DOI:10.11918/202510014

深度表示学习在环境污染物识别中的应用

张家彬,王 威,陈 川

(城乡水资源与水环境全国重点实验室(哈尔滨工业大学),哈尔滨 150090)

摘 要:为应对环境污染物检测成本高昂、依赖动物实验的问题,结构-活性关系(QSAR)模型已广泛用于环境中典型污染物(如持久性有机污染物、内分泌干扰物等)的分子活性与生态毒性预测。然而,传统QSAR方法在识别复杂环境样品中的特征污染物时,常受到"活性悬崖"与小样本数据稀缺的问题,导致模型泛化能力不足。针对这些问题,系统梳理了深度学习在环境分子建模中主流的训练范式,包括迁移学习、预训练-微调、自监督学习等策略,探讨其在特征提取、模型稳健性及跨任务泛化方面的理论基础与技术实现。文献分析与方法比较显示,这些训练范式在不同类型的自建环境数据集中,相较于传统机器学习方法,在小样本条件下展现出更强的泛化能力、跨任务适应性以及对未知化合物结构特征的识别优势。进一步提出了适用于环境污染物建模与风险评估的分子建模流程,并总结了模型评估体系的关键要点,为构建高泛化性、可解释的环境深度学习模型提供了系统化路径。该研究可为环境领域的模型开发提供结构化的技术指引与模型对比参考框架,促进对不同深度学习训练策略的系统性评估,为行业建立更加统一、可复现的智能建模标准提供支持,从而推动环境污染物筛查与风险评估向数据驱动、智能化方向发展。

关键词:环境污染物识别;深度表示学习;分子性质预测;结构-活性关系建模;分子生成
中图分类号:X52 **文献标志码**:A **文章编号**:0367-6234(2026)06-0170-15

Application of deep representation learning in environmental pollutant identification

ZHANG Jiabin, WANG Wei, CHEN Chuan

(State Key Laboratory of Urban-rural Water Resources and Environment (Harbin Institute of Technology), Harbin 150090, China)

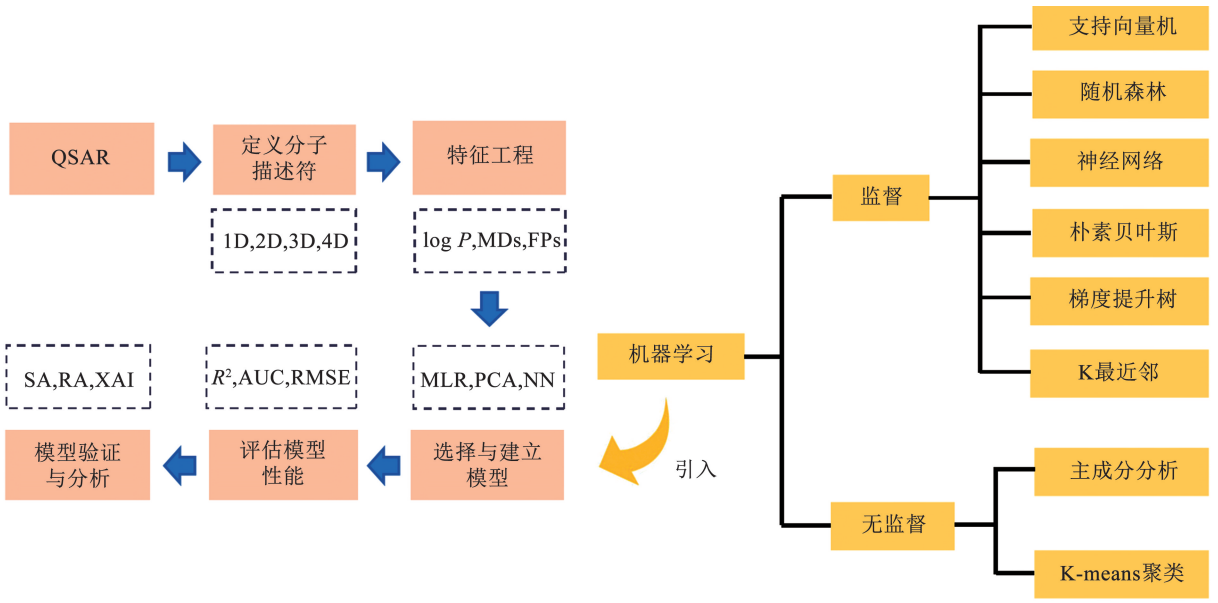
Abstract: To address the high costs of environmental pollutant detection and the reliance on animal experiments, quantitative structure-activity relationship (QSAR) models have been widely used in predicting the molecular activity and ecotoxicity of typical environmental pollutants (such as persistent organic pollutants and endocrine-disrupting chemicals). However, traditional QSAR methods often suffer from the "activity cliff" and the scarcity of small-sample data when identifying characteristic pollutants in complex environmental samples, resulting in insufficient generalization capability of the models. In view of these issues, this paper systematically reviewed the mainstream training paradigms of deep learning in environmental molecular modeling, including strategies such as transfer learning, pre-training and fine-tuning, and self-supervised learning, and it discussed their theoretical foundations and technical implementations in feature extraction, model robustness, and cross-task generalization. Literature analysis and method comparison show that compared with traditional machine learning methods, these training paradigms demonstrate stronger generalization capability, cross-task adaptability, and identification advantages for structural features of unknown compounds under small-sample conditions across different types of self-constructed environmental datasets. This paper further proposes a molecular modeling workflow suitable for environmental pollutant modeling and risk assessment and summarizes the key points of the model evaluation system, providing a systematic pathway for constructing highly generalizable and interpretable environmental deep learning models. This research can provide structured technical guidelines for model development in the environmental field and a reference framework for model comparison, promote the systematic evaluation of different deep learning training strategies, and offer support for the industry to establish more unified and reproducible intelligent modeling standards, thereby promoting the development of environmental pollutant screening and risk assessment toward data-driven and intelligent directions.

Keywords: environmental pollutant identification; deep representation learning; molecular property prediction; quantitative structure-activity relationship modeling; molecular generation

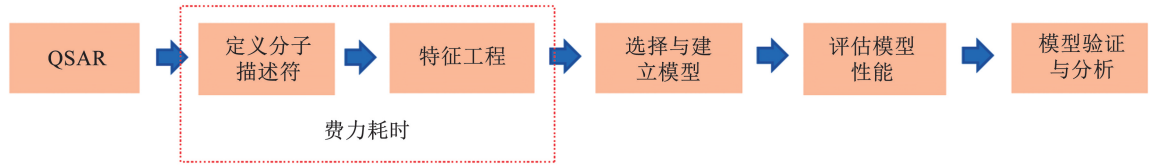
收稿日期:2025-10-11;**录用日期**:2025-11-25;**网络首发日期**:2026-01-06
网络首发地址:https://link.cnki.net/urlid/23.1235.T.20260106.1117.003
基金项目:国家自然科学基金(52400025);中国博士后科学基金面上项目(2024M754204);国家污泥安全处置与资源化工程研究中心科学基金(Z2024A001)
作者简介:张家彬(1996—),男,博士研究生;陈 川(1982—),男,教授,博士生导师
通信作者:陈 川,echo110244@126.com

随着环境污染问题的加剧，亟需高效识别和评估环境样品中的特征污染物，以支撑污染防控与生态风险管理。传统检测与毒性实验方法虽精确，但成本高、周期长、依赖动物实验，难以满足大规模筛查需求。近年来，基于数据驱动的分子性质预测方法迅速发展，逐渐成为传统实验的重要补充[1]，其通过构建分子指纹（Fingerprints，FP）或描述符（Molecular Descriptors，MD），建立结构与性质的映射关系[2]，无需湿实验即可预测新分子的环境行为特征，为污染物筛选和风险评估提供了新工具。该方向在分子科学中被称为定量构效关系（Quantitative Structure-Activity Relationship，QSAR）。自 1962 年由 Hansch 等[3]提出以来，QSAR 已历经从线性模型到集成学习方法的发展演进。随着机器学习技术的发展，随机森林（Random Forest，RF）、支持向量机（Support Vector Machine，SVM）、极端梯度提升（eXtreme Gradient Boosting，XGBoost）等方法被广泛引入后，显著提升了 QSAR 模型对非线性关系的拟合能力[4]，见图 1（a）。然而，在面对分子数据稀缺、污染物种类多样、先验知识有限等问题时，传统模型在特征选择、模型泛化等方面仍受制约，见图 1（b）。

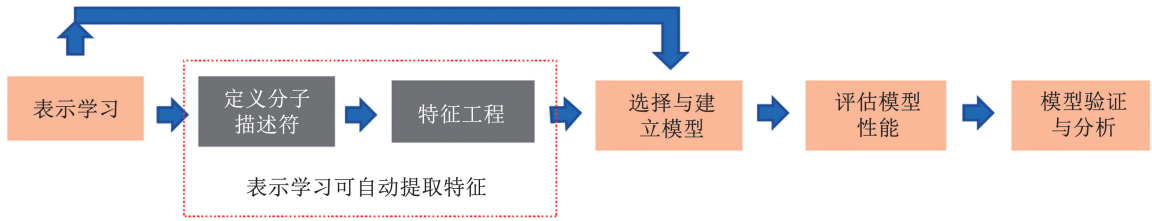
为解决这一问题，深度表示学习方法及其高阶的训练范式逐渐受到关注，其具备自动学习高层次特征的能力，可摆脱人工特征工程的限制，推动端到端建模方法的发展[5]。在这一背景下，预训练与迁移学习等范式已在分子科学、药物发现等领域取得重要进展[6]，正逐步应用于环境污染物识别、小样本毒性预测等场景中，见图 1（c）。



(a) 融合机器学习的QSAR建模流程



(b) QSAR模型构建的典型流程



(c) 基于深度表示学习的QSAR建模新路径

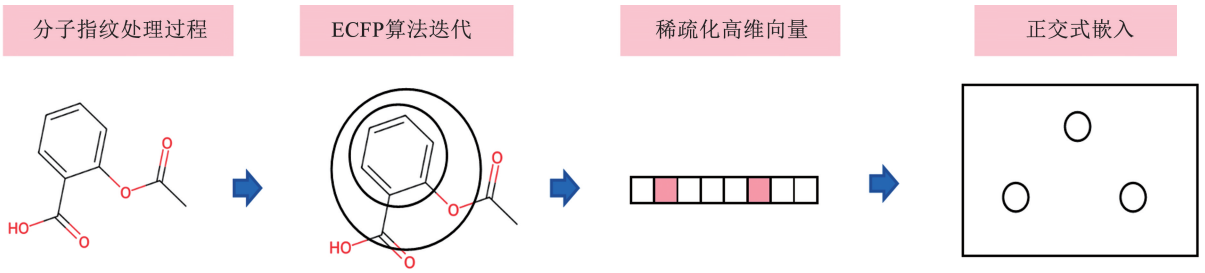
图 1　传统 QSAR 建模流程与深度表示学习建模流程的区别

Fig. 1　Differences between traditional QSAR modeling workflow and deep representation learning modeling workflow

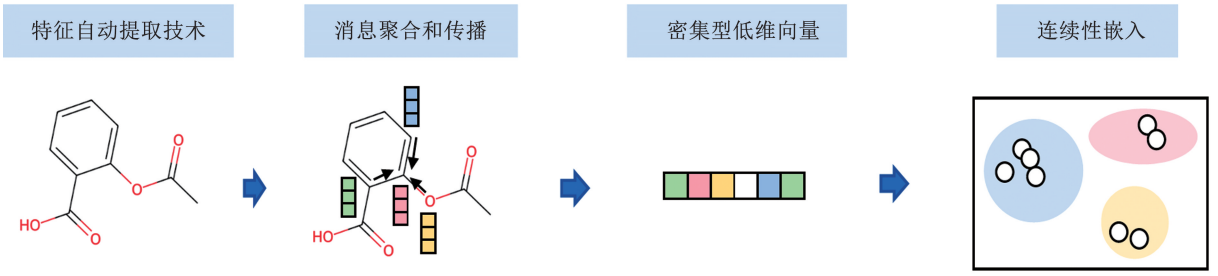
本研究简要回顾了特征工程与深度学习在化学分子建模中的应用进展与面临的挑战，重点介绍了以表示学习为核心的训练范式及其在环境特征污染物活性预测与分子生成中的典型实例，指出自监督学习、元学习等策略在小样本数据集场景下具有重要的泛化潜力。现有研究表明，这些方法在公共化学数据集及部分环境数据集中的应用表现出更优的泛化能力与模型稳健性，相较传统机器学习方法具有显著优势。另一方面，当前环境建模研究仍缺乏统一的模型构建与评估流程，不同研究间在模型架构与训练策略上的差异使横向比较与经验积累受到限制。针对这一问题，提出了一套系统化、结构化的建模流程框架，旨在为研究者提供可复用的建模路径，促进不同深度学习策略在多源环境数据中的有效迁移与比较，进一步提升模型开发效率与研究可复现性，为环境污染物筛查与风险评估的智能化发展提供理论支撑与方法参考。

1 特征工程与经典机器学习

基于经典机器学习的方法，如 RF、SVM、朴素贝叶斯（Naive Bayes，NB）、K 近邻（K-Nearest Neighbor，KNN）、决策树（Decision Trees，DT）以及 XGBoost 等模型，已在环境科学领域广泛应用于分子性质的预测任务，如毒性评估与可生物降解性预测等[7-8]。在数据预处理阶段，分子指纹与分子描述符因兼具良好的预测性能与一定的可解释性，在建模实践中广泛采用[9]。然而，此类分子表示方法在分子结构信息的提取方面存在一定局限[10]，难以充分刻画分子片段间的化学相似性，如图 2，进而在面对“活性悬崖”现象时[11]，可能限制模型的学习能力与泛化性能。此外，不同类型的描述符侧重提取分子的不同特征，尚无一种统一的表示方法可适用于所有预测任务。因此，往往需要投入大量精力对不同描述符进行评估、筛选，或尝试多种组合策略，以期获得更优的建模效果。



(a)基于分子指纹的特征构建与向量嵌入流程



(b) 基于特征自动提取技术的特征构建与向量嵌入流程

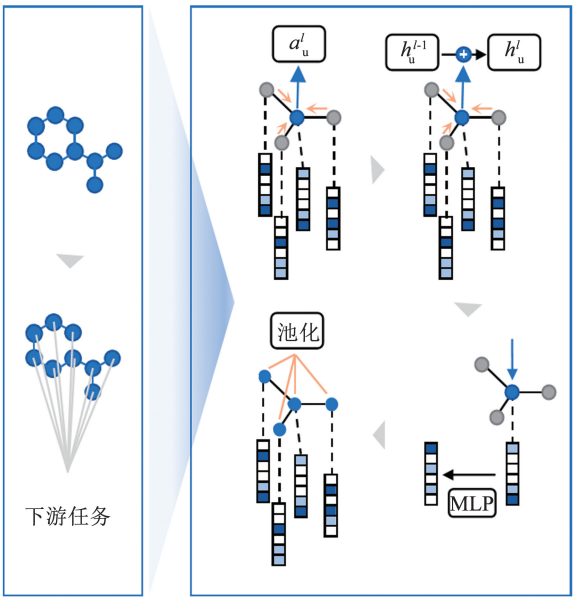
图 2 传统指纹方法与自动特征提取在分子表示中的应用对比

Fig. 2 Comparison of application between traditional fingerprint methods and automatic feature extraction in molecular representation

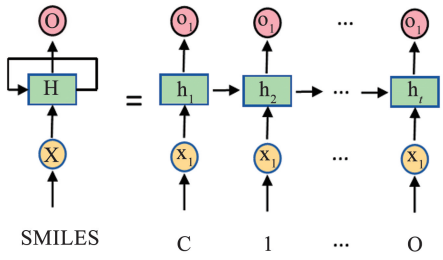
2 深度表示学习

相比之下，自动特征提取的编码器技术如图神经网络[12-13]、大语言模型[14]，变分自编码器[15]等因其直接从原始数据中学习紧凑且具有表达力的表示，无需手动选择或比较分子描述符，增强了下游模型的学习效果而成为新的一代技术。其中，图神经网络（Graph Neural Network，GNN）通过分子图的邻接矩阵不断更新并逐步聚合邻居节点的信息提取分子结构特征[16]；时序模型如 RNN 或大语言模型（Large Language Model，LLM）则对 SMILES 字符串进行 token 级编码，初始化词嵌入和位置编码后经过 Transformer 架构映射至高维空间[17]，并用于后续任务如分子性质预测[18-19]；变分自编码器

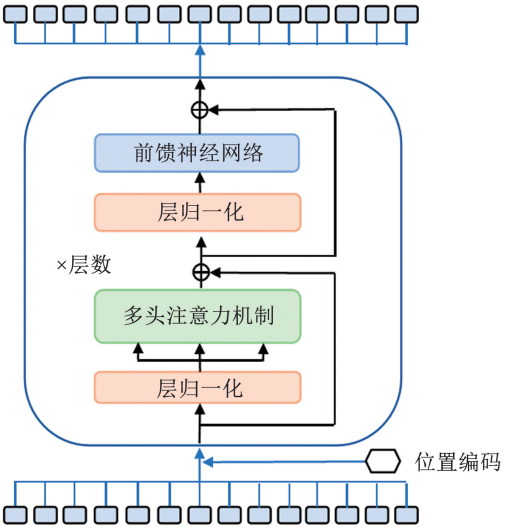
(Variational Autoencoder，VAE)常用于分子生成领域[20]，是学习连续、可控的潜在空间的代表性模型，适合需要从潜在空间采样、插值、优化的任务，如图 3。



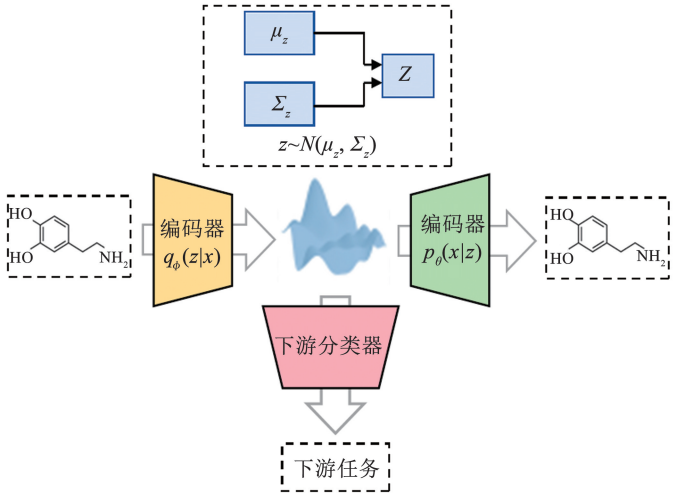
(a) 图神经网络GNN的学习过程



(b) 循环神经网络RNN的学习过程



(c) Transformer架构的学习过程



(d) 变分自编码器VAE的学习过程

图 3　深度表示学习技术的建模过程

Fig. 3　Modeling process of deep representation learning techniques

由不同编码器学习得到的分子表示通常表现为连续的高维稠密向量，尽管这些表示难以被人类直观解释，但其信息表达能力强，在数据充足的情况下能够有效捕捉分子结构与性质之间的复杂关系。已有研究表明，GNN 在多个与生物活性相关的分子性质预测任务中，相较于传统的人工设计分子描述符，表现出更高的预测精度[21]。

随后，数据科学家和人工智能研究人员基于这些深度学习的基础模型，在训练策略层面上优化创新，发展出了一系列高阶训练方法进一步增强了分子嵌入表示[22]，以进一步提高模型的性能。

3　高阶学习策略用于分子性质预测和生成效应识别方法

3.1　迁移学习

迁移学习是一种通过在源域 $D_s=\{X_s,P_s(X)\}$ 上预训练模型，并将所学知识迁移至目标域 $D_t=\{X_t,P_t(X)\}$ 以提升学习效率的方法，其优化目标可形式化为

$$\min_{\theta_T\leftarrow\theta_S} L_T(f_T(X_T;\theta_T),Y_T)$$

迁移学习又被称为预训练-微调范式[23]。尤其当目标域数据有限时，直接从零开始训练可能难以

获得泛化能力强的模型,因此,需要利用源任务获取相应的先验知识。利用该策略,迁移学习能够有效缓解小样本学习中的困难[24],如图 4(a)。

迁移学习的无监督部分不依赖人工标签,而是从数据本身构造训练任务[25],包括掩码学习、对比学习、重构学习等多种预训练技术。目前,迁移学习范式广泛地应用于环境分子性质预测[26-27]和药物发现领域[28-30]。如 Feinstein 等[31]开发了一种迁移学习模型,通过从公共数据集中检索大鼠口服实验测得的 LD50 的所有有机化合物,输入进 SelectiveNet 架构中进行训练,以预测 PFAS 的毒性。Li 等[32]从 ChEMBL 收集了一百万的分子,通过进行分子结构的自监督预训练,然后在 4 个下游数据集上进行微调,性能超过了经典的机器学习模型。Amabilino 等[33]提出先用 RNN 在大量分子上进行预训练,以学习 SMILES 语法,在此基础上进行迁移学习时,即使只使用 190 个分子的小数据集,也能有效生成分子。Morozov 等[34]利用从 UniProt 获得的 2 475 个有毒和 214 740 个无毒序列,对 ProteinBERT 模型进行了微调,仅基于氨基酸序列来识别有毒蛋白和肽。该模型在测试数据集上的 MCC 为 0.64,AUC 为 0.86,优于 CSM-Toxin 和 ToxinPred2。

3.2 强化学习

强化学习(RL)是一种通过与环境不断交互来学习策略的机器学习方法[35]。在该框架下,智能体(Agent)根据当前状态采取动作,并根据环境给予的奖励来调整其行为策略,以最大化长期累积收益,如图 4(b)。长期累积收益可表示为

$$J(\pi) = E_{\pi}\left[\sum_{t=0}^{\infty} r^{t} R(s_t, a_t)\right]$$

式中:π 表示策略,r 为折扣因子。

在分子科学领域,强化学习被广泛应用于基于 SMILES 或分子三维结构的生成模型,通过将目标性质,如药效、毒性、合成可行性等设计为奖励函数,引导模型生成符合要求的新化合物[36]。与传统的随机生成或规则驱动方法相比,RL 能够更高效地在化学空间中搜索高质量分子[37]。

强化学习的核心在于评价当前策略的优劣,即衡量当前状态下采取不同动作所带来的累积奖励期望。状态-值函数定义为

$$V^{\pi}(s) = E_{\pi}\left[\sum_{t=0}^{\infty} r^{t} R(s_t, a_t) \mid s_0 = s\right]$$

动作-值函数(Q 函数)定义为

$$Q^{\pi}(s,a) = E_{\pi}\left[\sum_{t=0}^{\infty} r^{t} R(s_t, a_t) \mid s_0 = s, a_0 = a\right]$$

在某一状态 s 处,选择不同的动作 a 会导致不同的累积奖励衡量,因此,策略优化通常通过 Q 函数来实现。如 Olivecrona 等[38]使用基于策略的深度强化学习调整序列生成模型 RNN,该方法可生成查询结构的类似物和生成对生物靶标有活性的化合物,生成目标化合物的概率可达到 95%。Zhou 等[39]提出 MolDQN 以实现分子的目标性质优化,该框架把分子修饰定义为一个马尔可夫决策过程,采用深度 Q-learning 算法寻找 Q 函数,可保证 100% 的化学有效性。在基准分子优化任务中,MolDQN 与其他几个最近发表的算法相比,实现了相当或更好的性能。Mazuz 等[40]基于 Transformer 架构开发出 Taiga 两段式模型,以预测下一个 token 为第一阶段训练,再使用策略梯度算法优化分子的 QED 性质。

3.3 元学习

元学习是一种“学习如何学习”的机器学习方法。其核心思想是让模型通过在支持集上更新参数,并在查询集上评估性能,从而获得“先验知识”,使得模型可以在面对新任务时,凭借少量样本迅速适应[37,41]。元学习过程一般分为两步,以 MAML 为例[42],首先,在内循环中,模型对每个支持集任务进行一次或多次梯度更新,即

$$\theta'_i = \theta - \alpha \nabla_{\theta} L^{S}_{T_i}(\theta)$$

再在查询集上评估这些适应后的参数,并对参数进行更新,即

$$\theta \leftarrow \theta - \beta \nabla_{\theta} \sum_{T_i \sim p(T)} L^{Q}_{T_i}(\theta'_i)$$

以最小化任务在查询集上的损失,如图 4(c)。

在分子性质预测任务中,元学习可有效应对冷启动问题或样本稀缺情形[43-45]。Schlender 等[46]建立了 The Bigger Fish 框架,该框架采用元学习策略来解决水生毒性中的 few-shot 问题。实验结果一致表明,知识共享技术优于单任务方法。考虑到灵活的暴露时间和化学品的广泛适用范围,该框架可以预测多种物种的毒性。Wang 等[47]引入了 PAR 框架,该框架采用元学习策略来解决分子性质预测中的 few-shot 问题。该框架在 GNN 编码器之上结合属性感知嵌入函数和自适应关系图学习模块来改进分子嵌入。实验结果一致表明,属性感知关系网络在小样本分子性质预测问题上取得了较好的效果。

3.4 对比学习

对比学习是一种自监督学习方法[48],其目标是在特征表示空间中学习到区分能力,使得语义相似的数据彼此接近,而不相似的数据彼此远离,如图 4(d)。对比学习不依赖人工标签,而是通过构造正样本对(positive pair)和负样本对(negative pair)

来训练编码器[49]。设样本 x_i 和其正样本 x'_i，通过编码器提取他们的表示 $z_i = f(x_i)$，$z'_i = f(x'_i)$，对比学习通过最小化以下 infoNCE 损失函数来优化模型：

$$L_i = -\log \frac{\exp(\mathrm{sim}(z_i, z'_i)/\tau)}{\sum_{j=1}^{N} \exp(\mathrm{sim}(z_i, z_j)/\tau)}$$

式中：sim(·,·)表示相似度函数，τ 为温度参数。

在分子特征学习中，Wang 等[50]引入分子图对比学习方法，通过改变潜在表示增强 GAT 模型在 15 个环境毒性终点数据集上的性能。Pinheiro 等[51]将分子图与 SMILES 序列作为对比学习对象进行预训练，并在 QM9 数据集上通过迁移学习分别实现了能量与电子性质预测误差的 44% 和 25% 降低。Yang 等[52]提出的 CACPP 方法克服了传统人工特征在抗癌肽预测中的不足，利用 TextCNN 提取高维序列特征，并通过对比学习获取更具判别力的表达，在多个基准数据集上超越现有最优方法。此外，Guo 等[53]开发的 CLERMS 框架用于 MS/MS 谱图的对比学习，不仅可区分不同化合物的谱图，还能揭示其结构相似性，在化合物识别和谱图聚类中均优于现有方法。

3.5　掩码学习

随着自注意力机制模块和 Transformer 架构的提出[54]，自监督学习作为训练语言模型的通用策略被广泛应用。掩码语言建模（Mask Language Modeling，MLM）已被广泛用于自然语言和生物序列自监督学习中，如 BERT 模型[55]就是通过双向序列上下文来恢复序列中被掩码的标记[56]。如前所述，这种方法可以利用大量未标注数据构建预训练模型，通过捕获语义信息，再对少量标注数据进行微调，以适应多种下游任务，如图 4(e)。这种“预训练-微调”模式（即迁移学习）在某些任务上优于传统的有监督训练方法。掩码语言建模的训练目标是最大化被掩盖位置上预测正确 token 的对数似然。

$$L_{\mathrm{MLM}} = -\sum_{i \in M} \log P(x_i \mid x_M) = -\sum_{i \in M} \log \hat{y}_i[x_i]$$

其中 M 表示被掩码的位置集合。

而在基于图结构分子任务中，预训练任务掩码节点预测的损失形式为

$$L_{\mathrm{mask-node}} = -\sum_{i \in M} \log P(y_i \mid G_M)$$

其中 G_M 为掩码后的分子图。

Ross 等[57]在 ZINC 和 Pubchem 数据库上各收集了 10 亿和 1.11 亿的分子进行掩码预训练 Transformer 编码器，在所有的 token 挑选 15% 进行随机的操作，在这些 token 中再依据选择 80% 的遮蔽，10% 随机更换和 10% 的保持不变，实验表明，根据该 Molformer 模型产生的嵌入在各基准数据集上超过了现有的基线模型。受到语言模型掩码的启发，Hu 等[58]在 Tox21、ToxCast、ClinTox 上比较了 4 种GNN 类型不同预训练掩码策略的节点水平和图水平。实验结果表明，经过上下文预测预训练的 GIN 模型是最具表现力的模型，在上述 3 个毒性数据集上的平均 ROC-AUC 分别为 4.1%、2.3% 和 14.6%。这些例子证明了基于自监督学习的预训练技术可以增强深度神经网络的学习和表达能力，包括参数初始化、模型正则化和泛化能力。

3.6　多模态学习

多模态学习是一种结合来自不同类型数据（模态）的信息进行学习的方法，能有效利用每种模态中的互补特征，以提升模型的表现力和泛化能力[59]。在分子领域，分子本身可以用多种不同的数据类型表示，如常见的 SMILES 字符串表示化学结构的线性序列，二维的分子图数据或基于光谱、图像或文本描述的其他模态[60]，如图 4(f)，因此，可定义 n 个模态的数据输入为

$$x_{(1)}, x_{(2)}, \cdots, x_{(n)}$$

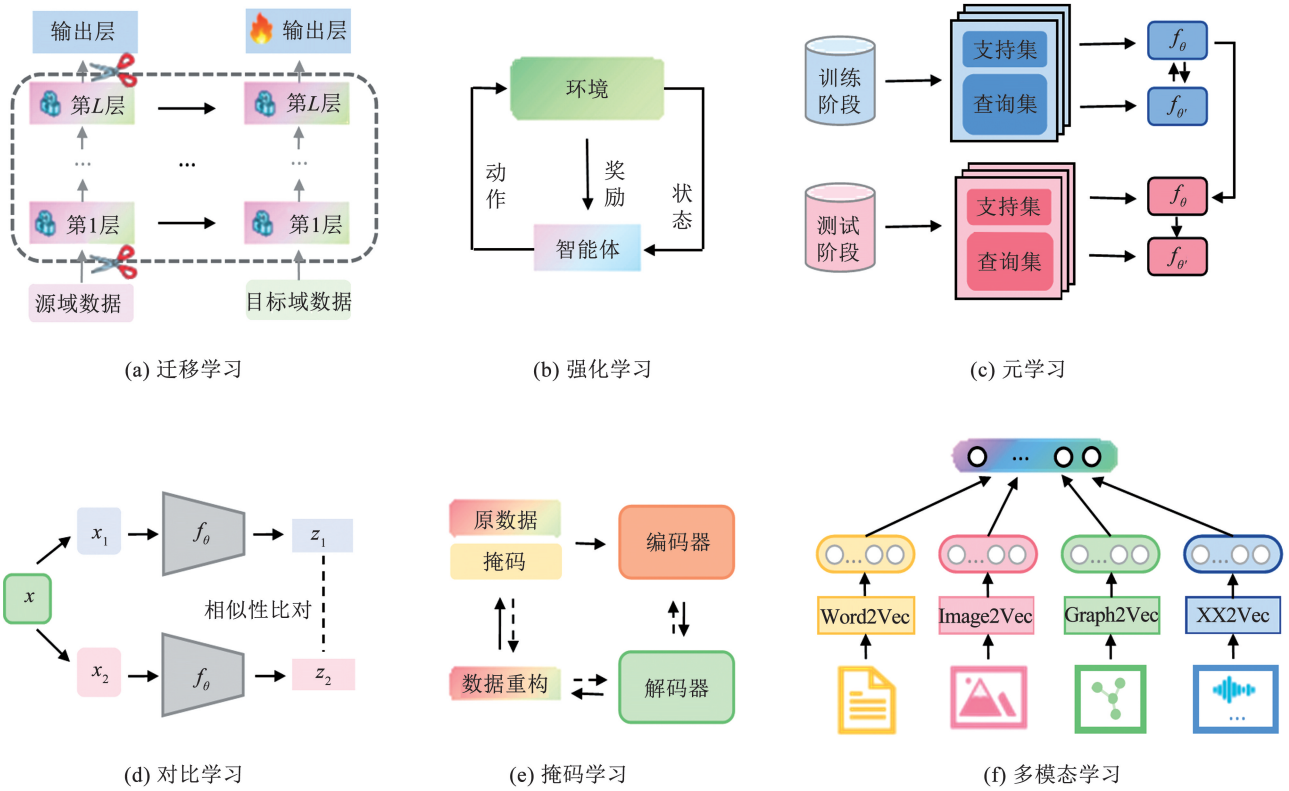
每个模态通过一个编码器映射到嵌入空间，即

$$z^{(1)} = f^{(i)}(x_{(i)})$$

以供下游模型进行整合[61]。

Rollins 等[62]融合了语言模型和图结构数据类型并提出了 MolPROP 模型，在 FreeSolv、ESOL、Lipo 等数据集中测试发现，该模型在小分子数据集和回归任务中表现良好，而在分类数据集中表现欠佳。Wang 等[63]基于分子图像数据和图结构搭建了 MolG 框架，使用自监督的训练方式有效地融合了两种分子表示形式的优势。Karim 等[64]把分子表示成字符串、图像和数字特征，并集成 3 种基础架构全连接层、卷积和时序网络，该方式在标准的毒性基准数据集上达到最先进的性能。

目前，这些方法已在分子性质预测领域取得显著进展，尤其在药物 ADMET 性能预测任务中展现出良好表现[65-66]，在缓解训练样本中存在的偏见、提升模型泛化能力以及应对数据稀缺等方面发挥了关键作用。然而，现有方法多聚焦于药物分子，尚缺乏针对环境分子独特特性和挑战的系统性方法开发。随着研究的深入，这些方法正逐步向环境化合物的预测任务拓展。



(a) 迁移学习 (b) 强化学习 (c) 元学习

(d) 对比学习 (e) 掩码学习 (f) 多模态学习

图 4 前沿训练策略在表示学习中的应用

Fig. 4 Application of advanced training strategies in representation learning

4 讨 论

4.1 负迁移效应

已有许多研究表明,迁移学习范式可以明显提升下游任务的性能[67]。但值得注意的是,迁移学习并不总是有效的。Hu 等[58]的实验表明,图级多任务学习中当使用的一些预训练任务与下游任务无关时,可能会产生负面影响。在他们的研究中,8 个分子数据集中有 2 个、40 个蛋白质预测任务中有13 个表现出了负迁移现象。针对这一问题,该文提出了对 GNN 不同层次进行训练的方法,并揭示了节点嵌入会加剧图级多任务学习中的负迁移。一种可能的解决方案是人为选择“真正相关”的监督预训练任务,并仅在这些任务上预训练 GNN。然而,这样的解决方案是非常昂贵的,因为选择相关的任务需要大量的领域专业知识,这似乎又回到了特征工程所面临的问题。因此,有研究提出在多任务预训练中动态调节各任务对模型的贡献,减少无关任务损失对梯度更新的负面影响,如 Zhou 等[68]在 Uni-Mol 中通过引入自适应损失加权机制来平衡对比任务、坐标回归任务等对模型训练的影响,有效减少负迁移。

4.2 模型解释性

尽管深度学习及其高阶训练策略在分子性质预测中展现出更高的预测能力,但其应用仍面临诸多挑战,亟需在未来加以克服[69]。首先,与传统 QSAR 方法相比,深度学习模型通常需要更多的计算资源,并且复杂的网络结构使得模型难以解释其预测结果[70]。这类模型具备从数据中学习复杂模式和关联的巨大潜力,但也可能在此过程中捕捉到难以察觉的隐藏偏差或错误[71]。因此,深度模型需要通过检查其决策过程是否与专业知识相符或是否能提供新的见解来进行解释与验证。

为解决“黑箱”模型的可解释性问题,研究者提出了多种可解释性方法[72],大体分为内在可解释性和事后解释性方法,前者在模型结构设计阶段引入透明机制,如注意力机制[73]、原子或键级贡献分解、基于规则的图网络等,使模型本身具备解释能力;后者是在模型训练完成后,通过分析输入与输出之间的关系推断模型决策依据。这类方法包括特征重要性分析如 SHAP[74]、LIME 等[75],梯度敏感性分析如 Grad-CAM[76-77]、Integrated Gradients[78-79] 等,扰动分析以及基于图结构的可视化解释等[80]。前者更关注结构透明性与相关性,后者则在不改变模型结构的前提下提供辅助性解释。

两种方法虽然可以在一定程度上提高模型的可解释性和可信度,但仍存在诸多局限。内在可解释模型通常需要简化特征交互或限制非线性表达能力,从而可能导致预测性能下降,难以捕捉分子性质

中高度非线性或高维的复杂关系。同时，内在可解释性方法也存在“表面可解释性”的风险[81]，其解释结果未必真实反映模型的内部决策机制，并且在不同数据集或任务间迁移时，其原有的解释结构可能失效或不再适用[82]。而事后解释性方法往往仅能对模型输出提供局部解释，无法全面揭示模型的决策过程[83]，且不同的方法所提供的解释结果存在差异[84]。更为关键的是，该类可解释性方法并不健壮，很容易被操纵以产生任何想要的解释[85]。

在实际应用中，可解释性方法的选择应根据任务特点与风险等级进行权衡。对于涉及化学品安全、毒性评估或其他高风险决策场景的研究，建议优先采用结构透明、原理清晰的“白盒”模型，如线性回归或基于规则的模型，以确保模型推理过程可被充分审查和验证。若任务对预测性能要求较高而安全性约束较弱，可考虑在高性能深度模型中引入可解释性机制，或结合事后解释方法以兼顾精度与可解释性。对于复杂场景，可采用多种解释方法或因果推断[86]进行交叉验证，从不同角度评估模型决策的一致性和稳定性，从而降低因单一解释造成误导的风险[87]。此外，在模型部署和应用过程中，应将解释结果与化学或生物学知识相对照，确保模型输出符合领域规律，避免“合理但错误”的解释被误用。总体而言，模型可解释性应被视为提高可信度与可验证性的核心环节，而非单纯的附加功能。

4.3　奥卡姆剃刀原理

奥卡姆剃刀原理是模型构建中非常重要的指导思想，强调在满足有效解释或预测数据的前提下，应优先选择结构更简单、参数更少的模型[88]，图 5 展示了在基准数据集上测试的常用分子性质预测模型的参数规模与性能对比。简化模型不仅有助于提高可解释性，还能降低对训练数据中噪声的敏感性，更容易泛化到未见样本，减少模型过拟合的风险，并降低模型训练和计算成本。在实际任务中，尤其需要通过对比实验和消融验证模型和模块的有效性，而非盲目增加模型复杂度；过度复杂的模型往往会导致过拟合，并在实际应用中难以推广。

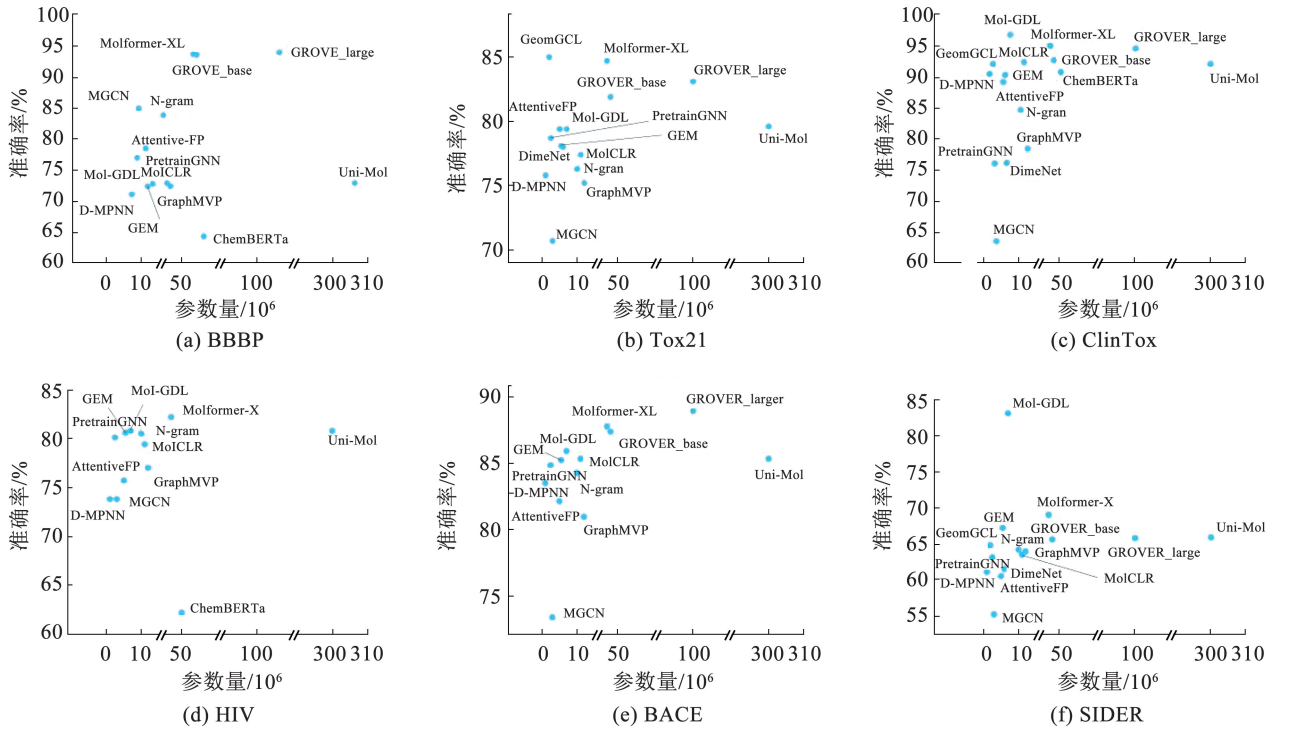


图 5　基础数据集中常用模型的参数与性能对比

Fig. 5　Comparison diagram of parameters and performance of commonly used models in benchmark datasets

随着深度学习技术的快速发展，模型复杂度呈指数级增长，并在其他领域展现出优异性能，但从实际应用角度来看，并不是所有问题都需要极深或极宽的神经网络[89]。相反，对于特征工程丰富、特征重要性明显或可以处理结构化数据的任务，如连续变量、离散变量、实验条件等，特征工程与经典的机器学习方法论仍然具有优势，应在充分实验的基础上进行比较，选择在性能和可解释性之间平衡更优的模型。尽管在人工智能领域，通过预训练模型直接在下游任务上进行微调已成为主流，但对于小样本数据集，尤其是生态或环境数据，仍需更多实证研究来证明该策略的有效性和可靠性。

此外,通过回顾以往人工智能赋能的环境科学研究可以发现,许多工作虽然基于自建数据集开展了模型训练,但目前仍普遍缺乏充分的同行间横向模型对比实验和系统性的模型架构分析,同时缺少基于基准数据集的泛化能力评估(表 1),从而限制了对模型架构有效性的验证(表 2)。模型的选择往往仅以预测精度或测试集准确率作为主要评估标准,缺乏对模型架构与数据分布之间适配性的系统研究,也缺乏对模型失效机制的理论分析。这种以性能指标为导向的建模思路,使得研究者在面对新的研究任务时,往往需要从零开始逐一尝试不同模型结构,难以提炼出具有迁移性和普适指导意义的建模规律,从而造成了研究资源与计算成本的浪费。针对这一问题,基于奥卡姆剃刀原理和 VC 维理论,提出了一套环境分子机器学习模型的规范化开发流程与准则(图 6),该流程旨在为环境科学领域的研究者提供在模型设计、选择与验证中的结构化指导[90],使模型开发不仅追求性能最优,更兼顾泛化性与可解释性。同时,通过该规范化建模框架,不同研究间的模型能够实现横向可比性与可复现性,从而逐步总结出适用于不同环境数据类型的建模规律,促进领域内模型开发由经验驱动向理论指导的转变。

表 1 常用的基准数据集介绍

Tab. 1 Introduction to commonly used benchmark datasets

数据集名称	数据描述	数据规模	任务性质
ESOL[91]	含一系列有机化合物的水溶解度实验数据,用于预测溶解性	2 874	回归
FreeSolv[92]	包含小分子在水中水合自由能的实验与计算值,用于能量预测	643	回归
Lipophilicity[1]	收录化合物在 pH 7.4 条件下的辛醇/水分配系数(log D) 实验值,用于疏水性建模	4 200	回归
HIV[2]	来源于药物治疗计划的抗 HIV 活性筛选数据集,用于预测抗病毒活性	约 4 万	分类
BACE[93]	含 β-分泌酶(BACE-1)抑制剂活性实验数据	1 514	分类
BBBP[94]	脑血屏障通透性(Blood-Brain Barrier Penetration)数据集	2 053	分类
Tox21[3]	含 12 个毒性通路的多标签分类数据集,用于环境毒理预测	7 833	分类
ToxCast[95]	大规模体外毒性筛查数据集,涵盖数千种化合物与多种生物终点	8 575	分类
SIDER[96]	收录上市药物及其不良反应的多标签分类数据集	1 429	分类
ClinTox[97]	包含临床毒性阳性/阴性药物对比,用于药物安全性分类	1 484	分类
Mutag[4]	MUTAG 是硝基芳香族化合物的集合,目标是预测其对鼠伤寒沙门氏菌的致突变性	188	分类
Muv[98]	基于 PubChem 筛选的高置信度化合物子集,用于在极度不平衡场景下评估模型的泛化与抗偏差能力	约 18 万	分类
PCBA[99]	PubChemBioAssay 数据的多任务分类集合,用于生物活性建模	约 87 万	分类
QM9[100]	小分子的量子化学性质数据,是能量与结构回归标准基准	约 13 万	回归

注:1. https://www.kaggle.com/datasets/markinsuff/lipophilicity; 2. https://wiki.nci.nih.gov/display/NCIDTPdata/AIDS+Antiviral+Screen+Data; 3. https://tripod.nih.gov/tox21/challenge/; 4. https://www.ijisrt.com/molecular-classification-with-graph-convolutionalnetworks-exploring-the-mutag-dataset-for-mutagenicity-prediction。

表 2 环境分子性质预测任务的模型性能对比

Tab. 2 Model performance comparison of environmental molecular property prediction tasks

任务	训练范式	模型评估	数据来源	最好的模型性能	文献
外源化学物质对病毒受体基因转录的影响	监督学习	SVM、RF、DNN、Gaussian Process、DT、GCN	自建数据集	GCN AUROC = 0.7	[101]
Tox21 数据集	监督学习	RF、LightGBM、DNN、GCN、GAT	Tox21	GCN AUROC = 0.734	[102]
化合物的急性毒性评估	监督学习	MPNN、GCN、GAT、PAGTNs、Attentive FP	自建数据集	Attentive FP R^2 = 0.65	[103]
化合物的 hERG 阻断特性	监督学习	RF、SVM、GCN	自建数据集	GCN AUROC = 0.952	[104]
突变预测	监督学习	SVM、GCN	Ames 诱变基准数据集	GCN AUROC = 0.83 ~ 0.87	[105]

表 2(续)

任务	训练范式	模型评估	数据来源	最好的模型性能	文献
持久性、生物蓄积性和有毒物质	监督学习	GAT、DCNN	自建数据集	GAT AUROC = 97.9，PR = 89.7，ACC = 93.8	[106]
生态毒理学	监督学习	LR、MLP、GraphSAGE	EnviroTox	GraphSAGE AUROC = 0.91，AUPRC = 0.95	[107]
化学雌激素活性	监督学习	RF、SVM、1D CNN	自建数据集	1D CNN 内部/外部测试准确率为 0.88，0.925	[108]
斑马鱼鱼仔的形态学分类	监督学习	TensorMask Mask R-CNN	自建数据集	Mask R-CNN 平均 precision > 0.93	[109]
有机化合物的急性毒性	监督学习	CNN-1D、CNN-2D、ANN、KNN、RF、MLR、SVR	自建数据集	CNN-1D 外部测试 R^2 = 0.7	[110]
公共数据集	监督学习	SVM、RF、ASNN、MLP、DNN、Multitask、CNN	Tox21	CNN AUC = 0.85	[111]
公共数据集	监督学习	LR、RF、SVM、DT、KNN、Smi2Vec-LSTM、Smi2Vec-BiGRU	MoleculeNet	Smi2Vec-BiGRU	[112]
公共数据集	监督学习	RF、SVM、KNN、SMILES2Vec、DeepTox、DeepAOT、BiGRU	ClinTox，Tox21，ToxB	BiGRU 测试集 AUROC = 0.94，0.95，0.91	[113]
门菌诱变性试验、抑制恶性疟原虫 Dd2、抑制丙型肝炎病毒	监督学习	NN、LSTM Logistic	自建数据集	LSTM AUROC > 0.9	[114]
水生生物急性和慢性化学毒性的准确预测	监督学习	ECOSAR v2.2、VEGA v1.1.5、T.E.S.T. v5.1.1.0、RoBERTa	自建数据集	RoBERTa 中位数误差因子为 2.66，鱼类和水生无脊椎动物为 2.82	[26]
分子毒性	监督学习	RF、MolBERT	自建数据集	MolBERT ACC = 0.86 ~ 0.87	[115]
公共数据集	掩码学习 监督学习	GraphSAGE、GPT_GNN、AttributeMask、ContextPred、InfoGraph、MoCL、GraphLoG、GraphCL、JOAO、MolCLR、G_Motif、MGSSL	MoleculeNet	HiMol 3-layers：AUROC = 71.8 5-layers：AUROC = 73.7	[116]
公共数据集	掩码学习、迁移学习	GCN、GraphSAGE、GIN、GAT	MoleculeNet	GIN AUROC = 74.2	[58]
公共数据集	半监督学习	KNN、NN、RF、SVM、XGBoost、GCN	Tox21	GCN AUROC = 0.757	[117]
公共数据集	多任务学习	RF、SVM、MTDNN	Tox21	MTDNN AUROC = 0.853	[118]
PFAS 毒性	迁移学习	DNN、GCN、GP、RF	EPA	Transfer-DNN R^2 = 0.9	[31]
多物种急性毒性	多任务学习	MT-DNN、ST-RF、ST-DNN、MTGCNN、DLCA	自建数据集	MT-DNN + MTGCNN R^2 = 0.57	[119]
公共数据集	掩码学习 迁移学习	RF、SVM、MGCN、D-MPNN、DimeNet、PretrainGNN、N-gram、MolCLR、GraphMVP-C、GeomGCL、GEM、ChemBERTa、Molformer	MoleculeNet	Molformer 8 个任务中 SOTA	[57]
公共数据集	对比学习 迁移学习	GCN、GIN、SchNet、MGCN、D-MPNN、ECFP4、N-Gram、MolCLR、HiMol、PretrainGNN、MOE	MoleculeNet	MOE 6 个任务中 SOTA	[120]

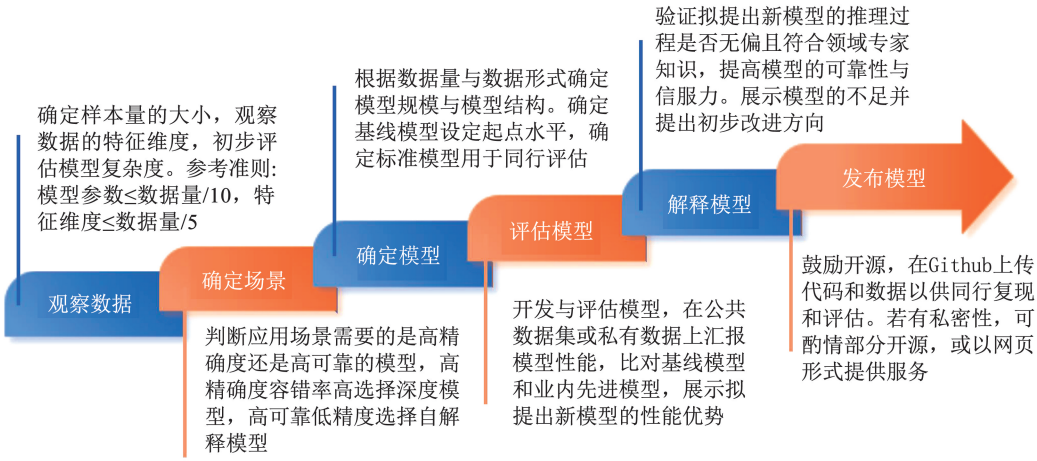


图 6 环境分子机器学习模型规范化开发流程与准则

Fig. 6 Standardized development workflow and guidelines for environmental molecular machine learning models

5 结语与展望

近年来，随着人工智能技术的飞速发展，高参数量、复杂结构的机器学习模型正逐步渗透至环境科学领域，在污染物识别、环境分子行为预测及绿色材料设计等关键任务中展现出深远的变革潜力。此类模型在解析复杂结构关系与提取深层环境特征方面具有显著优势，有望突破传统实验方法在成本、时效及伦理层面的瓶颈。然而，数据匮乏仍是制约环境智能建模实用性与泛化能力的主要问题。受限于环境样品的复杂性、污染物种类的多样性及低浓度特征，建模往往依赖于有限的实验或文献数据。在小样本约束下，数据驱动模型难以充分捕获底层的科学规律，预测结果往往缺乏统计稳定性与代表性，这与深度学习对大规模样本的依赖形成了固有矛盾。因此，小样本环境数据的高效建模已成为领域内的核心课题。现有研究表明，迁移学习与元学习通过知识迁移与经验复用，能够显著优化环境污染物性质预测与毒性分类任务。此外，深度学习模型在提升性能的同时，也面临高算力需求与“黑盒”不可解释性的双重挑战。鉴于预测结果直接关联下游环境决策，模型偏差在毒性预警等高风险任务中可能引发严重误判。因此，整合生成式任务、预训练模式、深度架构与可解释性方法，构建面向小样本的高效建模框架已成为当前研究的前沿方向[121]。在开发过程中，应秉持奥卡姆剃刀原则，结合系统化的建模流程进行架构与策略的横向对比，旨在为行业建立统一、可复现的智能建模规范。

值得注意的是，当前部分研究仍侧重于通用算法的迁移与微调，未能充分评估分布外样本、数据稀疏、高噪声标签及多模态特征等环境领域特有的现实挑战。同时，由于实验设计与预处理手段的差异，环境数据表现出显著的时空与技术异质性，增加了模型泛化难度并可能引入系统性偏差。因此，研究人员应保持积极且审慎的态度，兼顾模型性能、可解释性与数据复杂性，致力于解决环境分子预测中的关键科学问题，推动该领域向更具前瞻性与实用性的方向迈进。

参考文献

[1] KATRITZKY A R, LOBANOV V S, KARELSON M. QSPR: the correlation and quantitative prediction of chemical and physical properties from structure[J]. Chemical Society Reviews, 1995, 24(4): 279. DOI: 10.1039/CS9952400279

[2] FUJITA T, WINKLER D A. Understanding the roles of the "Two QSARs"[J]. Journal of Chemical Information and Modeling, 2016, 56(2): 269. DOI: 10.1021/acs.jcim.5b00229

[3] HANSCH C, MALONEY P P, FUJITA T, et al. Correlation of biological activity of phenoxyacetic acids with Hammett substituent constants and partition coefficients[J]. Nature, 1962, 194(4824): 178. DOI: 10.1038/194178b0

[4] MURATOV E N, BAJORATH J, SHERIDAN R P, et al. QSAR without borders[J]. Chemical Society Reviews, 2020, 49(11): 3525. DOI: 10.1039/d0cs00098a

[5] WALTERS W P, BARZILAY R. Applications of deep learning in molecule generation and molecular property prediction[J]. Acc Chem Res, 2021, 54(2): 263. DOI: 10.1021/acs.accounts.0c00699

[6] ASKR H, ELGELDAWI E, ABOUL ELLA H, et al. Deep learning in drug discovery: an integrative review and future challenges[J]. Artif Intell Rev, 2023, 56(7): 5975. DOI: 10.1007/s10462-022-10306-1

[7] JEONG J, CHOI J. Artificial intelligence-based toxicity prediction of environmental chemicals: future directions for chemical management applications[J]. Environmental Science & Technology, 2022, 56(12): 7532. DOI: 10.1021/acs.est.1c07413

[8] ZHAO Qiming, YU Yang, GAO Yuchen, et al. Machine learning-based models with high accuracy and broad applicability domains for screening PMT/vPvM substances[J]. Environmental Science & Technology, 2022, 56(24): 17880. DOI: 10.1021/acs.est.

2c06155
[9] HAGHIGHATLARI M, LI Jie, HEIDAR-ZADEH F, et al. Learning to make chemical predictions: the interplay of feature representation, data, and machine learning methods[J]. Chem, 2020, 6(7): 1527. DOI: 10.1016/j.chempr.2020.05.014
[10] WIGH D S, GOODMAN J M, LAPKIN A A. A review of molecular representation in the age of machine learning[J]. Wiley Interdisciplinary Reviews Computational Molecular Science, 2022, 12(5): e1603. DOI: 10.1002/wcms.1603
[11] STUMPFE D, BAJORATH J. Exploring activity cliffs in medicinal chemistry[J]. Journal of Medicinal Chemistry, 2012, 55(7): 2932. DOI: 10.1021/jm201706b
[12] COLEY C W, BARZILAY R, GREEN W H, et al. Convolutional embedding of attributed molecular graphs for physical property prediction[J]. Journal of Chemical Information and Modeling, 2017, 57(8): 1757. DOI: 10.1021/acs.jcim.6b00601
[13] DUVENAUDT D, MACLAURIN D, AGUILERA-IPARRAGUIRRE J, et al. Convolutional networks on graphs for learning molecular fingerprints[J]. Advances in Neural Information Processing Systems, 2015: 28. DOI:10.48550/arXiv.1509.09292
[14] CHITHRANANDA S, GRAND G, RAMSUNDAR B. Chem BERTa: large-scale self-supervised pretraining for molecular property prediction[J/OL]. arXiv, 2020, abs/2010.09885. DOI: 10.48550/arXiv.2010.09885
[15] GOMEZ-BOMBARELLI R, WEI J N, DUVENAUD D, et al. Automatic chemical design using a data-driven continuous representation of molecules[J]. Acs Central Science, 2018, 4(2): 268. DOI: 10.1021/acscentsci.7b00572
[16] KEARNES S, MCCLOSKEY K, BERNDL M, et al. Molecular graph convolutions: moving beyond fingerprints[J]. Journal of Computer-Aided Molecular Design, 2016, 30(8): 595. DOI: 10.1007/s10822-016-9938-8
[17] CHEN Yangyang, WANG Zixu, ZENG Xiangxiang, et al. Molecular language models: RNNs or transformer?[J]. Briefings in Functional Genomics, 2023, 22(4): 392. DOI: 10.1093/bfgp/elad012
[18] GRECHISHNIKOVA D. Transformer neural network for protein-specific de novo drug generation as a machine translation problem[J]. Scientific Reports, 2021, 11(1): 321. DOI: 10.1038/s41598-020-79682-4
[19] MENG Xin, WU Binglan, LIANG Yilei, et al. CIMG-BERT: pretraining bidirectional transformers with chemistry knowledge for molecular property prediction[C]//2022 2nd International Conference on Electronic Information Technology and Smart Agriculture (ICEITSA). Huaihua, China: IEEE, 2022: 201. DOI: 10.1109/ICEITSA57468.2022.00043
[20] YOSHIKAI Y, MIZUNO T, NEMOTO S, et al. A novel molecule generative model of VAE combined with Transformer for unseen structure generation[J/OL]. arXiv, 2024, abs/2402.11950. DOI: 10.48550/arXiv.2402.11950
[21] SHISHIR F S, HASIB K M, SAKIB S, et al. De novo drug property prediction using graph convolutional neural networks[C]// 2021 IEEE 9th Region 10 Humanitarian Technology Conference (R10-HTC). Bangalore, India: IEEE, 2021: 1. DOI: 10.1109/R10-HTC53172.2021.9641611
[22] BENGIO Y. Deep learning of representations for unsupervised and transfer learning[C]// Proceedings of ICML Workshop on Unsupervised and Transfer Learning. [S.l.]: JMLR Workshop and Conference Proceedings, 2012, 27: 17
[23] WEISS K, KHOSHGOFTAAR T M, WANG Dingding. A survey of transfer learning[J]. Journal of Big Data, 2016, 3: 1. DOI: 10.1186/s40537-016-0043-6
[24] ZHAO Wei. Research on the deep learning of the small sample data based on transfer learning[J]. AIP Conference Proceedings, 2017, 1864(1): 020018. DOI: 10.1063/1.4992835
[25] ZHUANG Fuzhen, QI Zhiyuan, DUAN Keyu, et al. A comprehensive survey on transfer learning[J]. Proceedings of the IEEE, 2020, 109(1): 43. DOI: 10.1109/JPROC.2020.3004555
[26] GUSTAVSSON M, KALL S, SVEDBERG P, et al. Transformers enable accurate prediction of acute and chronic chemical toxicity in aquatic organisms[J]. Science Advances, 2024, 10(10): eadk6669. DOI: 10.1126/sciadv.adk6669
[27] ZHANG Kunyang, SCHWALLER P, FENNER K. Predicting toxicity toward nitrifiers by attention-enhanced graph neural networks and transfer learning from baseline toxicity[J]. Environmental Science & Technology, 2025, 59(9): 4518. DOI: 10.1021/acs.est.4c12247
[28] CAI Chenjing, WANG Shiwei, XU Youjun, et al. Transfer learning for drug discovery[J]. Journal of Medicinal Chemistry, 2020, 63(16): 8683. DOI: 10.1021/acs.jmedchem.9b02147
[29] BUTEREZ D, JANET J P, KIDDLE S J, et al. Transfer learning with graph neural networks for improved molecular property prediction in the multi-fidelity setting[J]. Nature Communications, 2024, 15(1): 1517. DOI: 10.1038/s41467-024-45566-8
[30] ZHONG Shifa, ZHANG Yanping, ZHANG Huichun. Machine learning-assisted QSAR models on contaminant reactivity toward four oxidants: combining small data sets and knowledge transfer[J]. Environmental Science & Technology, 2021, 56(1): 681. DOI: 10.1021/acs.est.1c04883
[31] FEINSTEIN J, SIVARAMAN G, PICEL K, et al. Uncertainty-informed deep transfer learning of Perfluoroalkyl and Polyfluoroalkyl substance toxicity[J]. Journal of Chemical Information and Modeling, 2021, 61(12): 5793. DOI: 10.1021/acs.jcim.1c01204
[32] LI Xinhao, FOURCHES D. Inductive transfer learning for molecular activity prediction: Next-Gen QSAR Models with MolPMoFiT[J]. J Cheminform, 2020, 12(1): 27. DOI: 10.1186/s13321-020-00430-x
[33] AMABILINO S, POGANY P, PICKETT S D, et al. Guidelines for recurrent neural network transfer learning-based molecular generation of focused libraries[J]. J Chem Inf Model, 2020, 60(12): 5699. DOI: 10.1021/acs.jcim.0c00343
[34] MOROZOV V, RODRIGUES C H M, ASCHER D B. CSM-Toxin: a web-server for predicting protein toxicity[J]. Pharmaceutics, 2023, 15(2): 431. DOI: 10.3390/pharmaceutics15020431
[35] KAELBLING L P, LITTMAN M L, MOORE A W. Reinforcement learning: a survey[J]. Journal of Artificial Intelligence Research, 1996, 4(1): 237. DOI: 10.5555/1622737.1622748
[36] GOEL M, RAGHUNATHAN S, LAGHUVARAPU S, et al. MoleGuLAR: molecule generation using reinforcement learning with alternating rewards[J]. Journal of Chemical Information and Modeling, 2021, 61(12): 5815. DOI: 10.1021/acs.jcim.1c01341
[37] NEIL D, SEGLER M, GUASCH L, et al. Exploring deep recurrent

models with reinforcement learning for molecule design [C]// Proceedings of the International Conference on Learning Representations. 2018

[38] OLIVECRONA M, BLASCHKE T, ENGKVIST O, et al. Molecular de-novo design through deep reinforcement learning[J]. Journal of Cheminformatics, 2017, 9(1): 48. DOI: 10.1186/s13321-017-0235-x

[39] ZHOU Zhenpeng, KEARNES S, LI Li, et al. Optimization of molecules via deep reinforcement learning[J]. Sci Rep, 2019, 9(1): 10752. DOI: 10.1038/s41598-019-47148-x

[40] MAZUZ E, SHTAR G, SHAPIRA B, et al. Molecule generation using transformers and policy gradient reinforcement learning[J]. Sci Rep, 2023, 13(1): 8799. DOI: 10.1038/s41598-023-35648-w

[41] GUO Zhichun, ZHANG Chuxu, YU Wenhao, et al. Few-shot graph learning for molecular property prediction[C]// Proceedings of the Web Conference 2021. New York: ACM, 2021: 2559. DOI: 10.1145/3442381.3450112

[42] FINN C, ABBEEL P, LEVINE S. Model-agnostic meta-learning for fast adaptation of deep networks [C]// Proceedings of the 34th International Conference on Machine Learning. Sydney: PMLR, 2017: 1126

[43] KARIM A, RIAHI V, MISHRA A, et al. Quantitative toxicity prediction via meta ensembling of multitask deep learning models [J]. ACS Omega, 2021, 6(18): 12306. DOI: 10.1021/acsomega.1c01247

[44] ZHUANG Xiang, ZHANG Qiang, WU Bin, et al. Graph sampling-based meta-learning for molecular property prediction [J/OL]. arXiv, 2023, abs/2306.16780. DOI: 10.48550/arXiv.2306.16780

[45] NGUYEN C Q, KREATSOULAS C, BRANSON K M. Meta-learning GNN initializations for low-resource molecular property prediction[J/OL]. arXiv, 2020, abs/2003.05996. DOI: 10.48550/arXiv.2003.05996

[46] SCHLENDER T, VILJANEN M, VAN RIJN J N, et al. The bigger fish: a comparison of meta-learning QSAR models on low-resourced aquatic toxicity regression tasks [J]. Environmental Science & Technology, 2023, 57(46): 17818. DOI: 10.1021/acs.est.3c00334

[47] WANG Yaqing, ABUDUWEILI A, YAO Quanming, et al. Property-aware relation networks for few-shot molecular property prediction[J]. Advances in Neural Information Processing Systems, 2021, 34: 17441. DOI: 10.5555/3540261.3541595

[48] LIN Ye, YANG Xin, ZHANG Mingxuan, et al. CLSSATP: contrastive learning and self-supervised learning model for aquatic toxicity prediction[J]. Aquatic Toxicology, 2025, 279: 107244. DOI: 10.1016/j.aquatox.2025.107244

[49] LE-KHAC P H, HEALY G, SMEATON A F. Contrastive representation learning: a framework and review[J]. IEEE Access, 2020, 8: 193907. DOI:10.1109/ACCESS.2020.3031549

[50] WANG Haobo, LIU Wenjia, CHEN Jingwen, et al. Applicability domains based on molecular graph contrastive learning enable graph attention network models to accurately predict 15 environmental end points[J]. Environmental Science & Technology, 2023, 57(44): 16906. DOI: 10.1021/acs.est.3c03860

[51] PINHEIRO G A, DA SILVA J L F, QUILES M G. SMICLR: contrastive learning on multiple molecular representations for semisupervised and unsupervised representation learning [J]. Journal of Chemical Information and Modeling, 2022, 62(17): 3948. DOI: 10.1021/acs.jcim.2c00521

[52] YANG Xuetong, JIN Junru, WANG Ruheng, et al. CACPP: a contrastive learning-based siamese network to identify anticancer peptides based on sequence only [J]. Journal of Chemical Information and Modeling, 2024, 64(7): 2807. DOI: 10.1021/acs.jcim.3c00297

[53] GUO Hao, XUE Kebing, SUN Haiming, et al. Contrastive learning-based embedder for the representation of tandem mass spectra[J]. Analytical Chemistry, 2023, 95(20): 7888. DOI: 10.1021/acs.analchem.3c00260

[54] ACHIAM J, ADLER S, AGARWAL S, et al. GPT-4 technical report[J/OL]. arXiv, 2023, abs/2303.08774. DOI: 10.48550/arXiv.2303.08774

[55] JI Yanrong, ZHOU Zhihan, LIU Han, et al. DNABERT: pre-trained bidirectional encoder representations from transformers model for DNA-language in genome[J]. Bioinformatics, 2021, 37(15): 2112. DOI: 10.1093/bioinformatics/btab083

[56] DEVLIN J, CHANG M W, LEE K, et al. BERT: pre-training of deep bidirectional transformers for language understanding[C]// Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers). 2019: 4171

[57] ROSS J, BELGODERE B, CHENTHAMARAKSHAN V, et al. Large-scale chemical language representations capture molecular structure and properties[J]. Nature Machine Intelligence, 2022, 4(12): 1256. DOI:10.1038/s42256-022-00580-7

[58] HU Weihua, LIU Bowen, GOMES J, et al. Strategies for pre-training graph neural networks[J/OL]. arXiv, 2019, abs/1905.12265. DOI: 10.48550/arXiv.1905.12265

[59] BALTRUŠAITIS T, AHUJA C, MORENCY L P. Multimodal machine learning: a survey and taxonomy[J]. IEEE Transactions on Pattern Analysis Machine Intelligence, 2018, 41(2): 423. DOI: 10.1109/TPAMI.2018.2798607

[60] HU Jie, YIN Ran, PAN Yao, et al. Multimodal learning-assisted identification of effluent water quality and toxicity in wastewater treatment plants[J]. Environmental Science & Technology, 2025, 59(27): 13857. DOI: 10.1021/acs.est.5c04143

[61] HONG J, KWON H. Multimodal deep learning for chemical toxicity prediction and management[J]. Scientific Reports, 2025, 15(1): 19491. DOI: 10.1038/s41598-025-95720-5

[62] ROLLINS Z A, CHENG A C, METWALLY E. MolPROP: molecular property prediction with multimodal language and graph fusion[J]. J Cheminform, 2024, 16(1): 56. DOI: 10.1186/s13321-024-00846-9

[63] WANG Zhuoyuan, MI Jiacong, LU Shan, et al. Multimodal-learning for predicting molecular properties: a framework based on image and graph structures [J/OL]. arXiv, 2023, abs/2311.16666. DOI: 10.48550/arXiv.2311.16666

[64] KARIM A, SINGH J, MISHRA A, et al. Toxicity prediction by multimodal deep learning[C]//Pacific Rim Knowledge Acquisition Workshop. Cham: Springer International Publishing, 2019: 142. DOI: 10.1007/978-3-030-30639-7_12

[65] YANG Xin, WANG Yifei, BYRNE R, et al. Concepts of artificial intelligence for computer-assisted drug discovery [J]. Chemical Reviews, 2019, 119(18): 10520. DOI: 10.1021/acs.chemrev.

8b00728

[66] CHENG Feixiong, LI Weihua, LIU Guixia, et al. In silico ADMET prediction: recent advances, current challenges and future trends [J]. Current Topics in Medicinal Chemistry, 2013, 13(11): 1273. DOI: 10.2174/15680266113139990033

[67] TAN Chuanqi, SUN Fuchun, KONG Tao, et al. A survey on deep transfer learning[C]//International Conference on Artificial Neural Networks. Cham: Springer International Publishing, 2018: 270. DOI: 10.1007/978-3-030-01424-7_27

[68] ZHOU Gengmo, GAO Zhifeng, DING Qiankun, et al. Uni-Mol: a universal 3D molecular representation learning framework [C]// Proceedings of the International Conference on Learning Representations. 2023

[69] CHAKRABORTY S, TOMSETT R, RAGHAVENDRA R, et al. Interpretability of deep learning models: a survey of results[C]// Proceedings of the 2017 IEEE Smartworld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computed, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation (smartworld/SCALCOM/UIC/ATC/CBDcom/IOP/SCI). New York: IEEE, 2017: 1

[70] FEINBERG E N, SUR D, WU Zhenqi, et al. PotentialNet for molecular property prediction [J]. ACS Central Science, 2018, 4(11): 1520. DOI: 10.1021/acscentsci.8b00507

[71] YANG K, SWANSON K, JIN Wengong, et al. Analyzing learned molecular representations for property prediction [J]. Journal of Chemical Information and Modeling, 2019, 59(8): 3370. DOI: 10.1021/acs.jcim.9b00237

[72] ZHANG Yu, TIÑO P, LEONARDIS A, et al. A survey on neural network interpretability[J]. IEEE Transactions on Emerging Topics in Computational Intelligence, 2021, 5(5): 726. DOI: 10.1109/TETCI.2021.3100641

[73] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[J]. Advances in Neural Information Processing Systems, 2017, 30: 6000. DOI: 10.5555/3295222.3295349

[74] LUNDBERG S M, ERION G, CHEN H, et al. From local explanations to global understanding with explainable AI for trees [J]. Nat Mach Intell, 2020, 2(1): 56. DOI: 10.1038/s42256-019-0138-9

[75] RIBEIRO M T, SINGH S, GUESTRIN C. "Why should I trust you?" Explaining the predictions of any classifier [C]// Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York: ACM, 2016: 1135. DOI: 10.1145/2939672.2939778

[76] AGARWAL C, QUEEN O, LAKKARAJU H, et al. Evaluating explainability for graph neural networks[J]. Scientific Data, 2023, 10(1): 144. DOI: 10.1038/s41597-023-01974-x

[77] SELVARAJU R R, COGSWELL M, DAS A, et al. Grad-CAM: visual explanations from deep networks via gradient-based localization[C]// Proceedings of the IEEE International Conference on Computer Vision. Italy: IEEE, 2017: 618. DOI: 10.1109/ICCV.2017.74

[78] SANCHEZ-LENGELING B, WEI J, LEE B, et al. Evaluating attribution for graph neural networks [J]. Advances in Neural Information Processing Systems, 2020, 33: 5898. DOI: 10.5555/3495724.3496219

[79] SUNDARARAJAN M, TALY A, YAN Qiqi. Axiomatic attribution for deep networks [C]// Proceedings of the 34th International Conference on Machine Learning. Sydney: PMLR, 2017: 3319. DOI: 10.5555/3305890.3306024

[80] KAKKAD J, JANNU J, SHARMA K, et al. A survey on explainability of graph neural networks[J/OL]. arXiv, 2023, abs/2306.01958. DOI: 10.48550/arXiv.2306.01958

[81] LIPTON Z C. The mythos of model interpretability: in machine learning, the concept of interpretability is both important and slippery [J]. Queue, 2018, 16(3): 31. DOI: 10.1145/3236386.3241340

[82] YANG S C H, FOLKE T, SHAFTO P. Abstraction, validation, and generalization for explainable artificial intelligence[J]. Applied AI Letters, 2021, 2(4): e37. DOI: 10.1002/ail2.37

[83] HAO Yuxing, DUAN Zhihui, LIU Lizheng, et al. Development of an interpretable machine learning model for neurotoxicity prediction of environmentally related compounds [J]. Environ Sci Technol, 2025, 59(22): 11108. DOI: 10.1021/acs.est.5c03311

[84] BORDT S, FINCK M, RAIDL E, et al. Post-Hoc explanations fail to achieve their purpose in adversarial contexts[C]// Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency. New York: IEEE, 2022: 891

[85] SLACK D, HILGARD S, JIA E, et al. Fooling LIME and SHAP: adversarial attacks on post hoc explanation methods [C]// Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society. New York: IEEE, 2020: 180

[86] PEARL J. Causal inference in statistics: an overview[J]. Statistics Surveys, 2009, 3: 96. DOI:10.1214/09-SS057

[87] RUDIN C. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead [J]. Nature Machine Intelligence, 2019, 1(5): 206. DOI: 10.1038/s42256-019-0048-x

[88] JEFFERYS W H, BERGER J O. Ockham's razor and Bayesian analysis[J]. American Scientist, 1992, 80(1): 64

[89] BELKIN M, HSU D, MA S, et al. Reconciling modern machine-learning practice and the classical bias-variance trade-off [J]. Proceedings of the National Academy of Sciences, 2019, 116(32): 15849. DOI: 10.1073/pnas.1903070116

[90] ABU-MOSTAFA Y S, MAGDON-ISMAIL M, LIN H T. Learning from data[M]. New York: AMLBook, 2012: 1

[91] DUTTA A, KARMAKAR R. Estimating aqueous solubility directly from molecular structure using machine learning approach [C]// 2021 Fifth International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud)(I-SMAC). Palladam, India: IEEE, 2021: 467

[92] MOBLEY D L, GUTHRIE J P. FreeSolv: a database of experimental and calculated hydration free energies, with input files [J]. J Comput Aided Mol Des, 2014, 28(7): 711. DOI: 10.1007/s10822-014-9747-x

[93] SUBRAMANIAN G, RAMSUNDAR B, PANDE V, et al. Computational modeling of beta-secretase 1 (BACE-1) inhibitors using ligand based approaches[J]. J Chem Inf Model, 2016, 56(10): 1936. DOI: 10.1021/acs.jcim.6b00290

[94] MARTINS I F, TEIXEIRA A L, PINHEIRO L, et al. A Bayesian approach to in silico blood-brain barrier penetration modeling[J]. J Chem Inf Model, 2012, 52(6): 1686. DOI: 10.1021/ci300124c

[95] RICHARD A M, JUDSON R S, HOUCK K A, et al. ToxCast chemical landscape: paving the road to 21st century toxicology[J]. Chem Res Toxicol, 2016, 29(8): 1225. DOI: 10.1021/acs.

chemrestox. 6b00135

[96] KUHN M, LETUNIC I, JENSEN L J, et al. The SIDER database of drugs and side effects[J]. Nucleic Acids Res, 2016, 44(D1): D1075. DOI: 10.1093/nar/gkv1075

[97] GAYVERT K M, MADHUKAR N S, ELEMENTO O. A data-driven approach to predicting successes and failures of clinical trials [J]. Cell Chem Biol, 2016, 23(10): 1294. DOI: 10.1016/j. chembiol. 2016. 07. 023

[98] ROHRER S G, BAUMANN K. Maximum unbiased validation (MUV) data sets for virtual screening based on PubChem bioactivity data[J]. Journal of Chemical Information and Modeling, 2009, 49(2): 169. DOI: 10.1021/ci8002649

[99] TRAN-NGUYEN V K, JACQUEMARD C, ROGNAN D. LIT-PCBA: an unbiased data set for machine learning and virtual screening[J]. J Chem Inf Model, 2020, 60(9): 4263. DOI: 10.1021/acs. jcim. 0c00155

[100] RAMAKRISHNAN R, DRAL P O, RUPP M, et al. Quantum chemistry structures and properties of 134 kilo molecules[J]. Sci Data, 2014, 1: 140022. DOI: 10.1038/sdata. 2014. 22

[101] LIU Xian, GUO Yunhe, PAN Wenxiao, et al. Exogenous chemicals impact virus receptor gene transcription: insights from deep learning[J]. Environmental Science & Technology, 2023, 57(46): 18038. DOI: 10.1021/acs. est. 2c09837

[102] ZHANG Jin, NORINDER U, SVENSSON F. Deep learning-based conformal prediction of toxicity [J]. Journal of Chemical Information and Modeling, 2021, 61(6): 2648. DOI: 10.1021/acs. jcim. 1c00208

[103] KETKAR R, LIU Yue, WANG Hengji, et al. A benchmark study of graph models for molecular acute toxicity prediction [J]. International Journal of Molecular Sciences, 2023, 24(15): 11966. DOI: 10.3390/ijms241511966

[104] CHEN Yuanting, YU Xinxin, LI Weihua, et al. In silico prediction of hERG blockers using machine learning and deep learning approaches[J]. Journal of Applied Toxicology, 2023, 43(10): 1462. DOI: 10.1002/jat.4477

[105] LI Shimeng, ZHANG Li, FENG Huawei, et al. MutagenPred-GCNNs: a graph convolutional neural network-based classification model for mutagenicity prediction with data-driven molecular fingerprints[J]. Interdisciplinary Sciences: Computational Life Sciences, 2021, 13(1): 25. DOI: 10.1007/s12539 - 020 - 00407 - 2

[106] WANG Haobo, WANG Zhongyu, CHEN Jingwen, et al. Graph attention network model with defined applicability domains for screening PBT chemicals [J]. Environmental Science & Technology, 2022, 56(10): 6774. DOI: 10.1021/acs. est. 2c00765

[107] ANAND G, KONIUSZ P, KUMAR A, et al. Graph neural networks-enhanced relation prediction for ecotoxicology (GRAPE) [J]. Journal of Hazardous Materials, 2024, 472: 134456. DOI: 10.1016/j. jhazmat. 2024. 134456

[108] WANG Liguo, ZHAO Lu, LIU Xian, et al. Seppcnet: deeping learning on a 3D surface electrostatic potential point cloud for enhanced toxicity classification and its application to suspected environmental estrogens [J]. Environmental Science & Technology, 2021, 55(14): 9958. DOI: 10.1021/acs. est. 1c01228

[109] DONG Gongqing, WANG Nan, XU Ting, et al. Deep learning-enabled morphometric analysis for toxicity screening using zebrafish larvae[J]. Environmental Science & Technology, 2023, 57(46): 18127. DOI: 10.1021/acs. est. 3c00593

[110] DAGHIGHI A, CASANOLA-MARTIN G M, IDUOKU K, et al. Multi-endpoint acute toxicity assessment of organic compounds using large-scale machine learning modeling[J]. Environmental Science & Technology, 2024, 58(23): 10116. DOI: 10.1021/acs. est. 4c01017

[111] YUAN Qing, WEI Zhiqiang, GUAN Xu, et al. Toxicity prediction method based on multi-channel convolutional neural network[J]. Molecules, 2019, 24(18): 3383. DOI: 10.3390/molecules24183383

[112] LIN Xuan, QUAN Zhe, WANG Zhijie, et al. A novel molecular representation with BiGRU neural networks for learning atom[J]. Briefings in Bioinformatics, 2020, 21(6): 2099. DOI: 10.1093/bib/bbz125

[113] PENG Yuzhong, ZHANG Ziqiao, JIANG Qizhi, et al. Top: towards better toxicity prediction by deep molecular representation learning [C]// 2019 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). San Diego, CA, USA: IEEE, 2019: 318

[114] CHAKRAVARTI S K, ALLA S R M. Descriptor free QSAR modeling using deep learning with long short-term memory neural networks[J]. Frontiers in Artificial Intelligence, 2019, 2: 17. DOI: 10.3389/frai. 2019. 00017

[115] NORINDER U. Traditional machine and deep learning for predicting toxicity endpoints[J]. Molecules, 2022, 28(1): 217. DOI: 10.3390/molecules28010217

[116] ZANG Xuan, ZHAO Xianbing, TANG Buzhou. Hierarchical molecular graph self-supervised learning for property prediction [J]. Communications Chemistry, 2023, 6(1): 34. DOI: 10.1038/s42004 - 023 - 00825 - 5

[117] CHEN Jiarui, SI Y W, UN C W, et al. Chemical toxicity prediction based on semi-supervised learning and graph convolutional neural network [J]. Journal of Cheminformatics, 2021, 13(1): 93. DOI: 10.1186/s13321 - 021 - 00570 - 8

[118] LI Jincheng. Imbalanced toxicity prediction using multi-task learning and over-sampling[C]// 2020 International Conference on Machine Learning and Cybernetics (ICMLC). Adelaide, Australia: IEEE, 2020: 1. DOI: 10.1109/ICMLC51923. 2020. 9469546

[119] JAIN S, SIRAMSHETTY V B, ALVES V M, et al. Large-scale modeling of multispecies acute toxicity end points using consensus of multitask deep learning methods [J]. Journal of Chemical Information and Modeling, 2021, 61(2): 653. DOI: 10.1021/acs. jcim. 0c01164

[120] OLAYO-ALARCON R, AMSTALDEN M K, ZANNONI A, et al. Pre-trained molecular representations enable antimicrobial discovery[J]. Nature Communications, 2025, 16(1): 3420. DOI: 10.1038/s41467 - 025 - 58804 - 4

[121] TEVOSYAN A, KHONDKARYAN L, KHACHATRIAN H, et al. Improving VAE based molecular representations for compound property prediction[J]. J Cheminform, 2022, 14(1): 69. DOI: 10.1186/s13321 - 022 - 00648 - x

(编辑 刘 彤)