Our studies suggest an essential role of ML and data analysis in elucidating key structural and compositional factors that govern the mechanical properties of PI materials. By tailoring both molecular structure and filler composition, it may be possible to achieve PI materials with superior and targeted mechanical properties. This predictive approach provides a promising framework for advancing materials design and guiding experimental work in the development of PI and its composites.

## CONCLUSIONS

In this study, we present an effective ML-based approach for predicting and optimizing the mechanical properties of PIs and their composites. We developed six predictive models to evaluate various PI structures under different conditions, enabling the identification of several high-performance virtual PI structures and optimal filler types. By examining the substructures associated with high-performing PI materials, we highlighted key structural motifs that contribute significantly to tensile strength, modulus, and elongation at break. Note that the relationships between structure and property in polymer systems are relatively complex, and therefore, deep learning models, such as neural networks, may also be used to describe these complex relationships, which will be the focus of our future work.

Our findings show that the incorporation of fillers can greatly enhance the tensile strength and modulus of PIs, with particularly strong effects observed for fillers such as $SiO_2$ and GO. Notably, GO emerged as a versatile filler capable of simultaneously improving multiple mechanical properties, suggesting its potential for broader applications in PI composite design. Through our predictive model, we screened 800,000 virtual PI structures, identifying promising candidates with targeted mechanical properties. These virtual predictions may offer a roadmap for experimental synthesis and testing, potentially reducing the cost and time involved in PI material design.

Our current research focuses specifically on PI materials. Given the complexity and diversity of polymers, the model is not yet directly applicable to predicting the mechanical properties of all polymer types. However, with the inclusion of sufficient data points for other polymers, the framework can be readily extended to a broader range of materials. Another limitation of this study is the lack of experimental validation for the predicted structures, primarily due to challenges in synthesis and testing. To address this, the model and predicted structures are publicly available at https://github.com/Weilong-Hu/PI, enabling experimentalists to validate and refine the proposed candidates. Future efforts will focus on expanding the dataset, enhancing the generalizability of the model, and bridging the gap between computational predictions and experimental verification.

## DECLARATIONS

### Acknowledgments

We appreciate the financial support from the National Key R&D Program of China (no. 2022YFB3707303) and the National Natural Science Foundation of China (52293471). We would like to express our sincere gratitude to Professor Xiangling Ji at the State Key Laboratory of Polymer Physics and Chemistry, Changchun Institute of Applied Chemistry, Chinese Academy of Sciences, for providing valuable data and insights in the data mining aspect of this research. We are also grateful to the Network and Computing Center in Changchun Institute of Applied Chemistry for the hardware support.

### Authors' contributions

Data collection, Methodology, writing - original draft: Hu W
Data collection: Jing E
Investigation, supervision, writing - review and editing: Qiu H
Conceptualization, supervision, funding acquisition, writing - review and editing: Sun ZY

### Availability of data and materials

All the PI structures and their mechanical properties used for model training were manually mined from literature sources, and we showed the DOI of our data source literature in Supplementary Table 2.

### Financial support and sponsorship

This work was supported by the National Key R&D Program of China (no. 2022YFB3707303) and the National Natural Science Foundation of China (52293471).

### Conflicts of interest

All authors declared that there are no conflicts of interest.

### Ethical approval and consent to participate

Not applicable.

## Consent for publication

Not applicable.



## REFERENCES

1. Plis, E. A.; Engelhart, D. P.; Cooper, R.; Johnston, W. R.; Ferguson, D.; Hoffmann, R. Review of radiation-induced effects in polyimide. *Appl. Sci.* **2019**, *9*, 1999. DOI
2. Wu, X.; Shu, C.; Zhong, M.; et al. Irradiation tolerance of an optically transparent polyimide film under 1 MeV electron beam. *Appl. Surf. Sci.* **2022**, *583*, 152558. DOI
3. Lei, Y.; Zhang, L.; Zhou, L.; et al. Proton irradiation-induced changes in the tribological performance of polyimide composites. *Tribol. Int.* **2022**, *167*, 107427. DOI
4. Cherkashina, N.; Pavlenko, V.; Noskov, A. Synthesis and property evaluations of highly filled polyimide composites under thermal cycling conditions from -190 °C to +200 °C. *Cryogenics* **2019**, *104*, 102995. DOI
5. De, O. P. R.; Sukumaran, A. K.; Benedetti, L.; et al. Novel polyimide-hexagonal boron nitride nanocomposites for synergistic improvement in tribological and radiation shielding properties. *Tribol. Int.* **2023**, *189*, 108936. DOI
6. Kim, S. D.; Lee, B.; Byun, T.; et al. Poly(amide-imide) materials for transparent and flexible displays. *Sci. Adv.* **2018**, *4*, eaau1956. DOI PubMed PMC
7. Xu, Z.; Li, M.; Xu, M.; et al. Light extraction of flexible OLEDs based on transparent polyimide substrates with 3-D photonic structure. *Org. Electron.* **2017**, *44*, 225-31. DOI
8. Park, C. I.; Seong, M.; Kim, M. A.; et al. World's first large size 77-inch transparent flexible OLED display. *J. Soc. Info. Display.* **2018**, *26*, 287-95. DOI
9. Ke, T.; Kang, T.; Lee, C.; et al. Flexible OLED display with 620°C LTPS TFT and touch sensor manufactured by weak bonding method. *J. Soc. Info. Display*. **2020**, *28*, 392-400. DOI
10. Zhang, M.; Wang, L.; Xu, H.; Song, Y.; He, X. Polyimides as promising materials for lithium-ion batteries: a review. *Nanomicro. Lett.* **2023**, *15*, 135. DOI PubMed PMC
11. Li, M.; Sheng, L.; Xu, R.; et al. Enhanced the mechanical strength of polyimide (PI) nanofiber separator via PAALi binder for lithium ion battery. *Compos. Commun.* **2021**, *24*, 100607. DOI
12. Huang, X.; Liao, S.; Liu, Y.; Rao, Q.; Peng, X.; Min, Y. Design, fabrication and application of PEO/CMC-Li @PI hybrid polymer electrolyte membrane in all-solid-state lithium battery. *Electrochim. Acta.* **2021**, *389*, 138747. DOI
13. Parsaei, S.; Zebarjad, S. M.; Moghim, M. H. Fabrication and post-processing of PI/PVDF-HFP/PI electrospun sandwich separators for lithium-ion batteries. *Polym. Eng. Sci.* **2022**, *62*, 3641-51. DOI
14. Song, S.; Xu, X.; Lan, H.; et al. Design of co-cured multi-component thermosets with enhanced heat resistance, toughness, and processability via a machine learning approach. *Macromol. Rapid. Commun.* **2024**, *45*, e2400337. DOI
15. Ha, H. W.; Choudhury, A.; Kamal, T.; Kim, D. H.; Park, S. Y. Effect of chemical modification of graphene on mechanical, electrical, and thermal properties of polyimide/graphene nanocomposites. *ACS. Appl. Mater. Interfaces.* **2012**, *4*, 4623-30. DOI PubMed
16. Han, E.; Wang, Y.; Chen, X.; et al. Consecutive large-scale fabrication of surface-silvered polyimide fibers via an integrated direct ion-exchange self-metallization strategy. *ACS. Appl. Mater. Interfaces.* **2013**, *5*, 4293-301. DOI
17. Li, J.; Zhang, H.; Chen, J. Z. Y. Structural prediction and inverse design by a strongly correlated neural network. *Phys. Rev. Lett.* **2019**, *123*, 108002. DOI
18. Zhao, H.; Duan, P.; Li, Z.; et al. Unveiling the multiscale dynamics of polymer vitrimers via molecular dynamics simulations. *Macromolecules* **2023**, *56*, 9336-49. DOI
19. Gao, L.; Lin, J.; Wang, L.; Du, L. Machine learning-assisted design of advanced polymeric materials. *Acc. Mater. Res.* **2024**, *5*, 571-84. DOI
20. Zhang, K.; Gong, X.; Jiang, Y. Machine learning in soft matter: from simulations to experiments. *Adv. Funct. Mater.* **2024**, *34*, 2315177. DOI
21. Liu, L.; Li, Y.; Zheng, J.; Li, H. Expert-augmented machine learning to accelerate the discovery of copolymers for anion exchange membrane. *J. Membr. Sci.* **2024**, *693*, 122327. DOI
22. Dong, Q.; Xu, Z.; Song, Q.; Qiang, Y.; Cao, Y.; Li, W. Automated search strategy for novel ordered structures of block copolymers. *ACS. Macro. Lett.* **2024**, *13*, 987-93. DOI
23. Qiu, H.; Sun, Z. On-demand reverse design of polymers with PolyTAO. *npj. Comput. Mater.* **2024**, *10*, 1466. DOI
24. Lu, R.; Han, Y.; Hu, J.; et al. Deep learning model for precise prediction and design of low-melting point phthalonitrile monomers. *Chem. Eng. J.* **2024**, *497*, 154815. DOI
25. Qiu, H.; Zhao, W.; Pei, H.; Li, J.; Sun, Z. Highly accurate prediction of viscosity of epoxy resin and diluent at various temperatures utilizing machine learning. *Polymer* **2022**, *256*, 125216. DOI
26. Jumper, J.; Evans, R.; Pritzel, A.; et al. Highly accurate protein structure prediction with AlphaFold. *Nature* **2021**, *596*, 583-9. DOI

PubMed PMC
27. Xu, Y.; Liu, X.; Cao, X.; et al. Artificial intelligence: a powerful paradigm for scientific research. *Innovation* **2021**, *2*, 100179. DOI PubMed PMC
28. Qiu, H.; Liu, L.; Qiu, X.; Dai, X.; Ji, X.; Sun, Z. Y. PolyNC: a natural and chemical language model for the prediction of unified polymer properties. *Chem. Sci.* **2024**, *15*, 534-44. DOI PubMed PMC
29. Gao, H.; Struble, T. J.; Coley, C. W.; Wang, Y.; Green, W. H.; Jensen, K. F. Using machine learning to predict suitable conditions for organic reactions. *ACS. Cent. Sci.* **2018**, *4*, 1465-76. DOI PubMed PMC
30. 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
31. 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
32. Uddin, M. J.; Fan, J. Interpretable machine learning framework to predict the glass transition temperature of polymers. *Polymers* **2024**, *16*, 1049. DOI PubMed PMC
33. Qiu, H.; Qiu, X.; Dai, X.; Sun, Z. Design of polyimides with targeted glass transition temperature using a graph neural network. *J. Mater. Chem. C.* **2023**, *11*, 2930-40. DOI
34. Dong, X.; Wan, B.; Zheng, M. S.; et al. Dual-effect coupling for superior dielectric and thermal conductivity of polyimide composite films featuring "crystal-like phase" structure. *Adv. Mater.* **2024**, *36*, e2307804. DOI
35. Luo, G.; Huan, F.; Sun, Y.; Shi, F.; Deng, S.; Wang, J. Machine learning-based high-throughput screening for high-stability polyimides. *Ind. Eng. Chem. Res.* **2024**, *63*, 21110-22. DOI
36. Zhang, S.; He, X.; Xiao, P.; et al. Interpretable machine learning for investigating the molecular mechanisms governing the transparency of colorless transparent polyimide for OLED cover windows. *Adv. Funct. Mater.* **2024**, *34*, 2409143. DOI
37. Yue, T.; He, J.; Tao, L.; Li, Y. High-throughput screening and prediction of high modulus of resilience polymers using explainable machine learning. *J. Chem. Theory. Comput.* **2023**, *19*, 4641-53. DOI PubMed
38. Tao, L.; He, J.; Munyaneza, N. E.; et al. Discovery of multi-functional polyimides through high-throughput screening using explainable machine learning. *Chem. Eng. J.* **2023**, *465*, 142949. DOI
39. Lei, H.; Qi, S.; Wu, D. Hierarchical multiscale analysis of polyimide films by molecular dynamics simulation: investigation of thermo-mechanical properties. *Polymer* **2019**, *179*, 121645. DOI
40. 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
41. Campos, D.; Ji, H. IMG2SMI: translating molecular structure images to simplified molecular-input line-entry system. *arXiv* 2024;arXiv:2109.04202. Available from: https://doi.org/10.48550/arXiv.2109.04202. [accessed 21 Dec 2024]
42. Tao, L.; Varshney, V.; Li, Y. Benchmarking machine learning models for polymer informatics: an example of glass transition temperature. *J. Chem. Inf. Model.* **2021**, *61*, 5395-413. DOI PubMed
43. Chen, G.; Tao, L.; Li, Y. Predicting polymers' glass transition temperature by a chemical language processing model. *Polymers* **2021**, *13*, 1898. DOI PubMed PMC
44. 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*, 203. DOI
45. Zhang, Y.; Xu, X. Machine learning glass transition temperature of polyacrylamides using quantum chemical descriptors. *Polym. Chem.* **2021**, *12*, 843-51. DOI
46. Fan, Y.; Chen, M.; Zhu, Q. lncLocPred: predicting LncRNA subcellular localization using multiple sequence feature information. *IEEE. Access.* **2020**, *8*, 124702-11. DOI
47. Griffiths, W. E.; Hill, R. C. On the power of the *F*-test for hypotheses in a linear model. *Am. Stat.* **2022**, *76*, 78-84. DOI
48. Cakir, F.; He, K.; Bargal, S. A.; Sclaroff, S. Hashing with mutual information. *IEEE. Trans. Pattern. Anal. Mach. Intell.* **2019**, *41*, 2424-37. DOI PubMed
49. Xiao, X.; Zou, Y.; Huang, J.; et al. An interpretable model for landslide susceptibility assessment based on Optuna hyperparameter optimization and Random Forest. *Geomatics. Nat. Hazards. Risk.* **2024**, *15*, 2347421. DOI
50. Wan, Z.; Zhao, W.; Qiu, H.; et al. Data-driven exploration of polymer processing effects on the mechanical properties in carbon black-reinforced rubber composites. *Chin. J. Polym. Sci.* **2024**, *42*, 2038-47. DOI
51. Wan, Z.; Chen, S.; Feng, X.; Sun, Z. From processing to properties: enhancing machine learning models with microstructural information in polymer nanocomposites. *Compos. Commun.* **2024**, *51*, 102072. DOI