第 52 卷 第 6 期：2567-2579 高电压技术 Vol.52, No.6: 2567-2579
2026 年 6 月 30 日 High Voltage Engineering June 30, 2026
DOI: 10.13336/j.1003-6520.hve.20250988

基于定量构效关系的环保绝缘气体生物安全性预测评估与机理分析

叶凡超，肖　峰，田双双，张晓星
（湖北工业大学新能源及电网装备安全监测湖北省工程研究中心，武汉 430068）

摘　要：环保绝缘气体对于输配电装备绿色升级至关重要。尽管已有研究针对环保气体绝缘性能与物化性质展开预测建模，但生物安全性作为实际应用中的关键指标，其结构影响因素与其他性能参数显著不同，亟需构建独立的生物安全性预测模型来进行筛选和设计工作。为此，该研究基于经济合作与发展组织(Organization for Economic Co-operation and Development，OECD)指导原则，首次构建了碳氟类环保绝缘气体的急性吸入毒性定量-构效关系(structure-activity relationship，QSAR)模型。采用多种机器学习算法建模，并通过性能评估筛选出径向基函数核支持向量机(support vector machine with radial basis function kernel, SVM-RBF)与极端梯度提升(eXtreme gradient boosting, XGBoost)两个性能优异的模型，它们在训练集拟合能力、交叉验证稳定性和外部预测泛化性方面均表现突出。进一步引入 SHAP 方法，对关键分子描述符与生物毒性之间的关系进行了可解释性分析，揭示了影响生物安全性的核心结构特征。研究结果为碳氟类环保绝缘气体的分子设计与风险控制提供了理论依据和分析框架。
关键词：环保绝缘气体；LC50；QSAR；机器学习；SHAP

Predictive Assessment and Mechanistic Analysis of Eco-friendly Insulating Gas Biosafety Based on Quantitative Structure-activity Relationship

YE Fanchao, XIAO Feng, TIAN Shuangshuang, ZHANG Xiaoxing
(Hubei Engineering Research Center for Safety Monitoring of New Energy and Power Grid Equipment, Hubei University of Technology, Wuhan 430068, China)

Abstract：Eco-friendly insulating gases are essential for the green upgrading of power transmission and distribution systems. Although predictive modeling of insulation performance and physicochemical properties has been widely studied, biosafety, as a critical practical indicator, is governed by distinct structural factors and requires independent predictive models. In this work, following OECD guidelines, we developed the first QSAR models for the acute inhalation toxicity of fluorocarbon-based eco-friendly insulating gases. Through performance evaluation, support vector machine with radial basis function kernel (SVM-RBF) and eXtreme gradient boosting(XGBoost) were selected as two top-performing models, which demonstrated excellent performance in training set fitting capability, cross-validation stability, and external prediction generalizability. The SHapley additive explanations(SHAP) analysis was further applied to interpret the relationships between key molecular descriptors and toxicity, revealing the core structural features influencing biosafety. The results provide a theoretical basis and analytical framework for the molecular design and risk control of fluorocarbon-based eco-friendly insulating gases.
Key words：environmental-friendly insulation gas; LC50; QSAR; machine learning; SHAP

0　引言

气体绝缘设备是电力系统的核心装备，广泛采用 SF_6 作为绝缘和灭弧介质。SF_6 的温室效应潜势值(global warming potential, GWP)高达 CO_2 的 24 300 倍，且大气寿命长达 3 200 年，被《京都议定书》等条约列为重点管控气体。1994—2024 年间，大气中 SF_6 浓度增长了 317%，其中电力行业消耗的 SF_6 占全球总量的 80%[1]。碳氟类气体以其优异的强电负性与低 GWP 值，成为环保绝缘气体的首要选择[2-4]。

尽管关于 SF_6 替代气体的探索已经持续了数十年，但目前仍未找到能完全替代 SF_6 的环保型绝缘气体。这主要是由于用于高压电气设备的绝缘气体

基金资助项目：国家自然科学基金(52407170)；湖北省科学自然基金(2024AFB306)。
Project supported by National Natural Science Foundation of China (52407170), Natura Science Foundation of Hubei Province (2024AFB306).

需同时满足多项性能要求，如高电气强度、良好的热稳定性、低液化温度、温室效应小以及低毒性等，而这些性质往往彼此制约，难以在同一分子中实现协同优化[5]。在此背景下，基于分子层面的设计与筛选方法正逐渐成为研究重点。分子设计技术能够精确调控化合物的结构属性，相比传统依赖经验和试验的方式，这种方法可借助构效关系分析，从理论上揭示分子结构与绝缘性能、液化特性等关键参数之间的内在联系，从而实现高效筛选与性能定向优化，推动新型高性能环保气体的开发进程。

目前，环保绝缘气体分子的绝缘性能、液化温度和环保特性等性能的预测已经取得了一定的进展。Yu 等人基于静电势描述符和广义相互作用函数(generalized interaction properties function，GIPF)方法，通过量子化学优化分子结构，计算表面积、电势均值和方差等参数。最终建立了构效关系模型，在绝缘强度预测中表现出色[6]。王宝山等人利用 GIPF 参数建立了气体绝缘强度预测模型并通过分子杂化技术组合得到多种具有高绝缘强度的单一气体分子[7]。杨帅等人分析了电子域及电子概率密度对 GIPF 与气体绝缘强度相关性的影响，优化了 GIPF 参数计算方法，实现了气体绝缘领域对分子参数及 GIPF 的自定义[8]。Olivet 等人通过拟合实验数据建立了 SF_6 气体的力场模型，成功预测其蒸气压、沸点及输运特性，与实验结果高度一致[9]。Hou 等人基于高精度量子化学计算，构建了适用于 C_4F_7N 的全原子力场模型，准确预测了其气液相平衡、沸点及输运特性[10]。Betowski 等人采用密度泛函 (density functional theory，DFT) 及二阶 Møller-Plesset 微扰(second-order møller-plesset perturbation theory，MP2) 方法对 235 种气体的辐射效率进行计算，用以计算气体分子的 GWP 值，所得理论值与实测值高度一致[11]。

气体绝缘电气设备逐渐向人口密集区域安装，且设备内多充以 0.15~0.6 MPa 的高压绝缘气体介质，发生故障泄漏时易快速扩散且难以察觉，大多数环保绝缘气体为人工合成气体，其泄漏引发的暴露健康风险问题愈发受到关注。然而，用于预测生物安全性的结构特征与预测绝缘、液化温度等性能的参数并不相同，前者更侧重于分子活性与结构复杂性，后者则主要依赖极化率、电负性等物理电性参数。因此需建立独立的生物安全性预测模型。

目前仍缺乏有效的生物安全性预测方法对新型气体的生物安全性进行快速的早期评估。定量-构效关系(structure-activity relationship, QSAR)广泛应用于毒性预测等领域，通过构建分子结构与生物效应的定量关系模型，在药物开发、环境毒理学及新型材料筛选中表现出良好效果，为分子设计提供重要理论支撑[12-13]。Brand 等基于最高允许浓度(maximum permissible concentration，MAC) 毒性数据尝试建立绝缘气体毒性构效关系模型，但因数据稀缺未获实用结果[14]。Yu 等人基于 15 种 SF_6 替代气体的 MAC 毒性数据，提出以静电势描述符建立定义各气体与 SF_6 相对毒性的构效关系模型[15]。Chen 首次建立了基于氟代硝基多环芳烃(fluorinated nitro-polycyclic aromatic hydrocarbons，FNFPAHs)水生环境毒性的 QSAR 模型，为 FNFPAHs 的毒性风险评估提供了有效工具和机制解析[16]。Liu 利用机器学习构建 QSAR 模型精准预测微塑料毒性，发现粒径为关键因素，模型可用于微塑料风险评估[17]。Shuo Chen 采用简单 2D 描述符开发了全氟/多氟烷基物质(per-and polyfluoroalkyl substances，PFASs) 对鼠类急性口服毒性的 QSAR 和定量反应活性构效关系(quantitative reactivity activity structure activity relationship，q-RASAR)模型，模型通过验证并成功预测了大量未测试 PFASs 的毒性，为优先排序与毒性机制解析提供支持，可用于 PFASs 生态环境风险评估[18]。碳氟类气体分子结构具有其独特性，现有的 QSAR 模型难以直接适用于此类化合物的急性吸入毒性预测。

本文基于经济合作与发展组织(Organization For Economic Co-operation and Development，OECD)的开发原则对碳氟类环保绝缘气体急性吸入毒性开展了系统的 QSAR 建模研究。通过对多种机器学习算法的性能对比，筛选出最优算法构建预测模型，并利用可解释性方法(SHapley additive exPlanations，SHAP)方法分析模型中重要分子描述符的贡献，突破了传统黑箱模型的局限。研究深入探讨了分子描述符与毒性之间的机理关系，明确了影响碳氟类环保绝缘气体生物安全性的关键因素，为优化分子结构设计、生物安全性预测及气体的实际应用提供了科学依据和技术支持。

1 方法与数据集

1.1 数据集获取

本研究从 PubChem(https://pubchem.ncbi.nlm)数

据库中筛选获得 134 种碳氟类气体分子的急性吸入毒性数据小鼠 4 h 半数致死浓度(median lethal concentration，LC50)[19]，并对数据质量进行了严格把控，以确保其科学性与可信度。所构建数据集包含多类典型结构，分子主要由 C、F 及其他常见元素构成，涉及烃类、醇类、醛类、酮类、酯类、醚类、羧酸类、胺类、腈类以及卤代烃类。对于极少数公开数据仅提供大鼠 LC50 的情形，采用异速生长规律(Kleiber's law)的换算方法进行统一[20-22]。为便于毒性强度的量化与模型构建，原始的 LC50(μL/L)数值经负对数转换为 pLC50(−lgLC50)，其数值与毒性水平正相关。使用 pLC50 作为模型的预测终点，符合 OECD 组织对 QSAR 模型提出的原则 1“明确的预测终点”[23]。

1.2　分子特征描述符计算与清洗

本研究首先从 PubChem 数据库获取碳氟类气体分子的 3D 结构文件(SDF 格式)，并通过 ChemDraw 22.0 对其分子结构进行校核。随后，利用 E-Dragon 平台(http://www.vcclab.org)计算分子的结构、电子、几何、官能团与原子特性、统计与信息学特征以及整体性质等在内的 1 666 个分子描述符[24]。

对获得的描述符数据进行预处理时，首先剔除全部取值为零的描述符，以及方差低于设定阈值(0.1)的低信息特征。接着，计算剩余描述符间的皮尔逊相关系数矩阵，以衡量其线性相关性；当任意两个描述符的相关系数(R)大于 0.95 时，保留其中一个，删除冗余特征。经筛选后共移除 958 个描述符，保留 708 个用于进一步降维处理。

为提高模型训练效率并消除特征之间的量纲差异，对清洗后的数据进行 Z-score 标准化处理[25]。在此基础上，应用随机森林递归特征消除(random forest recursive feature elimination, RF-RFE)算法进行特征选择。该方法结合随机森林对特征重要性的评估结果，递归移除对模型贡献较低的特征，最终获得一个性能最优的特征子集[26]。

1.3　QSAR 模型

1.3.1　机器学习算法

机器学习在 QSAR 研究中得到了广泛应用，能够有效捕捉化学分子特征与其生物活性之间的非线性关系，从而构建具有较强预测能力的模型，提高分子筛选与优化的效率。本研究引入多种机器学习算法，对其在碳氟类环保绝缘气体毒性建模中的适用性与性能进行系统评估[27]，以满足 OECD 对 QSAR 模型提出的原则 2：“具备科学可解释性的算法模型”[23]。

选用的模型包括基于贝叶斯框架的稀疏学习方法相关向量机(relevance vector machine, RVM)，以及基于结构风险最小化原则的支持向量机(support vector machine, SVM)[28-29]。两者均采用径向基函数核(radial basis function kernel, RBF)和多项式核(polynomial kernel, Poly)提升对非线性特征的拟合能力。此外，本文还引入随机回归森林(random regression forest, RRF)[30]与极端梯度提升(eXtreme gradient boosting, XGBoost)算法[31]，进行建模分析。

1.3.2　模型的验证与评价

QSAR 模型的评价与验证是确保其可靠性与预测能力的关键环节，通常包括内部验证与外部验证的综合分析[32]。内部验证中，通过留一法交叉验证系数(Q^2_{LOO})与五折交叉验证系数(R^2_{CV5})评估模型稳健性。一般而言，一个可用的 QSAR 模型应满足拟合系数 R^2_{TR}>0.6，交叉验证系数 R^2_{CV5}>0.5。此外，当二者差值 $R^2_{\mathrm{TR}}-R^2_{\mathrm{CV5}}$<0.3，可以确保模型未出现过拟合，保证良好的泛化性能[33-35]。外部验证通过测试集决定系数 R^2_{TS} 评估模型对未知样本的预测能力。同时，训练集与测试集的平均绝对误差(mean absolute error, MAE)δ_{MAE} 与均方根误差(root mean square error, RMSE)δ_{RMSE} 用于量化预测值与真实值之间的偏差与误差，反映模型实际预测性能。其中，参数 δ_{MAETR}、δ_{RMSETR} 的下标 TR 表示训练集(training set)，δ_{MAETS}、δ_{RMSETS} 下标 TS 表示测试集(test set)，用于区分模型在不同数据集上的表现。

为进一步验证模型适用性，通过 Williams 图分析其应用域[36]。该方法结合杠杆值(h_i)与标准化残差(e_i)识别异常样本，借助阈值 h^*判断样本是否处于模型适用范围内，从而明确模型在化学空间中的可靠性与适用边界。符合 OECD 组织对 QSAR 模型提出的原则 3“明确的应用领域”与原则 4“科学的性能统计”[23]。

本文中所使用的各项性能指标计算式如下：

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

$$R^2_{\mathrm{CV5}}=1-\frac{\sum_{i=1}^{n}(y_i-\hat{y}_{\mathrm{CV}})^2}{\sum_{i=1}^{n}(y_i-\bar{y})^2} \tag{2}$$

$$\delta_{\mathrm{MAE}}=\frac{1}{n}\sum_{j=1}^{n}\left|u_j-\hat{u}_j\right| \tag{3}$$

$$\delta_{\mathrm{RMSE}}=\sqrt{\frac{1}{n}\sum_{j=1}^{n}\left(u_j-\hat{u}_j\right)^2} \tag{4}$$

$$Q^2_{\mathrm{LOO}}=1-\frac{\sum_{i=1}^{n}(y_i-\hat{y}_{i(-i)})^2}{\sum_{i=1}^{n}(y_i-\bar{y})^2} \tag{5}$$

$$e_i=\frac{y_i-\hat{y}_i}{s\sqrt{1-h_i}} \tag{6}$$

$$h_i=x_i^{\mathrm{T}}(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X})^{-1}\boldsymbol{x}_i \tag{7}$$

$$h^*=\frac{3(p+1)}{n} \tag{8}$$

式中：y_i 表示真实值；$\hat{y}_i$ 表示预测值；$\bar{y}$ 表示真实值的平均值；$\hat{y}_{\mathrm{CV}}$ 表示通过交叉验证得到的预测值；$\hat{y}_{i(-i)}$ 表示剔除第 i 个样本后模型的预测值；n 表示样本数量；s 表示模型的标准误差；$\boldsymbol{x}_i$ 表示第 i 个样本的输入特征；$\boldsymbol{X}$ 表示样本特征矩阵；p 表示模型中包含的特征数量。

1.4 基于 SHAP 方法的 QSAR 模型机理解释

SHAP 是一种源自博弈论的模型可解释性方法，用于定量评估各输入特征在机器学习预测中的独立贡献[37-39]。其理论基础为 Shapley 值，这一概念最初用于解决合作博弈中收益的公平分配问题。将其引入机器学习后，SHAP 可为每个特征分配明确的影响度，从而揭示模型的内部决策机制。SHAP 具体解释公式如下：

$$f(x)=\phi_0+\sum_{i=1}^{M}\phi_i \tag{9}$$

式中：$f(x)$ 表示模型对样本 x 的预测值，是目标变量的估计值；ϕ_0 表示基准值，表示无特征影响时模型的初始预测；ϕ_i 表示特征 i 的 SHAP 值，其代表特征 i 对预测结果的边际贡献。

SHAP 兼具全局和局部解释能力。一方面，可在整体层面分析模型对各特征的敏感性，明确关键驱动因素；另一方面，亦能对单一样本提供细粒度解释，展示各特征对预测结果的正负影响方向与程度[40]。这种解释机制有助于揭示特征结构与毒性响应之间的潜在关系，推动对复杂非线性模型预测路径的深入理解，从而增强模型在定量构效研究中的透明性与可信性。

在 QSAR 建模中，SHAP 的引入不仅提高了模型的可解释性，也为解析分子结构对毒性贡献的规律提供了量化依据，有助于支持机理推断与合理分子设计。符合 OECD 对 QSAR 模型提出的原则 5“科学的机理解释，如果可能的话”[23]。通过特征贡献的量化分析，可为毒性机制研究提供辅助判断与理论支撑，有助于建立更具科学依据的预测模型。

2 QSAR 建模分析

2.1 数据集分布与划分

在模型构建阶段，我们将数据集按照 4:1 的比例随机划分为训练集(training, TR)与测试集(test, TS)，数据分布如图 1 所示。可以观察到，pLC50 数据整体呈近似正态分布，表明毒性水平在样本间分布较为均匀。这种数据特征有助于避免局部样本密度过高或极端值主导模型训练，从而提高建模过程的稳定性与代表性。此外，近似正态的分布结构也有利于 QSAR 模型更准确地学习特征与毒性响应之间的潜在函数关系，降低模型对异常样本的敏感性，增强模型在未知样本上的泛化能力，减少过拟合风险，进一步确保预测的可靠性和实用性[41]。

在特征降维过程中，我们采用 RF-RFE 方法对经过预处理和初步清洗后的 708 个特征进行进一步筛选与优化。以十折交叉验证的均方根误差(δ_{RMSE})作为评估指标，分析了不同特征数量对模型性能的影响。如图 2 所示，随着特征数量的增加，δ_{RMSE} 值呈现出先急剧下降后逐渐趋于平稳的趋势。特征数量增加至 130 时，δ_{RMSE} 降至最低值 0.690 7，此时模型的拟合效果达到最佳。随着特征数量继续增加，δ_{RMSE} 略微上升，但整体趋势趋于稳定。

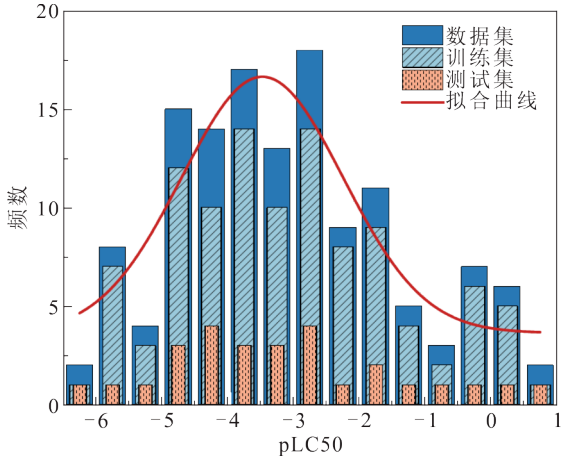


图 1 毒性数据 pLC50 分布情况

Fig.1 Distribution of toxicity data pLC50

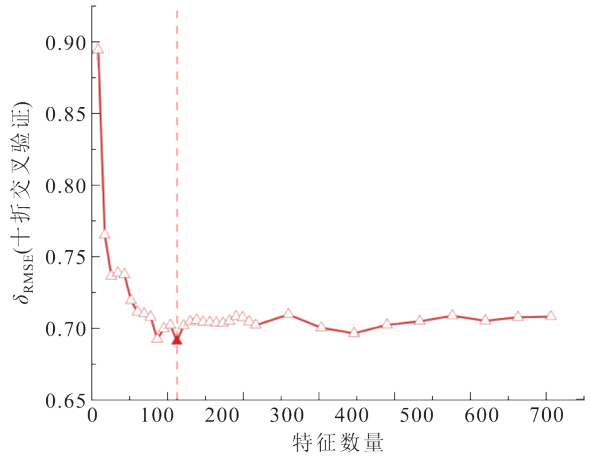


图 2　基于 RF-RFE 的分子描述符筛选过程(十折交叉验证)

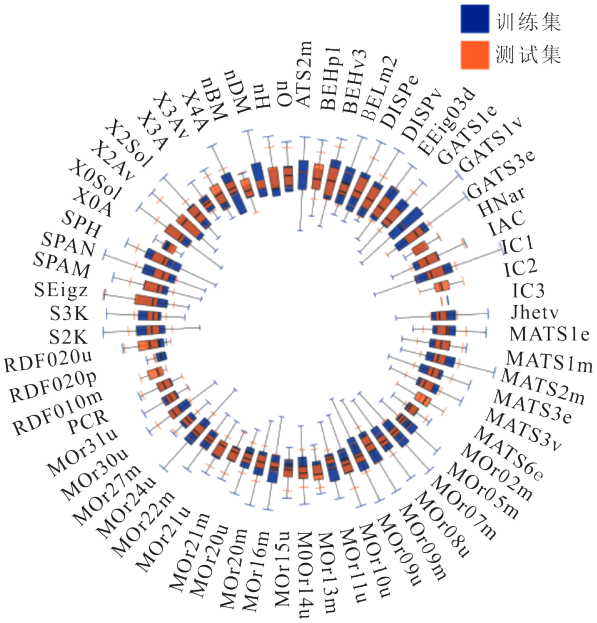
Fig.2　RF-RFE-based molecular descriptor screening process(ten-fold cross-validation)

在完成分子描述符的清洗以及训练集和测试集的划分后，我们进一步绘制了全部 130 种输入特征在训练集与测试集中的箱线图(如图 3 所示，图中分子描述符号定义见附录表 A1)，以比较两组数据在特征分布上的差异，并评估划分的合理性与代表性。从图中可以看出，大多数分子描述符在训练集与测试集之间的分布区间具有良好的一致性，分布范围、四分位区间及中位数位置整体一致，未出现系统性偏移，表明测试集样本在特征维度上覆盖了训练集的主要化学信息范围。这说明本文的数据划分策略在保证测试集独立性的同时，也较好地保留了其对整体化学空间的代表性，有助于模型性能的客观评估和泛化能力的验证。

2.2　QSAR 算法性能对比

在筛选出 130 个关键分子描述符并完成训练集与测试集划分后，构建了基于 RVM、SVM、RRF 和 XGBoost 的碳氟类气体急性吸入毒性 QSAR 回归模型。其中，RVM 与 SVM 分别结合 RBF 与 Poly 核函数开展对比分析，探讨核函数选择对模型性能的影响。模型超参数通过网格搜索法优化[42]，结合交叉验证确定最优配置(见表 1)。随后，依据内部验证与外部验证结果，对各模型性能进行系统评估(见表 2)，以全面检验其预测能力与稳健性。

如图 4 所示，从训练集的拟合效果来看，XGBoost 在捕捉数据规律方面表现最为出色(R^2_{TR}=0.979 2，δ_{RMSETR}=0.238 9)，其次是 SVM-RBF(R^2_{TR}=0.971 0，δ_{RMSETR}=0.282 0)，两者在训练阶段均表现出较高的精度。SVM-Poly 和



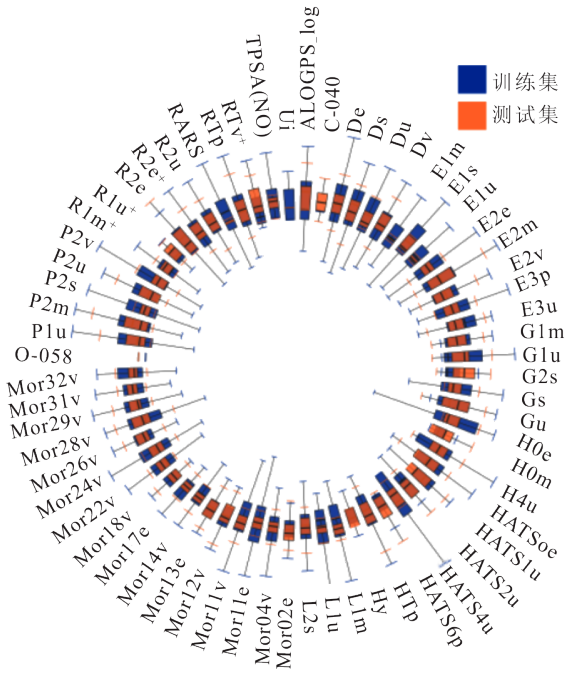


图 3　选定分子描述符在不同数据集中的分布情况

Fig.3　Distribution of selected molecular descriptors across data sets

RVM-Poly 的拟合能力也较为稳定，但相较略逊，而 RVM-RBF 与 RRF 在建模过程中对训练数据的拟合能力相对较弱。

进一步比较模型的泛化性能时，引入交叉验证决定系数(R^2_{CV5})、留一法交叉验证(Q^2_{LOO})并使用测试集决定系数(R^2_{TS})作为外部验证指标。结果显示，SVM-RBF 在多个指标上均优于其他算法(R^2_{CV5}=0.722 9，Q^2_{LOO}=0.744 6，R^2_{TS}=0.753 2)，具有较强的泛化能力和稳定性。XGBoost 虽然在训练拟

表 1 基于网格搜索的模型最佳超参数

Table 1 Optimal hyperparameters for models based on grid search

模型	最佳超参数
SVM-RBF	C=25.0, gamma=0.001
SVM-Poly	C=60.0, gamma=0.001
RVM-RBF	gamma=0.001, init_alpha=0.01
RVM-Poly	gamma= 0.002 5, init_alpha=0.01
XGBoost	learning_rate=0.02, max_depth= 3, n_estimators=300, colsample_bytree=0.5, min_child_weight=5, subsample= 0.5, gamma=0.01, reg_alpha= 0.7, reg_lambda= 0.1
RRF	max_depth=10, min_samples_leaf=1, min_samples_split=2, n_estimators=200, Step=0.01, threshold= 0.01

注：C 表示正则化系数；gamma 是核函数宽度参数；init_alpha 指初始学习率；learning_rate 控制学习速率；max_depth 限定最大树深度；n_estimators 设置基学习器数量；colsample_bytree 和 subsample 分别表示特征采样比例和样本采样比例；min_child_weight 规定叶节点最小权重和；reg_alpha 与 reg_lambda 对应 L1 和 L2 正则化系数；min_samples_leaf 和 min_samples_split 分别定义叶节点最小样本数和节点分裂最小样本数；Step 表示参数更新步长；threshold 为决策阈值。

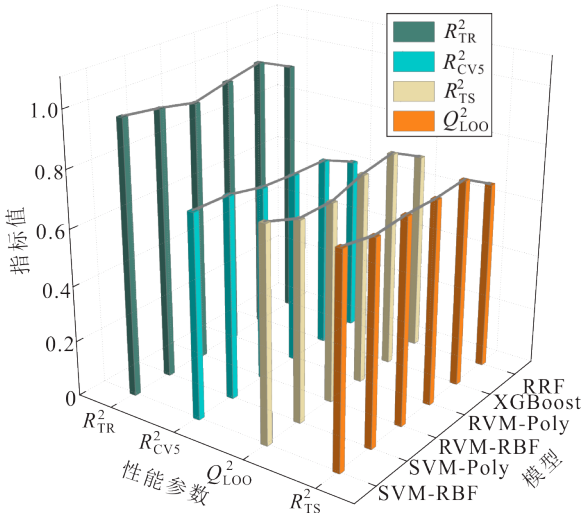


图 4 不同算法模型内外验证性能指标对比

Fig.4 Comparison of performance metrics for internal and external validation of different algorithmic models

合方面领先，但其泛化指标略低于 SVM-RBF (R^2_{CV5}=0.680 9，Q^2_{LOO}=0.735 0，R^2_{TS}=0.766 4)，仍展现出良好的外推能力。相比之下，SVM-Poly、RVM 系列和 RRF 在泛化表现上有所下降，适用范围相对受限。

同时，各模型的拟合系数与交叉验证系数之差 ($R^2_{TR}-R^2_{CV5}$)均小于 0.3，说明模型未出现明显过拟合，具备一定的稳健性。结合各项指标，SVM-RBF 在预测准确性和鲁棒性方面均衡发展，适合用于广泛的样本外推；而 XGBoost 则在拟合复杂模式方面表现突出，适用于特征维度较高的建模任务。因此 SVM-RBF 和 XGBoost 可以作为碳氟类气体生物安全性预测的核心算法，为后续研究提供重要参考。二者的预测响应图如图 5 所示，可为不同应用需求提供模型选择的依据。

表 2 模型性能对比

Table 2 Model performance comparison

模型	R^2_{TR}	R^2_{CV5}	δ_{RMSETR}	δ_{MAETR}	R^2_{TS}	δ_{RMSETS}	δ_{MAETS}	Q^2_{LOO}
SVM-RBF	0.971 0	0.722 9	0.282 0	0.206 7	0.753 2	0.714 1	0.576 1	0.744 6
SVM-Poly	0.951 6	0.719 5	0.303 5	0.218 6	0.706 3	0.778 9	0.639 5	0.719 5
RVM-RBF	0.924 2	0.691 1	0.456 1	0.383 0	0.706 6	0.778 6	0.665 0	0.728 5
RVM-Poly	0.953 7	0.683 0	0.356 6	0.298 2	0.745 7	0.724 8	0.572 3	0.726 0
XGBoost	0.979 2	0.680 9	0.238 9	0.171 1	0.766 4	0.694 7	0.603 8	0.735 0
RRF	0.922 6	0.624 5	0.402 1	0.330 2	0.702 7	0.783 7	0.615 2	0.666 3

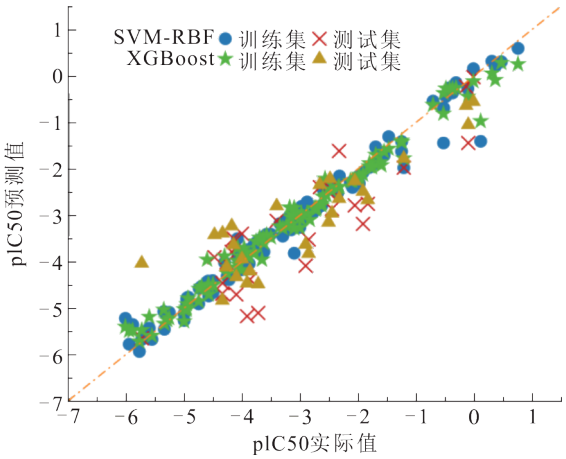


图 5 SVM-RBF、XGBoost 模型预测响应图

Fig.5 Response plot of SVM-RBF and XGBoost model prediction

为进一步验证 SVM-RBF 与 XGBoost 模型的稳定性与预测可靠性，我们利用 Williams 图对模型的应用域(applicability domain, AD)进行了分析，如图 6 所示。在分析中，标准化残差的阈值为±3，杠杆值阈值(h^*)为 2.932 8，分别用于识别预测异常值和高杠杆样本。结果表明，两种模型中均有 4 个样本位于 AD 范围之外，约占样本总数的 2.99%，其余 97.01%的样本均落在应用域内，说明模型整体具有较好的稳定性和适用性。

进一步分析异常分子发现，AD 之外的样本主要由于标准化残差较大所致，部分样本还具有结构特征明显偏离训练数据分布的特点，如异构体之间的毒性差异较大、极低毒性值或复杂的分子骨架结

构等，这些因素可能导致模型对其预测偏差较大。尽管存在少量异常样本，但整体来看，SVM-RBF 与 XGBoost 模型在大多数样本中的表现稳定，具备良好的泛化能力与实际应用潜力。

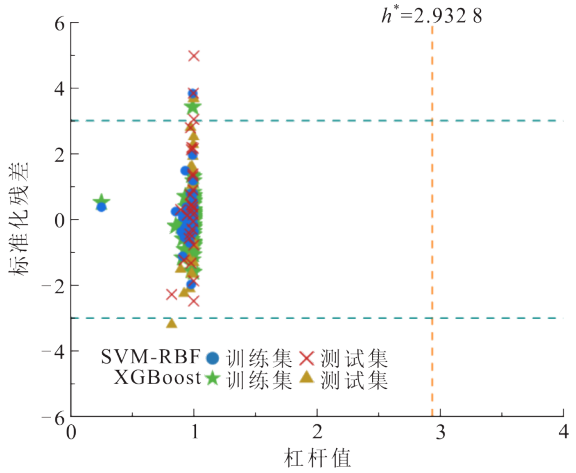


图 6　基于杠杆值与标准化残差的 Williams 模型应用域分析图

Fig.6　Analytical plot of the application domain of the Williams model based on leverage values and standardized residuals

3　SHAP 可解释性分析

3.1　QSAR 模型机理分析

为了进一步探讨最佳模型所揭示的毒性机理，对 SVM-RBF 和 XGBoost 两个模型进行了 SHAP 值分析，评估各分子描述符对模型输出的平均贡献。具体而言，计算了每个特征 SHAP 值绝对值的平均值作为其贡献度，并据此进行排序，结果如图 7 所示。在两个模型中，排名前 25 的特征贡献显著，表明它们在毒性预测中具有重要作用。

进一步比较显示，两种模型在前 25 个特征中均包含 6 个相同的描述符，且在各自模型中排名靠前，说明这些特征在不同算法中都展现出良好的稳定性和解释能力，是影响碳氟类气体毒性的关键因素。需要指出的是，这些共同特征涵盖了结构、电子及三维空间等多个分子属性维度，体现了模型从多角度揭示毒性机理的能力，增强了模型结果的稳健性和科学性。

这些描述符包括表示双键总数的 nDB(number of double bonds)，体现空间中原子电子密度相关性的三维结构描述符 Mor11e(第 11 阶 3D-MoRSE 描述符，按电负性加权)，衡量分子间范德华力分散作用能量的 DISPe(位移能量描述符)，反映异原子电子拓扑特性的 Jhetv(Balaban 型异原子拓扑指数)，描述局部原子属性分布的 HATS1u(未加权一阶 H-自相关描述符)，以及反映分子三维形状复杂度的拓扑结构参数 S3K(三阶 Shannon 熵描述符，基于 Kier 指数)。这些特征所涉及的结构、电子和空间信息共同揭示了毒性变化的潜在机制，也为理解碳氟类气体毒性的分子来源提供了有力支持。

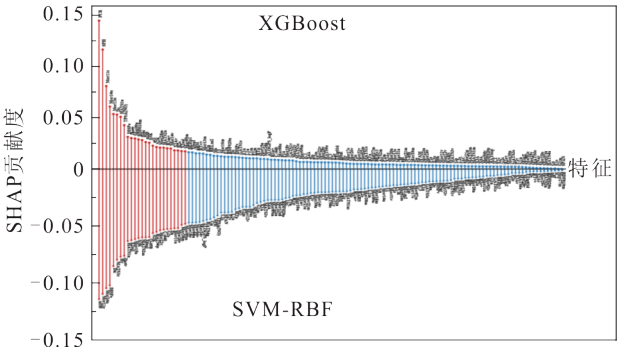
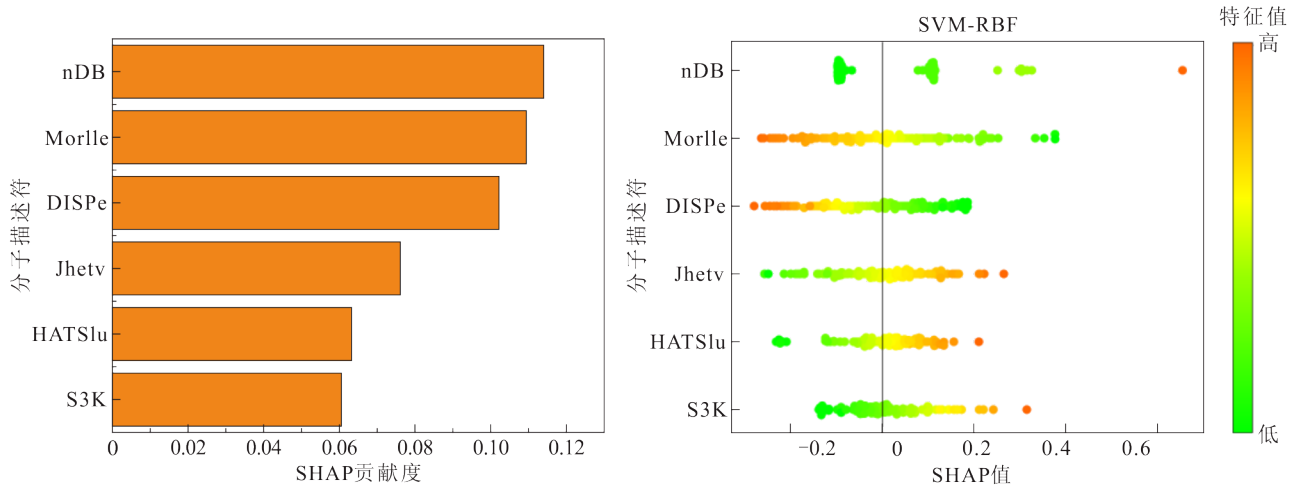


图 7　分子描述符在模型中 SHAP 贡献度排序

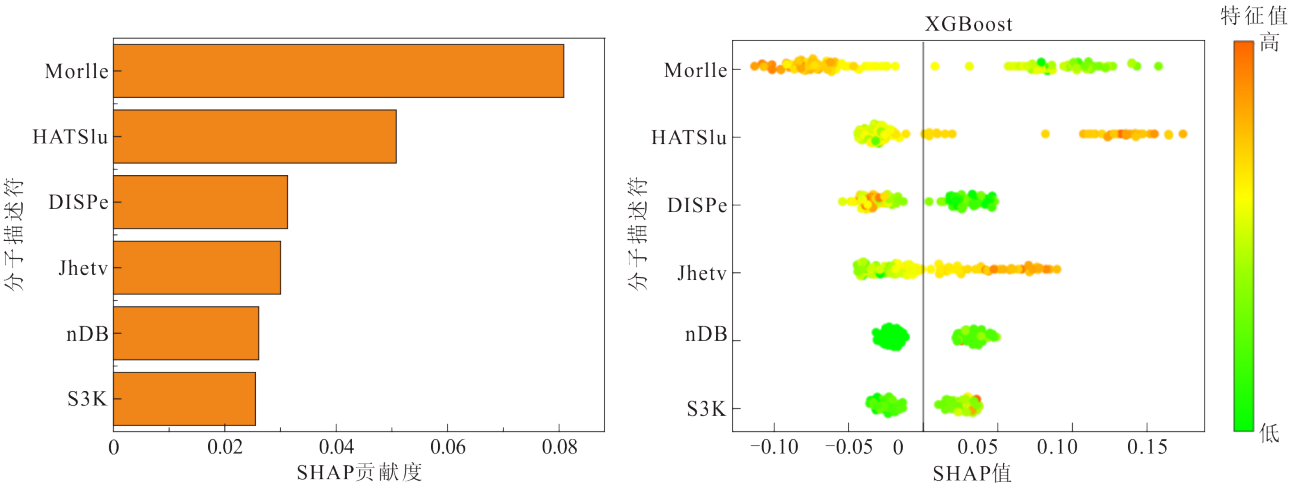
Fig.7　Molecular descriptors in the model SHAP contribution value ordering

为了深入理解关键分子描述符与毒性预测之间的关系，我们进一步结合 SHAP 蜂群图和特征贡献度条形图对 SVM-RBF 和 XGBoost 模型的解释结果进行了分析。如图 8 所示，SHAP 蜂群图揭示了各个描述符对模型预测的整体影响，点的颜色代表特征值大小，横向分布反映特征值变化对毒性的正负贡献。两个模型在 6 个核心描述符上表现出高度一致性，这些特征在 SHAP 值排序中也位列前列，如图 8 的条形图所示。

从蜂群图可见，nDB(分子中双键总数)与毒性呈显著正相关，在两个模型中均有较高的重要性(SVM-RBF 中为 0.114 14，XGBoost 中为 0.026 13)。双键的存在通常影响分子的化学反应性，使其更容易与生物靶点发生作用，从而增强毒性；这一点与双键在分子活性中的潜在作用机制相符[43-45]。Mor11e 与毒性呈负相关，且在两个模型中均为贡献度最大的特征之一(0.109 47 和 0.080 70)，说明电子密度分布和三维几何结构的稳定性在毒性降低中起重要作用。DISPe 同样表现出明显负相关性，SVM-RBF 和 XGBoost 模型中的贡献度分别为 0.102 19 和 0.031 21，其增大通常反映分子疏水性增强，从而减少与生物分子的相互作用。Jhetv 值较高时对应的 SHAP 值通常为正，表明其与毒性呈正相关，可能与氟等异原子分布密度及电子活性有关，在两个模型中也具有较强的解释力。HATS1u 和



(a) SVM-RBF模型关键分子描述符的SHAP贡献值排序图和SHAP蜂群图



(b) XGBoost模型关键分子描述符的SHAP贡献值排序图和SHAP蜂群图

图 8 重点分子描述符 SHAP 蜂群图及其贡献条形图

Fig.8 Focused molecular descriptor SHAP swarm map and its contribution bar graphs

S3K 也均呈现出与毒性增强相关的趋势，前者反映局部极性或电负性差异对生物相互作用的潜在影响，后者则表明复杂形状结构可能增强与靶标分子的匹配性。这些描述符在两个模型中虽存在重要性排序的细微差异，但整体趋势一致，体现出模型对关键结构-毒性关系的稳定识别能力。

3.2 模型预测过程

为了进一步验证模型的实际应用能力，我们对近年来新开发的碳氟类环保绝缘气体的吸入毒性进行了预测，并通过图 9 中的 SHAP 决策路径图以 XGBoost 模型为例，直观展示模型的预测过程与实际数据的符合程度，从而评估模型的准确性与可靠性。在所预测的分子中，C_4F_7N 属于训练集样本，而 CF_3SO_2F 为未出现在训练集中的新型环保气体。由于模型训练阶段采用了标准化处理，实际输出时已通过反向标准化还原为真实毒性值。

从预测结果来看，CF_3SO_2F 和 C_4F_7N 的 pLC50 预测值分别为−3.431 625 8 和−3.167 758 2。其中，C_4F_7N 的参考值为小鼠 4 h pLC50(−3.139 879~−3.070 037)，预测误差仅为 0.014 410~0.170 53，偏差较小，说明模型在该类已知物种数据下具备良好的拟合能力与可靠性[46]。相比之下，CF_3SO_2F 的参考值来自大鼠实验(−3.602 059~−3.531 478)，由于物种差异，小鼠通常对毒性更敏感，因此预测值略高于文献值也符合生理规律。因此，这两个典型分子的结果表明模型在不同数据来源下均表现出合理的趋势判断能力，验证了其预测的科学性[47]。SHAP 决策路径图进一步揭示了模型在预测过程中各个特征的贡献路径，使得模型决策过程更加透明可解释。

综合来看，SVM-RBF 和 XGBoost 模型不仅具备较强的拟合能力，也展现出良好的泛化性能，能够为现有碳氟类分子的生物安全性评估及新分子的

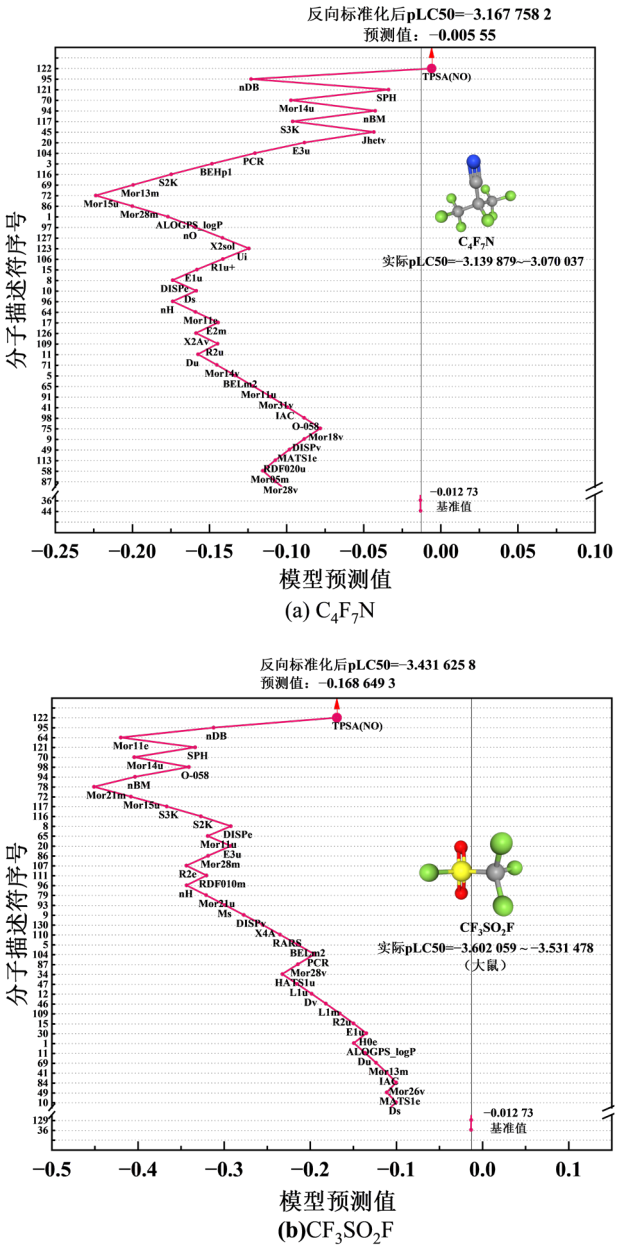


图 9　XGBoost 模型 SHAP 决策路径图

Fig.9　XGBoost model SHAP decision path map

预测筛选提供有效支持，对推动环保型绝缘气体的开发具有重要意义。

通过 SHAP 决策路径图对重点描述符的进一步分析发现，碳氟类气体分子的生物安全性主要受分子结构复杂性、极性疏水性特征、亲活性结构以及局部几何信息等多维因素的协同影响。在结构复杂性与拓扑特性方面，相关描述符如 S3K、IAC、X4A 等，体现了分子在空间上的分支状况、连接形式与整体拓扑稳定性。较为复杂的拓扑结构通常意味着更强的空间识别能力与靶点结合选择性，可能增强其毒性；而拓扑稳定性增强则有助于降低分子活性。

极性与疏水性相关描述符在图中也呈现出较高的重要性，例如 TPSA(NO)、ALOGPS_logP、X2sol 等，这些指标表征了分子的极性表面积、脂溶性、水溶性及其在生物体系中的分布倾向。极性与疏水性的平衡直接影响分子是否能够顺利穿过细胞膜、是否易于在靶点处富集，从而影响其潜在生物安全性。此外，nDB、nBM、E3u 等描述符反映了分子中亲活性结构单元的存在，如双键、芳香环、电子效应中心等，这些结构常为生物反应活性位点，易于参与亲电子或亲核反应，进而影响分子的代谢稳定性和生物安全性强度。

在几何与局部结构信息方面，Mor 系列中的 Mor28 v、Mor13 m，以及 RDF020u、HATS1u 等描述符揭示了分子在原子层面的空间排布、电荷分布特征以及局部结构特异性。这些细节性结构信息往往决定了分子与特定生物大分子间是否能形成精确的相互作用，是毒性机制中不可忽视的组成部分。此外，nO、nH 等原子组成类描述符所体现的官能团数量，也在一定程度上反映了分子的亲水性、代谢通路以及与受体结合的能力，对毒性表现形成间接但显著的影响。

通过 SHAP 全局解释分析与重点样本分析来看，碳氟类气体分子的毒性来源于结构与性质间的多维耦合机制。通过对决策路径图中关键描述符的分析，我们可以更明确地识别出提高生物安全性的潜在策略，如适当简化分子的拓扑结构、减少高反应活性基团(如双键和芳香环)、优化极性与疏水性比例，以及削弱局部高密度反应中心等方向。这些认识不仅为现有碳氟类绝缘气体的生物安全性评估提供理论基础，也为未来新型分子的绿色设计与筛选提供了科学支撑。

4　结论

1）通过比较多种机器学习算法的建模性能，最终构建了基于 SVM-RBF 和 XGBoost 的 QSAR 模型，并系统评估其预测能力与稳健性。SVM-RBF 模型的 R^2_{TR} =0.971 0，R^2_{CV5} =0.722 9，Q^2_{LOO} =0.744 6，R^2_{TS} =0.753 2；XGBoost 模型的 R^2_{TR} =0.979 2，R^2_{CV5} =0.680 9，Q^2_{LOO} =0.735 0，R^2_{TS} =0.766 4。两个模型的应用域覆盖率均达 97.01%，表明其在合理的化学空间内具备良好的外推性能和鲁棒性，可为新型碳氟类环保气体分子的生物安全性预测与筛选提供可靠支持。

2）基于 SHAP 方法对模型进行解释性分析，识别出多个对毒性具有显著贡献的关键分子描述符。其中，nDB、Jhetv、HATS1u、S3K 等与毒性呈正相关，Mor11e 与 DISPe 则与毒性呈负相关，揭示了这些结构特征在调控分子毒性过程中的潜在机制。

3）综合 SHAP 全局变量重要性排序与单分子毒性形成机制的深入解析发现，碳氟类气体分子的毒性来源于其分子结构复杂性、拓扑稳定性、极性疏水性特征、亲活性中心以及几何与局部结构信息等多种因素的耦合效应。具有较高拓扑复杂度或活性结构单元(如双键)往往提升分子与靶点的结合能力，从而降低生物安全性；而适当的极性与疏水性平衡以及几何结构的简化则有助于生物安全性的提升。因此，从分子设计角度出发，控制拓扑复杂性、减少活性位点、优化分子理化性质，是提升碳氟类环保绝缘气体生物安全性、实现绿色低毒分子筛选的可行策略。

附录见本刊网络版(http://hve.epri.sgcc.com.cn)。

参考文献 References

[1] SIMMONDS P G, RIGBY M, MANNING A J, et al. The increasing atmospheric burden of the greenhouse gas sulfur hexafluoride (SF_6)[J]. Atmospheric Chemistry and Physics, 2020, 20(12): 7271-7290.

[2] 陈圣，林莘，高克利，等. C_4F_7N/CO_2 混合气体在直流与工频电压下的绝缘特性[J]. 高电压技术，2026，52(1)：243-251.
Chen Shen, LIN Xin, GAO Keli, et al. Insulation characteristics of C_4F_7N/CO_2 gas mixtures under DC and power frequency voltages[J]. High Voltage Engineering, 2026, 52(1): 243-251.

[3] 唐炬，唐博文，陈钇江，等. 环保气体 C_4F_7N 绝缘设备内有害分解产物吸附材料研究进展[J]. 高电压技术，2025，51(8)：3876-3887.
TANG Jun, TANG Bowen, CHEN Yijiang, et al. Research progress on adsorbent materials for harmful by-products in environmental gas C_4F_7N insulated equipment[J]. High Voltage Engineering, 2025, 51(8): 3876-3887.

[4] 叶凡超，湛寡，田双双，等. $C_4F_7N/CO_2/O_2$ 混合气体 PD 作用下气体和固体副产物生成特性[J]. 高电压技术，2023，49(11)：4507-4516.
YE Fanchao, ZHAN Ao, TIAN Shuangshuang, et al. Gas and solid by-product generation properties of $C_4F_7N/CO_2/O_2$ gas mixture under partial discharge[J]. High Voltage Engineering, 2023, 49(11): 4507-4516.

[5] 张咪，高克利，侯华，等. SF_6 替代绝缘气体的虚拟筛选与分子设计综述[J]. 高电压技术，2023，49(7)：2816-2830.
ZHANG Mi. GAO Keli, HOU Hua, et al. Review on computational screening and molecular design of replacement gases for SF_6[J]. High Voltage Engineering, 2023, 49(7): 2816-2830.

[6] YU X J, HOU H, WANG B S. Prediction on dielectric strength and boiling point of gaseous molecules for replacement of SF_6[J]. Journal of Computational Chemistry, 2017, 38(10): 721-729.

[7] 侯华，余小娟，周文俊，等. 绝缘气体介电强度的构效关系[J]. 高等学校化学学报，2018，39(11)：2477-2484.
HOU Hua, YU Xiaojuan, ZHOU Wenjun, et al. Theoretical investigations on the structure-activity relationship to the dielectric strength of the insulation gases[J]. Chemical Journal of Chinese Universities, 2018, 39(11): 2477-2484.

[8] 刘关平，杨帅，张闹闹，等. 基于电子定域化函数表面分析的气体介质绝缘强度预测[J]. 高电压技术，2022，48(6)：2208-2214.
LIU Guanping, YANG Shuai, ZHANG Naonao, et al. Prediction on electrical strength of gaseous medium based on electron location function surface analysis[J]. High Voltage Engineering, 2022, 48(6): 2208-2214.

[9] OLIVET A, VEGA L F. Optimized molecular force field for sulfur hexafluoride simulations[J]. The Journal of Chemical Physics, 2007, 126(14): 144502.

[10] HOU H, WANG B S. An optimized force field for vapor-liquid equilibria and molecular dynamics simulations of eco-friendly dielectric fluid perfluoronitriles[J]. The Journal of Physical Chemistry B, 2021, 125(17): 4465-4475.

[11] BETOWSKI D, BEVINGTON C, ALLISON T C. Estimation of radiative efficiency of chemicals with potentially significant global warming potential[J]. Environmental Science & Technology, 2016, 50(2): 790-797.

[12] TROPSHA A, ISAYEV O, VARNEK A, et al. Integrating QSAR modelling and deep learning in drug discovery: the emergence of deep QSAR[J]. Nature Reviews Drug Discovery, 2024, 23(2): 141-155.

[13] CHERKASOV A, MURATOV E N, FOURCHES D, et al. QSAR modeling: where have you been? Where are you going to?[J]. Journal of Medicinal Chemistry, 2014, 57(12): 4977-5010.

[14] BRAND K P. Dielectric strength, boiling point and toxicity of gases-different aspects of the same basic molecular properties[J]. IEEE Transactions on Electrical Insulation, 1982, EI-17(5): 451-456.

[15] YU X J, HOU H, WANG B S. Structure-activity relationship and molecular design for discovery of environmentally sustainable dielectric gases[C]//2020 IEEE Electrical Insulation Conference (EIC). Knoxville, USA: IEEE, 2020: 124-127.

[16] CHEN S, SUN G H, FAN T J, et al. Ecotoxicological QSAR study of fused/non-fused polycyclic aromatic hydrocarbons (FNFPAHs): assessment and priority ranking of the acute toxicity to *Pimephales promelas* by QSAR and consensus modeling methods[J]. Science of the Total Environment, 2023, 876: 162736.

[17] LIU C Z, ZONG C, CHEN S, et al. Machine learning-driven QSAR models for predicting the cytotoxicity of five common microplastics[J]. Toxicology, 2024, 508: 153918.

[18] CHEN S, FAN T J, ZHANG N, et al. The oral acute toxicity of per-and polyfluoroalkyl compounds (PFASs) to *Rat* and *Mouse*: a mechanistic interpretation and prioritization analysis of untested PFASs by QSAR, q-RASAR and interspecies modelling methods[J]. Journal of Hazardous Materials, 2024, 480: 136071.

[19] KIM S, CHEN J, CHENG T J, et al. PubChem 2025 update[J]. Nucleic Acids Research, 2025, 53(D1): D1516-D1525.

[20] SCHNEIDER K, OLTMANNS J, HASSAUER M. Allometric principles for interspecies extrapolation in toxicological risk assessment-empirical investigations[J]. Regulatory Toxicology and Pharmacology, 2004, 39(3): 334-347.

[21] CAO Q M, YU J, CONNELL D. New allometric scaling relationships and applications for dose and toxicity extrapolation[J]. International Journal of Toxicology, 2014, 33(6): 482-489.

[22] NAIR A B, JACOB S. A simple practice guide for dose conversion between animals and human[J]. Journal of Basic and Clinical Pharmacy, 2016, 7(2): 27.

[23] BARBER C, HEGHES C, JOHNSTON L. A framework to support the application of the OECD guidance documents on (Q)SAR model validation and prediction assessment for regulatory decisions[J]. Computational Toxicology, 2024, 30: 100305.
[24] TETKO I V, GASTEIGER J, TODESCHINI R, et al. Virtual computational chemistry laboratory-design and description[J]. Journal of Computer-Aided Molecular Design, 2005, 19(6): 453-463.
[25] DEMIRCIOĞLU A. The effect of feature normalization methods in radiomics[J]. Insights into Imaging, 2024, 15(1): 2.
[26] SINGH D, SINGH B. Investigating the impact of data normalization on classification performance[J]. Applied Soft Computing, 2020, 97: 105524.
[27] WU Z X, ZHU M F, KANG Y, et al. Do we need different machine learning algorithms for QSAR modeling? A comprehensive assessment of 16 machine learning algorithms on 14 QSAR data sets[J]. Briefings in Bioinformatics, 2021, 22(4): bbaa321.
[28] TIPPING M E. Sparse Bayesian learning and the relevance vector machine[J]. Journal of Machine Learning Research, 2001, 1: 211-244.
[29] DARNAG R, MINAOUI B, FAKIR M. QSAR models for prediction study of HIV protease inhibitors using support vector machines, neural networks and multiple linear regression[J]. Arabian Journal of Chemistry, 2017, 10: S600-S608.
[30] SHERIDAN R P. Using random forest to model the domain applicability of another random forest model[J]. Journal of Chemical Information and Modeling, 2013, 53(11): 2837-2850.
[31] ÇELIK F K, DOĞAN S, KARADUMAN G. Drug-induced torsadogenicity prediction model: an explainable machine learning-driven quantitative structure-toxicity relationship approach[J]. Computers in Biology and Medicine, 2024, 182: 109209.
[32] CONSONNI V, BALLABIO D, TODESCHINI R. Comments on the definition of the Q^2 parameter for QSAR validation[J]. Journal of Chemical Information and Modeling, 2009, 49(7): 1669-1678.
[33] TROPSHA A. Best practices for QSAR model development, validation, and exploitation[J]. Molecular Informatics, 2010, 29(6/7): 476-488.
[34] WANG Y K, CHEN X B. A joint optimization QSAR model of fathead minnow acute toxicity based on a radial basis function neural network and its consensus modeling[J]. RSC Advances, 2020, 10(36): 21292-21308.
[35] DE P, KAR S, AMBURE P, et al. Prediction reliability of QSAR models: an overview of various validation tools[J]. Archives of Toxicology, 2022, 96(5): 1279-1295.
[36] HAMADACHE M, BENKORTBI O, HANINI S, et al. A quantitative structure activity relationship for acute oral toxicity of pesticides on rats: validation, domain of application and prediction[J]. Journal of Hazardous Materials, 2016, 303: 28-40.
[37] LI Z Q. Extracting spatial effects from machine learning model using local interpretation method: an example of SHAP and XGBoost[J]. Computers, Environment and Urban Systems, 2022, 96: 101845.
[38] ZHANG J Y, MA X L, ZHANG J L, et al. Insights into geospatial heterogeneity of landslide susceptibility based on the SHAP-XGBoost model[J]. Journal of Environmental Management, 2023, 332: 117357.
[39] HUANG J H, WEN H J, HU J W, et al. Deciphering decision-making mechanisms for the susceptibility of different slope geohazards: a case study on a SMOTE-RF-SHAP hybrid model[J]. Journal of Rock Mechanics and Geotechnical Engineering, 2025, 17(3): 1612-1630.
[40] VAN DEN BROECK G, LYKOV A, SCHLEICH M, et al. On the tractability of SHAP explanations[J]. Journal of Artificial Intelligence Research, 2022, 74: 851-886.
[41] ZHOU Z H. Open-environment machine learning[J]. National Science Review, 2022, 9(8): nwac123.
[42] RADZI S F M, KARIM M K A, SARIPAN M I, et al. Hyperparameter tuning and pipeline optimization via grid search method and tree-based AutoML in breast cancer prediction[J]. Journal of Personalized Medicine, 2021, 11(10): 978.
[43] ALABUGIN I V. Stereoelectronic effects: a bridge between structure and reactivity[M]. Hoboken, USA: John Wiley & Sons, 2016.
[44] ZHENG Y Z, DENG G, CHEN D F, et al. The influence of C2=C3 double bond on the antiradical activity of flavonoid: different mechanisms analysis[J]. Phytochemistry, 2019, 157: 1-7.
[45] LI M M, GUAN E Q, BIAN K. Structure elucidation and toxicity analysis of the degradation products of deoxynivalenol by gaseous ozone[J]. Toxins, 2019, 11(8): 474.
[46] ZHANG X X, YE F C, LI Y, et al. Acute toxicity and health effect of perfluoroisobutyronitrile on mice: a promising substitute gas-insulating medium to SF_6[J]. Journal of Environmental Science and Health, Part A, 2020, 55(14): 1646-1658.
[47] ZHANG L, PENG R C, HUANG Y J, et al. Toxic study on the new eco-friendly insulating gas trifluoromethanesulfonyl fluoride: a substitute for SF_6[J]. Sustainability, 2022, 14(9): 5239.



叶凡超
1993—，男，博士，讲师，硕导
主要从事高压电气设备绝缘状态监测与评估、新型环保绝缘气体等研究工作
E-mail: yefanchao@hbut.edu.cn

YE Fanchao
Ph.D.



肖　峰
2002—，男，硕士生
主要从事新型环保绝缘气体的研究工作
E-mail: xiaofeng49712811@163.com

XIAO Feng



张晓星(通信作者)
1974—，男，博士，教授，博导
主要从事高压电气设备绝缘状态监测与评估、新型传感技术和环保绝缘气体等研究工作
E-mail: xiaoxing.zhang@outlook.com

ZHANG Xiaoxing
Ph.D., Professor
Corresponding author

收稿日期　2025-07-02　在线出版日期　2025-08-04　　编辑　曹昭君

附录 A

表 A1　E-Drogon 平台分子描述符定义

Table A1　Definition of molecular descriptors as calculated by the E-Dragon platform

序号	分子描述符	定义	序号	分子描述符	定义
1	ALOGPS_logP	ALOGPS 预测的辛醇/水分配系数(logP)	66	Mor11v	MoRSE 11(体积加权)
2	ATS2m	滞后 2 的 Broto-Moreau 自相关(质量加权)	67	Mor12v	MoRSE 12(体积加权)
3	BEHp1	最高路径阶 1 的 Burden 特征值(极化率加权)	68	Mor13e	MoRSE 13(电负性加权)
4	BEHv3	最高路径阶 3 的 Burden 特征值(范德华体积加权)	69	Mor13m	MoRSE 13(质量加权)
5	BELm2	最低路径阶 2 的 Burden 特征值(质量加权)	70	Mor14u	MoRSE 14(未加权)
6	C-040	C 原子特定拓扑片段计数	71	Mor14v	MoRSE 14(体积加权)
7	De	电负性距离指数	72	Mor15u	MoRSE 15(未加权)
8	DISPe	电负性离散度	73	Mor16m	MoRSE 16(质量加权)
9	DISPv	范德华体积离散度	74	Mor17e	MoRSE 17(电负性加权)
10	Ds	电拓扑状态离散度	75	Mor18v	MoRSE 18(体积加权)
11	Du	未加权离散度指数	76	Mor20m	MoRSE 20(质量加权)
12	Dv	体积离散度指数	77	Mor20u	MoRSE 20(未加权)
13	E1m	第一 WHIM 轴(质量加权)	78	Mor21m	MoRSE 21(质量加权)
14	E1s	第一 WHIM 轴(电状态加权)	79	Mor21u	MoRSE 21(未加权)
15	E1u	第一 WHIM 轴(未加权)	80	Mor22m	MoRSE 22(质量加权)
16	E2e	第二 WHIM 轴(电负性加权)	81	Mor22v	MoRSE 22(体积加权)
17	E2m	第二 WHIM 轴(质量加权)	82	Mor24u	MoRSE 24(未加权)
18	E2v	第二 WHIM 轴(体积加权)	83	Mor24v	MoRSE 24(体积加权)
19	E3p	第三 WHIM 轴(极化率加权)	84	Mor26v	MoRSE 26(体积加权)
20	E3u	第三 WHIM 轴(未加权)	85	Mor27m	MoRSE 27(质量加权)
21	EEig03d	第 3 个边邻接特征值(距离加权)	86	Mor28m	MoRSE 28(质量加权)
22	G1m	第一全局 WHIM 对称性(质量加权)	87	Mor28v	MoRSE 28(体积加权)
23	G1u	第一全局 WHIM 对称性(未加权)	88	Mor29v	MoRSE 29(体积加权)
24	G2s	第二全局 WHIM 对称性(电状态加权)	89	Mor30u	MoRSE 30(未加权)
25	GATS1e	滞后 1 Geary 自相关(电负性加权)	90	Mor31u	MoRSE 31(未加权)
26	GATS1v	滞后 1 Geary 自相关(体积加权)	91	Mor31v	MoRSE 31(体积加权)
27	GATS3e	滞后 3 Geary 自相关(电负性加权)	92	Mor32v	MoRSE 32(体积加权)
28	Gs	全局对称性指数(电状态加权)	93	Ms	平均电拓扑状态
29	Gu	全局对称性指数(未加权)	94	nBM	多重键数
30	H0e	滞后 0 GETAWAY 指数(电负性加权)	95	nDB	双键数
31	H0m	滞后 0 GETAWAY 指数(质量加权)	96	nH	氢原子数
32	H4u	滞后 4 GETAWAY 指数(未加权)	97	nO	氧原子数
33	HATS0e	滞后 0 HATS 自相关(电负性加权)	98	O-058	O 原子特定拓扑片段计数
34	HATS1u	滞后 1 HATS 自相关(未加权)	99	P1u	第一主成分方向指数(未加权)
35	HATS2u	滞后 2 HATS 自相关(未加权)	100	P2m	第二主成分方向指数(质量加权)
36	HATS4u	滞后 4 HATS 自相关(未加权)	101	P2s	第二主成分方向指数(电状态加权)
37	HATS6p	滞后 6 HATS 自相关(极化率加权)	102	P2u	第二主成分方向指数(未加权)
38	HNar	芳香原子数	103	P2v	第二主成分方向指数(体积加权)
39	HTp	总 GETAWAY 指数(极化率加权)	104	PCR	主成分回归指数
40	Hy	疏水性指数	105	R1m+	R 型 WHIM 指数 1(质量加权)
41	IAC	信息含量指数	106	R1u+	R 型 WHIM 指数 1(未加权)
42	IC1	1 阶信息含量	107	R2e	R 型 WHIM 指数 2(电负性加权)
43	IC2	2 阶信息含量	108	R2e+	R 型 WHIM 指数 2 增强版(电负性加权)
44	IC3	3 阶信息含量	109	R2u	R 型 WHIM 指数 2(未加权)
45	Jhetv	异原子 Balaban 指数(体积加权)	110	RARS	相对芳香性比表面积
46	L1m	第一 WHIM 轴长度(质量加权)	111	RDF010m	RDF 1.0Å(质量加权)
47	L1u	第一 WHIM 轴长度(未加权)	112	RDF020p	RDF 2.0Å(极化率加权)
48	L2s	第二 WHIM 轴长度(电状态加权)	113	RDF020u	RDF 2.0Å(未加权)
49	MATS1e	滞后 1 Moran 自相关(电负性加权)	114	RTp	R 型总指数(极化率加权)
50	MATS1m	滞后 1 Moran 自相关(质量加权)	115	RTv+	R 型总指数增强版(体积加权)
51	MATS2m	滞后 2 Moran 自相关(质量加权)	116	S2K	二阶 Kier 形状指数

续表 A1

序号	分子描述符	定义	序号	分子描述符	定义
52	MATS3e	滞后 3 Moran 自相关(电负性加权)	117	S3K	三阶 Kier 形状指数
53	MATS3v	滞后 3 Moran 自相关(体积加权)	118	SEigZ	Z 加权谱特征值
54	MATS6e	滞后 6 Moran 自相关(电负性加权)	119	SPAM	分子平均跨度
55	Mor02e	MoRSE 02(电负性加权)	120	SPAN	分子最大跨度
56	Mor02m	MoRSE 02(质量加权)	121	SPH	球形度指数
57	Mor04v	MoRSE 04(体积加权)	122	TPSA(NO)	N 和 O 贡献的拓扑极性表面积
58	Mor05m	MoRSE 05(质量加权)	123	Ui	不饱和指数
59	Mor07m	MoRSE 07(质量加权)	124	X0A	0 阶连接指数(平均型)
60	Mor08u	MoRSE 08(未加权)	125	X0sol	0 阶溶液连接指数
61	Mor09m	MoRSE 09(质量加权)	126	X2Av	2 阶平均连接指数(体积加权)
62	Mor09u	MoRSE 09(未加权)	127	X2sol	2 阶溶液连接指数
63	Mor10u	MoRSE 10(未加权)	128	X3A	3 阶连接指数(平均型)
64	Mor11e	MoRSE 11(电负性加权)	129	X3Av	3 阶平均连接指数(体积加权)
65	Mor11u	MoRSE 11(未加权)	130	X4A	4 阶连接指数(平均型)