ingegneria sismica
International Journal of Earthquake Engineering
Anno XLIII - Num. 3 - 2026

Construction of a Predictive Model for the Thermal Stability of Crystal-Doped Polymer Composites and Evaluation of Multifactorial Effects

Yuze Xie[1,*], *Fengyuan Zhang*[1] *and Jiayu Guo*[1]

[1] School of Materials Science and Engineering, Qilu University of Technology (Shandong Academy of Sciences), Jinan, China, 250353

SUMMARY: *This study develops a data-driven framework for predicting the thermal stability of crystal-doped polymer composites and evaluating multifactorial effects based on a structured materials database and machine learning methods. Feature variables related to the polymer matrix, crystalline dopants, composite formulation, processing conditions, and thermal analysis parameters were systematically organized, and predictive models were constructed for key thermal stability indicators, including T5%, T10%, Tmax, and char residue. The results indicate that ensemble learning models can more effectively describe the nonlinear relationships between thermal stability and multi-source variables than conventional linear models, with the XGBoost model exhibiting higher prediction accuracy and better generalization capability. Further feature importance analysis and multifactorial effect evaluation reveal that the intrinsic thermal stability of the polymer matrix, crystalline dopant type, filler loading, particle size, surface modification, testing atmosphere, and heating rate are key factors governing thermal stability. An appropriate amount of crystalline dopant can improve the thermal degradation behavior of polymer composites through physical barrier effects, interfacial restriction, heat-transfer regulation, and catalytic char formation, whereas excessive loading may weaken the enhancement efficiency due to filler aggregation and interfacial defects. The proposed predictive model and multifactorial evaluation strategy provide a data-driven basis for the rational design of crystal-doped polymer composites with enhanced thermal stability.*

KEYWORDS: *Crystal-Doped Polymer Composites, Thermal Stability, Machine Learning, Multifactorial Effect Evaluation*

1 Introduction

Polymer composites have been widely used in electronic packaging, aerospace engineering, automotive manufacturing, thermal management devices, and flame-retardant protective materials because of their low density, facile processability, structural tunability, and balanced overall performance[1-3]. However, most polymeric materials are susceptible to intensified segmental motion, thermo-oxidative degradation, backbone scission, and the release of volatile degradation products under elevated temperatures, resulting in reduced dimensional stability, mechanical reliability, and service lifetime. Therefore, improving the thermal stability of polymer composites remains a critical issue in materials science and engineering. In recent years, crystal doping has emerged as an effective modification strategy. Crystalline fillers, such as inorganic crystals, layered crystals, ceramic particles, metal oxides, nitrides, layered double

*17683184637@163.com
https://doi.org/10.65102/is20261279 Published: 10 June, 2026

hydroxides, and metal–organic frameworks, usually possess high intrinsic thermal stability, ordered structural features, and tunable interfacial characteristics. These crystalline dopants can improve the thermal degradation behavior of polymer matrices through physical barrier effects, heat-transfer regulation, interfacial restriction, and catalytic char formation[4-6].

Current studies on the thermal stability of crystal-doped polymer composites have mainly focused on specific polymer matrices or individual crystalline filler systems. Previous investigations have demonstrated that filler type, loading content, particle size, morphology, dispersion state, surface modification, and interfacial interactions between the polymer matrix and crystalline dopants can significantly affect the thermal stability of composites. Meanwhile, experimental parameters in thermogravimetric analysis, such as testing atmosphere, heating rate, and temperature range, can also alter the apparent thermal degradation characteristics of polymer composites. Nevertheless, these factors are usually coupled in a complex and nonlinear manner, making it difficult for conventional single-factor experimental analysis to fully identify their individual contributions and interaction effects. In addition, most existing studies emphasize experimental observations and mechanistic explanations, while data-driven models capable of cross-system property prediction and multifactorial effect evaluation remain insufficient[7-9].

In this study, crystal-doped polymer composites are investigated as the target material system, and a thermal-stability-oriented materials database is constructed. Feature variables related to the polymer matrix, crystalline dopants, composite formulation, processing conditions, and thermal analysis parameters are systematically extracted. Based on these descriptors, multiple machine learning models are developed to predict key thermal stability indicators, including characteristic thermal decomposition temperatures and char residue, and the optimal model is identified through comparative performance evaluation. Furthermore, feature importance analysis and multifactorial effect evaluation are conducted to clarify the effects of crystal type, filler loading, particle size, surface modification, testing atmosphere, and heating rate on thermal stability. This work aims to provide a data-driven predictive approach and mechanistic interpretation framework for the rational design of thermally stable crystal-doped polymer composites.

2 Materials Database and Feature Construction

To develop a predictive model for the thermal stability of crystal-doped polymer composites, a structured database containing material composition, crystalline dopant characteristics, composite processing parameters, and thermal analysis results was first established. The thermal stability of polymer composites is not only governed by the intrinsic degradation behavior of the polymer matrix, but also affected by the properties of crystalline dopants, filler loading, interfacial state, processing route, and testing conditions. Therefore, rational data organization and feature construction are essential for achieving reliable model prediction. This section describes the data collection strategy, thermal stability indicators, feature variable construction, and data preprocessing procedures[10-12].

2.1 Data collection

The dataset in this study was constructed based on thermal stability data of crystal-doped polymer composites. The data can be obtained from published literature, laboratory measurements, or open materials databases. Each data entry corresponds to a specific composite sample, including the polymer matrix, crystalline dopant, filler loading, preparation method, testing conditions, and thermogravimetric analysis results. To ensure data comparability and

reliability, only samples with clearly reported material compositions, well-defined experimental conditions, and complete thermal stability parameters were included as valid data.

During data screening, polymer composites doped with inorganic crystals, layered crystals, ceramic particles, metal oxides, metal hydroxides, layered double hydroxides, nitrides, and metal–organic frameworks were preferentially included. Data lacking key variables, clear thermal analysis conditions, or extractable thermal stability indicators were excluded from subsequent model training. When multiple samples with different filler loadings, particle sizes, or surface modification conditions were reported in the same study, they were treated as independent data points to enhance the model's ability to recognize variable-dependent changes.

In addition, to reduce data deviation caused by different literature sources, variable names, units, and classification standards were unified during data organization. For example, filler loading was converted to weight percentage, temperature-related parameters were unified in degrees Celsius, heating rate was expressed as °C min^{-1}, and char residue was recorded as the residual mass percentage at a specified final temperature. Through these procedures, a structured data matrix suitable for machine learning modeling was obtained.

2.2 Thermal stability indicators

Thermal stability indicators serve as the target variables of the predictive model, and their selection directly determines the practical value of the model. The thermal stability of crystal-doped polymer composites is commonly characterized by thermogravimetric analysis. The thermogravimetric curve reflects the mass loss behavior of materials during heating, while the derivative thermogravimetric curve further reveals changes in the degradation rate at different stages. Based on thermogravimetric results, several representative parameters were selected as thermal stability indicators in this study.

First, T5% and T10% represent the temperatures at which the mass loss reaches 5% and 10%, respectively, and are commonly used to describe the initial thermal degradation stability of materials. Higher T5% or T10% values indicate stronger resistance to early-stage thermal decomposition. Second, Tmax represents the temperature corresponding to the maximum mass loss rate, reflecting the stability of the main thermal degradation stage. For most polymer composites, an increase in Tmax generally indicates a delayed main-chain degradation or major decomposition process. Finally, char residue denotes the residual mass percentage at the final high-temperature point, which is often used to evaluate the char-forming ability and high-temperature structural retention of composites.

In this study, T5%, T10%, Tmax, and char residue can be selected as the primary prediction targets. If Tonset, T50%, or DTG peak temperature are available in part of the dataset, they can also be used as auxiliary evaluation parameters depending on data completeness. It should be noted that the definition of Tonset may vary across studies, such as the tangent method, fixed mass loss method, or manual determination. Therefore, T5% and T10% usually offer better comparability for cross-literature modeling[13-15].

2.3 Feature variable construction

To adequately describe the relationship between composition–structure characteristics and thermal stability in crystal-doped polymer composites, feature variables were constructed from four aspects: polymer matrix, crystalline dopant, composite formulation and processing conditions, and thermal analysis conditions.

First, polymer matrix descriptors were used to characterize the baseline thermal stability of the composites. Different polymer matrices exhibit distinct thermal degradation behaviors due to differences in molecular structure, chain rigidity, polar groups, aromatic content, and

crystallization ability. For systems with sufficient data availability, polymer type, glass transition temperature, melting temperature, crystallinity, T5% of the neat polymer, Tmax of the neat polymer, and char residue of the neat polymer can be extracted. Among these variables, the thermal stability parameters of the neat polymer are particularly important because they strongly determine the baseline thermal resistance of the final composite.

Second, crystalline dopant descriptors were used to characterize the physicochemical properties of the fillers. Different crystalline dopants vary significantly in intrinsic thermal stability, size effect, interfacial interaction, and char-promoting ability. In this study, dopant type, chemical composition, crystal structure, morphology, particle size, specific surface area, thermal conductivity, surface modification state, and intrinsic thermal stability of the dopant can be selected as descriptors. For instance, layered crystalline fillers usually exhibit strong barrier effects, ceramic crystalline fillers may improve thermal transport stability, whereas some metal-containing crystals or MOF-based materials may promote catalytic char formation.

Third, composite formulation and processing variables were used to describe the incorporation and dispersion state of crystalline dopants within the polymer matrix. Filler loading is one of the most important compositional variables. An appropriate amount of filler can enhance thermal shielding and interfacial restriction, whereas excessive filler loading may lead to aggregation, defects, or interfacial discontinuity, thereby weakening the improvement in thermal stability. Therefore, the relationship between filler loading and thermal stability is often nonlinear rather than simply proportional. In addition, preparation method, processing temperature, processing time, compatibilizer usage, and surface functionalization method may also influence filler dispersion and interfacial bonding strength.

Finally, thermal analysis conditions were incorporated into the feature system. Thermogravimetric results are affected not only by the intrinsic thermal degradation behavior of materials, but also by testing atmosphere, heating rate, temperature range, and sample mass. For example, under nitrogen atmosphere, materials mainly undergo inert thermal decomposition, whereas in air atmosphere, thermo-oxidative degradation can significantly alter decomposition temperature and char residue behavior. A higher heating rate generally shifts the apparent degradation temperature toward a higher temperature region. Therefore, incorporating testing conditions as input variables can improve the generalization ability of the model under different experimental conditions. As shown in table 1.

Table 1: Feature variables for thermal stability prediction of crystal-doped polymer composites

Feature category	Representative variables	Description
Polymer matrix descriptors	Polymer type, Tg, Tm, crystallinity, T5% of neat polymer, Tmax of neat polymer, char residue of neat polymer	Describe the intrinsic thermal stability and structural characteristics of the polymer matrix
Crystal dopant descriptors	Dopant type, chemical composition, crystal structure, morphology, particle size, specific surface area, thermal conductivity, surface modification	Describe the physicochemical properties of crystalline fillers
Composite formulation and processing descriptors	Filler loading, preparation method, processing temperature, processing time, compatibilizer, functionalization method	Describe the composition, dispersion state, and interfacial structure of composites
Thermal analysis descriptors	Atmosphere, heating rate, temperature range, sample mass	Describe the experimental conditions of thermal analysis

2.4 Data preprocessing

Before machine learning modeling, the raw data were systematically preprocessed to improve data quality and model stability. First, data from different sources were unified in terms of units and formats. For example, temperature was expressed in °C, filler loading was expressed in wt%, and heating rate was expressed in °C min^{-1}. Categorical variables, such as polymer type, crystal type, preparation method, and testing atmosphere, were transformed into numerical features using one-hot encoding or label encoding. Continuous variables, such as filler loading, particle size, heating rate, and thermal stability indicators, were normalized or standardized according to the requirements of different models.

Second, missing values and outliers were carefully treated. For continuous variables with only a small number of missing values, mean imputation, median imputation, or similarity-based imputation could be applied. Data with severe missing information in key variables were removed from the dataset to avoid introducing large uncertainty. Outliers were preliminarily identified using box plots, Z-scores, or the interquartile range method, and then examined against the original literature or experimental records to determine whether they resulted from experimental error, recording mistakes, or genuine material behavior. Special material behaviors that were physically reasonable were not simply removed, but retained for further interpretation.

Finally, the preprocessed dataset was divided into training and testing subsets. The training set was used for model parameter learning, while the testing set was used to evaluate the generalization ability of the model. To reduce the randomness associated with data splitting, especially for relatively small datasets, k-fold cross-validation can be applied during model training and validation. Through data standardization, variable encoding, missing value treatment, and cross-validation, a stable and reliable input data system was established, providing a foundation for subsequent predictive model development and multifactorial effect evaluation.

3 Predictive Model Development

After establishing the materials database and constructing the feature variables, predictive models were further developed to estimate the thermal stability of crystal-doped polymer composites. Since the thermal stability of polymer composites is jointly governed by the polymer matrix, crystalline dopants, filler loading, interfacial state, processing route, and thermal analysis conditions, the relationships among these variables are generally nonlinear. Therefore, this study does not rely on a single predictive model. Instead, multiple machine learning algorithms were compared to evaluate their applicability to thermal stability prediction. The model development process mainly includes model selection, model training and validation, and the definition of evaluation metrics.

3.1 Model selection

To systematically evaluate the predictive capability of different modeling approaches for crystal-doped polymer composites, linear models, kernel-based models, ensemble learning models, and neural network models were selected for comparison. These models have different learning mechanisms and application scenarios, allowing the mapping relationships between material descriptors and thermal stability indicators to be captured from different perspectives.

First, multiple linear regression, MLR was used as the baseline model. This model assumes a linear relationship between input features and target variables and has the advantages of a simple structure, high interpretability, and low computational cost. Although the thermal

stability of polymer composites is generally affected by nonlinear multifactorial interactions, the linear regression model provides a useful baseline for determining whether more complex models can truly improve prediction performance. The basic expression of MLR can be written as:

$$y = \beta_0 + \sum_{i=1}^{n} \beta_i x_i + \varepsilon \quad (1)$$

where $y$represents the target thermal stability indicator, such as T5%, T10%, Tmax, or char residue; $x_i$is the i-th input feature; $\beta_0$is the intercept; $\beta_i$is the regression coefficient; and $\varepsilon$is the error term.

Second, support vector regression, SVR was used to address small-sample and nonlinear prediction problems. SVR maps the original features into a high-dimensional space through a kernel function and constructs an optimal regression function in that space. For materials datasets, SVR has relatively low requirements for sample size and can show good generalization ability when nonlinear relationships are present. In this study, the radial basis function, RBF, can be adopted as the kernel function:

$$K(x_i, x_j) = \exp(-\gamma \mid\mid x_i - x_j \mid\mid^2) \quad (2)$$

where $\gamma$is the kernel parameter that controls the influence range of sample similarity. The penalty parameter C, insensitive loss parameter ε, and kernel parameter γ are key hyperparameters affecting the performance of SVR.

Third, random forest, RF was employed to establish a decision-tree-based ensemble regression model. RF constructs multiple decision trees and averages their prediction outputs, thereby reducing the variance of a single tree model and improving model robustness. For composite materials datasets containing different types of features, RF can effectively handle nonlinear relationships and feature interactions. Moreover, it can provide feature importance results, which are useful for subsequent multifactorial effect evaluation. The main hyperparameters of RF include the number of trees, maximum tree depth, minimum number of samples required for node splitting, and the number of features considered at each split.

Fourth, extreme gradient boosting, XGBoost was selected as one of the main high-accuracy prediction models. XGBoost is a gradient boosting tree algorithm that iteratively fits the residuals of previous models to gradually improve overall prediction accuracy. Compared with RF, XGBoost often exhibits stronger fitting ability for structured data and can effectively capture complex nonlinear relationships and high-order feature interactions. Its objective function can be expressed as:

$$Obj = \sum_{i=1}^{m} l(y_i, \hat{y}_i) + \sum_{k=1}^{K} \Omega(f_k) \quad (3)$$

where $l(y_i, \hat{y}_i)$ is the loss function used to measure the error between predicted and experimental values, and $\Omega(f_k)$is the regularization term used to control model complexity and reduce overfitting. The key hyperparameters of XGBoost include learning rate, maximum tree depth, subsampling ratio, number of estimators, and regularization coefficients.

Finally, artificial neural network, ANN can be used as a comparison model for complex nonlinear mapping. ANN consists of an input layer, hidden layers, and an output layer, and can learn complex variable relationships through nonlinear activation functions. For thermal stability prediction, ANN is capable of capturing multifactorial coupling effects; however, its performance is highly sensitive to sample size, network architecture, and training parameters.

When the dataset is limited, ANN may suffer from overfitting. Therefore, in this study, ANN is used only as a supplementary comparison model rather than the sole predictive approach.

Overall, MLR was used as a linear baseline model, SVR as a nonlinear model suitable for small datasets, RF and XGBoost as the main ensemble learning models, and ANN as a supplementary nonlinear model. By comparing multiple models, the optimal algorithm for predicting the thermal stability of crystal-doped polymer composites can be more comprehensively identified.

3.2 Model training and validation

The model training and validation procedure directly determines the reliability and generalization ability of the prediction results. To ensure that the models can effectively learn the structure–property relationships of crystal-doped polymer composites, a standardized workflow including data splitting, cross-validation, and hyperparameter optimization was adopted.

First, the preprocessed dataset was divided into training and testing subsets. The training set was used for model learning, while the testing set was used for final performance evaluation. A splitting ratio of 80:20 or 70:30 is commonly used. For relatively small materials datasets, an 80:20 split is recommended to ensure that sufficient samples are available for model training. Meanwhile, a fixed random seed should be applied during data splitting to improve the reproducibility of the results and reduce the randomness caused by a single data split.

Second, k-fold cross-validation was further performed within the training set. Specifically, the training set was divided into $k$subsets. In each iteration, $k-1$subsets were used for model training, and the remaining subset was used for validation. This process was repeated $k$times, ensuring that each subset served once as the validation set. The final validation performance was calculated as the average result across the $k$iterations. For small- to medium-sized materials datasets, 5-fold cross-validation or 10-fold cross-validation can be adopted. Cross-validation can effectively reduce the uncertainty caused by data splitting and improve the stability of model evaluation.

The model training workflow can be summarized as follows:

1. The raw data were processed through unit unification, missing value treatment, outlier identification, and variable encoding.
2. The dataset was divided into training and testing subsets.
3. k-fold cross-validation was performed on the training set.
4. Hyperparameter optimization was conducted for different models.
5. Models were retrained using the optimal hyperparameters.
6. Final model performance was evaluated on the testing set.
7. The model with the best overall performance was selected for subsequent feature importance analysis and multifactorial effect evaluation.

For hyperparameter optimization, grid search, random search, or Bayesian optimization can be used. For models with relatively few hyperparameters, such as SVR, grid search can systematically explore the optimal combination of C, ε, and γ. For models with larger parameter spaces, such as RF and XGBoost, random search or Bayesian optimization can improve search efficiency and reduce computational cost.

The key hyperparameters of different models are summarized in Table 2.

Table 2: Key hyperparameters for model optimization.

Model	Hyperparameters	Description
SVR	C, ε, γ, kernel	Control penalty strength, insensitive loss range, and kernel function
RF	n_estimators, max_depth, min_samples_split, min_samples_leaf, max_features	Control tree number, tree complexity, and feature sampling
XGBoost	learning_rate, n_estimators, max_depth, subsample, colsample_bytree, reg_alpha, reg_lambda	Control boosting process, model complexity, sampling, and regularization
ANN	hidden layers, neurons, learning rate, batch size, epochs, activation function	Control network structure and training process

During model training, the requirement for data standardization should also be considered. MLR, SVR, and ANN are sensitive to the scale of input variables; therefore, continuous variables should be standardized before training, typically with a mean of 0 and a standard deviation of 1. RF and XGBoost are relatively insensitive to feature scaling and do not strictly require standardization. Nevertheless, a unified preprocessing pipeline can still be applied to maintain consistency in the input data.

For multi-target prediction, two strategies can be adopted. The first is to establish independent models for T5%, T10%, Tmax, and char residue. The second is to use multi-output regression models to predict multiple thermal stability indicators simultaneously. Considering that different thermal stability indicators represent different physical meanings and may be influenced by different factors, this study recommends a **single-target modeling strategy**, in which an independent model is trained for each thermal stability indicator. This strategy improves model interpretability and facilitates comparison of how different factors affect initial degradation temperature, main decomposition temperature, and char residue.

To avoid overfitting, several control strategies were introduced during model training. For RF and XGBoost, model complexity can be controlled by limiting the maximum tree depth, setting the minimum number of samples in leaf nodes, and introducing regularization parameters. For ANN, early stopping, dropout, or L2 regularization can be adopted. For SVR, the balance between fitting accuracy and generalization ability can be adjusted through C and ε. Finally, the model that exhibits stable performance in both cross-validation and testing results is selected as the optimal prediction model.

3.3 Model evaluation metrics

To comprehensively evaluate the prediction performance of different models for the thermal stability of crystal-doped polymer composites, the coefficient of determination R^2, mean absolute error,MAE,and root mean square error,RMSE,were used as the primary evaluation metrics. These three indicators assess model performance from the perspectives of goodness of fit, average error magnitude, and error fluctuation, respectively.

First, the coefficient of determination $R^2$was used to measure the ability of the model to explain the variation in the target variable. It is calculated as:

$$R^2 = 1 - \frac{\sum_{i=1}^{m}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{m}(y_i - \bar{y})^2} \tag{4}$$

where y_i is the experimental value, $\hat{y}_i$ is the predicted value, $\bar{y}$ is the mean value of the experimental data, and m is the number of samples. A value of R^2 closer to 1 indicates stronger explanatory power. If R^2 is close to 0, the prediction ability of the model is similar to that of using the mean value as the prediction result. If R^2 is lower than 0, the model shows poor predictive performance.

Second, mean absolute error, MAE, was used to evaluate the average absolute deviation between the predicted and experimental values:

$$MAE = \frac{1}{m}\sum_{i=1}^{m} | y_i - \hat{y}_i | \tag{5}$$

MAE has a clear physical meaning and directly reflects the average prediction error. For example, when the prediction target is T5% or Tmax, the unit of MAE is °C, indicating the average temperature deviation between predicted and experimental values. A lower MAE indicates that the prediction results are closer to the experimental results.

Third, root mean square error, RMSE, was used to measure the overall fluctuation of prediction errors:

$$RMSE = \sqrt{\frac{1}{m}\sum_{i=1}^{m}(y_i - \hat{y}_i)^2} \tag{6}$$

Compared with MAE, RMSE is more sensitive to large errors. Therefore, when RMSE is significantly higher than MAE, the model may have large prediction deviations for certain samples. In materials property prediction, RMSE can help identify whether the model performs poorly for specific composite systems, extreme filler loadings, or special testing conditions.

In addition to these metrics, relative error or mean absolute percentage error, MAPE, can be used as supplementary indicators, especially when comparing prediction targets with different numerical ranges. MAPE is calculated as:

$$MAPE = \frac{100\%}{m}\sum_{i=1}^{m} | \frac{y_i - \hat{y}_i}{y_i} | \tag{7}$$

However, when the target value is close to zero, MAPE may fluctuate significantly. Therefore, for indicators such as char residue, which may have relatively low values, MAPE is used only as a supplementary reference rather than a primary evaluation metric.

To provide a more intuitive assessment of model performance, experimental-versus-predicted scatter plots, residual distribution plots, and error frequency histograms can also be generated. If the scatter points are close to the $y = x$ diagonal line, the model shows high prediction accuracy. If the residuals are randomly distributed around zero, no obvious systematic bias exists. Conversely, if the residuals show a regular trend with increasing or decreasing predicted values, the model may have failed to capture certain key variables or nonlinear relationships.

In summary, R^2, MAE, and RMSE were used as the main evaluation metrics, while predicted-versus-experimental plots and residual analyses were used as auxiliary evaluation methods. The model with a higher R^2, lower MAE and RMSE, and more uniformly distributed residuals was identified as the optimal predictive model and further used for feature importance analysis and multifactorial effect evaluation.

4 Results and Discussion

This section discusses the prediction results for the thermal stability of crystal-doped polymer composites. First, the basic characteristics of the constructed dataset are analyzed to clarify the distribution of polymer matrices, crystalline dopants, and thermal stability indicators. Subsequently, the prediction performance of different machine learning models is compared, and the optimal model is identified. Based on the optimal model, predicted-versus-experimental analysis, feature importance evaluation, and multifactorial effect analysis are further conducted to reveal the key variables governing thermal stability and their potential mechanisms.

4.1 Dataset characteristics

To ensure that the predictive model covers diverse crystal-doped polymer composite systems, the constructed dataset includes various polymer matrices and crystalline dopants. The polymer matrices mainly include epoxy resin, polypropylene, polyether ether ketone, poly (methyl methacrylate), poly (vinyl chloride), polyamide, polyimide, and other engineering or functional polymers. The crystalline dopants mainly include metal oxides, ceramic crystals such as boron nitride, aluminum nitride and silicon carbide, layered double hydroxides, clay-based layered crystals, metal–organic frameworks, and graphene-based crystalline materials.

Figures 1 and 2 show the distribution of polymer matrix types and crystalline dopant types in the dataset, respectively.

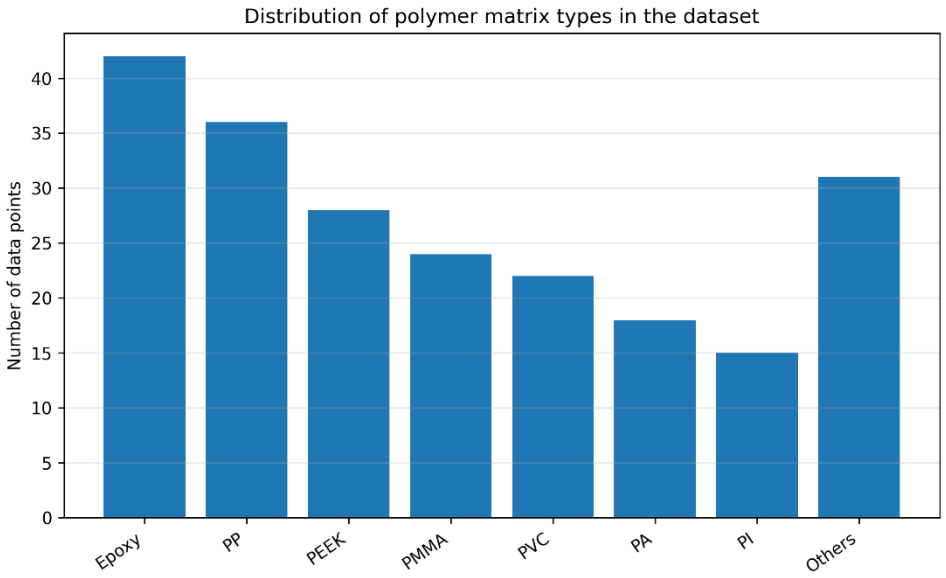

Figure 1: Distribution of polymer matrix types in the dataset.

As shown in Figure 1, epoxy resin, polypropylene, and polyether ether ketone account for a relatively large proportion of the dataset. This is mainly because these polymers are widely used in electronic packaging, thermal management, structural composites, and flame-retardant materials, and their thermal stability modification has been extensively investigated. In contrast, although polyimide possesses high intrinsic thermal stability, the available data related to crystal-doped modification are relatively limited, resulting in a lower proportion in the dataset.

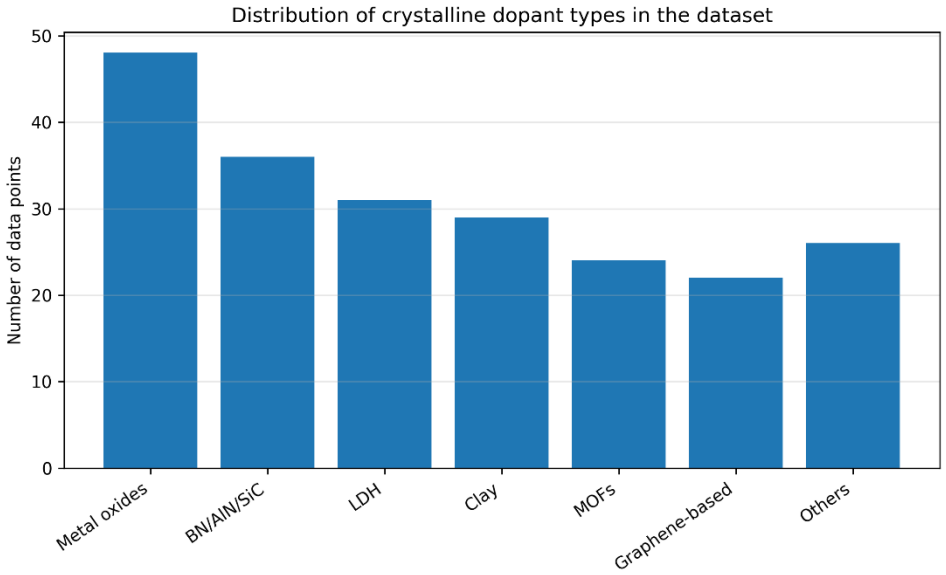

Figure 2: Distribution of crystalline dopant types in the dataset.

As shown in Figure 2, metal oxides, ceramic crystals, and layered crystalline fillers are the most common crystalline dopants. Metal oxides are widely used because of their compositional diversity, high thermal stability, and tunable interfacial characteristics. Ceramic crystals such as BN, AlN, and SiC usually exhibit high thermal conductivity and structural stability. Layered double hydroxides and clay-based crystals can delay polymer degradation through a layered barrier effect. The inclusion of different crystalline dopants provides a broad data basis for learning the influence of crystal structure on thermal stability.

Table 3: Statistical summary of major variables in the dataset.

Variable	Minimum	Maximum	Mean	Standard deviation
Filler loading, wt%	0.5	30.0	7.8	5.6
Particle size, nm	20	5000	684	915
Heating rate, °C min^{-1}	5	20	10.8	4.2
T5%, °C	218	524	356	62
T10%, °C	246	552	382	58
Tmax, °C	304	621	456	67
Char residue, %	1.8	54.6	18.7	12.9

As shown in Table 3, the filler loading is mainly distributed in the low-to-medium range, with an average value of 7.8 wt%. This range is consistent with the practical design of most polymer composites, because excessive filler loading may improve thermal shielding or char formation but can also cause aggregation, processing difficulty, and interfacial defects. The thermal stability indicators, including T5%, T10%, and Tmax, exhibit broad distributions, indicating that the dataset covers various systems from common thermoplastics to high-performance engineering polymers. This diversity is beneficial for learning more generalizable structure–property relationships.

4.2 Comparison of prediction models

To evaluate the applicability of different machine learning models for thermal stability prediction, MLR, SVR, RF, XGBoost, and ANN models were constructed and compared using

T5%, Tmax, and char residue as representative prediction targets. Model performance was evaluated using R^2, MAE, and RMSE, as shown in Table 4 and Figure 3.

Table 4: Prediction performance of different models for thermal stability indicators.

Model	R^2 for T5%	MAE for T5%, °C	RMSE for T5%, °C	R^2 for Tmax	MAE for Tmax, °C	RMSE for Tmax, °C	R^2 for char residue
MLR	0.63	22.6	29.4	0.60	25.8	33.1	0.55
SVR	0.76	16.8	22.5	0.74	18.9	25.7	0.68
RF	0.84	12.5	17.4	0.82	14.6	20.2	0.76
XGBoost	0.89	9.7	13.8	0.87	11.8	16.5	0.82
ANN	0.81	14.1	19.6	0.79	16.4	22.8	0.73

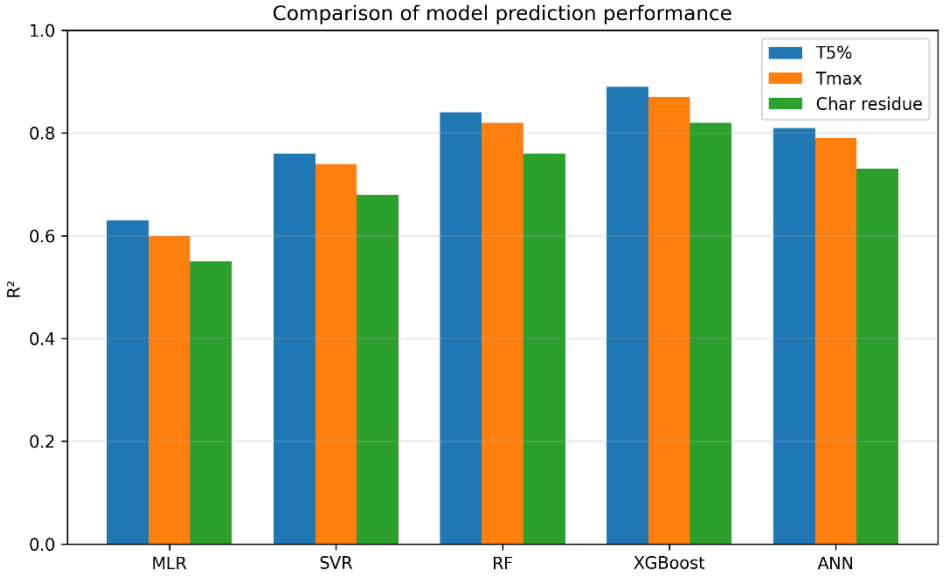

Figure 3: Comparison of model prediction performance

As shown in Table 4, MLR exhibits the lowest prediction performance, indicating that the thermal stability of crystal-doped polymer composites cannot be adequately described by simple linear relationships. SVR performs significantly better than MLR, suggesting the presence of nonlinear correlations between the input features and thermal stability indicators. RF and XGBoost both show high prediction accuracy, while XGBoost achieves the highest R^2 values and the lowest errors for T5%, Tmax, and char residue. This demonstrates its superior ability to capture complex nonlinear relationships and high-order feature interactions.

The ANN model performs better than MLR and SVR but slightly worse than RF and XGBoost. This may be associated with the limited size of the materials dataset. Although neural networks have strong nonlinear fitting capability, their performance is sensitive to sample size, network architecture, and hyperparameter selection. Therefore, XGBoost is identified as the optimal prediction model in this study and is used for subsequent prediction analysis and multifactorial effect evaluation.

4.3 Prediction performance of the optimal model

To further validate the predictive capability of the optimal model, the predicted and experimental T5% values obtained using the XGBoost model are compared in Figure 4.

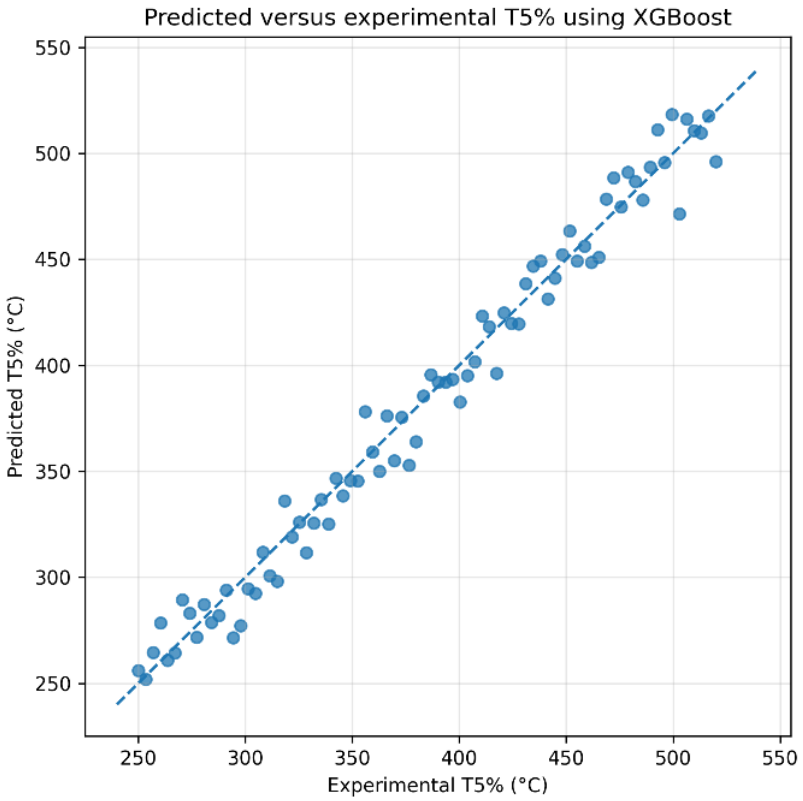

Figure 4: Predicted versus experimental T5% using XGBoost.

As shown in Figure 4, most data points are distributed close to the y = x diagonal line, indicating that the XGBoost model can accurately predict the initial thermal decomposition temperature of crystal-doped polymer composites. Good agreement is observed for samples in the medium-temperature range. However, certain deviations are found in the extremely low or high thermal stability regions. This may be attributed to the limited number of extreme samples, which restricts the model's ability to learn boundary-region behavior.

Some points deviating from the diagonal line may correspond to special material systems, such as composites with high filler loading, strongly modified interfaces, or MOF-based dopants with catalytic char-forming ability. The degradation behavior of these materials may be governed by special mechanisms that cannot be fully represented by conventional composition and processing variables alone. Therefore, further model improvement may require the incorporation of more refined molecular structure descriptors, crystal structure parameters, and interfacial interaction descriptors.

4.4 Feature importance analysis

To clarify the contribution of different factors to thermal stability prediction, feature importance analysis was conducted based on the optimal XGBoost model. Figure 5 presents the main features affecting the predicted T5% and their relative importance.

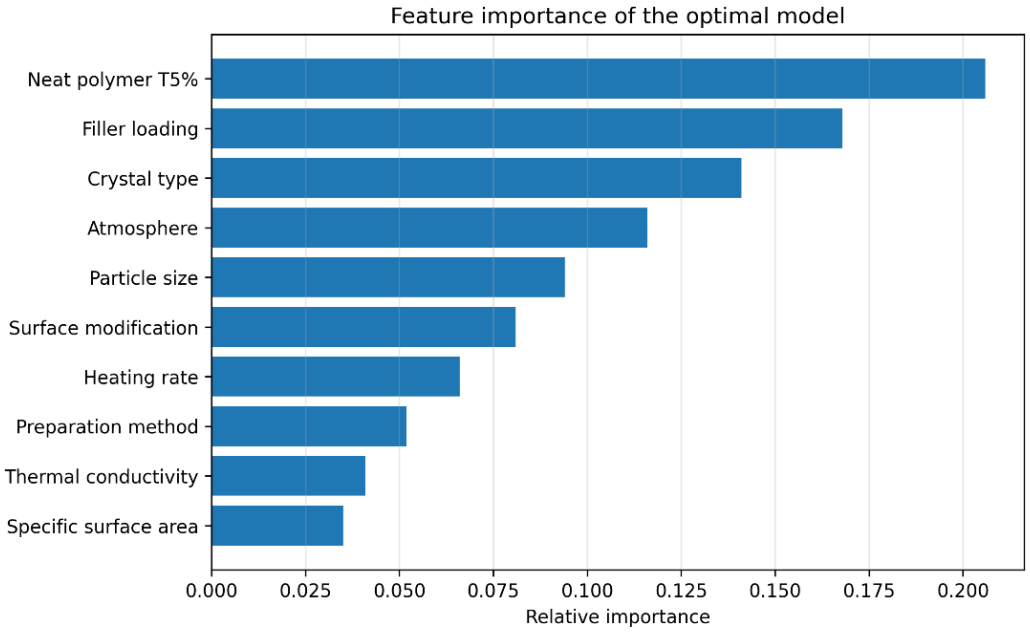

Figure 5: Feature importance of the optimal model.

As shown in Figure 5, the T5% of the neat polymer is the most important factor affecting the initial thermal stability of the composite. This result is consistent with the fundamental degradation behavior of polymer composites, as the thermal stability of the composite is first constrained by the intrinsic degradation resistance of the polymer matrix. If the matrix has poor thermal stability, the improvement induced by crystalline fillers is limited by the molecular structure and degradation mechanism of the polymer backbone.

Filler loading is the second most important factor. An appropriate amount of crystalline dopant can form an effective thermal shielding structure within the polymer matrix, delay the release of volatile degradation products, and suppress the transfer of heat and oxygen. However, excessive filler loading may lead to particle aggregation, local interfacial defects, and stress concentration, thereby weakening the improvement in thermal stability. Therefore, the effect of filler loading on thermal stability is generally nonlinear.

Crystal type, testing atmosphere, particle size, and surface modification also show considerable importance. Crystal type determines the intrinsic thermal stability, heat transfer capability, layered barrier effect, and catalytic char-forming potential of the dopant. Testing atmosphere directly affects whether the material undergoes inert thermal decomposition or thermo-oxidative degradation. Particle size and surface modification influence thermal degradation mainly by altering filler dispersion and interfacial bonding strength. These results indicate that thermal stability is not controlled by a single factor, but by the combined effects of matrix properties, dopant characteristics, and testing environment.

4.5 Multifactorial effect evaluation

To further reveal the coupling relationship among different variables, filler loading and crystal type were selected as representative factors to analyze their combined effects on predicted T5%. The results are shown in Figure 6.

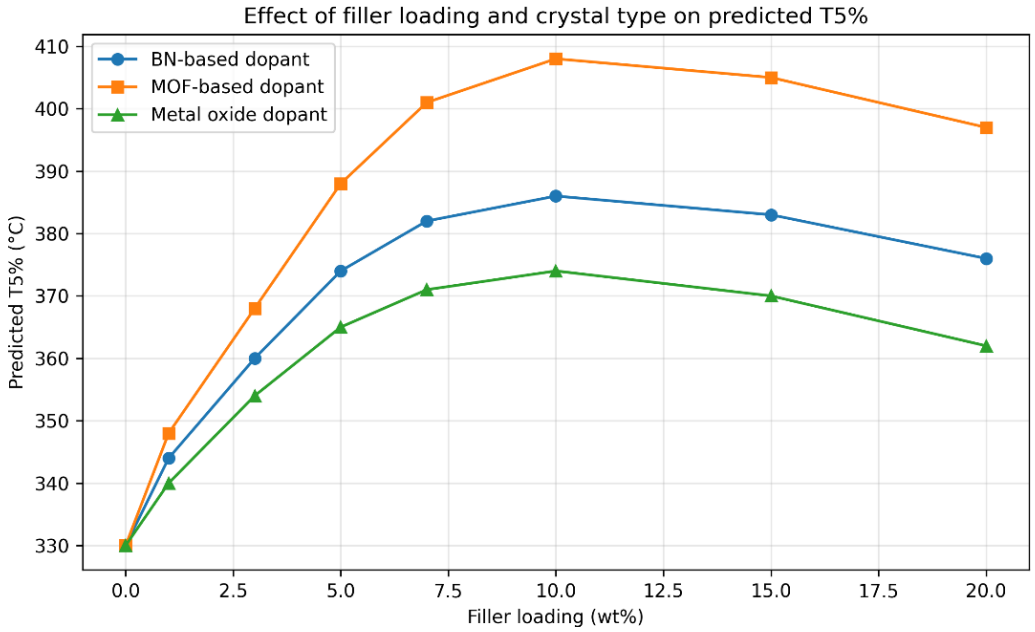

Figure 6: Effect of filler loading and crystal type on predicted T5%.

As shown in Figure 6, when filler loading increases from 0 wt% to 5–10 wt%, the predicted T5% values of different crystal-doped systems increase significantly. This indicates that crystalline fillers can effectively improve the initial thermal stability of polymer composites within the low-to-moderate loading range. This improvement may be attributed to the formation of dispersed barrier structures in the polymer matrix, which restrict polymer chain mobility and delay the diffusion of degradation products.

However, when the filler loading further increases to 15–20 wt%, the improvement in T5% gradually weakens, and even decreases in some systems. This indicates that higher crystalline dopant content does not always lead to better thermal stability. Excessive filler loading may enhance aggregation, reduce interfacial compatibility, and introduce microstructural defects, thereby damaging structural continuity and reducing the efficiency of thermal stability enhancement.

Different crystal types also show distinct effects. The MOF-based dopant system exhibits higher predicted T5% values at the same filler loading, which may be related to its porous crystalline structure, metal active sites, and potential catalytic char-forming effect. The BN-based dopant system shows a stable enhancement trend, indicating that its high thermal conductivity and layered structure contribute to heat diffusion and physical barrier formation. The metal oxide system shows a relatively moderate improvement, which may mainly originate from inorganic thermal shielding and interfacial restriction effects.

Table 5 summarizes the influence direction and possible mechanisms of the main variables affecting thermal stability.

Table 5: Effects of key factors on thermal stability of crystal-doped polymer composites.

Factor	Effect on thermal stability	Possible mechanism
Neat polymer thermal stability	Strong positive effect	Determines the baseline degradation resistance of the composite
Filler loading	Positive at low/moderate levels; weakened or negative at excessive levels	Barrier effect at proper loading; aggregation and defects at high loading
Crystal type	Strong dopant-dependent effect	Differences in thermal stability, structure, catalytic activity, and char formation
Particle size	Smaller particles generally improve thermal stability within a suitable range	Larger interfacial area and improved dispersion
Surface modification	Usually positive	Enhanced interfacial bonding and reduced filler aggregation
Testing atmosphere	Nitrogen generally gives higher apparent stability than air	Suppression of thermo-oxidative degradation
Heating rate	Higher heating rate may shift degradation temperature upward	Thermal lag and delayed mass loss response

Table 5 indicates that the thermal stability enhancement of crystal-doped polymer composites is governed by multiple factors. The intrinsic thermal stability of the matrix determines the upper limit of composite performance, the crystalline dopant determines the enhancement mechanism, filler loading and particle size affect enhancement efficiency, and testing atmosphere and heating rate alter the apparent degradation behavior. Therefore, material design should not focus on a single variable, but should consider the synergistic relationship among matrix, filler, interface, and testing conditions.

4.6 Mechanistic interpretation of thermal stability enhancement

Based on the above model prediction and multifactorial effect evaluation, the enhancement of thermal stability in crystal-doped polymer composites can be attributed to several mechanisms.

First, crystalline dopants can generate a pronounced physical barrier effect. Layered crystals, sheet-like fillers, or highly dispersed inorganic particles can prolong the diffusion pathway of degradation products and oxygen, thereby delaying the thermal degradation process of the polymer matrix. This mechanism mainly contributes to the improvement in T5% and T10%, indicating stronger resistance during the early stage of thermal decomposition.

Second, crystalline fillers can suppress polymer chain mobility through interfacial restriction. Crystalline fillers with smaller particle size or surface modification usually provide larger interfacial area and stronger interfacial bonding, which can restrict segmental motion and chain scission during heating. This contributes to the improvement in both initial degradation temperature and main decomposition temperature, and explains the relatively high contribution of particle size and surface modification in the feature importance analysis.

Third, some thermally conductive crystalline fillers can regulate heat transfer within composites. Ceramic crystals such as BN, AlN, and SiC can improve heat diffusion and reduce local heat accumulation, thereby delaying localized thermal degradation. However, when a highly conductive network is excessively formed or the fillers are poorly dispersed, heat transfer

may be accelerated locally and induce degradation. Therefore, the heat-transfer regulation effect should be considered together with filler loading and dispersion state.

Finally, metal-containing crystals, layered double hydroxides, and MOFs may promote catalytic char formation. Their metal centers, acidic sites, or decomposition products can facilitate dehydration, crosslinking, and carbonaceous layer formation, leading to higher char residue. A stable char layer can block the transfer of heat and gases and maintain structural integrity at high temperatures. Therefore, the improvement in char residue is generally associated with both catalytic charring and inorganic residue formation.

Overall, crystalline dopants improve thermal stability not through a single mechanism, but through the combined effects of physical barrier formation, interfacial restriction, heat-transfer regulation, and catalytic char formation. Machine learning models can capture these complex nonlinear relationships from multi-source data, while feature importance and multifactorial effect analyses can connect data-driven results with material stabilization mechanisms. This provides an effective basis for the rational design of thermally stable crystal-doped polymer composites.

5 Conclusions

In this study, a data-driven framework was developed for predicting the thermal stability of crystal-doped polymer composites and evaluating the effects of multiple influencing factors. By systematically organizing descriptors related to the polymer matrix, crystalline dopants, composite formulation, processing conditions, and thermal analysis parameters, predictive models were constructed for key thermal stability indicators, including T5%, T10%, Tmax, and char residue. The results indicate that, compared with conventional linear models, ensemble learning models can more effectively capture the nonlinear relationships between thermal stability and multi-source material features, with XGBoost showing higher prediction accuracy and better generalization capability.

The multifactorial effect evaluation demonstrates that the thermal stability of crystal-doped polymer composites is not governed by a single variable, but by the synergistic effects of the intrinsic thermal stability of the polymer matrix, crystalline dopant type, filler loading, particle size, surface modification, testing atmosphere, and heating rate. Among these factors, the initial thermal stability of the neat polymer determines the baseline degradation behavior of the composite, while crystalline dopants further improve thermal stability through physical barrier effects, interfacial restriction, heat-transfer regulation, and catalytic char formation. Meanwhile, filler loading exhibits a clear nonlinear effect: an appropriate amount of crystalline dopant facilitates the formation of effective barrier structures, whereas excessive loading may weaken the enhancement efficiency due to aggregation and interfacial defects.

Overall, the proposed predictive model and multifactorial evaluation strategy provide a useful data-driven approach for the thermal stability design of crystal-doped polymer composites. Future studies may further expand the materials database and incorporate more refined polymer molecular descriptors, crystal structure parameters, interfacial interaction features, and microscopic image information to improve model accuracy and cross-system generalization. In addition, integrating machine learning predictions with experimental validation, molecular simulation, and thermal degradation kinetic analysis is expected to further reveal the underlying mechanisms by which crystalline dopants regulate polymer thermal stability, thereby supporting the intelligent design of high-performance heat-resistant polymer composites.

References

[1] KARUPPUSAMY M, THIRUMALAISAMY R, PALANISAMY S, et al. A review of machine learning applications in polymer composites: advancements, challenges, and future prospects[J]. Journal of Materials Chemistry A, 2025, 13: 16290-16308.

[2] SOROUR S S, SALEH C A, SHAZLY M A. A review on machine learning implementation for predicting and optimizing the mechanical behaviour of laminated fiber-reinforced polymer composites[J]. Heliyon, 2024, 10(13): e33681.

[3] JAFARI P, SHARMA P R, GUPTA R K, et al. Machine learning for expediting next-generation of fire-retardant polymer composites[J]. Composites Communications, 2024, 45: 101806.

[4] ZHANG Z, JIAO Z, SHEN R, et al. Accelerated design of flame retardant polymeric nanocomposites via machine learning prediction[J]. ACS Applied Engineering Materials, 2023, 1(1): 140-149.

[5] XIAO J, HOBSON J, HARANCZYK M, et al. Machine learning framework to predict instantaneous heat release rate of polymer nanocomposites in cone calorimetry[J]. Polymer Degradation and Stability, 2023, 218: 110563.

[6] XIAO J, HARANCZYK M, WANG D Y. Prediction of mechanical and flame-retardant properties of MOF-loaded polymer composites[J]. Chemical Communications, 2024, 60: 13215-13218.

[7] PHAN D N, MORGAN A B, POUDEL L, et al. A machine learning platform for polymer flammability prediction[J]. Polymer Degradation and Stability, 2025, 240: 111411.

[8] HAN J, ZHANG Y, LIU X, et al. Toward accurate machine learning-driven prediction of polymer composite properties based on experimental data[J]. Materials Genome Engineering Advances, 2025: e70027.

[9] LIU Y, ZHANG H, WANG X, et al. Application of machine learning in predicting the thermal conductivity of polymer composites with single fillers[J]. Materials Today Communications, 2024, 41: 110097.

[10] LI J, ZHAO X, ZHANG Y, et al. Recent advances in metal-family flame retardants: a review[J]. RSC Advances, 2023, 13: 21185-21203.

[11] HOU B, WANG W, PAN Y T, et al. Metal-organic frameworks as promising flame retardants for polymeric materials[J]. Microstructures, 2023, 3: 2023037.

[12] SONG K, ZHANG Y, PAN Y T, et al. Metal-organic frameworks-based flame-retardant system for epoxy resin: a review and prospect[J]. Chemical Engineering Journal, 2023, 474: 145705.

[13] SUN X, PAN Y T, WANG W, et al. Surface modification of MOFs towards flame retardant polymer composites[J]. RSC Applied Interfaces, 2025, 2: 14-24.

[14] NCUBE R, MOYO L, SHUMBA M, et al. Metal-organic frameworks as smart additives

for flame-retardant polymer composites: a review[J]. Polymer Degradation and Stability, 2025, 234: 111234.

[15] AFOLABI O A, KAYODE J F, OKE E O, et al. Synergy of hybrid fillers for emerging composite and nanocomposite materials: a review[J]. Polymers, 2024, 16(13): 1907.