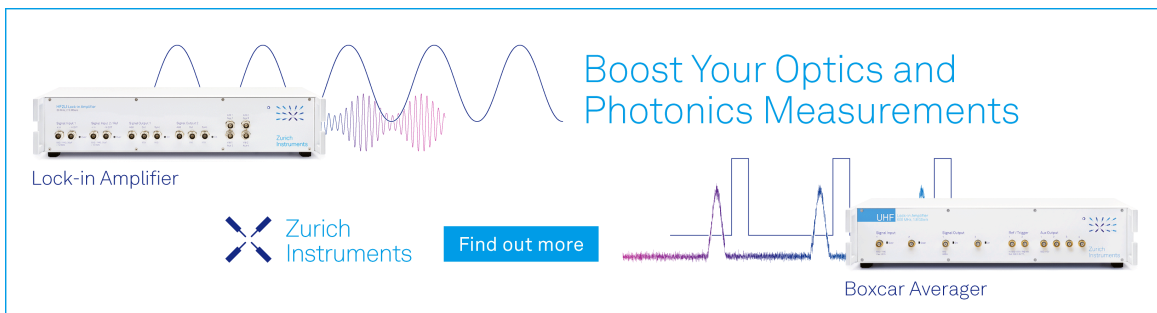


RESEARCH ARTICLE | JUNE 21 2023

Advancing flame retardant prediction: A self-enforcing machine learning approach for small datasets

Special Collection: [Accelerate Materials Discovery and Phenomena](#)Cheng Yan ; Xiang Lin ; Xiaming Feng ; Hongyu Yang ; Patrick Mensah; Guoqiang Li *Appl. Phys. Lett.* 122, 251902 (2023)<https://doi.org/10.1063/5.0152195>

CrossMark



Boost Your Optics and Photonics Measurements

Lock-in Amplifier

 Zurich Instruments

[Find out more](#)

Boxcar Averager

Advancing flame retardant prediction: A self-enforcing machine learning approach for small datasets

Cite as: Appl. Phys. Lett. **122**, 251902 (2023); doi: [10.1063/5.0152195](https://doi.org/10.1063/5.0152195)

Submitted: 28 March 2023 · Accepted: 2 June 2023 ·

Published Online: 21 June 2023



View Online



Export Citation



CrossMark

Cheng Yan,^{1,a)} Xiang Lin,² Xiaming Feng,^{2,a)} Hongyu Yang,² Patrick Mensah,¹ and Guoqiang Li³

AFFILIATIONS

¹Department of Mechanical Engineering, Southern University and A&M College, Baton Rouge, Louisiana 70813, USA

²College of Materials Science and Engineering, Chongqing University, 174 Shazhengjie, Shapingba, Chongqing 400044, China

³Department of Mechanical and Industrial Engineering, Louisiana State University, Baton Rouge, Louisiana 70803, USA

Note: This paper is part of the APL Special Collection on Accelerate Materials Discovery and Phenomena.

^{a)}Authors to whom correspondence should be addressed: cheng.yan@sus.edu and fengxm@cqu.edu.cn

ABSTRACT

Improving the fireproof performance of polymers is crucial for ensuring human safety and enabling future space colonization. However, the complexity of the mechanisms for flame retardant and the need for customized material design pose significant challenges. To address these issues, we propose a machine learning (ML) framework based on substructure fingerprinting and self-enforcing deep neural networks (SDNN) to predict the fireproof performance of flame-retardant epoxy resins. Our model is based on a comprehensive understanding of the physical mechanisms of materials and can predict fireproof performance and eliminate the needs for properties descriptors, making it more convenient than previous ML models. With a dataset of only 163 samples, our SDNN models show an average prediction error of 3% for the limited oxygen index (LOI). They also provide satisfactory predictions for the peak of heat release rate PHR and total heat release (THR), with coefficient of determination (R^2) values of 0.87 and 0.85, respectively, and average prediction errors less than 17%. Our model outperforms the support vector model SVM for all three indices, making it a state-of-the-art study in the field of flame retardancy. We believe that our framework will be a valuable tool for the design and virtual screening of flame retardants and will contribute to the development of safer and more efficient polymer materials.

Published under an exclusive license by AIP Publishing. <https://doi.org/10.1063/5.0152195>

Polymer materials are widely used in modern society due to their high performance and low cost. However, their flammability remains a significant issue, posing a threat to human life and property. Currently, epoxy resins are commonly used in construction, automotive, electronics, and aerospace industries, and they are highly flammable. In addition, flame retardant additives have been developed to improve their flame retardancy, which is essential for human space colonization. That is, the spacecraft cabin environment, with its enriched oxygen atmosphere, is highly susceptible to fire. Traditional development of flame-retardant polymers relies on experiments or modeling, both of which have limitations. Experiment-based material design, constrained by scientists' domain knowledge, is costly and time-consuming. Analytical^{1,2} and numerical^{3,4} modeling approaches, while helpful, involve simplifications or are time-consuming. Therefore, developing an efficient predictive model for fireproof epoxy resins is highly desirable.

Fortunately, the recently developed machine learning (ML) methods provide a promising tool for this intractable problem. Essentially, the majority of engineering problems can be considered as tasks to formulate governing equations and subsequently solve them. However, these equations are difficult to establish, not to mention to solve. On the other hand, ML can directly explore structure-property relation through fitting, eliminating the need to derive governing equations and solve them. Furthermore, the dramatic advances in hardware significantly arouse the interest for researchers to leverage ML. Since 1990s,^{5,6} investigators began to design or predict polymer performance with ML techniques. To date, ML has obtained some gratifying achievements in the design of multiple types of polymers, such as the prediction of electron affinity (EA) and ionization based on graph convolutional neural network (GCNN),⁷ shape memory properties prediction based on dual-convolutional-model framework,⁸ and transfer learning-variational autoencoder.⁹ So far, a couple of excellent review

TABLE I. Hyperparameters adopted in the training of the SDNN model.

Hyperparameters names	Values or item
Ratio between training data and test data	80/20
Batch size	32
Learning rate	0.01
Number of filter 1–2	64,64
Activation function for 1–8 hidden layers	Relu
Activation function for 9th hidden layer	Linear
Neuron number in hidden layers 1–9	4160, 4160, 4160, 4160, 4160, 1024, 512, 512, 256, 128
Random state in LOI model, PHRR model, and THR model	6, 49, 7

papers have been published in this field,^{10–15} and a comprehensive framework to design polymer materials with ML approach was proposed by Yan and Li.¹¹

In the field of flame retardancy, ML also began to cut a figure and a few papers have been published. For example, the heat release properties of flammable fiber-reinforced polymer laminates were predicted using the algorithms of multiple linear regression (MLR) and Bayesian regularized artificial neural network with Gaussian prior (BRANNGP).¹⁶ Chen *et al.* leveraged posynomial and other four ML approaches (conventional linear regression, nonlinear artificial neural networks, and their combination of Lasso, Ridge, and ANN) to optimize the flame retardancy of polymer nanocomposites.^{17,18} Chen Z *et al.* applied properties descriptor and regression to design organic phosphorus-containing flame retardants composite.¹⁹ However, the usage of ML in analyzing and exploring flame retardants epoxy resins has never been done before. Furthermore, the approaches that all the four papers^{16–18} applied are not easy enough to implement. To be specific, all they leveraged are the quantitative structure properties or fraction of certain components for a specified flame retardant to analyze the properties, which possess an apparent limitation. That is, extensive computational resources could be needed to obtain properties descriptors. In addition, the fraction or molar ratio of flame retardant often plays significant role for fireproofing, but it was also omitted by all the previous studies.^{16–19} Meanwhile, previous studies have focused on developing flame retardant composite, whereas current trend in research is to utilize grafting techniques to achieve fireproofing. Thus, developing a ML framework that does not rely on descriptors, but account for fraction and can handle flame retardant polymer, is highly desired, which is one of the motivations and contributions of this paper.

Furthermore, in order to better predict the properties of flame retardants, we must establish an approach that aligns with its chemical mechanisms, which previous models fallen short of addressing.^{16–19} Historically, various flame retardants have been developed to enhance polymer fire resistance, with different mechanisms based on their chemical elements and structures.²⁰ Flame retardants can be grouped as brominated, phosphorus-based, nitrogen-based, and metallic compound,²¹ each functioning differently. Brominated flame retardants rely on gaseous phase mechanisms, where bromine radicals form HBr, consuming high-energy H and OH radicals, stabilizing the fire

tetrahedron, and suppressing combustion. Phosphorus-based flame retardants exhibit mixed mechanisms.²² Phosphate flame retardants primarily involve condensed phase mechanisms, where phosphates degrade into acids, promoting intumescent char layer formation, which suppresses heat and fuel/oxygen transfer. Some phosphorus-containing structures, like hypophosphite²³ and 9,10-dihydro-9-oxa-10-phosphaphenanthrene-10-oxide (DOPO),²⁴ combine gaseous and condensed phase mechanisms, employing both free radical trapping and catalytic charring. Overall, chemical elements in flame retardants primarily improve fire retardancy, while molecular structures play promotive and supplemental roles. Thus, in this study, we decompose molecules structure into elements and substructures for feature identification, making our approach a suitable choice for predicting property of flame retardants. Moreover, as the performance of fireproofing can be influenced by the varying weights of different substructures, we have developed a self-enforcing deep neural network (SDNN) that can effectively account for these difference in weight. This is the other main contribution of this paper.

The aim of this study is to develop an ML framework that is able to reasonably predict and optimize the performance of flame retardant based on its structures. The paper is organized as follows. First, we introduce the detailed fingerprint method and mathematical framework of the SDNN model. Following that, we investigate the data distribution and validate the performance of two models using collected experimental data. Finally, we derive significant conclusions from our findings.

As exhibited in Fig. 1, the pipeline of prediction for the performance of flame retardants can be divided into four steps.

- (1) Fingerprinting (1 → 2). In this step, 2D chemical structures are fingerprinted by Morgan fingerprinting and one-hot, which will lead to a high-dimensional binary vector.
- (2) Enforcing mapping (2 → 3). The binary vector will be input into component 1 of the ML model for further mapping with an enforcing method. The mapping will be another latent vector.
- (3) Formation of new input (3,4 → 5). The new latent vector will be combined with the input 2 (molar ratio) to form a new input 5.
- (4) Prediction (6 → 7). The new input will be continuously fed into component 2 of the ML model (5 → 6) to predict the performance of flame retardant.

In this study, we collected a total of three types of datasets (LOI, PHRR, and THR), belonging to 56 different flame retardants. In combination with different molar ratios, we found a total of 163 combinations for LOI, 131 combinations for THR, and 126 combinations for PHRR (see the supplementary material). The flame retardants include small molecules and macromolecules, and their corresponding molar ratio is considered. The detailed method to handle molar ratio is shown in supplementary material S1.

In this stage, we utilized substructure fingerprinting composed of two methods, i.e., substructure decomposition and one-hot encoding. The substructure decomposition is based on Morgan fingerprinting.²⁵ The main idea can be divided into three steps. In the first step, distinct integers are assigned to each atom in the chemical structure [see Fig. 2(a)]. Second, every atom is iteratively updated by gradually enlarging the radius of bond. For example, when radius = 0, only each element itself is identified. When radius = 1, only the single bonds that

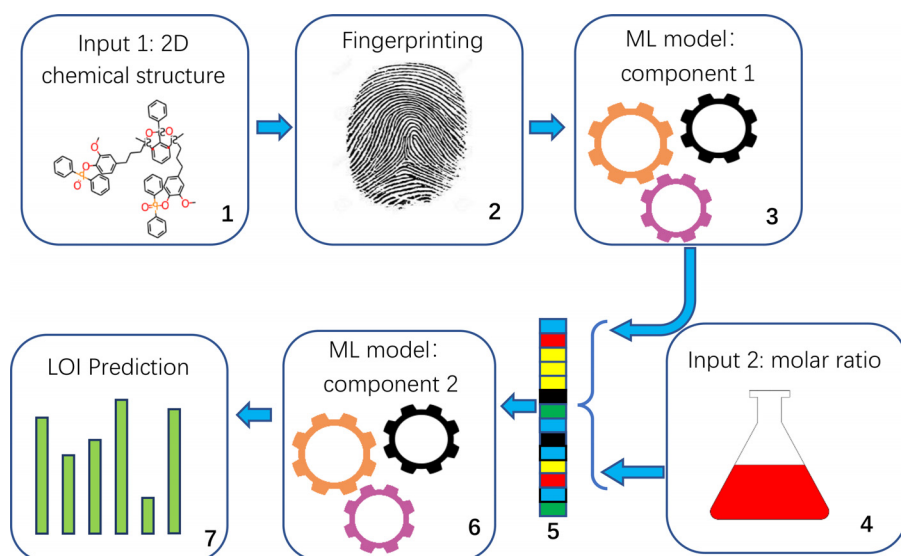


FIG. 1. Pipeline of prediction for the performance of flame retardants.

neighbor the core atom are considered. Finally, all substructures can be found by this method with gradually expanded radius. For the flame retardant EGN-Si/P in Fig. 2, there are a total of 103 substructures. In this process, the repetitive substructures will be deleted. Through studying a set of molecules, many substructures can be found, and every unique substructure will be assigned an integer, it can be performed by open sourced software package RDKit.²⁶ Next, we adopted one-hot to further convert these integer vectors into binary vectors. One example of this one-hot can be found in supplementary material S2.

As mentioned earlier, both elements and substructures contribute to the fireproofing performance and their respective role carry different weights. By realizing this, we leveraged a SDNN model. The main idea is inspired from self-attention mechanism.²⁷ As shown in Fig. 3, the binary vector representing molecule structure interacts with the vector derived from one component of the DNN. This interaction generates a new vector that is subsequently fed into the other component of SDNN for further processing and computation. Specifically, the binary vector $\mathbf{v}_1$ is initially fed into four fully connected layers (FL1–FL4) for learning. This process results in a vector $\mathbf{v}_2$ and reads

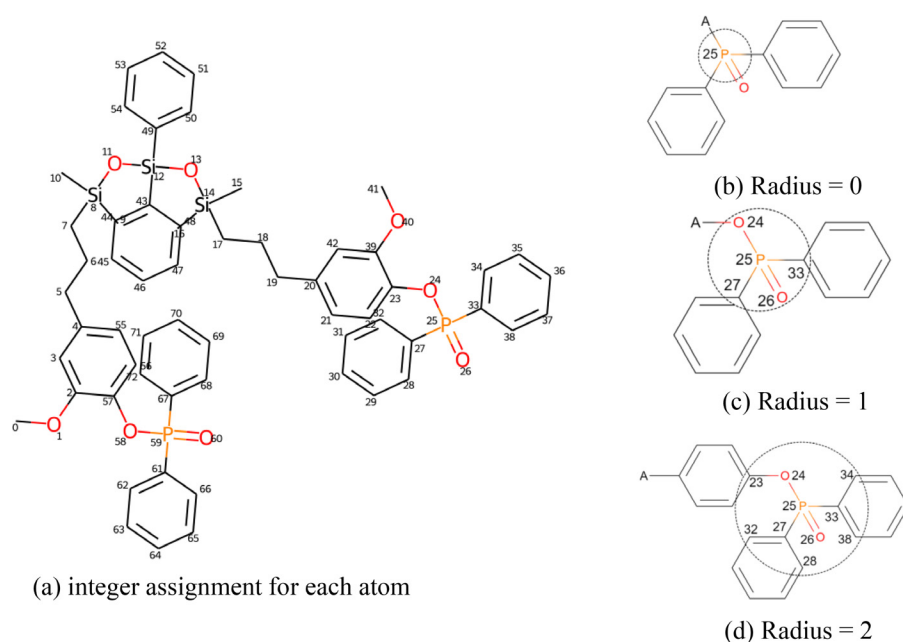


FIG. 2. Iteration process for the molecule of a flame retardant EGN-Si/P. (a) Integer assignment for each atom. (b)–(d) Involved elements when radius = 0, 1, and 2, respectively.

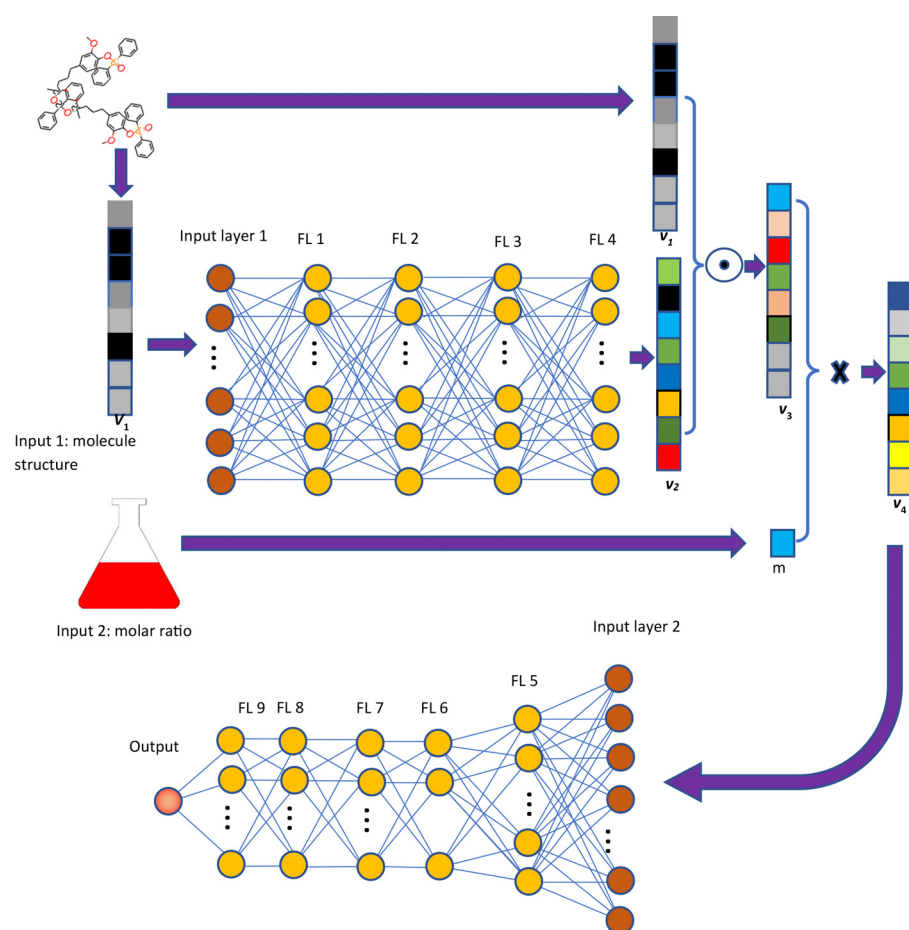


FIG. 3. Basic pipeline structures for the SDNN. FL represents fully connected layer. The model hyperparameters is shown in Table I.

$$\mathbf{v}_2 = h^{(4)} = z^{(4)}(\mathbf{w}^{(4)}\mathbf{x}^{(3)} + \mathbf{b}_4), \quad (1)$$

where $\mathbf{h}^{(4)}$ indicates the output from 4th fully connected layer. $\mathbf{w}^{(4)}$ and $\mathbf{b}_4$ denote the weight and bias tensors in fourth layer, respectively. $\mathbf{x}^{(3)}$ is the output from third layer. z^4 represents a rectified linear (ReLU) function in fourth layer, which is a piecewise function and reads

$$z^{(4)}(x) = \begin{cases} 0 & \text{if } x \leq 0 \\ x & \text{if } x > 0. \end{cases} \quad (2)$$

This activation function is also applied for fully connected layers FL 2–FL 8. This vector $\mathbf{v}_2$ will enhance the binary input $\mathbf{v}_1$ (fingerprinted from molecule structures) by endowing different weights through a dot product as

$$\mathbf{v}_3 = \mathbf{v}_1 \cdot \mathbf{v}_2. \quad (3)$$

Subsequently, $\mathbf{v}_3$ serves as an augmented input, which is passed to the subsequent layer for further processing. After that, $\mathbf{v}_2$ interact with the molar ratio m by

$$\mathbf{v}_4 = \mathbf{v}_3 \cdot m. \quad (4)$$

The SDNN model in this study is essentially a type of deep neural network (DNN). The aim of the DNN is to solve a problem of minimum, i.e., minimize the compressive error between ground truth and

predictions. The input of n th layer of the DNN can be mathematically expressed as

$$\mathbf{h}^{(n)} = z^{(n)}(\boldsymbol{\theta}^{(n)}\mathbf{x}^{(n-1)}), \quad (5)$$

where $\boldsymbol{\theta}$ is the updatable tensor that is composed of the weight tensor $\mathbf{w}$ and bias tensor $\mathbf{b}$, i.e., $\boldsymbol{\theta} = [\mathbf{w}, \mathbf{b}]^T$.

The activation function in ninth layer (last layer) is a linear function, which reads

$$z^{(9)}(x) = x. \quad (6)$$

Its objective function (loss function) can be represented by first-order approximation on high-dimensional Taylor theorem as

$$L(\boldsymbol{\theta}_{i+1}) = L(\boldsymbol{\theta}_i) + [\nabla L(\boldsymbol{\theta}_i)]^T (\boldsymbol{\theta}_{i+1} - \boldsymbol{\theta}_i) + O\|\boldsymbol{\theta}_{i+1} - \boldsymbol{\theta}_i\|. \quad (7)$$

Given a positive constant α (learning rate) and let $\boldsymbol{\theta}_{i+1} - \boldsymbol{\theta}_i = -\alpha \nabla L(\boldsymbol{\theta}_i)$, then the loss function Eq. (7) can be written as

$$L(\boldsymbol{\theta}_{i+1}) = L(\boldsymbol{\theta}_i) - \alpha \|\nabla L(\boldsymbol{\theta}_i)\|^2 + O\|\boldsymbol{\theta}_{i+1} - \boldsymbol{\theta}_i\|. \quad (8)$$

Therefore, the loss function continually reduces with iterations. In this study, we employed mean absolute value as loss function, which can be written as

$$L(\theta) = \frac{\sum_i^i |h^{(9)}(\theta) - y_i|}{i}, \quad (9)$$

where i represents the total number of ground truth in a batch. $h^{(9)}$ denotes the output of 9th fully connected layer. The activation function for the ninth fully connected layer is a identify function

$$z^{(9)}(x) = x. \quad (10)$$

The dataset is split into two segments: 80% is allocated for training data, while the remaining 20% is reserved for testing purposes. In order to thoroughly check the performance of our model, we employed 3 evaluation metrics, i.e., coefficient of determination (R^2) and mean average percentage error (MAPE) and percentage of correct point (PCP). The first two are defined as

$$R^2 = 1 - \frac{\sum_{i=0}^n (y_i - \hat{y}_i)^2}{\sum_{i=0}^n (y_i - \bar{y}_i)^2}, \quad (11)$$

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right|, \quad (12)$$

where $\hat{y}_i$, $\bar{y}_i$, and y_i are prediction outcome, average ground truth, and ground truth, respectively; and n is the total number of sample in the training data. As for the metrics “percentage of correct point (PCP),” we employed the word “correct” to indicate the difference between the prediction and the ground truth is within a bearable margin. Here, we established different PCPs for the three metrics. Namely, LOI has the largest dataset and smaller number range (the numbers range from 22 to 60), while PHRR and THR possess smaller datasets and larger number range (0–80). Therefore, we assign a smaller PCP for LOI and a larger PCP for PHRR and THR. Specifically, we consider the percentage error for LOI within 10% as “correct” and the percentage error for PHRR and THR within 20% is deemed correct, respectively.

From the scatterplot for the ground truth and prediction (Fig. 4), we found that the attained generally satisfactory performance for LOI, PHRR, and THR. Also, it can be clearly seen from Table II that the R^2 of test data for LOI, PHRR, and THR are 0.86, 0.87, and 0.85, respectively. This suggests that at least 85% of variances for the indices can be explained by the features from substructure fingerprinting. Simultaneously, both the MAPE for the test data of PHRR and THR fall within the acceptable range (17% and 15%, respectively). Notably, the model demonstrates best performance in predicting LOI, wherein a mere 3% MAPEs, and 96.97% errors are below 10% in the test data. Given that there are only 100–200 data points in each database, we believe that the chosen fingerprinting effectively capture the essential features of elements and substructures. Moreover, these results that our design for the ML model design is reasonable.

Meanwhile, the reliability of the SDNN model is evident. Generally, the difference in MAPE between the training data and test data should not exceed 20% to avoid overfitting. As shown in Table II, the difference in MAPE between the training data and test data for LOI is merely 2%, indicating a well-fitted model without overfitting occurs. Additionally, the differences of MAPE between the training

data and test data of PHRR and THR are both 10%, which are considered satisfactory results, given the small training dataset.

The primary errors can be attributed to two sources. First, the collected dataset is not exhaustive. Due to the inherent randomness and small volume of collected data, none of the three datasets conform to either normal distribution or uniform distribution. This can be roughly observed through the histograms in Fig. 5. It can also be validated by D’Agostino and Pearson’s test. We discovered that the measure of departure from normality, K^2 , for all the three datasets is 71.07, 11.346, and 15.976, respectively. Consequently, all the three databases deviate from normal distribution. Second, the datasets in the original references are, to some extent, artificially skewed by researchers. As reported Yan *et al.*,⁹ nearly all the reported results in references do not conform to Gaussian random distribution. This is because, although experimental results are supposed to yield a Gaussian distribution, researchers often report selected results, implying that only specific flame retardants were reported. For example, only the flame retardants with excellent fireproofing index were reported, while the major part (the flame retardants with moderate and weak fireproofing index) was involuntarily ignored. In particular, flame retardants with a LOI below 25 are more likely be omitted by the researchers because of their poor performance. Conversely, if the identified flame retardant has a LOI greater than 30, there is a high likelihood that it would be reported. This trend can be partially observed from histogram of LOI [Fig. 5(a)]. Despite the bias database, the model still shows strong performance across the three evaluation metrics, leading us to conclude that our SDNN model is reliable.

We also predict three indices by support vector machine (SVM) with same fingerprinting method (see supplementary material S3). The prediction results of SVM for the three indices are exhibited in Fig. S3 and Table S1. Upon comparing with the two model performances (see Tables II and S1), the SVM model shows inferior results for every index in the test data. This variation primarily stems from the distinct models utilized in each approach. Essentially, the two ML models in this study can be viewed as a mathematical optimization method based on curve fitting. We can examine this problem by considering the form of fitting functions and their respective number of trainable parameters. We investigated the number of fitting parameters in SDNN and SVM models, separately. First, the SDNN model has a total of 34 696 909 parameters, while the SVM model has 200–400 trainable parameters ($2 \times$ number of training data points). Given that the fireproofing parameters are relatively complex, the model with more parameters could be effective. Specifically, the model with more parameters is more flexible and adaptable, enabling them to capture non-linear relationships and intricate patterns within the data. This increased flexibility allows them to better model complex problems that may involve numerous interacting factors or dependencies. Second, the fitting function in two ML techniques is distinct: the SVM model presumed an exponential function, while the SDNN employed multiple piecewise functions (Relu). For this specific problem, the piecewise functions could be more suitable to describe the nature of fireproofing performance. Therefore, these two reasons contribute to the significant superiority of the SDNN model.

Simultaneously, the overfitting in the SVM model is notably more severe than in the SDNN model. Specifically, the R^2 differences between training and test data for the three indices (0.44, 0.36, and 0.28) are significantly larger than those of the SDNN model (0.12, 0.06, and 0.07). Concurrently, the SVM’s MAPE differences exceed

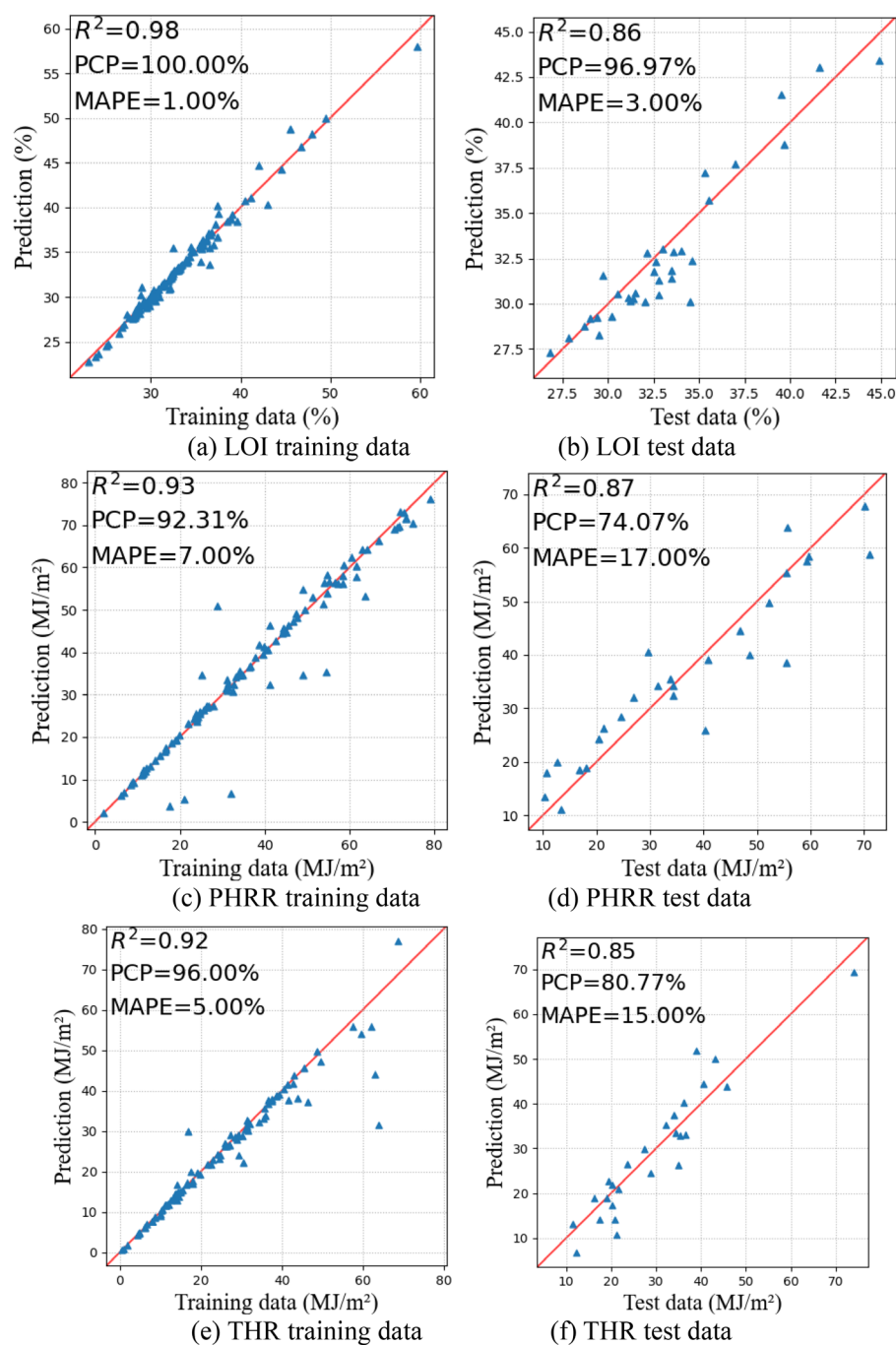


FIG. 4. Scatter plot of structure decomposition + SDNN model ($M_{SD+SDNN}$) predictions vs ground truth: (a) LOI training data; (b) LOI test data; (c) PHRR training data; (d) PHRR test data; (e) THR training data; and (f) THR test data.

those of the SDNN model (4% vs 2%, 29% vs 10%, and 15% vs 10%, respectively).

In order to delve deeper into the performance of our model, we employed conventional property descriptors for fingerprinting, as comprehensively outlined in Section S4. Our analysis highlights that this model, when utilizing property descriptors, yields comparable results for LOI prediction to those achieved with the $M_{SD+SDNN}$

approach. However, it exhibits a marked reduction in performance when applied to the other two predictive scenarios. These observations lead us to conclude that, while the property descriptor-based model demonstrates proficiency in predicting certain aspects of fireproof performance, it exhibits limitations when tasked with predictions for smaller datasets. This underlines the necessity for a more robust approach when dealing with limited data.

TABLE II. Prediction of structure decomposition + SDNN for LOI, PHRR, and THR.

Index Metrics	LOI (%)		PHRR (kW/m ²)		THR (MJ/m ²)	
	Training data	Test data	Training data	Test data	Training data	Test data
R ²	0.98	0.86	0.93	0.87	0.92	0.85
PCP(%)	100	96.97	92.31	74.07	96.00	80.77
MAPE(%)	100	3	7	17	5	15

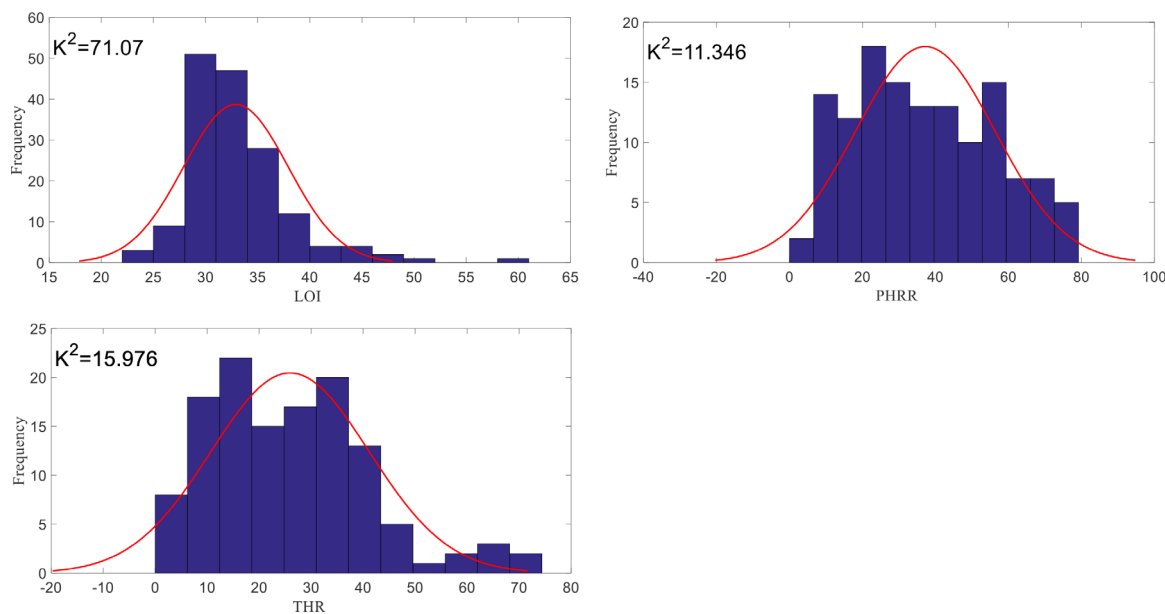


FIG. 5. Histograms for LOI, PHRR, and THR. K^2 represents the deviations from normality.

In this paper, based on our understanding for structure-property relation for flame retardants, we proposed a ML framework by employing substructure fingerprinting and enhanced SDNN model. Through this framework, we are able to provide satisfactory predictions for the three important indices of flame retardants on basis of small datasets. Comparing with previous works, this work can directly predict the unknown or *de novo* flame retardant without requiring the quantifications of their properties; thus, it provides distinct application advantages over previous methods. Moreover, it can work as an effective tool for fine-tuning the properties of flame retardants, as well as enabling inverse design for their development. However, it is noted that, while the prediction of flame retardants for epoxy resin has been conducted in this study, the modeling on relationship between flame retardants and polymer matrices is needed to fully establish the prospects of machine learning in developing flame retardants, which will be also comprehensively studied in our future works. In addition, as mentioned above, the fireproofing performance still possesses a complex mechanism, which influences the convergence of prediction. In the future, we anticipate collecting more data and integrating the effects of catalyst, resulting in a more practical model.

See the supplementary material that contains the details for calculation of the molar ratio for small molecules and macromolecules, example of one hot fingerprinting, SVM prediction, and property descriptors-based prediction.

This work was supported by the U.S. National Science Foundation under Grant No. OIA-1946231 and the Louisiana Board of Regents for the Louisiana Materials Design Alliance (LAMDA), Faculty Fellow Program (FFP)+Summer Undergraduate Research Funding (SURF) sponsored by NASA & Louisiana Board of Regents, under Grant No. NASA(2023)-SURF-01, and the National Science Foundation under Grant No. 1736136.

AUTHOR DECLARATIONS

Conflict of Interest

The authors have no conflicts to disclose.

Author Contributions

Cheng Yan and Xiang Lin contributed equally to this work.

Cheng Yan: Conceptualization (lead); Investigation (lead); Methodology (lead); Software (lead); Writing – original draft (lead).

Xiang Lin: Data curation (lead); Investigation (equal). **Xiaming Feng:** Supervision (equal); Writing – review & editing (equal). **Hongyu Yang:** Supervision (equal). **Patrick Mensah:** Resources (equal). **Guoqiang Li:** Validation (equal); Writing – review & editing (equal).

DATA AVAILABILITY

The data that support the findings of this study are available within the supplementary material.

REFERENCES

- ¹S. D. Watt, J. E. J. Staggs, A. C. McIntosh, and J. Brindley, *Fire Saf. J.* **36**, 421 (2001).
- ²J. E. J. Staggs, *Fire Saf. J.* **32**, 221 (1999).
- ³P. Patel, T. R. Hull, A. A. Stec, and R. E. Lyon, *Polym. Adv. Technol.* **22**, 1100 (2011).
- ⁴C. Vovelle, J. L. Delfau, M. Reuillon, J. Bransier, and N. Laraqui, *Combust. Sci. Technol.* **53**, 187 (1987).
- ⁵V. Venkatasubramanian, K. Chan, and J. M. Caruthers, *Comput. Chem. Eng.* **18**, 833 (1994).
- ⁶V. Venkatasubramanian, K. Chan, and J. M. Caruthers, *J. Chem. Inf. Comput. Sci.* **35**, 188 (1995).
- ⁷M. Aldeghi and C. W. Coley, *Chem. Sci.* **13**, 10486 (2022).
- ⁸C. Yan, X. Feng, C. Wick, A. Peters, and G. Li, *Polymer* **214**, 123351 (2021).
- ⁹C. Yan, X. Feng, and G. Li, *ACS Appl. Mater. Interfaces* **13**, 60508 (2021).
- ¹⁰W. Sha, Y. Li, S. Tang, J. Tian, Y. Zhao, Y. Guo, W. Zhang, X. Zhang, S. Lu, Y. Cao, and S. Cheng, *InfoMat* **3**, 353 (2021).
- ¹¹C. Yan and G. Li, *Adv. Intell. Syst.* **5**, 2200243 (2022).
- ¹²C. Yan and G. Li, in *Encyclopedia of Materials: Plastics and Polymers* (Elsevier, 2021), Vol. 2, pp. 267–279.
- ¹³M. M. Cencer, J. S. Moore, and R. S. Assary, *Polym. Int.* **71**, 537 (2022).
- ¹⁴L. Chen, G. Pilania, R. Batra, T. D. Huan, C. Kim, C. Kuenneth, and R. Ramprasad, *Mater. Sci. Eng., R* **144**, 100595 (2021).
- ¹⁵A. J. Gormley and M. A. Webb, *Nat. Rev. Mater.* **6**, 642 (2021).
- ¹⁶H. T. Nguyen, K. T. Q. Nguyen, T. C. Le, L. Soufeiani, and A. P. Mouritz, *Compos. Sci. Technol.* **215**, 109007 (2021).
- ¹⁷F. Chen, J. Wang, Z. Guo, F. Jiang, R. Ouyang, and P. Ding, *ACS Appl. Mater. Interfaces* **5**, 53425 (2021).
- ¹⁸F. Chen, L. Weng, J. Wang, P. Wu, D. Ma, F. Pan, and P. Ding, *Compos. Sci. Technol.* **231**, 109818 (2023).
- ¹⁹Z. Chen, B. Yang, N. Song, T. Chen, Q. Zhang, C. Li, J. Jiang, T. Chen, Y. Yu, and L. X. Liu, *Chem. Eng. J.* **455**, 140547 (2022).
- ²⁰B. W. Liu, H. B. Zhao, and Y. Z. Wang, *Adv. Mater.* **34**, 2107905 (2022).
- ²¹S. T. Lazar, T. J. Kolibaba, and J. C. Grunlan, *Nat. Rev. Mater.* **5**, 259 (2020).
- ²²M. S. Özer and S. Gaan, *Prog. Org. Coat.* **171**, 107027 (2022).
- ²³L. A. Savas, F. Hacıoglu, M. Hancer, and M. Dogan, *Polym. Bull.* **77**, 291 (2020).
- ²⁴K. A. Salmeia and S. Gaan, *Polym. Degrad. Stab.* **113**, 119 (2015).
- ²⁵D. Rogers and M. Hahn, *J. Chem. Inf. Model* **50**, 742 (2010).
- ²⁶G. Landrum (2013). “RDKit: Open-Source Cheminformatics Software,” GitHub <https://www.rdkit.org/>.
- ²⁷A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, *Adv. Neural Inf. Process. Syst.* **30**, 5998 (2017).