



STRUCTURE-PROPERTY RELATIONSHIPS AND MACHINE LEARNING PREDICTION OF GLASS TRANSITION TEMPERATURE IN POLYMER MATERIALS

Yogeshwari Hardas

Assistant professor, Department of CSE (AI & ML), A P Shah Institute of Technology, Thane,
ydhardas@apsit.edu.in

Received:-02/06/2025, Revised:-10/08/2026, Accepted:-05/09/2026, Published:-25/09/2026

ABSTRACT

Polymer properties such as structural rigidity, chain flexibility, composition and intermolecular properties affect the glass transition temperature (T_g). Accurate T_g prediction can aid in polymer screening and minimize reliance on lengthy experimental testing. This study aimed to characterize structure-property relationships in polymer materials and develop interpretable machine-learning models for predicting experimental T_g from composition-weighted molecular descriptors. Analyses were based on 666 experimental observations of 386 polymer systems. A total of ten composition-weighted molecular descriptors were calculated, and six were retained after correlation and multicollinearity screening. The partitioning using polymer-system-awareness, five-fold grouped cross-validation, and an independent test were used to evaluate elastic net, random forest, gradient boosting, and support vector regression. Ring count showed the strongest positive correlation with T_g ($r = 0.7851$), while Fraction Csp³ showed the strongest negative correlation ($r = -0.7574$). Gradient Boosting achieved the lowest cross-validated RMSE of 32.33 K and attained $R^2 = 0.8697$, RMSE = 31.16 K, and MAE = 23.40 K on the held-out test set. Permutation importance identified ring count as the dominant predictor, followed by rotatable-bond count and Fraction Csp³. These results establish that interpretable structural descriptors together with polymer-system-aware machine learning can accurately predict the T_g and help to understand the relevant structure-property relationships.

KEYWORDS: *glass transition temperature, polymer informatics, machine learning, molecular descriptors, structure-property relationships*

INTRODUCTION

The glass transition temperature (T_g) is one of the key thermal properties of polymeric materials that represents the temperature at which the relatively rigid glassy state of the polymer changes to a state with enhanced segmental molecular mobility. Its location is important in terms of dimensional stability, mechanical response, processing conditions and the temperature range in which a polymer can be used reliably. The thermal behavior thus plays a pivotal role in the choice and engineering of polymeric materials in various operating conditions (Pielichowski et al., 2022). There is no single molecular feature that is truly responsible for T_g . The thermomechanical behavior is affected by several factors that are related to polymer backbone structure, intermolecular interactions, chain mobility, side groups, and related structural features (Saba & Jawaideh, 2018). Chain dynamics and macroscopic material response can be further modified by changing the molecular interactions, such as in polymer systems with dynamic covalent bonds (Chakma & Konkolewicz, 2019). Predicting T_g accurately is also difficult because of the molecular basis of T_g . Among chemically similar polymers, there can be significant differences with regard to chain packing, conformational freedom, structural rigidity and local interactions. The glass transition behaviour may also be influenced by the physical environment and molecular constraints, and bulk polymer, confined polymer and interfacial glass transition regions may have different behaviour (Napolitano et al., 2017). It is also important to interpret the polymer structure when assessing polymer performance, considering other temperature-dependent phenomena that may coexist, such as crystallization (Fambri & Migliaresi, 2022). Computational methods offer the possibility to explore larger chemical spaces and to find structural trends related to T_g , but experimental measurement is still essential. One pathway for estimating polymer glass transitions has been molecular simulation that has been used for a long time. Changes in the behaviour of polymers with temperature can be simulated using molecular dynamics, although the estimates of T_g are dependent on the procedure used in the simulation, cooling protocol, timescale, and the criteria used to determine the transition (Wu, 2017). The ever-growing number of polymers whose structures are known from experiment has led to a complementary pathway through statistical learning. Rather than explicitly modelling all of the molecular interactions, a machine-learning model can learn relationships between structural representations and measured polymer properties. The relevance of data-driven structure–property analysis has thus grown in both polymer property prediction and identification of structural factors responsible for material behaviour (Amamoto, 2022).

The use of machine learning to model polymer properties has recently been shown to be possible. Nonlinear models, in particular, convolutional neural networks, have been applied to make quantitative mappings straight from chemical structure to polymer properties, demonstrating the capacity of nonlinear models for learning chemically informative representations (Miccio and Schwartz, 2020). Machine-learning-guided screening has also been used to link chemical structure with desired thermal properties in wider candidate spaces to discover high-temperature polymers (Tao et al., 2021). In the case of glass-transition prediction, structure-derived information has been used to devise machine learning-based models for conjugated polymers, where it has been shown that appropriate representations and learning algorithms can allow the prediction of T_g from the chemical structure (Alesadi et al., 2022). How the polymer chemistry is represented numerically is critical to the quality of such models. Molecular descriptors are numerical values derived from the structural features of a molecule and are related to properties including molecular size, topology, polarity, hydrogen bonding ability, ring content and conformational properties. They can be used for predictive modeling and interpretation since individual descriptors still have a good physicochemical meaning compared to many high-dimensional representations. Polymer informatics reviews have pointed out that the incorporation of molecular descriptors with machine learning is increasingly employed in polymer design and prediction of properties (Zhao et al., 2023). The general advancement of polymer science also reveals growing use of data-driven approaches to learn complex polymer structure–processing–property relations (Ge et al., 2025). Besides reproducing measured properties, machine learning can be used to obtain more information about a product, such as identifying its shape. Computational frameworks can help explore and design novel molecules and polymers, though there are limitations: the quality of the data, representation of the molecule and polymer, and the strategy for validation, as well as the capacity to generalize beyond known chemical structures (Chen et al., 2020). This is especially important for T_g , as structurally similar polymers and several different polymer compositions can appear in a set of data. Randomly shuffling such observations into model development and model evaluation subsets may not give as robust evidence of performance on truly unseen polymer systems.

An outstanding challenge remains, then, that the experimentally measured T_g values be coupled to interpretable structural descriptors, comparative machine-learning models, and polymer-system-aware validation in a single framework of structure–property relationships. Previous studies have demonstrated that

the computation of T_g is feasible, but relatively little focus has been given to the study of interpretable composition-weighted descriptors and the control of structural overlap during training and testing, as well as to characterization of nonlinear responses of the descriptors to T_g changes. Considering these factors can help to establish a better link between the predictability of the model and the molecular properties being modeled.

The primary goal of this work was to characterize the distribution of glass transition temperatures for a series of polymer materials that were tested experimentally, establish relationships between the polymer structural descriptors that can be interpreted and the composition characteristics and experimental glass transition temperature, and build and test machine-learning models to predict the glass transition temperature from **polymer structural characteristics**.

2. METHODOLOGY

2.1 Research Design

Relationships between polymer structural properties and experimental glass transition temperature (T_g) were analysed using a quantitative computational design, and machine-learning models were developed to predict polymer T_g . The analysis included data preparation, molecular descriptor generation, structure–property analysis, feature screening, model development, grouped cross-validation, independent testing, and model interpretation. Experimental T_g was taken as the continuous dependent variable, and composition-weighted molecular descriptors were used as the independent variables. Grouping units were maintained in polymer systems in order to avoid the same system being present in the training and testing subsets.

2.2 Data Source

The data used were taken from an open dataset from Manav et al., 2026. The experimental file included the identities of polymers, structural representation, compositions of the components, and experimental T_g values. The analysis was based on experimental observations only, and records with enumerated or predicted T_g values were not included in the analytical data set.

2.3 Data Preparation and Molecular Descriptor Generation

The experimental records were checked for duplication, missing values, validity of T_g and consistency of composition. No secondary-polymer information missing from homopolymer records was regarded as the absence of analytical data and was instead considered as a structural characteristic. The structural strings were pre-processed to contain polymer connection markers, which were then replaced by attachment points that handle wildcards. The following 10 descriptors were calculated for each polymer component: the number of rotatable bonds, the number of rings, the number of aromatic rings, the number of heavy atoms, the number of heteroatoms, the number of hydrogen-bond acceptors, the number of hydrogen-bond donors, molecular weight, topological polar surface area, and Fraction Csp³.

If the copolymer had several components, the composition-weighted values were calculated from the component-level descriptors using the primary-component fraction. If the polymer was a homopolymer, the primary-component descriptors were kept as such. The procedure resulted in the structural variables that were used in further analysis.

2.4 Structure–Property and Feature Analysis

The results of the experiments were summarized using mean, standard deviation, median, minimum, and maximum values overall and for each type of polymer. Linear relationships between structural descriptors and T_g were quantified using Pearson correlation, and monotonic relationships were quantified using Spearman rank correlation. The predictor redundancy was explored by the pairwise Pearson correlation and variance inflation factors (VIFs). Pairwise Pearson correlations were calculated among the structural descriptors, and highly correlated descriptor pairs with an absolute correlation of at least 0.80 were screened to reduce predictor redundancy. Six descriptors were selected: heteroatom count, ring count, rotatable-bond count, hydrogen-bond donor count, topological polar surface area, and Fraction Csp³.

2.5 Machine-Learning Development and Validation

T_g prediction using Elastic Net, Random Forest, Gradient Boosting and support vector regression models was evaluated. The data were partitioned by polymer system, and no systems from the training set were present in the held-out test set. To compare models and tune their hyperparameters, five-fold grouped cross-validation was implemented inside the training data set. However, the values of the descriptors were used directly in the tree-based models, while for Elastic Net and support vector regression, scaling was incorporated. Model selection was done by cross-validation RMSE, and then evaluation was performed on the held-out test set.

2.6 Model Evaluation and Interpretation

The coefficient of determination (R^2), the root mean squared error (RMSE), and the mean absolute error (MAE) were used to assess the performance of the model. The model that was trained from the grouped cross-validation was then tested on the independent test set. Pearson and Spearman correlations and residual analysis were also used to investigate the observed and predicted T_g values. Permutation importance was estimated to determine the contribution of each structural descriptor to the prediction. Partial-dependence analysis was used to explore the model-predicted relationships of the three most important descriptors over the range of the descriptors.

3. RESULTS

3.1 Dataset Characteristics and Experimental Glass Transition Temperature

Analytical data were exported to the final dataset for analysis, which included 666 experimental observations and 386 unique polymer systems, including 301 homopolymer and 365 copolymer observations. Experimental T_g, polymer composition, structural representations and derived molecular descriptors were included in the dataset. The primary polymer component (x1) was used to represent the composition, varying from 0 to 1. The following 10 composition-weighted structural descriptors were first created from the polymer structures: molecular weight, count of heavy atoms, count of hetero atoms, count of rings, count of aromatic rings, count of rotatable bonds, count of hydrogen-bond donors, count of hydrogen-bond acceptors, topological polar surface area, Fraction Csp³. The six descriptors were selected for predictive analysis after redundancy screening. Experimental T_g ranged from 130 to 685 K, with an overall mean of 337.29 ± 91.48 K and a median of 327.50 K. Homopolymers had a mean T_g of 367.29 K compared with 312.56 K for copolymers. Table 1 presents the overall and by polymer type numerical distribution of the experimental T_g values.

Table 1. Descriptive characteristics of experimental glass transition temperature

Group	N	Mean T _g (K)	SD T _g (K)	Median T _g (K)	Minimum T _g (K)	Maximum T _g (K)
Overall	666	337.29	91.48	327.50	130	685
Copolymer	365	312.56	43.40	314.00	215	390
Homopolymer	301	367.29	120.90	362.00	130	685

The experimental T_g distribution for the two polymer types, including the broader T_g range found between homopolymers and the narrower concentration of copolymer observations, is plotted in Figure 1.

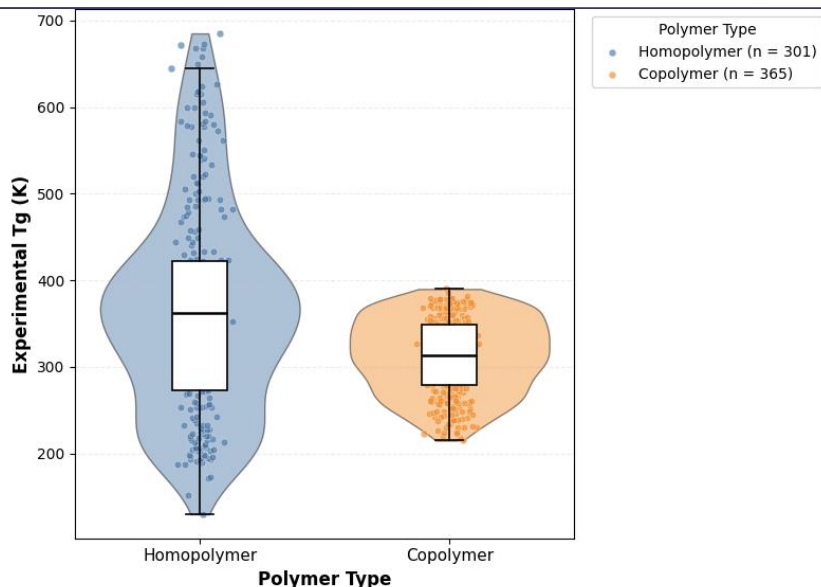


Figure 1. Distribution of experimental glass transition temperature by polymer type.

3.2 Polymer Structural Descriptors and Tg Relationships

The associations between experimental Tg and the composition-weighted structural descriptors varied in magnitude and direction. Composition-weighted ring count showed the strongest positive association with Tg (Pearson $r = 0.7851$; Spearman $\rho = 0.6741$), whereas composition-weighted Fraction Csp3 showed the strongest negative association (Pearson $r = -0.7574$; Spearman $\rho = -0.6610$). The topological polar surface area ($r = 0.5143$) and the number of heteroatoms ($r = 0.4261$) showed a positive correlation with Tg. The other coefficients for hydrogen-bond donor count were weaker. Rotatable-bond count had Pearson $r = -0.0526$ and Spearman $\rho = -0.2491$.

Table 2. Relationships between polymer structural descriptors and experimental glass transition temperature

Structural descriptor	Pearson r	Spearman ρ
Composition-Weighted Ring Count	0.7851	0.6741
Composition-Weighted Fraction Csp3	-0.7574	-0.6610
Composition-Weighted Topological Polar Surface Area	0.5143	0.1964
Composition-Weighted Heteroatom Count	0.4261	0.2066
Composition-Weighted Hydrogen-Bond Donor Count	0.1850	0.1328
Composition-Weighted Rotatable Bond Count	-0.0526	-0.2491

Initially, there were seven pairs of descriptors with absolute Pearson correlations of ≥ 0.80 . During the redundancy reduction, the following parameters were dropped: molecular weight, count of aromatic rings, count of hydrogen-bond acceptors, and count of heavy atoms. The resulting six-predictor dataset presented here was composed of heteroatom count, count of rings, count of rotatable bonds, count of hydrogen-bond donors, topological polar surface area and Fraction Csp3. The maximum absolute value of the inter-descriptor correlation was 0.7688, and the VIF values ranged from 1.285 to 3.754.

3.3 Training–Testing Partition and Cross-Validated Model Performance

A total of 499 observations of 308 polymer systems were assigned to training and 167 observations of 78 polymer systems to independent testing, with no overlap in polymer systems. Since complete polymer systems, instead of individual records, were assigned to partitions, the resulting observation proportions were 74.92% and 25.08%, respectively. Mean Tg was 336.82 ± 93.14 K in training data and 338.70 ± 86.59 K in testing data. There were 251 copolymer and 248 homopolymer observations in the training partition and 114 and 53 observations in the test partition. The training partition was used with 5-fold cross-validation. The tuned cross-

validation results are presented in Table 3, with Gradient Boosting attaining the lowest mean RMSE of 32.3296 K. Random Forest produced a mean RMSE of 35.2977 K, followed by SVR at 41.8774 K and Elastic Net at 44.0100 K. Gradient Boosting was therefore pre-selected for subsequent evaluation on the independent test partition.

Table 3. Grouped five-fold cross-validation performance of tuned machine-learning models

Model	Mean CV RMSE (K)	SD RMSE (K)	Minimum Fold RMSE (K)	Maximum Fold RMSE (K)
Gradient Boosting	32.3296	5.5342	26.7327	38.9953
Random Forest	35.2977	5.6437	29.6398	42.5895
SVR	41.8774	6.1919	34.4322	49.1644
Elastic Net	44.0100	4.5873	38.8879	50.5029

3.4 Held-Out Predictive Performance

For the 167-observation independent test set, the pre-selected Gradient Boosting model gave results of $R^2 = 0.8697$, RMSE = 31.1643 K, and MAE = 23.3959 K. Observed and predicted Tg values on the held-out test set had Pearson $r = 0.9380$ and Spearman $\rho = 0.8845$ (both $p < 0.001$). The difference between the grouped cross-validation and held-out RMSE was -1.1653 K. The mean residual, calculated as observed minus predicted Tg, was 8.3495 K, with a median absolute error of 19.4672 K, as shown in Figure 2.

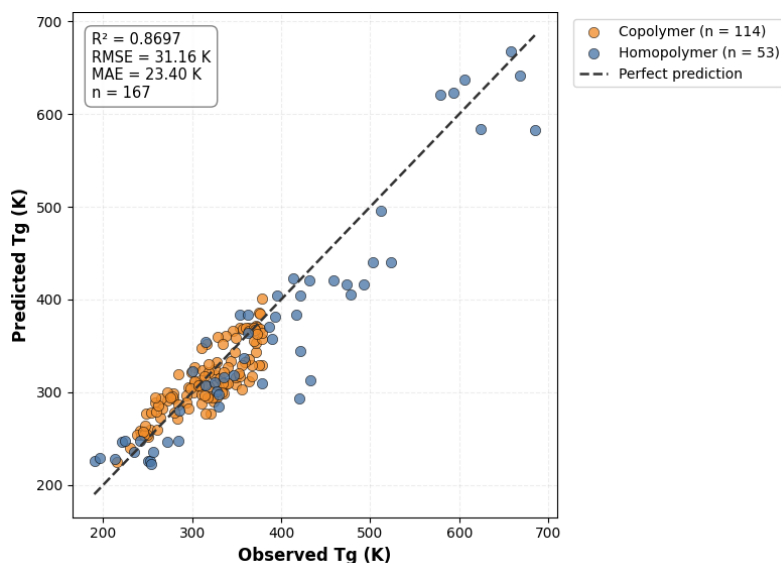


Figure 2. Observed versus predicted glass transition temperature for the pre-selected Gradient Boosting model on the held-out polymer-system test set. The dashed diagonal indicates perfect prediction.

The remaining performance of all four tuned algorithms is presented in Table 4. Random Forest yielded $R^2 = 0.8706$, RMSE = 31.0480 K, and MAE = 23.2049 K. Gradient Boosting yielded $R^2 = 0.8697$, while SVR and Elastic Net yielded R^2 values of 0.7902 and 0.7765, respectively. However, the Gradient Boosting model was still the one selected as the best model since it was the one that was previously selected by the grouped cross-validation results, instead of the test results.

Table 4. Predictive performance on the held-out polymer-system test set

Model	R ²	RMSE (K)	MAE (K)
Random Forest	0.8706	31.0480	23.2049
Gradient Boosting	0.8697	31.1643	23.3959
SVR	0.7902	39.5367	29.7110
Elastic Net	0.7765	40.8121	33.2314

3.5 Structural Descriptor Importance and Partial Dependence

The contribution of each of the descriptors retained to Gradient Boosting predictions was quantified by permutation analysis. Table 5 shows that composition-weighted ring count contributed most to the RMSE increase with a mean RMSE increase of 72.5138 K and relative importance of 49.81%. Rotatable-bond count ranked second at 30.0805 K (20.66%), followed by Fraction Csp3 at 22.5277 K (15.48%). The RMSEs were 9.4339 K, 7.4659, and 3.5469 for the topological polar surface area, heteroatom count, and hydrogen-bond donor count, respectively.

Table 5. Permutation importance of structural descriptors for the Gradient Boosting model

Structural descriptor	Mean RMSE increase (K)	SD RMSE increase (K)	Relative importance (%)
Composition-Weighted Ring Count	72.5138	3.0785	49.81
Composition-Weighted Rotatable Bond Count	30.0805	1.8360	20.66
Composition-Weighted Fraction Csp3	22.5277	1.3372	15.48
Composition-Weighted Topological Polar Surface Area	9.4339	1.3184	6.48
Composition-Weighted Heteroatom Count	7.4659	1.4361	5.13
Composition-Weighted Hydrogen-Bond Donor Count	3.5469	0.7313	2.44

Partial-dependence profiles were created for the three descriptors that had the greatest permutation importance. The partial dependence of T_g increased with increasing composition-weighted ring count and ranged from −40.63 to 208.78 K (Figure 3A), while the partial dependence decreased with an increase in the composition-weighted rotatable bond count and ranged from a maximum of 54.30 K to a minimum of −63.67 K (Figure 3B). Figure 3C shows a non-monotonic profile as a function of the composition-weighted Fraction Csp3, ranging from −62.79 to 18.27 K.

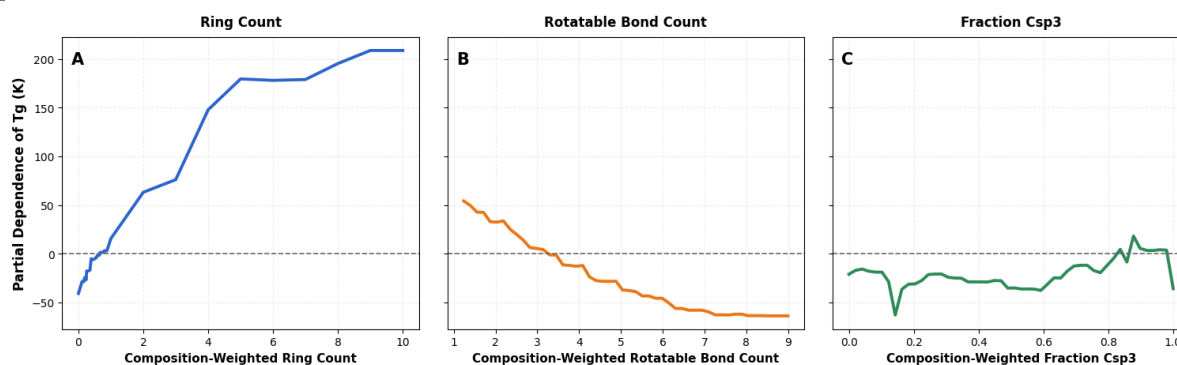


Figure 3. Partial dependence of key polymer structural descriptors in the Gradient Boosting model: (A) composition-weighted ring count; (B) composition-weighted rotatable bond count; and (C) composition-weighted Fraction Csp3.

4. DISCUSSION

These results show a good correlation between polymer structure and glass transition behaviour. The 666 observations of T_g were significantly different; the mean T_g of homopolymer being higher than that of copolymers, and the standard deviation was larger for homopolymers than for copolymers. The difference reflects the thermal diversity of polymers, and the extensive homopolymer family is an indication of the greater structural diversity within the homopolymer family. Retained composition-weighted descriptors, which reflect the molecular structure, were most associated with T_g: the strongest positive association was with the number of rings, while Fraction Csp3 had a strong inverse correlation. These trends suggest that there is a close correlation between structural rigidity and aromatic/cyclic character (higher T_g) and aliphatic character (lower T_g) in the chemical space explored. The same result was obtained from the multivariable analysis, for which the six descriptors that remained after removing highly correlated ones were able to retain useful structural information. The descriptors that had the highest correlation with the data and highest permutation importance were "Ring count" and "Rotatable-bond count", although the latter did not have as strong a Pearson correlation with the data, suggesting it contributed in a nonlinear or interaction way. This was confirmed by partial-dependence profiles, which showed that as the number of rings increased, the predicted T_g increased dramatically, while as the number of rotatable bonds increased, the predicted T_g decreased consistently. In fact, Fraction Csp3 showed a non-monotonic effect, meaning that it varied as a function of the structural context rather than having a consistent effect throughout the continuum. The predictive results also revealed that nonlinear ensemble models were able to model the T_g better than the regularized linear model. Gradient Boosting achieved the lowest RMSE during grouped cross-validation and was selected before evaluation of the held-out test set. Its independent test performance remained strong, with $R^2 = 0.8697$, RMSE = 31.16 K, and MAE = 23.40 K. Random Forest produced nearly identical held-out performance, with $R^2 = 0.8706$, RMSE = 31.05 K, and MAE = 23.20 K, but Gradient Boosting remained the pre-selected model because model selection was based on grouped cross-validation. The consistency of the cross-validation results and held-out RMSE further suggests that the polymer-system-aware split gave a consistent RMSE for the unseen polymer systems.

The advantage of a nonlinear and descriptor-based method for T_g prediction has been confirmed by previous studies. As seen here, the held-out performance of the Random Forest models is competitive with the performance observed in this case for the estimation of polymer T_g (Yu et al., 2025). Cheminformatics-based and molecular-dynamics-informed approaches have also shown the ability to translate polymer structure to predictive information for glass transition behavior (Karuth et al., 2021). Further, T_g prediction in donor-acceptor polymers has demonstrated that specific structural changes can be correlated with predictable changes in thermal response, further establishing the need for chemically interpretable structural variables (Zhang et al., 2020). Other polymer informatics projects also support the use of machine learning to explore structure-property relationships over large chemical spaces of polymers. While the more traditional trial-and-error approach to polymer design remains accepted, molecular design algorithms have proven successful in the identification of polymers with desired thermal properties, showing that computational screening can be valuable beyond trial-and-error experiments (Wu et al., 2019). As for Gradient Boosting, Random Forest, SVR, and Elastic Net, there are performance differences in this analysis where the polymer T_g was used as the target variable, which were also found in benchmark studies (Tao et al., 2021), with performance differences also observed depending on representation, learning algorithm and validation strategy. The benefits of data-driven

models in capturing complex relationships that are hard to represent with simple equations have also been reported in earlier work for the prediction of the Tg of polymers using machine learning (Zhang & Xu, 2020). In the context of polymer design, interpretability is becoming more and more important, especially when predictive models are used to assist with polymer design. QSPR-based machine-learning analysis of polyimide Tg has shown that molecular descriptors can provide both predictive capability and structural interpretation (Zhang et al., 2024). Methods combining machine learning and quantum-chemical data have also been used to enhance the understanding of polymers' thermal behavior by introducing chemically relevant features (Hickey et al., 2024). Interpretable machine-learning frameworks have also highlighted the importance of determining which descriptors contribute to Tg prediction rather than considering predictive accuracy alone (Uddin & Fan, 2024). Less complex machine-learning models have also yielded helpful estimates of Tg, which justify the use of descriptor-based workflows when calculation time and transparency are key considerations (Fatriansyah et al., 2024).

The results apply to polymer screening and preliminary materials selection. Composition-weighted ring count, rotatable-bond count and Fraction Csp3 are easily interpretable measures of the structural features that vary with Tg. The modeling framework can thus facilitate fast evaluation of potential polymer systems prior to the more resource-consuming experimental characterization. The conservative assessment of predictive performance using polymer-system-aware validation is also due to the fact that chemically related systems are not used in the model development and test sets.

A few limitations should be acknowledged. The analysis was limited to the chemical space spanned by the available experimental records, and the composition-weighted descriptors reduced features like detailed intermolecular interactions, architecture and monomer sequence. The conditions under which the process was carried out, and factors depending on the measurement that can affect Tg, were also not explicitly modeled. The framework should therefore be tested in the future on larger external polymer datasets, with the use of more powerful structural and physicochemical representations, and on the possibility of improving the prediction power of the framework while maintaining a clear and easily interpretable structure–property relationship that integrates interpretable descriptors with more advanced molecular representations.

5. CONCLUSION

Measurable structural features, such as ring content, conformational flexibility, and Fraction Csp3, had a strong correlation with the polymer glass transition temperature. These trends were found to be most significant with composition-weighted ring count and Fraction Csp3, with the latter having negative associations with thermal transition behavior, across 666 experimental observations. Six interpretable descriptors were retained after redundancy reduction to be used for machine-learning analysis. The cross-validated performance was the highest for Gradient Boosting and was chosen in advance of independent testing. For polymer systems not in the training set, it showed $R^2 = 0.8697$, RMSE = 31.16 K and MAE = 23.40 K, which are very good predictive results on unseen polymer systems. The importance of the remaining features was further confirmed by permutation importance, which found that ring count was the most important feature, followed by rotatable-bond count and Fraction Csp3. Our results demonstrate the usefulness of molecular descriptors and machine learning that takes into account the polymer system for predicting Tg. The framework offers a useful foundation for structure–property analysis, candidate screening and polymer data-driven design without losing the physicochemical interpretability.

REFERENCES

1. World Hel. Alesadi, A., Cao, Z., Li, Z., Zhang, S., Zhao, H., Gu, X., & Xia, W. (2022). Machine learning prediction of glass transition temperature of conjugated polymers from chemical structure. *Cell Reports Physical Science*, 3(6).
2. Amamoto, Y. (2022). Data-driven approaches for structure-property relationships in polymer science for prediction and understanding. *Polymer Journal*, 54(8), 957-967.
3. Chakma, P., & Konkolewicz, D. (2019). Dynamic covalent bonds in polymeric materials. *Angewandte Chemie*, 131(29), 9784-9797.
4. Chen, G., Shen, Z., Iyer, A., Ghumman, U. F., Tang, S., Bi, J., ... & Li, Y. (2020). Machine-learning-assisted de novo design of organic molecules and polymers: opportunities and challenges. *Polymers*, 12(1), 163.
5. Fambri, L., & Migliaresi, C. (2022). Crystallization and thermal properties. *Poly (Lactic Acid) Synthesis, Structures, Properties, Processing, Applications, and End of Life*, 135-151.
6. Fatriansyah, J. F., Linuwih, B. D. P., Andreano, Y., Sari, I. S., Federico, A., Anis, M., ... & Jaafar,

- M. (2024). Prediction of glass transition temperature of polymers using simple machine learning. *Polymers*, 16(17), 2464.
7. Ge, W., De Silva, R., Fan, Y., Sisson, S. A., & Stenzel, M. H. (2025). Machine learning in polymer research. *Advanced Materials*, 37(11), 2413695.
 8. Hickey, K., Feinstein, J., Sivaraman, G., MacDonell, M., Yan, E., Matherson, C., ... & Picel, K. (2024). Applying machine learning and quantum chemistry to predict the glass transition temperatures of polymers. *Computational Materials Science*, 238, 112933.
 9. Karuth, A., Alesadi, A., Xia, W., & Rasulev, B. (2021). Predicting glass transition of amorphous polymers by application of cheminformatics and molecular dynamics simulations. *Polymer*, 218, 123495.
 10. Manav, B., Afzal, M. A. F., Alex, C., Andrea, B., & Mathew, H. (2026). Glass Transition Prediction of Binary Copolymers Across Large Chemical Spaces Using Machine Learning and Physics-Based Modeling [Dataset]. Zenodo. <https://doi.org/10.5281/zenodo.19242815>
 11. Miccio, L. A., & Schwartz, G. A. (2020). From chemical structure to quantitative polymer properties prediction through convolutional neural networks. *Polymer*, 193, 122341.
 12. Napolitano, S., Glynos, E., & Tito, N. B. (2017). Glass transition of polymers in bulk, confined geometries, and near interfaces. *Reports on Progress in Physics*, 80(3), 036602.
 13. Pielichowski, K., Njuguna, J., & Majka, T. M. (2022). Thermal degradation of polymeric materials. Elsevier.
 14. Saba, N., & Jawaid, M. (2018). A review on thermomechanical properties of polymers and fibers reinforced polymer composites. *Journal of industrial and engineering chemistry*, 67, 1-11.
 15. Tao, L., Chen, G., & Li, Y. (2021). Machine learning discovery of high-temperature polymers. *Patterns*, 2(4).
 16. Tao, L., Varshney, V., & Li, Y. (2021). Benchmarking machine learning models for polymer informatics: an example of glass transition temperature. *Journal of Chemical Information and Modeling*, 61(11), 5395-5413.
 17. Uddin, M. J., & Fan, J. (2024). Interpretable machine learning framework to predict the glass transition temperature of polymers. *Polymers*, 16(8), 1049.
 18. Wu, C. (2017). Re-examining the procedure for simulating polymer T_g using molecular dynamics. *Journal of Molecular Modeling*, 23(9), 270.
 19. Wu, S., Kondo, Y., Kakimoto, M. A., Yang, B., Yamada, H., Kuwajima, I., ... & Yoshida, R. (2019). Machine-learning-assisted discovery of polymers with high thermal conductivity using a molecular design algorithm. *Npj Computational Materials*, 5(1), 66.
 20. Yu, X., Fang, Z., & Wu, F. (2025). Random forest based approach for predictions of glass transition temperatures in polymers. *Polymer Engineering & Science*, 65(11), 6238-6247.
 21. Zhang, S., Alesadi, A., Selivanova, M., Cao, Z., Qian, Z., Luo, S., ... & Gu, X. (2020). Toward the prediction and control of glass transition temperature for donor–acceptor polymers. *Advanced Functional Materials*, 30(27), 2002221.
 22. Zhang, T., Wang, S., Chai, Y., Yu, J., Zhu, W., Li, L., & Li, B. (2024). Prediction and Interpretability Study of the Glass Transition Temperature of Polyimide Based on Machine Learning with Quantitative Structure–Property Relationship (T_g-QSPR). *The Journal of Physical Chemistry B*, 128(36), 8807-8817.
 23. Zhang, Y., & Xu, X. (2020). Machine learning glass transition temperature of polymers. *Heliyon*, 6(10).
 24. Zhao, Y., Mulder, R. J., Houshyar, S., & Le, T. C. (2023). A review on the application of molecular descriptors and machine learning in polymer design. *Polymer Chemistry*, 14(29), 3325-3346.