Polymer Degradation and Stability 250 (2026) 112207

Contents lists available at ScienceDirect

Polymer Degradation and Stability

journal homepage: www.journals.elsevier.com/polymer-degradation-and-stability

Data-driven prediction and non-compensatory assessment of flame retardancy in phosphorus-containing flame-retardant epoxy resin

Qiong Tan, De-Yi Wang *

IMDEA Materials Institute, C/Eric Kandel, 2 Getafe, 28906, Madrid, Spain

ARTICLE INFO

Keywords:
Flame retardancy
Prediction
Machine learning
Non-compensatory assessment
Epoxy resin

ABSTRACT

Flame retardancy is essential in epoxy resin (EP) applications. Phosphorus-containing flame retardants (P-FRs) have been recognized as one of the most promising halogen-free additives for improving the fire resistance of EP. However, rapid and low-cost screening of high-efficiency EP/P-FR composites and systematic assessment of their flame-retardant performance remain challenging. In this work, a data-driven strategy is developed for the prediction and comprehensive assessment of flame retardancy in EP/P-FR systems. Polymer structural descriptors, formulation, and condition-related features are jointly considered and integrated through feature fusion. Key flame retardancy evaluation metrics, including the UL-94 vertical burning rating and limiting oxygen index (LOI), are predicted using machine-learning models. The proposed models achieve a high accuracy of 0.81 for UL-94 classification and an R^2 of 0.79 (MAE=1.21%) for LOI prediction on the test set. Feature importance analysis quantifies the contributions of structural and formulation variables, revealing key factors governing flame-retardant performance. Furthermore, a non-compensatory evaluation framework is proposed for polymer flame retardancy assessment. The predicted UL-94 vertical burning rating and LOI values are incorporated as core evaluation inputs within a unified decision framework. Based on this framework, a six-state performance space is constructed, and four flame-retardant semantic levels, denoted as L-1 to L-4, representing excellent, good, moderate, and poor flame retardancy performance, respectively, with suffixes a and b indicating UL-94-prioritized and LOI-prioritized decision strategies. This work not only provides a practical strategy for the discovery and design of high-performance EP/P-FR composites, circumventing extensive trial-and-error, but also enables a more scientific and comprehensive non-compensatory evaluation of flame retardancy in such systems.

1. Introduction

Epoxy resin (EP), one of the most important polymers, has been widely employed in areas like construction, automotive, electronics, and aerospace industries, etc. [1]. In Europe, EP market has exceeded 410.60 kg tonnes per year. However, their inherent flammability can result in rapid burning under fire conditions, accompanied by the release of large amounts of heat and toxic smoke, which limits its further use in applications with high safety requirements [2,3]. Incorporating flame retardants (FRs) into EP was approved as an efficient approach to improve their fire safety., [4–6]. Phosphorus-based flame retardants (P-FRs) are among the most employed halogen-free systems for epoxy resins [7,8]. Traditionally, the development and optimization of efficient FRs is a labor-intensive and costly process, driven by cycles of informed design, synthesis, characterization, and property testing, which are time-consuming, costly, and highly sensitive to operating conditions [9–11]. Additionally, the complexity and diversity of polymer material systems and formulations makes it difficult to establish reliable structure-performance relationships through conventional experimental exploration alone [12]. The ability to predict the flame-retardant performance of EP from molecular structure and formulation parameters has been desired for decades to accelerate materials development.

In recent years, data-driven methods, particularly machine learning (ML) techniques, have shown considerable potential in predicting flame retardancy of polymers [13–18]. ML constructs end-to-end predictive models that can capture the complex relationship of flame-retardant performance prediction [19,20]. Based on this, the researchers have developed ML models to predict the common flammability properties, including LOI, UL-94 vertical burning rating, and peak heat release rate (PHRR) [21–27]. Xiao et al. reported a ML framework to predict heat release rate curve, simulating the whole measurement in cone calorimeter test (CCT) [21]. Wang et al. proposed an interpretable flame

* Corresponding author.
E-mail address: deyi.wang@imdea.org (D.-Y. Wang).

https://doi.org/10.1016/j.polymdegradstab.2026.112207
Received 10 March 2026; Received in revised form 20 April 2026; Accepted 8 May 2026
Available online 9 May 2026

retardancy quantitative model that can achieve highly accurate flame retardancy prediction [28]. Li et al. constructed ML models for LOI, PHRR, and UL-94 vertical burning rating for epoxy resin [26]. However, most existing predictive models are developed for specific flame-retardant system and thus not directly applicable to other FR-based EP composites. In addition, most studies focused on predicting individual flame-retardant metrics or presenting multiple predicted indicators in parallel, while only limited attention has been paid to how such predictions can be interpreted and translated into a unified and comparable evaluation of overall flame-retardant performance. Consequently, there is a strong need for new approaches that can reduce experimental effort while enabling systematic, interpretable, and integrated evaluation of polymer flame-retardant performance.

In this work, we propose a data-driven framework for the prediction and comprehensive evaluation of flame-retardant performance of EP/P-FR composites. ML models that integrate structural and formulation-related features were developed to predict key evaluation indicators, including UL-94 vertical burning ratings and LOI values. These predicted results are incorporated into a non-compensatory evaluation framework, in which LOI values and UL-94 vertical burning rating are treated as two core evaluation dimensions. Within this framework, a flame-retardant state space consisting of six performance combinations is constructed, and four flame retardancy levels with clear physical and engineering meaning are defined.

2. Methodology

2.1. Overview of the data-driven flame retardancy evaluation framework

In this work, the EP/P-FRs system is adopted as a representative case for developing and validating the proposed data-driven framework, as it provides sufficient experimental diversity and coverage in the available literature.

The data-driven flame retardancy evaluation framework was proposed to achieve the prediction of key flammability properties, such as LOI and UL-94 vertical burning rating and enable a comprehensive evaluation of flame-retardant performance. The overall workflow of this study is illustrated in Fig. 1. Experimental data and published literature on EP composites containing P-FRs since 2003 were first collected to construct the dataset. The dataset was then dealt with feature engineering, including data cleaning, data normalization, one-hot encoding, and multi-source feature extraction. Key features related to P-FR molecular structure, formulation parameters, and other condition-related factors were subsequently fused to generate a unified feature representation. A trained ML model was then constructed to predict key flammability indicators, including the LOI and UL-94 vertical burning rating. Based on the predicted LOI values and UL-94 vertical burning rating ratings, a non-compensatory evaluation strategy was further applied to integrate individual flammability metrics into a comprehensive assessment of flame-retardant performance.

2.2. Feature representation and fusion

Feature engineering is a critical step in establishing accurate and reliable ML models, as it enables flame-retardant polymer composites to be comprehensively represented by integrating structural, formulation, and other condition-related features into a unified feature space. As illustrated in Fig. 2, a multi-source feature engineering strategy was adopted to comprehensively capture the structural characteristics of P-FRs as well as formulation and condition-related information.

For structural feature extraction, the molecular structures of P-FRs were first encoded using SMILES representations. Then, molecular descriptors were calculated using the RDKit toolkit to characterize global molecular properties. In parallel, atomic-level features were extracted to capture local chemical environments. To transform these atom-level features into molecule-level descriptors, atom-level descriptors were aggregated using the mean, standard deviation, maximum, and minimum values. In addition, count-based features, including atom-type distributions and hybridization-state distributions, were calculated. The aggregated atom-derived descriptors were subsequently merged with molecular descriptors using a unique molecular identifier, without applying additional weighting at this stage.

In addition to structural descriptors, condition-related features, such as matrices, synergist, flame retardants loading, were incorporated to reflect practical composite properties. These features were normalized and encoded to ensure compatibility with structure-based descriptors.

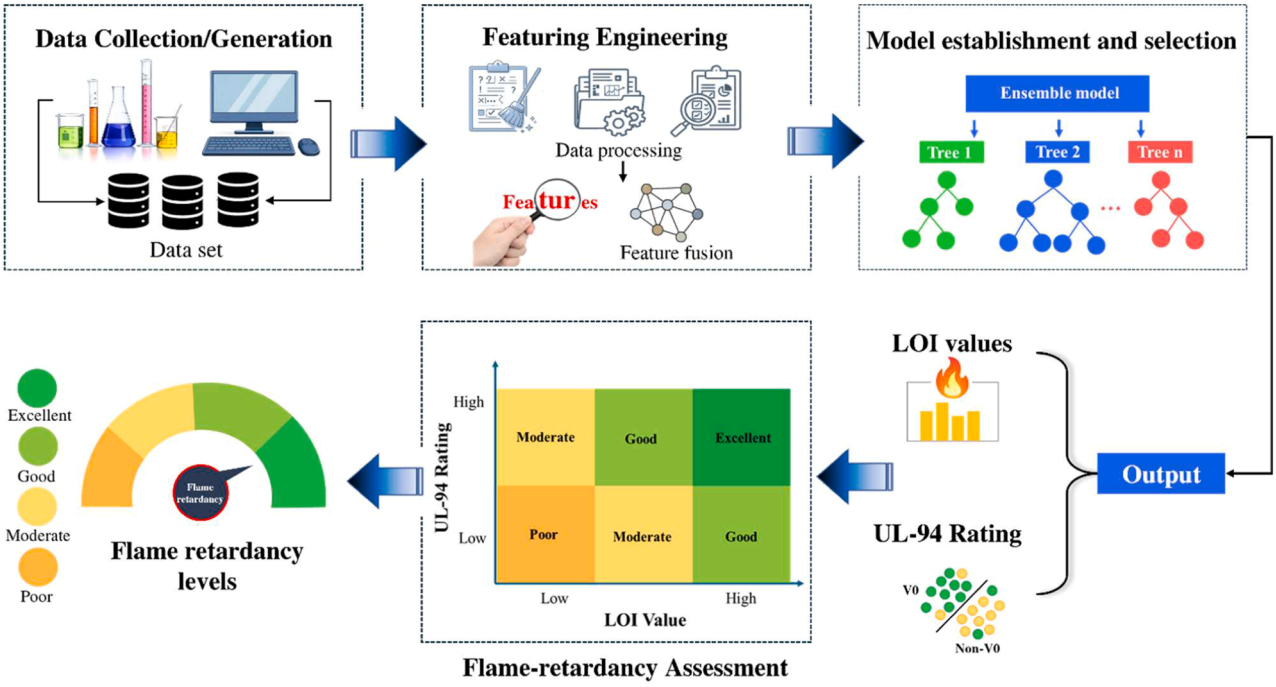

Fig. 1. General schematic diagram of flame-retardant performance prediction and assessment.

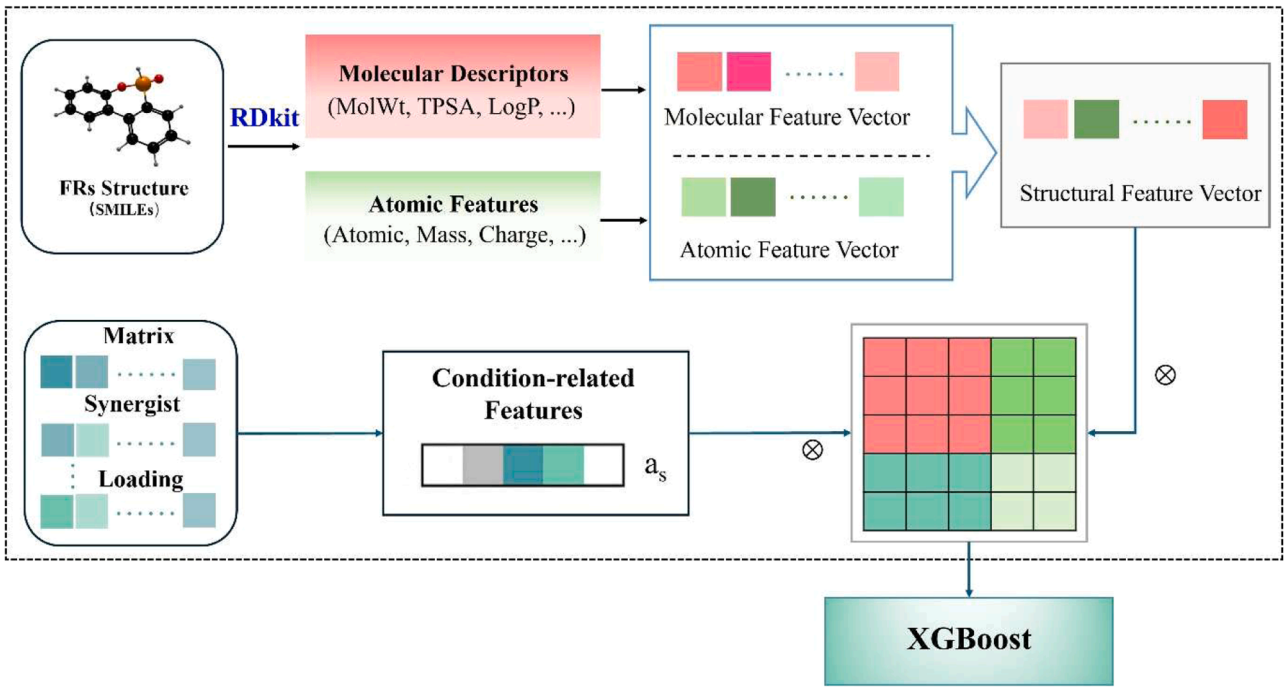

Fig. 2. Schematic diagram of feature engineering.

Finally, a feature-level fusion strategy based on aggregation and concatenation was applied to integrate structural descriptors and condition-related descriptors into a unified feature representation. Specifically, structural descriptors and condition-related features were combined by matching samples using their corresponding identifiers (SMILES), with all feature groups treated equally, to construct the final feature matrix. The resulting fused feature matrix was used as input for the ML model, enabling the learning of complex, nonlinear relationships between multi-scale features and flame-retardant performance.

2.3. Machine learning models for UL-94 vertical burning rating and LOI prediction

Recent advances in ML techniques have given rise to the possibility of identifying underlying connections or patterns within complex chemical systems in modern chemistry research [29]. ML techniques can be mainly categorized into three types: supervised learning, unsupervised learning, and semi-supervised learning. Most supervised-learning problems can be categorized as either "classification" problems or "regression" problems. The problem of predicting flame-retardant performance of epoxy resin can be potentially solved by combining

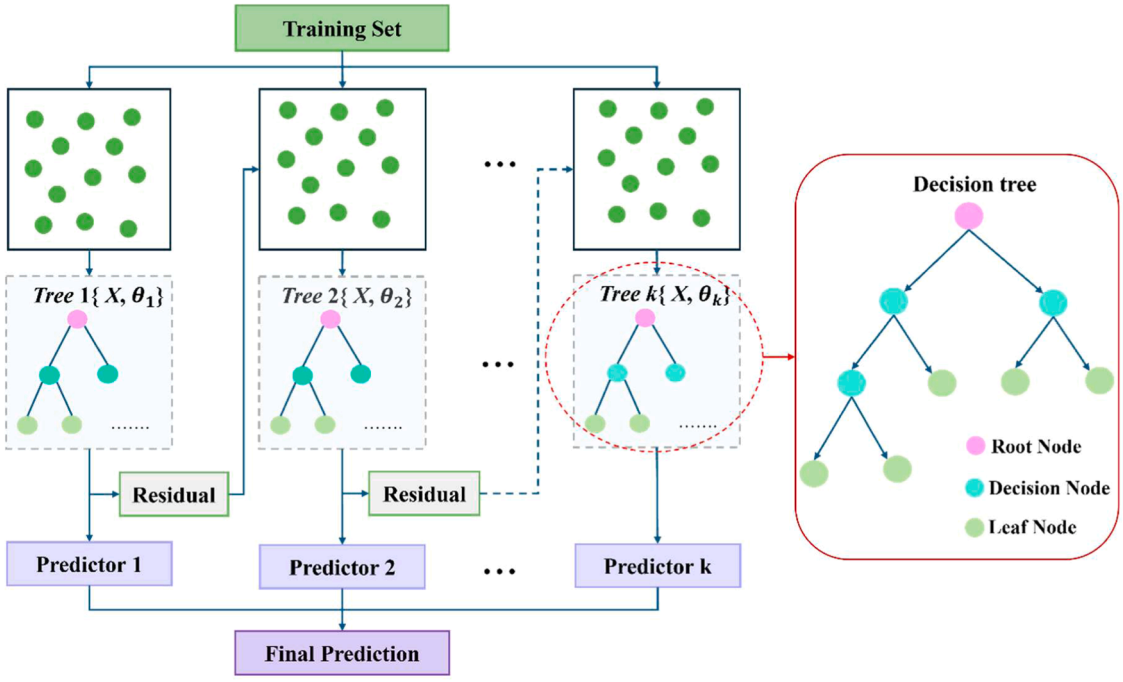

Fig. 3. Schematic diagram of XGBoost algorithm.

high-quality datasets with appropriate ML algorithms. For key flammability properties of EP, the prediction of LOI values corresponds to a supervised regression task, while UL-94 vertical burning ratings represent a supervised classification problem. Based on the analysis and comparison of various machine-learning models [30], the XGBoost algorithm was adopted for both regression and classification predictions in this study.

The XGBoost algorithm, proposed in 2016 [31], is a decision tree–based ensemble machine-learning algorithm developed from gradient boosting decision trees (GBDT), as shown in Fig. 3. It can handle both regression and classification tasks by sequentially constructing decision trees to minimize prediction errors.

Compared with conventional ML algorithms, XGBoost incorporates regularization terms into its objective function to control model complexity and reduce overfitting. In addition, XGBoost supports efficient parallel computation, robust handling of missing values, and flexible objective function design, making it well suited for modeling complex, nonlinear relationships in heterogeneous datasets. The objective function of the XGBoost model is defined as follows [32]:

$$Obj(\theta) = \sum_{i=1}^{n} l(y_i, \widehat{y}_i) + \sum_{k=1}^{k} \Omega(f_k) \tag{1}$$

Where $l(y_i, \widehat{y}_i)$ is used to represent the loss error, Ω denotes the regularization term which is introduced into the objective function to control the complexity of the model and prevent the model from overfitting, f_k is the model of the *k-th* tree.

In the above expression, the specific expression of the regularization term is:

$$\Omega(f_k) = \gamma T + \frac{1}{2}\lambda \sum_{j=1}^{T} {\omega_j}^2 \tag{2}$$

where γ is the complexity cost by additional leaf; λ is the regularization parameter; T is the number of leaf nodes in the tree; ω_j is the weight of the j-th leaf node. XGBoost has an ensemble of K regression trees whose individual prediction outcomes are denoted as $f_k(x_i)$. The final prediction is obtained as the sum of outputs of all the trees.

$$\widehat{y}_i = \sum_{k=1}^{K} f_k(x_i) \tag{3}$$

To smoothly minimize the objective function, the following transformations are needed:

$$\widehat{y}_i^{(t)} = \widehat{y}_i^{(t-1)} + f_t(x_i) \tag{4}$$

Accordingly, the objective function of the Kth tree can be represented as:

$$Obj(f_t) = \sum_{i=1}^{n} l\left(y_i, \widehat{y}_i^{(t-1)} + f_t(x_i)\right) + \Omega(f_t) \tag{5}$$

The objective function can be expanded by Taylor's second order expansion:

$$Obj(f_t) = \sum_{i=1}^{n} \left[l\left(y_i, \widehat{y}_i^{(t-1)}\right) + g_i f_t(x_i) + \frac{1}{2} h_i {f_t}^2(x_i) \right] + \Omega(f_t) \tag{6}$$

where $g_i = \partial_{\widehat{y}_i^{(t-1)}} l\left(y_i, \widehat{y}_i^{(t-1)}\right)$ and $h_i = \partial^2_{\widehat{y}_i^{(t-1)}} l\left(y_i, \widehat{y}_i^{(t-1)}\right)$.

The objective function can be further simplified:

$$Obj^{(t)} = \sum_{j=1}^{T} \left[G_j \omega_j + \frac{1}{2}\left(H_j + \lambda\right) {\omega_j}^2 \right] + \gamma T \tag{7}$$

Here, $G_j = \sum_{i \in I_j} g_i$ and $H_j = \sum_{i \in I_j} h_i$.

Taking the derivative of ω_j in the objective function, we obtain:

$$\omega_j = -\frac{G_j}{H_j + \lambda} \tag{8}$$

At this point, the objective function can be computed by:

$$Obj(\theta) = -\frac{1}{2} \sum_{j=1}^{T} \frac{{G_j}^2}{H_j + \lambda} + \gamma T \tag{9}$$

The hyperparameters of the XGBoost model were optimized using Optuna with a Tree-structured Parzen Estimator (TPE) sampler. The search space and corresponding optimal values for each hyperparameter for the UL-94 classification and LOI prediction models are summarized in Table S2 and Table S3, respectively. The optimization covered key parameters, including learning rate, maximum depth, subsample ratio, colsample_bytree and regularization coefficients (λ and α).

3. Results and discussion

3.1. Dataset representation

In this study, a total of 510 EP composite samples incorporating P-FRs were used to train and test the model. For each sample, both UL-94 rating and LOI value are included. Based on available data, a wide range of factors can influence LOI and UL-94 vertical burning ratings such as EP matrix, synergist and P-FRs structure are collected. These factors can generally be categorized into three main groups: (a) formulation descriptors, (b) P-FRs composition and loading, and (c) P-FRs structural descriptors, as shown in Table 1.

Before feeding the data into an ML model, data preprocessing and feature engineering are necessary to make all the values of features understood easily by its algorithm. Duplicate entries, inconsistent records, and samples with missing critical information were first removed. Discrete features, such as matrix type, curing agent, and phosphorus location were converted into numerical vectors using one-hot encoding, while numerical descriptors such as additive loading were normalized to eliminate scale effects.

The molecular structures of P-FRs were converted into the SMILES string. These SMILES representations were then transformed into molecular descriptors (e.g., MaxEStateIndex, NumHAcceptor, the numbers of Ring etc.), and atom descriptors (e.g., the numbers of C, H, N, etc.) using RDKit package (version 2025. 09.5) in Python (version 3.10.19). Starting with 108 features, we preprocessed the data by removing any perfectly correlated features and any constant variables, resulting in 94 retained descriptors. From this pool, 92 features were used for both the UL-94 and LOI models. Although the majority of features are shared, the two models differ by two descriptors, leading to slightly different feature sets. The details of these features are provided in Table S1. Subsequently, the molecular descriptors, atomic descriptors, along with the formulation descriptors, and P-FRs loading descriptors were aligned by sample ID and merged column-wise to construct a unified input matrix, which served as the comprehensive feature database for subsequent ML model training and evaluation.

3.2. Prediction performance of UL-94 vertical burning rating and LOI

3.2.1. UL-94 vertical burning rating prediction

The distribution of UL-94 vertical burning ratings in the EP/P-FR dataset was given in Fig. 4(a). The numbers of NR, V2, V1, and V0

Table 1
The summary of dataset size and descriptors.

Number of samples	Descriptor type	Number of Descriptors
510	Formulation	2
	P-FRs composition and loading	11
	P-FRs structure	81

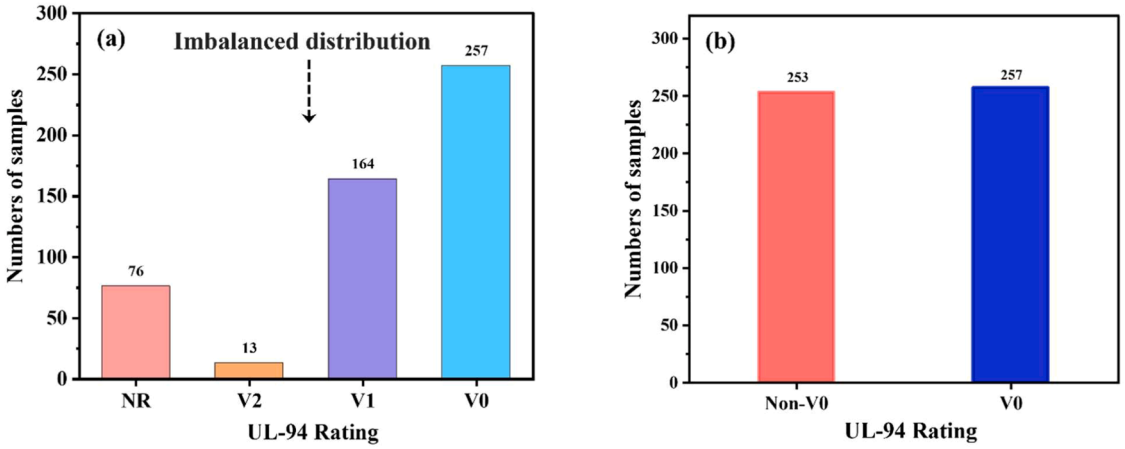

Fig. 4. Distribution of UL-94 vertical burning ratings in collected dataset: (a) 4 categories and (b) 2 categories.

were 76, 13, 164, and 257, respectively, with an uneven overall distribution. Notably, the number of NR and V2 categories are significantly underrepresented, which can be attributed to a reporting bias in the literature, where only the good performance of the flame retardants is more likely to be published. This imbalance can cause problems with various data-driven activities, such as categorization and prediction. To address the class imbalance issue, the original multi-class UL-94 vertical burning classification problem was simplified into a binary classification task. In practical flame retardancy evaluation, whether a material achieves the V0 rating is commonly regarded as a critical criterion. Therefore, the V0/non-V0 distinction is more relevant for practical decision-making. Accordingly, samples rated as V0 were assigned to the positive class, whereas all other categories (V1, V2, and NR) were grouped into the negative class. Fig. 4(b) illustrates the results and shows a more equitable representation of positive and negative instances, potentially leading to improved outcomes in predicting UL-94 vertical burning rating.

To assess the predictive performance of the UL-94 vertical burning classification model, four performance metrics, including Accuracy, F1-Score, Precision, and Recall were employed for the binary classification task [33]. Accuracy is the ratio of correctly predicted samples to the total samples, which can show the overall performance of the predictive model. Precision is defined as the ratio of positive predictions that are correct. The recall is the ratio of actual positive results that are predicted correctly. The F1-score is the harmonic mean of precision and recall. Their calculation formulas are given below:

$$Accruacy = \frac{TP + TN}{TP + FP + TN + FN} \tag{10}$$

$$Precision = \frac{TP}{TP + FP} \tag{11}$$

$$Recall = \frac{TP}{TP + FN} \tag{12}$$

$$F1 - score = \frac{2 \times Precision \times Recall}{Precision + Recall} \tag{13}$$

where TP is true positive, the number of V0 that were correctly predicted as positive. TN is truly negative, the number of non-V0 that were correctly predicted as negative. FP is false positive, the number of non-V0 that were incorrectly predicted as positive. FN is false negative, the number of V0 that were incorrectly predicted as negative.

The dataset was randomly partitioned into 80% training and 20% testing. The performance metrics of the UL-94 vertical burning classification model are summarized in Table 2. Notably, the UL-94 vertical burning model's accuracy, precision, recall, and F1 score were 0.814, 0.833, 0.784 and 0.808 respectively. Fivefold cross-validation was further performed to evaluate the robustness of the model. The cross-validation accuracy was 0.71± 0.05, which is lower than the single hold-out test result. This difference can be attributed to the inherent complexity of the UL-94 classification task and the limited size of the dataset. Future work will focus on expanding the dataset and incorporating more physically meaningful features to further improve the model's generalization ability.

Table 2
The accuracy, precision, recall and F1 score of the UL-94 vertical burning rating predictive model.

Accuracy of the training set	Accuracy of the test set	Precision	Recall	F1 score
0.988	0.814	0.833	0.784	0.808

The confusion matrix serves as a formalized method for evaluating ML models, summarizing the numbers of correct and incorrect predictions. Fig. 5 presents the confusion matrices for the XGBoost-based UL-94 vertical burning binary classification model. As shown in Fig. 5, the model could recall 205 samples from 209 V0 samples and achieved 198 correct predictions among the 199 samples predicted as non-V0 in the training set. For the testing set, 43 negative and 40 positive samples were correctly predicted, while 8 false positives and 11 false negatives were observed. Misclassifications occur in both V0 and non-V0 classes, indicating that the model does not exhibit a strong bias toward either class.

To evaluate the prediction accuracy of the XGBoost algorithm used for the proposed model, four commonly used machine learning algorithms, Light Gradient Boosting Machine (LightGBM), Multilayer Perceptron (MLP), Random Forest (RF), and Support Vector Machine (SVM), were selected for comparison purposes. Table 3 presents the experimental results obtained from the different algorithms using the same datasets. Among them, XGBoost achieved the best overall performance, with the highest accuracy (0.81) and F1-score (0.81), indicating a favorable balance between precision and recall. LightGBM showed comparable performance, with both accuracy and F1-score close to 0.78. MLP model shows a high recall of 0.84 but a lower precision of 0.70, suggesting a tendency to overpredict the positive class, whereas RF demonstrates the opposite trend, achieving a precision of 0.84 but a lower recall of 0.73. The SVM model yields consistently lower scores across all metrics (approximately 0.63). The superior performance of XGBoost and LightGBM suggests that gradient boosting methods are more effective in capturing the nonlinear decision patterns underlying UL-94 vertical burning rating classification.

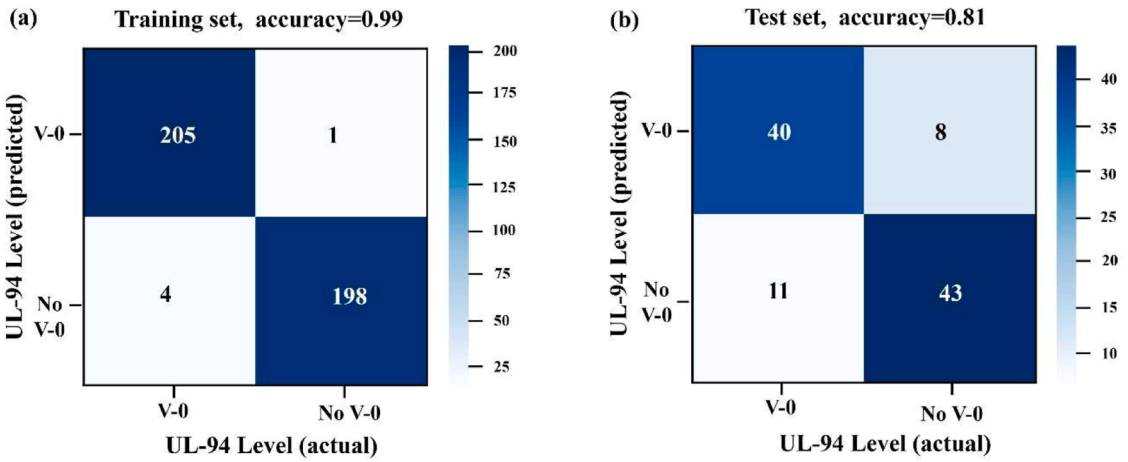

Fig. 5. Confusion matrix related to XGBoost classification performance on the (a) train and (b) test set.

Table 3
The accuracy, precision, recall and F1 score of the UL-94 vertical burning rating predictive model.

Model	Accuracy of the test set	Precision	Recall	F1 score
XGBoost	0.81	0.83	0.78	0.81
LightGBM	0.80	0.80	0.80	0.80
MLP	0.77	0.7	0.84	0.79
RF	0.79	0.84	0.73	0.78
SVM	0.63	0.63	0.63	0.63

3.2.2. LOI prediction

In this work, the LOI prediction model was developed based on the same set of 510 EP composite samples. The addition levels of these flame retardants range from 0.09% to 32%. The distribution of LOI values in EP matrix dataset was given in Fig. 6(a). As can be observed, the LOI values were not normally distributed or uniformly distributed, with most samples concentrated in a narrow range (27–38%). Such graphical plots provide important context for interpreting subsequent regression results, particularly when non-normality or data concentration may influence model behavior.

Fig. 6(b) shows the LOI distributions corresponding to the different UL-94 vertical burning rating with the numbers of NR, V2, V1, and V0 being 76, 13, 164, and 257, respectively. The median LOI for the V0 rating is approximately 33, and when the LOI of the flame-retardant EP reaches a certain value of 30.8, the data corresponding to the 75th NR indicate that the possibility of the flame-retardant EP being NR is lower. However, a noticeable overlap is observed around the intermediate LOI range, where materials with similar LOI values may fall into different UL-94 vertical burning categories. Further, the Spearman correlation coefficient between LOI and UL-94 vertical burning ratings was calculated, yielding a value of $\rho \approx 0.348$. This indicates that there is a weak monotonic relationship between LOI value and UL-94 vertical burning rating. These results demonstrate that LOI and UL-94 vertical burning rating characterize different aspects of flame-retardant behavior and cannot be used interchangeably.

In this study, the performance of LOI value prediction models was evaluated by three indices: mean absolute error (MAE), root mean square error (RMSE), and coefficient of determination (R^2). Their calculation formulas are given below:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i^{act} - y_i^{pre}| \tag{14}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i^{act} - y_i^{pre})^2} \tag{15}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i^{act} - y_i^{pre})^2}{\sum_{i=1}^{n}(y_i^{act} - \overline{y_i})^2} \tag{16}$$

where, n is the number of observations, y_i^{pre} is the predicted value, and y_i^{act} is the actual value, $\overline{y_i}$ is the average of observed values. The lower values of RMSE and MAE indicate better predictive accuracy of

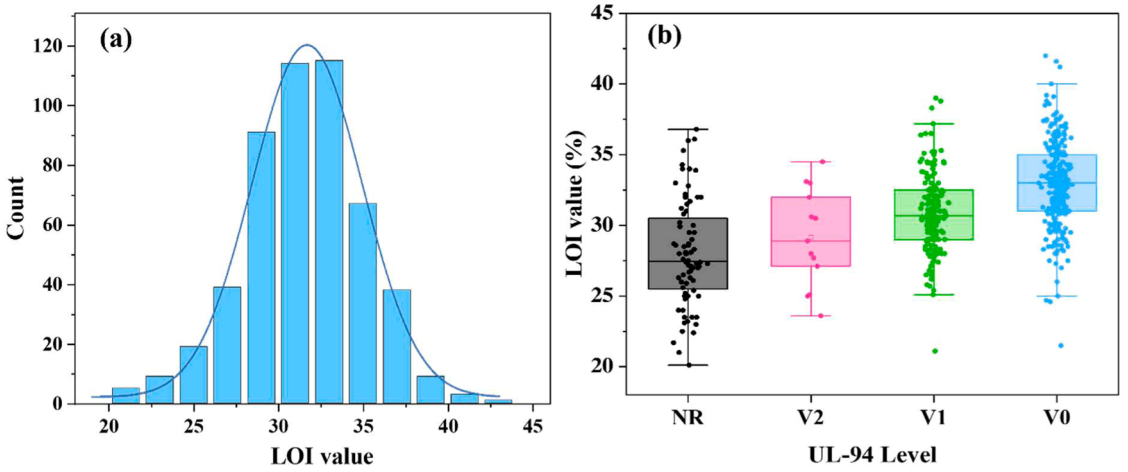

Fig. 6. (a) Distribution of LOI values in collected dataset; (b) LOI values distribution corresponding to different UL-94 vertical burning ratings.

the regression models. R^2 is a coefficient ranging from 0 to 1, and a larger R^2 indicates a more effective prediction model.

The evaluation results of XGBoost algorithms were shown in Fig. 7. In Fig. 7(a), the predicted LOI values are plotted against the actual LOI values. Generally, the predicted values tend to be close to their corresponding actual values, indicating that the XGBoost model can effectively learn the relationship between LOI and the molecular structure of P-FRs, formulation parameters, and other condition-related descriptors. The R^2 scores for the training and test sets are 0.995 and 0.731, respectively. The corresponding RMSE and MAE values are 2.029 and 1.383, respectively. After five-fold cross-validation by default, the R^2 score for the test set is 0.67 ± 0.08. The gap between training and test performance is consistent with the observations in UL-94 classification and can be attributed to the limited dataset size and the uneven distribution of sample. As shown in Fig. 7(b), larger prediction errors are observed in both the low-LOI region (approximately 20–25) and the high-LOI region (approximately 35–40). In the low-LOI region, the predicted values tend to be higher than the experimental values, whereas in the high-LOI region, the predicted values tend to be lower. This systematic deviation is likely due to the limited number of samples in these extreme ranges (19 low-LOI and 13 high-LOI samples), which restricts the model's ability to learn accurate boundary behaviors.

To mitigate the systematic underestimation and overestimation observed at the low- and high- LOI regions, a residual-based piecewise bias correction strategy was introduced as a post-processing step to refine the XGBoost predictions. According to the corrected LOI is given by:

$$LOI = LOI_{pre} + \Delta_b \tag{17}$$

Where, Δ_b is the correction term, which is estimated only from the training data. The correction term Δ_b can be carried out according to Eqs. (18).

$$\Delta_b = \frac{1}{n_b}\sum_{i\in b}\delta_i \tag{18}$$

For each bin b, the prediction residual δ_i is computed:

$$\delta_i = LOI_{i,act} - LOI_{i,pre} \tag{19}$$

In this study, the training samples are grouped into 4 bins: [20,25), [25,30), [30,35), [35,40).

Fig. 8 presents the performance of XGBoost model after applying the residual-based piecewise bias correction. The higher coincidence between the fitting curve and the 'X = Y' dashed line indicates a better prediction accuracy. As shown in Fig. 8(a), the corrected XGBoost predictions are closer to the actual LOI values than those of the original model. Notably, the R^2 reached 0.79 when introducing residual-based piecewise bias experimental values than correction. Furthermore, RMSE and MAE values decreased to 1.64 and 1.21 respectively. Furthermore, the prediction errors are substantially reduced in both the low-LOI region and the high-LOI region as shown in Fig. 8(b).

Based on the same input and output, XGBoost, k-Nearest Neighbors (KNN), RF, MLP, LightGBM and Support Vector Regression (SVR) are used to construct models for LOI values prediction. Fig. 9 shows the R^2 and MAE for the seven compared algorithms. As shown in Fig. 9(a), the proposed method, XGBoost, demonstrated the highest R^2 among all compared methods. This is followed by RF, LightGBM KNN, SVR and MLP, with values of 0.78, 0.75, 0.71, 0.70 and 0.61 respectively. This indicates that ensemble tree-based models exhibit the strongest explanatory capability for the LOI prediction task. Consistent trends are observed in Fig. 9(b), where XGBoost also achieves the lowest MAE (1.21), followed by RF (1.26) and LightGBM (1.28). In contrast, KNN, MLP and SVR exhibit higher MAE values, and they are 1.51, 1.44 and 1.44, respectively. Overall, the results demonstrate that ensemble learning models, particularly gradient boosting models, are more suitable for LOI values prediction in complex flame-retardant EP systems.

3.3. Feature importance analysis

3.3.1. UL-94 vertical burning rating prediction

The SHapley Additive exPlanations (SHAP) method is a powerful visualization tool to interpret the contribution of each feature in a machine learning model [34]. In addition to quantifying feature importance, SHAP can reveal non-monotonic relationships between input features and target properties, providing insights into which features are most relevant and actively contribute to the predicted UL-94 vertical burning rating.

Fig. 10 illustrates the SHAP feature importance of the XGBoost model for UL-94 vertical burning classification based on the test set. For clarity, only the features with importance on the top 10 are listed. Features with higher SHAP values are indicative of a greater influence on the output. As shown in Fig. 10, the flame-retardant loading (P-FRs loading) emerges as the most significant feature, exerting its dominant role in improving flame-retardant performance. The second influential factor is phosphorus content in the flame-retardant EP (P-). Higher values of P- consistently contribute positively to the prediction of V0. The positive effect of P-, which is consistent with the well-established role of phosphorus in promoting char formation and enhancing condensed-phase flame-retardant mechanisms. In contrast, secondary features such as Synergist and O- content demonstrate moderate but non-negligible effect. It is important to note that the synergist descriptor is encoded as a discrete categorical variable, where different synergist types are represented by distinct integer labels rather than continuous physicochemical quantities. Therefore, the observed SHAP importance of the synergist

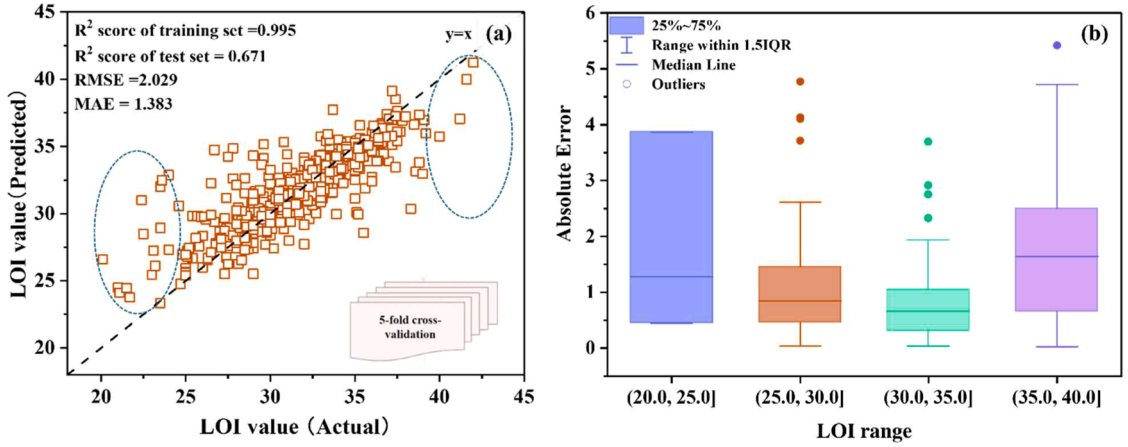

Fig. 7. (a) Scatter plots between the predicted and actual LOI; (b) Absolute error of the test set for different LOI range.

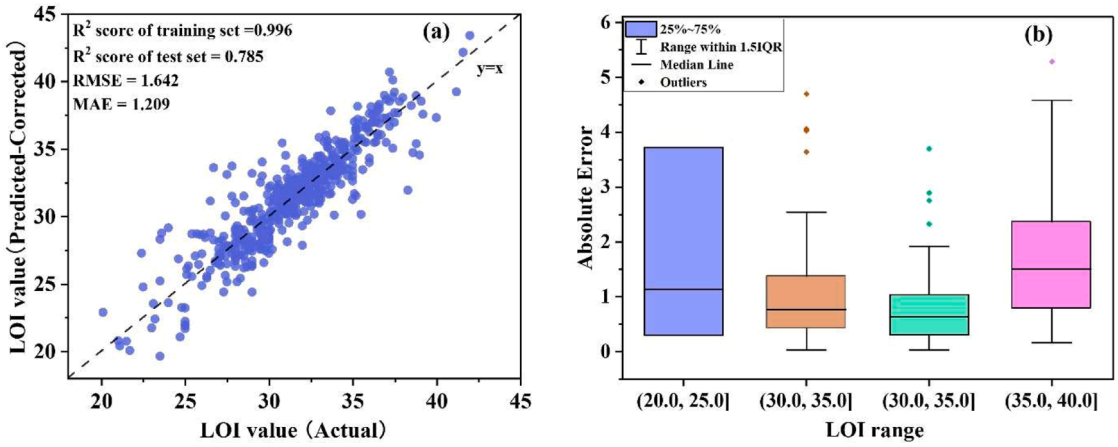

Fig. 8. Summary results of LOI prediction after residual-based piecewise bias correction: (a) Scatter plots between the predicted and actual LOI; (b) Absolute error of the test set for different LOI range.

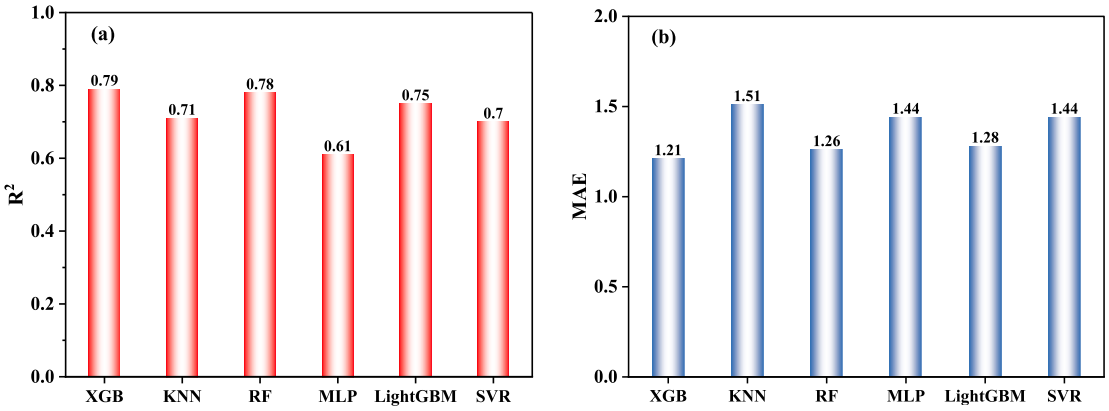

Fig. 9. Comparison of the six algorithms for LOI predictive model (a) R^2 and (b) MAE.

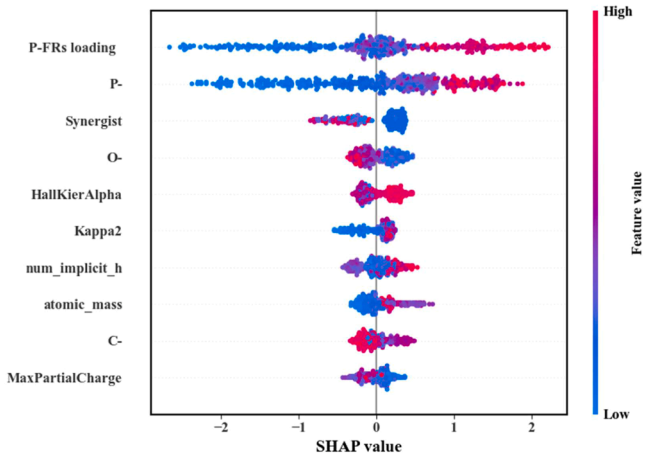

Fig. 10. SHAP feature importance of the UL-94 vertical burning rating predictive model.

feature primarily reflects the model's ability to distinguish between different synergist categories, rather than implying a monotonic or quantitative relationship between the numerical label and UL-94 performance. For the O- content, no clear monotonic relationship with UL-94 performance is observed. Its SHAP values are distributed on both positive and negative sides with a narrow spread, indicating a context-dependent effect. This suggests that oxygen-containing functionalities may play dual roles, either enhancing char formation or promoting the generation of volatile combustible species, depending on the chemical environment. For further insight into these relationships, SHAP dependence plots for the top five most important features are provided in Figures S1. Structural descriptors, such as HallKierAlpha, Kappa2, and num_implicit_h, collectively capture information related to molecular topology, geometry, and bonding environments. Among them, HallKierAlpha shows a slight shift toward positive SHAP contributions at higher values, whereas Kappa2 and num_implicit_h remain centered around zero with mixed contributions, indicating the absence of a clear monotonic relationship with flame-retardant performance. Overall, the SHAP analysis demonstrates that UL-94 vertical burning performance is governed by a synergistic interplay between flame-retardant loading, elemental composition, electronic structure, and global molecular topology, rather than by a single dominant factor.

3.3.2. LOI prediction

Fig. 11 demonstrates the prominent features influencing the XGBoost model predictions for LOI based on the test samples. As shown in Fig. 11, P-, P-FRs loading, and synergist type, exhibit the highest contributions to LOI prediction, highlighting their dominant roles in governing flame-retardant behavior. Specifically, higher P- and P-FRs loading are generally associated with positive SHAP values, indicating an enhancement of LOI, which is consistent with the well-established role of phosphorus in promoting char formation and suppressing flame propagation. Similar to the interpretation in the UL-94 SHAP analysis, the synergist descriptor was encoded as a categorical variable; therefore,

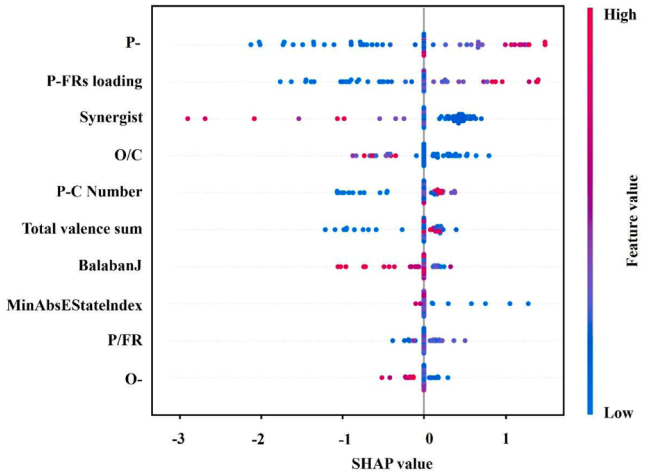

Fig. 11. SHAP feature importance of the LOI predictive model.

its SHAP importance should be understood as reflecting the distinction among different synergist types, rather than implying a monotonic or quantitative relationship between the assigned numerical label and LOI. In addition to formulation-related parameters, several molecular structure descriptors, including the O/C ratio, the number of P-C bonds, total valence sum, Balaban index, and MinAbsEStateIndex, also exhibit notable influence. These features reflect the importance of elemental composition and molecular topology in determining LOI. For example, higher O/C ratios tend to show positive contributions to LOI, which may be related to the presence of oxygen-containing functional groups that facilitate dehydration and char formation during thermal decomposition. However, the mixed color distribution observed for several structural descriptors suggests non-monotonic relationships between molecular structure and LOI, highlighting the complexity of structure-property interactions in flame-retardant systems. The SHAP dependence plots of the top five most important features are presented in Figure S2.

3.4. *Flame-retardant performance evaluation and validation*

As discussed above, the LOI and UL-94 vertical burning ratings are important indicators for evaluating flame retardancy. However, neither LOI nor UL-94 vertical burning rating alone can fully characterize the overall flame-retardant performance [35]. On this basis, a non-compensatory evaluation framework is introduced to integrate these two key indicators. Fig. 12 presents a non-compensatory flame retardancy evaluation matrix, in which an overall flame retardancy level can be directly determined based on the results of LOI values and UL-94 vertical burning ratings. In this matrix, the flame-retardant performance is classified into four discrete levels: Levels 1(L-1), Levels 2 (L-2), Levels 3 (L-3), Levels 4(L-4), with subcategories L-2a, L-2b, L-3a, and L-3b, as illustrated by the color-coded scheme.

The vertical axis corresponds to the UL-94 vertical burning rating, reflecting the self-extinguishing behavior and dripping characteristics of materials under vertical burning conditions, while the horizontal axis represents LOI, which quantitatively characterizes the minimum oxygen concentration required to sustain combustion. In this study, the UL-94 vertical burning rating results are categorized into V0 and non-V0, representing whether the materials meet the highest vertical burning classification. The LOI values are divided into three ranges, below 21%, 21–27%, and above 27%, corresponding to low, moderate, and high intrinsic flame-retardant efficiency, respectively. The LOI thresholds of 21% and 27% are selected based on their common use in flame retardancy studies, where 21% approximates the oxygen concentration in air and 27% is widely regarded as a practical criterion for highly flame-retardant polymeric materials [36].

Based on these classifications, the non-compensatory evaluation matrix assigns an overall flame retardancy level to each composite. Importantly, two evaluation strategies, denoted as a and b, are introduced to reflect different decision logics. In practical engineering applications, particularly in electronic and electrical systems, achieving the V0 is frequently regarded as a mandatory criterion for product qualification. Accordingly, in a strategy, priority is assigned to the UL-94 vertical burning rating, such that materials achieving the V0 are first grouped into a higher category than non-V0 materials and are then further differentiated according to their LOI values. Conversely, in materials development, formulation optimization, and mechanism-oriented studies, LOI is often considered a more fundamental parameter, as it quantitatively characterizes the intrinsic flammability of the polymer matrix and is less sensitive to specific test configurations. Therefore, the suffixes a and b denote two distinct decision logics within the same overall flame retardancy level, corresponding to classification strategies prioritizing UL-94 vertical burning rating (L-2a, L-3a) and LOI (L-2b, L-3b), respectively.

Table 4 presents a comparison between the experimentally reported values and the predictions obtained from the proposed model for both UL-94 vertical burning rating and LOI, along with the corresponding flame retardancy level classification. The three literature cases presented in Table 4 were not included in the 510-sample dataset used for model training and testing. They were used as independent external validation samples. As representative examples reported in the literature [22,37,38], the predicted UL-94 vertical burning ratings and LOI values in good agreement with the reported experimental data. Across these independent cases, the predicted UL-94 vertical burning ratings (V0 or non-V0) are fully consistent with the corresponding experimental outcomes. The predicted LOI values also agree well with the reported experimental results (the absolute deviation is within 1 percentage point). For instance, in case 1, an epoxy composite containing 11 wt%

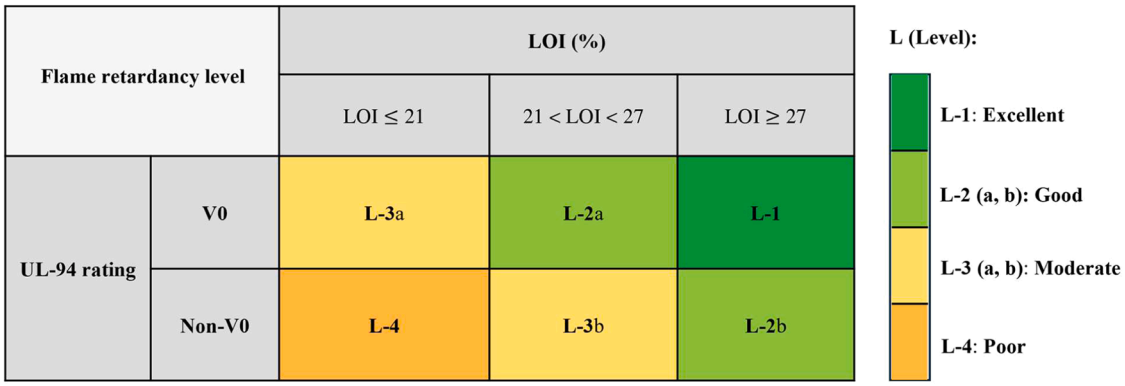

Fig. 12. Flame retardancy grading matrix.

Table 4
Comparison of experimental and predicted UL-94 vertical burning rating, LOI, and flame-retardancy level.

Cases	Reference	UL-94 rating	LOI	Flame retardancy level
Case 1	Chen et al., 2023 [22]	V0*	32.7*	L1
	Predictions (this work)	V0	32.1	L1
Case 2	Liu et al., 2026 [37]	V0*	31.7*	L1
	Predictions (this work)	V0	32.1	L1
Case 3	Liu et al., 2026 [38]	non-V0*	27.6*	L2-b
	Predictions (this work)	non-V0	28.6	L2-b

* : Experimental results.

BDOPO exhibits a LOI of 32.7 and achieves V0 rating. The present models predict V0 rating and an LOI value of 32.1%. In case 2, the epoxy composite containing 7 wt% DVB shows a V0 rating and LOI of 31.7%, and the predicted results are V0 and 32.1%. In case 3, the epoxy composite containing 3 wt% DOPR reported non-V0 rating and LOI of 27.6%, which are consistently predicted as non-V0 and 28.6%. Importantly, in all cases, the reported and predicted results lead to the same flame retardancy classification. These results demonstrate that the proposed framework can provide accurate prediction of key flammability properties and enable reliable classification of flame retardancy levels without additional combustion experiments, which lays a foundation for screening and design of high-performance EP in the future.

This study proposed a data-driven framework for the prediction and comprehensive evaluation of flame-retardant performance based on a dataset of P-FRs epoxy systems, which may limit its applicability to other material systems. Future work will focus on expanding the dataset and exploring more advanced models to enhance its general applicability and robustness.

In addition, the LOI thresholds used in this study are based on commonly adopted values in flame retardancy research. Further refinement of LOI thresholds settings, in combination with industry standards and practical requirements, should be taken into consideration.

4. Conclusions

This work presents a data-driven and non-compensatory evaluation framework for the prediction and comprehensive assessment of flame-retardant performance in EP/P-FR systems. By integrating structural, formulation, and other condition-related features through feature fusion, ML models were constructed to predict key flame retardancy metrics, including UL-94 vertical burning ratings and LOI values. The proposed models achieved reliable predictive performance, with an accuracy of 0.81 for UL-94 vertical burning classification and an R^2 of approximately 0.79 for LOI prediction on the test set.

Through interpretable machine-learning analysis, the relative importance and contributions of structural and formulation-related features were quantitatively revealed, enabling systematic identification of the key factors governing polymer flame-retardant behavior. Furthermore, based on the predicted UL-94 vertical burning and LOI values, a non-compensatory decision-making methodology was further established. A representative case study confirmed the consistency between model predictions and reported experimental results.

Benefiting from the integration of predictive modeling and decision analysis, the proposed framework provides a more scientific and comprehensive strategy for polymer flame retardancy assessment. This work not only reduces the reliance on extensive experimental trial-and-error but also offers a promising data-driven paradigm for the discovery and rational design of high-performance EP/P-FR composites.

CRediT authorship contribution statement

Qiong Tan: Writing – original draft, Visualization, Software, Methodology, Investigation, Formal analysis, Data curation. **De-Yi Wang:** Writing – review & editing, Supervision, Conceptualization.

Declaration of competing interest

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

Funding sources

This work was supported by the European Union under the Marie Skłodowska-Curie Actions Postdoctoral Fellowships (HORIZON-MSCA-2024-PF-01), Project 101205162 - FireDesign.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.polymdegradstab.2026.112207.

Data availability

Data will be made available on request.

References

[1] X. Mi, N. Liang, H. Xu, J. Wu, Y. Jiang, B. Nie, D. Zhang, Toughness and its mechanisms in epoxy resins, Prog. Mater. Sci. 130 (2022) 100977, https://doi.org/10.1016/j.pmatsci.2022.100977.
[2] M. Yu, Z. Dai, M. He, Y. Huang, D. Wang, J. Shi, H. Gong, X. Han, A phosphorus/silicon hybrid curing agent for epoxy resin: synergistically enhanced thermal resistance, flame retardancy, and toughness, Chem. Eng. J. 519 (2025) 164924, https://doi.org/10.1016/j.cej.2025.164924.
[3] G. Huang, Y. Pan, L. Liu, P. Song, R. Yang, Metal-organic frameworks and their derivatives for sustainable flame-retardant polymeric materials, Adv. Nanocompos. 2 (2025) 1–14, https://doi.org/10.1016/j.adna.2024.10.001.
[4] Y. Wang, L. Liu, L. Ma, J. Yuan, L. Wang, H. Wang, F. Xiao, Z. Zhu, Transparent, flame retardant, mechanically strengthened and low dielectric EP composites enabled by a reactive bio-based P/N flame retardant, Polym. Degrad. Stabil. 204 (2022) 110106, https://doi.org/10.1016/j.polymdegradstab.2022.110106.
[5] Y. Ai, L. Xia, F. Pang, Y. Xu, H. Zhao, R. Jian, Mechanically strong and flame-retardant epoxy resins with anti-corrosion performance, Compos. Part B-Eng. 193 (2020) 108019, https://doi.org/10.1016/j.compositesb.2020.108019.
[6] J. Capricho, B. Fox, N. Hameed, Multifunctionality in epoxy resins, Polym. Rev. 60 (2020) 1–41, https://doi.org/10.1080/15583724.2019.1650063.
[7] X. Wang, Y. Hu, L. Song, W. Xing, H. Lu, P. Lv, G. Jie, Flame retardancy and thermal degradation mechanism of epoxy resin composites based on a DOPO substituted organophosphorus oligomer, Polymer 51 (2010) 2435–2445, https://doi.org/10.1016/j.polymer.2010.03.053.
[8] Q. Yang, J. Wang, X. Chen, S. Yang, S. Huo, Q. Chen, P. Guo, X. Wang, F. Liu, W. Chen, P. Song, H. Wang, A phosphorus-containing tertiary amine hardener enabled flame retardant, heat resistant and mechanically strong yet tough epoxy resins, Chem. Eng. J. 468 (2023) 143811, https://doi.org/10.1016/j.cej.2023.143811.
[9] X. Mu, X. Li, C. Liao, H. Yu, Y. Jin, B. Yu, L. Han, L. Chen, Y. Chun, L. Song, Y. Hu, Phosphorus-fixed stable interfacial nonflammable gel polymer electrolyte for safe flexible lithium-ion batteries, Adv. Funct. Mater. 32 (2022) 2203006, https://doi.org/10.1002/adfm.202203006.
[10] Nelson G.L. Fire and polymers: an overview. 1995.
[11] B. Zou, Z. Qian, Z. Zhang, W. Xing, S. Zhong, J. Li, F. Chu, Y. Chun, J. Chen, Y. Hu, Quantitative assessment of whether phosphorus-based flame retardants are optimizing or degrading the fire hazard of aircraft carbon fiber/epoxy composites, Compos. Part B-Eng. 279 (2024) 111413, https://doi.org/10.1016/j.compositesb.2024.111413.
[12] Z. Chen, B. Yang, N. Song, Y. Liu, F. Rong, X. Zhang, T. Chen, Q. Zahng, J. Jiang, T. Chen, Y. Hu, L. Li, Machine learning-enabled rational design of organic flame retardants for enhanced fire safety of epoxy resin composites, Compos. Commun. 44 (2023) 101756, https://doi.org/10.1016/j.coco.2023.101756.
[13] W. Ge, R. De Silva, Y. Fan, S. Sisson, M. Stenzel, Machine learning in polymer research, Adv. Mater. 37 (2025) 2413695, https://doi.org/10.1002/adma.202413695.

[14] P. Jafari, R. Zhang, S. Huo, Q. Wang, J. Yong, M. Hong, Ravinesh Deo, H. Wang a, P. Song, Machine learning for expediting next-generation of fire-retardant polymer composites, Compos. Commun. 45 (2024) 101806, https://doi.org/10.1016/j.coco.2023.101806.
[15] T. Fu, Y. Wang, Design and development of fire-safety materials in artificial intelligence era, Acc. Mater. Res. 6 (2025) 544–549. https://pubs.acs.org/doi/full/10.1021/accountsmr.5c00065.
[16] J. Zhuang, A. Midgley, Y. Wei, Q. Liu, D. Kong, X. Huang, Machine-learning-assisted nanozyme design: lessons from materials and engineered enzymes, Adv. Mater. 36 (2024) 2210848, https://doi.org/10.1002/adma.202210848.
[17] P. Kalhor, N. Jung, S. Bräse, C. Wöll, M. Tsotsalas, P. Friederich, Functional material systems enabled by automated data extraction and machine learning, Adv. Funct. Mater. 34 (2024) 2302630, https://doi.org/10.1002/adfm.202302630.
[18] X. Chen, M. Nian, F. Zhao, Y. Ma, J. Yao, S. Wang, X. Chen, D. Li, M. Fang, Artificial intelligence for the discovery of safe and effective flame retardants, Environ. Sci. Technol. 59 (2025) 7187–7199, https://doi.org/10.1021/acs.est.4c14787.
[19] W. Ma, L. Li, Y. Zhang, M. Li, N. Song, P. Ding, Active learning-based generative design of halogen-free flame-retardant polymeric composites, J. Mater. Inf. 5 (2025) 35, https://doi.org/10.20517/jmi.2025.09.
[20] D. Phan, A. Morgan, L. Poudel, R. Bhowmik, A machine learning platform for polymer flammability prediction, Polym. Degrad. Stabil. 240 (2025) 111411, https://doi.org/10.1016/j.polymdegradstab.2025.111411.
[21] J. Xiao, J. Hobson, M. Haranczyk, D-Y Wang, Machine learning framework to predict instantaneous heat release rate of polymer nanocomposites in cone calorimetry, Polym. Degrad. Stabil. 218 (2023) 110563, https://doi.org/10.1016/j.polymdegradstab.2023.110563.
[22] Z. Chen, B. Yang, N. Song, T. Chen, Q. Zhang, C. Li, J. Jiang, T. Chen, Y. Hu, L. Liu, Machine learning-guided design of organic phosphorus-containing flame retardants to improve the limiting oxygen index of epoxy resins, Chem. Eng. J. 455 (2023) 140547, https://doi.org/10.1016/j.cej.2022.140547.
[23] C. Yan, X. Lin, X. Feng, H. Yang, P. Mensah, G. Li, Advancing flame retardant prediction: a self-enforcing machine learning approach for small datasets, Appl. Phys. Lett. 122 (2023) 251902, https://doi.org/10.1063/5.0152195.
[24] E. Verret, A. Collin, S. Duquesne, Optimization of flame retardant polypropylene via machine learning, J. Phys.: Conf. Ser. 2885 (2024) 012017, https://doi.org/10.1088/1742-6596/2885/1/012017.
[25] H. Nguyen, K. Nguyen, T. Le, L. Soufeiani, A.P. Mouritz, Predicting heat release properties of flammable fiber-polymer laminates using artificial neural networks, Compos. Sci. Technol. 215 (2021) 109007, https://doi.org/10.1016/j.compscitech.2021.109007.
[26] J. Li, B. Zou, A. Bekeshev, M. Akhmetova, R. Orynbassar, X. Wang, Y. Hu, Practical machine learning model selection and interpretation for organophosphorus flame retardancy in Epoxy resin, Polym. Degrad. Stabil. 234 (2025) 111209, https://doi.org/10.1016/j.polymdegradstab.2025.111209.
[27] Z. Chen, C. He, K. Wang, F. Rong, L. Xiang, Z. Zuo, C. Wang, X. Yang, Y. Guo, J. Jiang, Y. Yu, Machine learning-driven molecular generation for accelerated screening of high-performance flame retardants in epoxy resin composites, Chem. Eng. J. 516 (2025) 163946, https://doi.org/10.1016/j.cej.2025.163946.
[28] R. Wang, T. Fu, Y. Yang, X. Wang, Y. Wang, Deeper insights into flame retardancy of polymers by interpretable, quantifiable, yet accurate machine-learning model, Polym. Degrad. Stabil. 230 (2024) 110981, https://doi.org/10.1016/j.polymdegradstab.2024.110981.
[29] C. Liu, E. Fujita, Y. Katsura, Y. Inada, A. Ishikawa, R. Tamura, K. Kimura, R. Yoshida, Machine learning to predict quasicrystals from chemical compositions, Adv. Mater. 33 (2021) 2102507, https://doi.org/10.1002/adma.202170284.
[30] J. Aniceto, B. Zezere, C. Silva, Prediction of diffusion coefficients in aqueous systems by machine learning models, J. Mol. Liq. 405 (2024) 125009, https://doi.org/10.1016/j.molliq.2024.125009.
[31] T. Chen, XGBoost: A Scalable Tree Boosting System, Cornell University, New York, 2016.
[32] Z. Hong, M. Tao, L. Liu, M. Zhao, C. Wu, An intelligent approach for predicting overbreak in underground blasting operation based on an optimized XGBoost model, Eng. Appl. Artif. Intell. 126 (2023) 107097, https://doi.org/10.1016/j.engappai.2023.107097.
[33] J. Novakovic, A. Veljovic, S. Ilic, Z. Papic, M. Tomovic, Evaluation of classification models in machine learning, Theory Appl. Math. Comput. Sci. 7 (2017) 39–46.
[34] S. Lundberg, S. Lee, A unified approach to interpreting model predictions, Adv. Neural Inf. Process. Syst. 30 (2017) 4768–4777, https://doi.org/10.48550/arXiv.1705.07874.
[35] C. McCoy, S. Stoliarov, Experimental characterization and modeling of boundary conditions and flame spread dynamics observed in the UL-94V test, Combust. Flame. 225 (2021) 214–227, https://doi.org/10.1016/j.combustflame.2020.10.054.
[36] L. Chen, Y. Wang, A review on flame retardant technology in China. Part I: development of flame retardants, Polym. Adv. Technol. 21 (2010) 1–26, https://doi.org/10.1002/pat.1550.
[37] Y. Liu, Q. Xie, M. Huang, X. Qu, J. Liu, X. Li, W. Wei, Synthesis of a P/N/B-containing halogen-free modifier for high-efficiency flame retardancy, reinforcement and toughening of epoxy thermosets, Appl. Polym. Sci. 143 (2026) e70496, https://doi.org/10.1002/app.70496.
[38] Y. Liu, S. Zheng, F. Zhou, H. Yuan, Phosphorus-containing resveratrol-based flame retardants derived from esterification reactions for flame-retardant epoxy resins, Giant 27 (2026) 100385, https://doi.org/10.1016/j.giant.2026.100385.