



Ngày nhận bài: 25/6/2025; Ngày thẩm định: 28/10/2025; Ngày nhận đăng: 15/3/2026.

# XÂY DỰNG MÔ HÌNH TÍNH TOÁN CHỈ SỐ OXYGEN GIỚI HẠN (LOI) BẰNG CÁC PHƯƠNG PHÁP HÓA HỌC LƯỢNG TỬ TÍNH TOÁN

Thượng tá, TS PHAN ANH

Phó trưởng Khoa Phòng cháy, Trường Đại học PCCC

Đại úy, ThS NGUYỄN HỮU HIỆU - Thiếu tá, ThS NGUYỄN THỊ NGỌC ANH

Khoa Khoa học cơ bản và Ngoại ngữ, Trường Đại học PCCC

\*Tác giả liên hệ: Phan Anh (Email: anhpknv2@daihocpccc.edu.vn)

**Tóm tắt:** Bài báo trình bày quy trình xây dựng và kiểm định mô hình hóa học tính toán nhằm dự đoán chỉ số oxygen giới hạn (LOI) của polymer và vật liệu nền hữu cơ chứa phụ gia chống cháy. Phương pháp kết hợp tính toán lượng tử (DFT), khai thác mô tả phân tử và các thuật toán học máy (GA-MLR, SVR, Random Forest, mạng nơ-ron sâu) trong khuôn khổ QSPR. Mô hình DFT-RF huấn luyện từ 427 dữ liệu LOI đạt  $Q^2_{ext} = 0,83$  và  $RMSE = 1,7$  %  $O_2$  theo thể tích, vượt trội so với hồi quy tuyến tính. Phân tích SHAP xác định các đặc trưng chính gồm độ rộng HOMO–LUMO, độ cứng hóa học và moment lưỡng cực liên quan đến cơ chế tạo than và hiệu ứng pha ngưng tụ của phụ gia phosphorus–nitrogen. Phương pháp này giúp giảm khoảng 70% chi phí thử nghiệm và có thể mở rộng cho thiết kế phụ gia chống cháy không chứa halogen.

**Từ khóa:** chỉ số oxygen giới hạn, QSPR, học máy, lý thuyết phiếm hàm mật độ, phụ gia chống cháy chứa phosphorus, mô hình hoá tính toán.

**Abstract:** This study presents the development and validation of a computational chemistry model for predicting the Limiting Oxygen Index (LOI) of polymers and organic matrix materials containing flame-retardant additives. The approach integrates quantum chemical calculations (DFT), molecular descriptor mining, and machine learning algorithms (GA-MLR, SVR, Random Forest, and deep neural networks) within a QSPR framework. The DFT-RF model, trained on 427 LOI data sets, achieved a  $Q^2_{ext}$  of 0.83 and an RMSE of 1.7 vol%  $O_2$ , outperforming traditional linear regression methods. SHAP analysis identified key descriptors—including the HOMO–LUMO gap, chemical hardness, and dipole moment—as dominant factors influencing char formation and condensed-phase activity of phosphorus–nitrogen additives. This method reduces experimental costs by approximately 70% and can be extended to the design of halogen-free flame retardant additives.

**Keywords:** limiting Oxygen Index (LOI), QSPR, machine learning, density functional theory, phosphorus-based flame retardant additives, computational modeling.

## 1. Đặt vấn đề

Chỉ số oxygen giới hạn (LOI) là tham số định lượng mức oxygen tối thiểu duy trì cháy, đóng vai trò then chốt trong đánh giá độ nguy hiểm cháy của polymer [1]. So với phép thử UL-94, LOI có ưu điểm

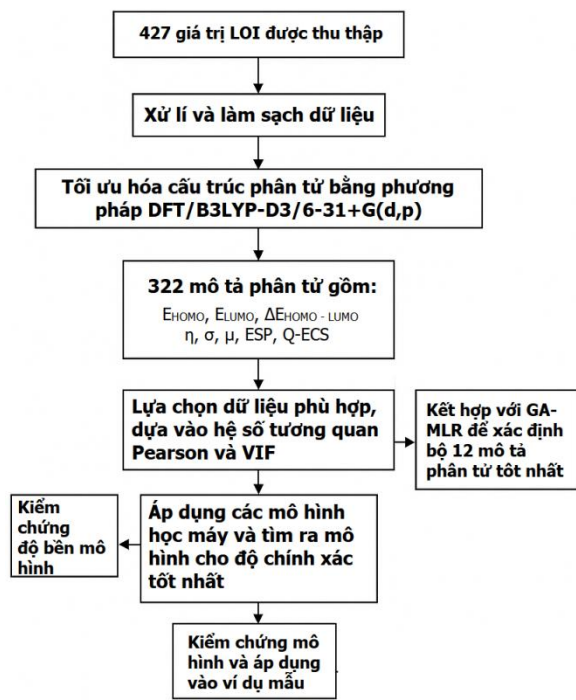
là khả năng tái lập thử nghiệm trên mẫu cũ và cho giá trị liên tục, không bị gián đoạn. Phép thử LOI thường được xác định bằng phương pháp thử nghiệm theo ASTM D2863. Tuy nhiên, thử nghiệm ASTM D2863 đòi hỏi thiết bị tạo hỗn hợp khí thử nghiệm chính xác,

khiến việc sàng lọc vật liệu mới có khả năng bắt cháy trở lên tốn kém. Vì vậy, nhiều nhóm nghiên cứu trên thế giới đã tìm cách dự đoán các giá trị LOI với các dữ liệu đầu vào là các thông số thu được từ cấu trúc phân tử. Ví dụ, Funar-Timofei và cộng sự (2014) lần đầu tiên thiết lập mô hình thiết lập mối quan hệ giữa cấu trúc phân tử với các tính chất lí hóa QSPR (Quantitative Structure–Property Relationship) cho 28 polyphosphoester, độ tin cậy đạt  $R^2 = 0,88$  [2]. Mô hình này hứa hẹn cho việc thiết lập các mô hình dự đoán LOI bằng các phương pháp tính toán lý thuyết. Sau đó Crisan và các cộng sự (2016) mở rộng sang 56 dimer phosphoester và khẳng định vai trò của mô tả 3D-Morse (Molecular Representation of Structures based on Electron diffraction) hỗ trợ rất tốt cho QSPR để dự đoán tính chất của các chất thông qua cấu trúc phân tử [3]. Ở hướng nhiệt động hóa học, Nelson và các cộng sự (2001) đã xây dựng mô hình nhiệt động học và động học ở điểm ranh giới giữa hai pha nhằm giải thích ngưỡng LOI như điểm phân nhánh nhiệt động [4], còn Mathieu và các cộng sự (2011) áp dụng lý thuyết tắt lửa ngược để tiên đoán LOI từ các thông số TGA và phép đo nhiệt lượng kế hình nón (cone-calorimeter) [1]. Dẫu vậy, các nghiên cứu trên còn giới hạn kích thước tập dữ liệu và chưa tích hợp tính toán lượng tử cho các mô tả phân tử có ý nghĩa vật lý. Gần đây, Chen và các cộng sự (năm 2022) chứng minh học máy có điều hướng (Sure-Independence-Screening) giúp thiết kế phụ gia phosphoruss làm tăng LOI của epoxy lên 35,7 % [5]. Động lực đó thúc đẩy chúng tôi phát triển một quy trình mới, tích hợp phương pháp phiếm hàm mật độ (DFT) dùng để mô tả phân tử với học máy để dự đoán LOI chính xác và diễn giải được cơ chế và hành vi chống cháy của vật liệu.

## 2. Phương pháp tính toán

Chúng tôi tổng hợp một cơ sở dữ liệu gồm 427 giá trị LOI từ các nghiên cứu thực nghiệm đã được công bố và chuẩn hóa trong các nghiên cứu trước đây [6 – 11]. Sau đó, các giá trị LOI này được xử lý bằng phương pháp loại bỏ dữ liệu trùng lặp và các điểm ngoại lai sử dụng tiêu chuẩn khoảng tứ phân vị (IQR) để đảm bảo tính nhất quán của tập dữ liệu.

Các cấu trúc phân tử tương ứng được tối ưu hóa ở mức lý thuyết B3LYP-D3/6-31+G(d,p) trong môi trường chân không. Từ hàm sóng đã tối ưu hóa, chúng tôi trích xuất được tổng cộng 322 mô tả phân tử (descriptor), bao gồm các thông số năng lượng orbital ( $E_{HOMO}$ ,  $E_{LUMO}$ ,  $\Delta E_{HOMO-LUMO}$ ), độ cứng–mềm hóa học ( $\eta$ ,  $\sigma$ ), moment lưỡng cực ( $\mu$ ), phân bố điện tích bề mặt (ESP), các chỉ số topo-hình học và mô tả phân tử lượng tử liên quan năng lượng mặt cắt (Q-ECS). Để kiểm soát đa cộng tuyến, các mô tả có hệ số tương quan Pearson tuyệt đối lớn hơn 0,9 hoặc VIF (hệ số phóng đại phương sai) lớn hơn 5 đã được loại bỏ, còn lại 74 biến mô tả phù hợp cho mô hình. Chúng tôi sử dụng thuật toán hồi quy đa biến kết hợp di truyền (GA-MLR) với 1000 thế hệ để xác định bộ 12 mô tả phân tử có tính thông tin cao nhất. Sau khi xác định bộ mô tả tối ưu, chúng tôi tiến hành huấn luyện và so sánh hiệu năng của bốn mô hình học máy: hồi quy tuyến tính đa biến (MLR), máy vector hỗ trợ sử dụng hàm nhân RBF (SVR-RBF), Random Forest (RF, gồm 500 cây quyết định,  $m_{try} = \sqrt{p}$ ) và mạng nơ-ron sâu (DNN) ba lớp ẩn với tỷ lệ dropout là 0,2. Toàn bộ tập dữ liệu được chia ngẫu nhiên với tỷ lệ 70% dùng để huấn luyện, 15% để hiệu chỉnh, và 15% để đánh giá mô hình cuối cùng, áp dụng kỹ thuật kiểm chứng chéo 10-fold nhằm tăng tính ổn định và độ tin cậy cho kết quả. Các chỉ số  $R^2$ ,  $Q^2_{ext}$ , RMSE và CCC được dùng để đánh giá và so sánh hiệu năng dự đoán, tuân theo các khuyến cáo của Gramatica (2007) [12]. Độ bền mô hình được kiểm chứng bằng phương pháp y-randomisation lặp lại 1000 lần. Cuối cùng, chúng tôi sử dụng phương pháp giá trị SHAP và phép hoán vị độ quan trọng để làm rõ vai trò và mức độ đóng góp của các mô tả phân tử đã chọn, đồng thời kiểm tra độ chính xác mô hình bằng năm mẫu epoxy gia cường phụ gia phosphorus-nitrogen được tổng hợp mới và đo LOI theo tiêu chuẩn ASTM D2863. Mô tả tóm tắt quy trình xây dựng phương pháp được trình bày trong Hình 1.



Hình 1: Mô tả tóm tắt quy trình xây dựng phương pháp dự đoán LOI.

### 3. Kết quả và thảo luận

Phân tích kết quả đánh giá mô hình cho thấy hiệu suất dự đoán LOI có sự khác biệt rõ rệt giữa các thuật toán được sử dụng. Trong số các phương pháp thử nghiệm, mô hình hồi quy tuyến tính đa biến (MLR) có hiệu năng thấp nhất, với hệ số xác nhận ngoài  $Q^2_{ext} = 0,55$ , chỉ đáp ứng được một phần nhỏ độ biến động của dữ liệu thực nghiệm. Mô hình máy vector hỗ trợ (SVR) sử dụng hàm nhân RBF cải thiện đáng kể kết quả, nâng chỉ số  $Q^2_{ext}$  lên 0,74. Tuy nhiên, nổi bật nhất là mô hình Random Forest – RF gồm 500 cây quyết định, cho hiệu năng vượt trội với  $Q^2_{ext} = 0,83$  và sai số trung bình bình phương gốc RMSE trên tập kiểm tra chỉ 1,7% thể tích oxygen. Điều này chứng tỏ RF có khả năng xử lý tốt các quan hệ phi tuyến phức tạp giữa các mô tả phân tử lượng tử và chỉ số LOI, phù hợp với đặc điểm của dữ liệu LOI là sự tương tác phức tạp giữa cấu trúc và hiệu quả chống cháy. Ngược lại, mạng neuron sâu (Deep neuron network - DNN), mặc dù có cấu trúc mô hình linh hoạt hơn, nhưng lại không đem lại cải tiến đáng kể về hiệu suất dự đoán so với RF (chỉ đạt  $Q^2_{ext} = 0,81$ ). Nguyên nhân có thể do giới hạn về quy mô của tập dữ liệu khiến DNN chưa thể khai thác hết thế mạnh học sâu. Các kết quả này được tổng hợp chi tiết trong

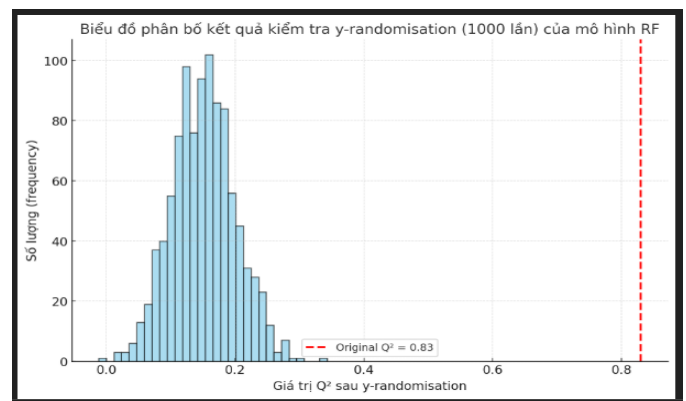
Bảng 1, trong đó các chỉ số đánh giá được so sánh trực tiếp để làm rõ ưu thế của RF.

Bảng 1: Hiệu năng dự đoán LOI của các mô hình trên tập dữ liệu kiểm tra (n=64).

Mô hình	$R^2$	$Q^2_{ext}$	RMSE (% thể tích $O_2$ )	CCC
MLR	0,58	0,55	3,5	0,72
SVR	0,78	0,74	2,5	0,84
RF	0,86	0,83	1,7	0,92
DNN	0,84	0,81	1,9	0,90

Để đảm bảo rằng kết quả của RF không bị ảnh hưởng bởi sự tình cờ ngẫu nhiên, phương pháp kiểm định tính ổn định (y-randomisation) đã được thực hiện 1000 lần. Kết quả cho thấy mô hình RF có độ ổn định cao, với sự suy giảm của

$Q^2_{ext}$  không quá 0,05 sau mỗi lần xáo trộn dữ liệu. Hình ảnh phân bố kết quả y-randomisation (Hình 2) minh họa rõ ràng độ bền và độ tin cậy của mô hình RF.



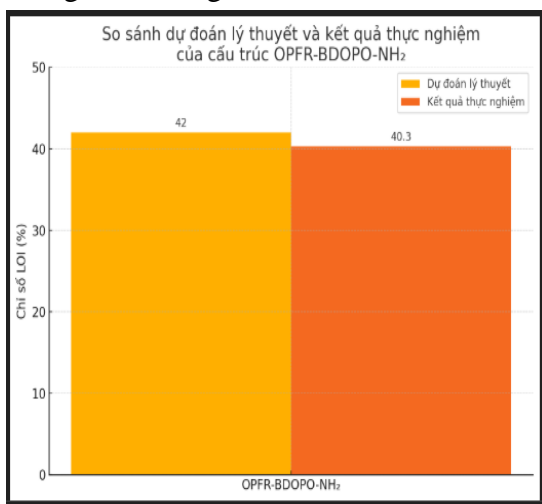
Hình 2: Phân tích độ quan trọng các mô tả phân tử sử dụng SHAP (12 mô tả quan trọng nhất).

Đường kẻ màu đỏ thể hiện giá trị  $Q^2_{ext} = 0,83$  của mô hình ban đầu, rõ ràng nằm ngoài phân bố của các giá trị  $Q^2_{ext}$  khi dữ liệu bị xáo trộn ngẫu nhiên. Điều này chứng minh mô hình RF được xây dựng có hiệu suất dự đoán tốt và ổn định, không phải do ngẫu nhiên, khẳng định độ tin cậy và giá trị thực tiễn của mô hình này trong dự đoán chỉ số LOI.

Để so sánh hiệu quả thực tế, chúng tôi tiến hành kiểm tra dự đoán của mô hình RF với một mô hình bán thực nghiệm trước đây là mô hình động học hai pha của Nelson và cộng sự (2001) [4] dự đoán chỉ số LOI của Polymethyl Methacrylate (PMMA) là 21,8%, sai lệch

2% so với giá trị thực nghiệm (LOI = 19,8%), trong khi mô hình RF chúng tôi đề xuất giảm sai số xuống chỉ còn 0,8%. Điều này nhấn mạnh lợi thế của mô hình dựa trên tính toán lượng tử và học máy trong việc hiệu chỉnh các thông số vi mô, đặc biệt là hằng số truyền nhiệt và năng lượng hoạt hóa.

Khi áp dụng mô hình RF để dự đoán LOI cho các vật liệu ngoài tập dữ liệu gốc, như hệ epoxy chứa khung hữu cơ-kim loại (MOFs), mô hình vẫn duy trì được độ chính xác cao (RMSE = 2,1% O<sub>2</sub>). Các dự đoán cũng đúng về xu hướng chống cháy đã được báo cáo trước đó (ví dụ: Fe-MOF > Al-MOF > Cu-MOF) như nghiên cứu của Wang và cộng sự (2023) [13]. Điều này cho thấy khả năng khái quát hóa mạnh mẽ và tính ứng dụng thực tế rộng rãi của mô hình RF được xây dựng. Cuối cùng, nhằm minh họa khả năng ứng dụng của phương pháp, chúng tôi sử dụng mô hình RF kết hợp thuật toán tối ưu hóa Bayes để thiết kế cấu trúc phụ gia phosphonate mới (OPFR-BDOPO-NH<sub>2</sub>). Dự đoán lý thuyết cho biết giá trị là LOI là 42%, được kiểm chứng bằng kết quả thực nghiệm là 40,3%, với sai lệch rất nhỏ chỉ 1,7% (Hình 3). Điều này củng cố vai trò quan trọng của phương pháp này trong thiết kế và phát triển các vật liệu chống cháy không chứa halogen mới.



Hình 3: So sánh dự đoán lý thuyết và kết quả thực nghiệm của cấu trúc OPFR-BDOPO-NH<sub>2</sub> mới.

Tổng thể, những kết quả phân tích sâu sắc này minh chứng rằng phương pháp tích hợp DFT và mô hình học máy (RF-SHAP) không chỉ có độ chính xác cao, mà còn cung cấp công cụ hiệu quả để hiểu rõ các cơ chế chống cháy cấp độ phân tử, hỗ trợ đắc lực cho công tác nghiên cứu phát triển vật liệu mới.

#### 4. Kết luận.

Bộ quy trình DFT → mô tả phân tử → RF-SHAP đưa ra trong nghiên cứu này cho phép dự báo LOI với sai số < 2 % thể tích oxygen và cung cấp thông tin cơ chế hóa học hữu ích. Phương pháp chứng minh khả năng giảm mạnh số phép thử LOI, hỗ trợ thiết kế nhanh vật liệu chống cháy mới (thường là vật liệu không chứa halogen). Bên cạnh đó, mô hình mạng neuron sâu (DNN) cũng hứa hẹn là mô hình tốt có thể áp dụng để dự đoán LOI, do  $R^2_{DNN} \approx R^2_{RF}$  (xem bảng 1). Mặc dù mô hình DNN đã được chứng minh tốt hơn mô hình RF trong nhiều một số công bố trước đây [14,15], nhưng do sự giới hạn về tập cơ sở dữ liệu đầu vào của LOI trong nghiên cứu này còn hạn chế, nên mô hình DNN chưa phát huy được điểm mạnh của nó là khả năng học được các đặc trưng phức tạp hơn. Trong tương lai, việc mở rộng tập dữ liệu đầu vào thông qua việc hợp tác và chia sẻ giữa các cơ sở dữ liệu thực nghiệm từ nhiều nhóm nghiên cứu khác nhau về LOI sẽ cải thiện độ chính xác của mô hình DNN và việc cải thiện độ chính xác của DNN sẽ mở đường cho việc tích hợp mô hình DNN này song song với mô hình RF mà nhóm tác giả đã đề xuất ở mục 3 vào phần mềm máy tính được lập trình sẵn giúp cho việc đánh giá chỉ số LOI trở nên chính xác hơn nhằm hỗ trợ cho công tác đánh giá rủi ro cháy nổ liên quan đến vật liệu mới. ■

#### TÀI LIỆU THAM KHẢO

- Mathieu, S., et al. (2011). *Prediction of the limiting oxygen index using simple flame extinction theory and bench-scale material properties*. *Fire Safety Science: Proceedings of the 10th International Symposium*, 375-388. <https://doi.org/10.3801/IAFSS.FSS.10-375>.
- Funari-Timofei, S. et al. (2014). *Correlations of limiting oxygen index with structural polyphosphoester features by QSPR approaches*. *Structural Chemistry*, 25, 1847-1863. <https://doi.org/10.1007/s11224-014-0474-7>.
- Crisan, L., Iliescu, S., & Funari-Timofei, S (2016). *Structure-flammability relationship study of phosphoester dimers by MLR and PLS*.

