

人工智能范式有机聚合物功能材料的研究与设计进展

刘秀敏^{1,2}, 叶财超¹

(1. 南方科技大学材料科学与工程系, 广东 深圳 518055; 2. 南京理工大学化学与化工学院, 南京 210094)

摘 要: 有机聚合物材料的高维度、非线性的化学空间制约了新材料的研发效率, 而人工智能(AI)通过数据驱动模型揭示构效关系, 正推动该领域进入高效设计的新阶段。本文系统综述了 AI 驱动聚合物功能材料研究的新范式。首先, 针对数据瓶颈, 阐述了高通量计算、机器学习与标准化数据库的协同策略; 其次, 针对结构表示难题, 创新性引入“功能基元”概念, 发展了聚合基元指纹(PUFp)与聚合物基元图(PU-Graph)等可解释特征编码方法; 进而总结了从宏观工艺参数到微观结构编码的机器学习预测模型进展, 并重点剖析了 AI 在有机半导体、光伏及电介质材料中的典型运用案例, 展现了“数据-模型”协同驱动范式转型的现状。本文针对有机聚合物功能材料专用数据库的构建与结构编码方法的梳理, 也为第四范式聚合物材料研发提供了数据及方法支撑。最后, 展望了 AI 与材料大数据深度融合的前景, 包括逆向设计、实验-模型闭环集成等关键方向。可以预见, 随着这些问题的突破, AI 将从辅助工具演变为聚合物材料创新的核心引擎。

关键词: 人工智能范式; 有机聚合物功能材料; 功能基元; 数据驱动; 结构-性能关系

中图分类号: O6 文献标志码: A 文章编号: 0454-5648(2026)07-2380-17

网络出版时间: 2026-04-10



材料作为构成物品、器件及其他产品的物质基础, 其每一次革新都推动着生产力与科技文明的跨越。在知识经济时代, 材料与能源、信息并列为现代科技的 3 大支柱, 具有不可替代的战略地位。有机聚合物材料作为材料科学的重要分支, 以其结构可调、性能多样、加工简便和成本较低的优势, 在能源、电子、医疗、航天等关键领域得到广泛应用^[1-3]。有机聚合物材料由一种或多种结构单元(单体)通过化学键重复连接形成, 其分子量范围通常为 $10^3 \sim 10^7$ 。单体作为聚合物的“结构基元”, 其化学组成、结构及排列方式的微小变化, 可通过重复单元、分子链、聚集态至宏观功能的层层传递, 显著影响最终材料性能^[4-5]。当有机聚合物具备特定物理化学性质(如导电、介电、光电、生物相容等), 可满足特定功能需求时, 即被称为有机聚合物功能材料(图 1)^[6-7]。例如, 聚 3,4-乙烯二氧噻吩: 聚苯乙烯磺酸盐(PEDOT:PSS)通过共轭主链与离子复合结构, 实现高电导率(1000~5000 S/cm)与良好水溶性的兼顾, 成为经典导电聚合物^[8]; 有机太阳能电池活性层材料[如聚 3-乙

基噻吩: 苯基 C61 丁酸甲酯(P3HT:PCBM)]依托其宽光谱吸收、匹配能级与溶解性, 不断提升光电转换效率^[9]; 柔性热电材料则结合高电导、低热导与柔韧性, 逐步替代传统无机热电体系^[10]。尽管聚合物功能材料已渗透至众多技术领域, 但其多尺度复杂行为仍使精准设计面临巨大挑战^[11-13]。

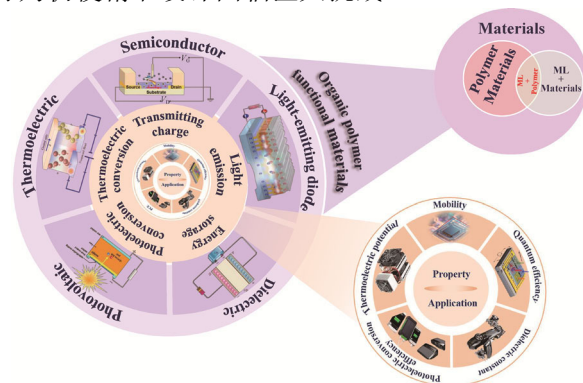


图 1 具有重复单元组成和功能性能的有机聚合物功能材料的应用领域

Fig. 1 Application fields of organic polymer functional materials with repeating unit composition and functional performance

收稿日期: 2025-12-30。 修订日期: 2026-01-25。

基金项目: 国家自然科学基金面上项目(52473235); 江苏省研究生科研与实践创新计划(KYCX25_0718)。

第一作者: 刘秀敏(1996—), 女, 博士研究生。

通信作者: 叶财超(1988—), 男, 博士, 副教授。

Received date: 2025-12-30. Revised date: 2026-01-25.

First author: LIU Xiumin (1996—), female, Doctoral candidate.

E-mail: 1540353710@qq.com

Correspondent author: YE Caichao (1988—), male, Ph.D., Associate Professor.

E-mail: yecc@sustech.edu.cn

自 20 世纪 90 年代以来,随着计算能力的提升,基于第一性原理、密度泛函理论(DFT)、分子动力学(MD)等模拟方法在材料研究中日益普及,推动了材料研发范式从实验试错向计算驱动的转变^[14-16]。然而,面对能源转型与器件微型化对材料性能的更高要求(如高介电常数、高迁移率、高能量转换效率等),传统材料计算方法在处理聚合物这类高维、非线性体系时仍显不足,难以系统揭示其构效关系与功能机制。在此背景下,“材料基因组计划”(MGI)倡导的数据驱动研究范式应运而生,促进人工智能与材料科学的深度融合,标志着材料研发进入以数据密集为特征的“第四范式”^[17-20]。机器学习(ML)作为人工智能的重要分支,能够从海量实验与模拟数据中学习隐藏规律,为材料研究提供了新视角与新工具。目前,其已在多孔材料、合金、聚合物电介质与光伏材料等体系展现出优异的性能预测与设计能力^[21-27]。特别值得指出的是,功能基元(Functional Unit, FU)作为连接微观化学结构与宏观性能的桥梁,为理解与设计聚合物功能材料提供了有力框架。功能基元指材料中承载特定物理化学功能的最小结构单元,其种类、组合与排列方式直接决定材料最

终性能^[28-31]。该基元具有不可替代性,若替换为其他结构基元,则会导致对应功能丧失或显著衰减。功能基元可以包括特定的原子构型、分子相互作用以及表面和界面上的局部结构。与金属、陶瓷和半导体中的晶胞不同,功能基元简化了结构与性能之间的关系,为大数据驱动的新范式提供了一座桥梁。在晶体与有序材料中,基于功能基元的调控已取得一系列突破,如通过引入非线性光学活性基元设计高性能光学晶体^[32],或借助二维纳米基元的有序组装实现超快离子传输的储能薄膜(图 2)^[33]。在聚合物体系中,功能基元同样具有关键作用:通过设计具有特定电子结构、极性或空间构象的基元,可系统调节聚合物的光电、介电、力学等性能^[34]。人工智能(AI)的引入,可进一步揭示功能基元与宏观性能之间的定量关联,使得材料设计从传统的经验试错转向数据驱动的“基元-结构-性能”精准匹配。例如,定量结构-性能关系(QSPR)模型即常以结构描述符(含基元特征)作为输入,预测聚合物性质^[35-39]; Barnett 等^[40]通过基元拓扑与路径指纹构建机器学习模型,筛选出具有高渗透性能的新型聚合物,体现了功能基元在数据驱动设计中的价值。

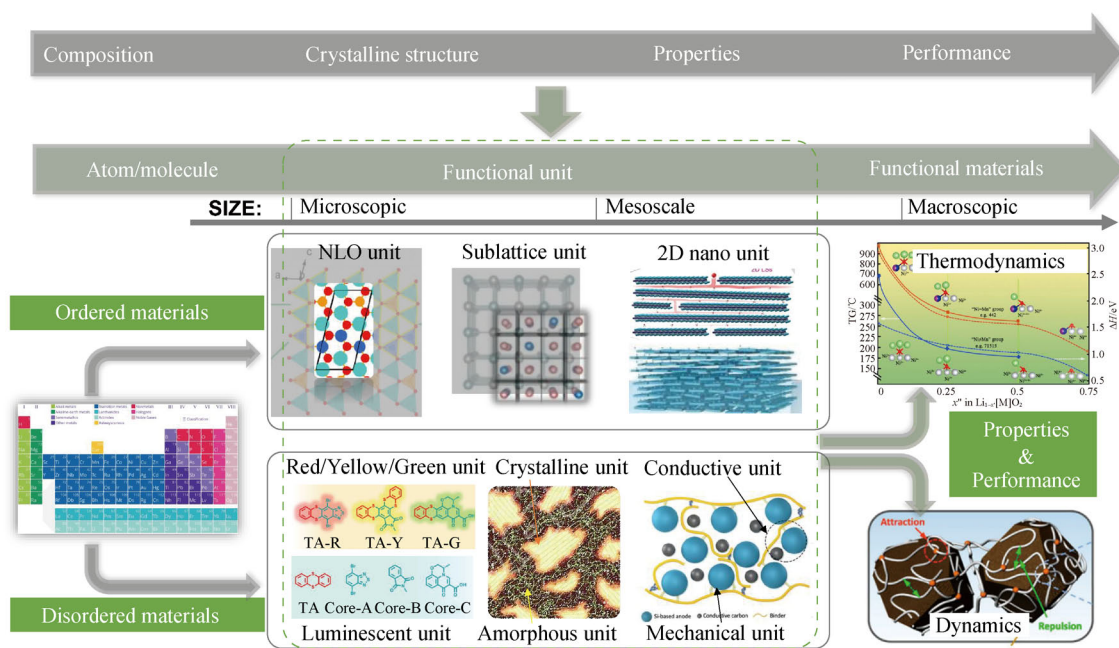


图 2 “功能基元与结构”对新型材料设计与性能调节的探索^[32-34]

Fig. 2 Exploration for novel material design and performance regulation by “functional units and architectures”^[32-34]

尽管机器学习在聚合物性能预测与设计方面已取得进展,但仍面临若干挑战:数据规模与质量受限于合成与表征成本;聚合物分子表示与特征工程仍需研发更有效的编码方法;模型可解释性不足制

约机理洞察;实验-计算闭环尚未完全打通等。这些问题是当前该领域研究的前沿与难点。与已有综述多聚焦于算法框架不同,本综述在梳理 AI 应用于聚合物材料研究一般流程的同时,结合科研范式变迁,

系统探讨 AI 在有机聚合物功能材料领域的进展。本文将从数据库构建、特征表示方法、模型可解释性等方面展开, 总结机器学习在半导体、光伏、介电等功能聚合物中的预测与设计案例, 最后展望利用 AI 数据科学和先进计算方法实现高效有机聚合物功能材料设计的发展前景。

1 有机聚合物功能材料研究的范式演进与人工智能(AI)的引入

材料科学研究历经 4 类范式的演进(图 3): 经验

范式、理论范式、计算范式与数据驱动范式。传统聚合物材料的研发长期依赖实验试错与性能表征, 这一过程资源消耗量大、周期漫长。以热力学为代表的第二范式虽奠定了理论基础, 但在处理广阔而复杂的化学空间时仍显局限。随着理论完善与计算技术发展, 聚合物研究迈入计算模拟阶段, DFT、MD、蒙特卡洛方法以及连续介质模型等手段逐渐得到广泛应用。尽管如此, 传统建模方法在刻画聚合物功能材料从分子尺度到宏观性能之间的跨尺度关联机制方面, 依然存在明显不足。

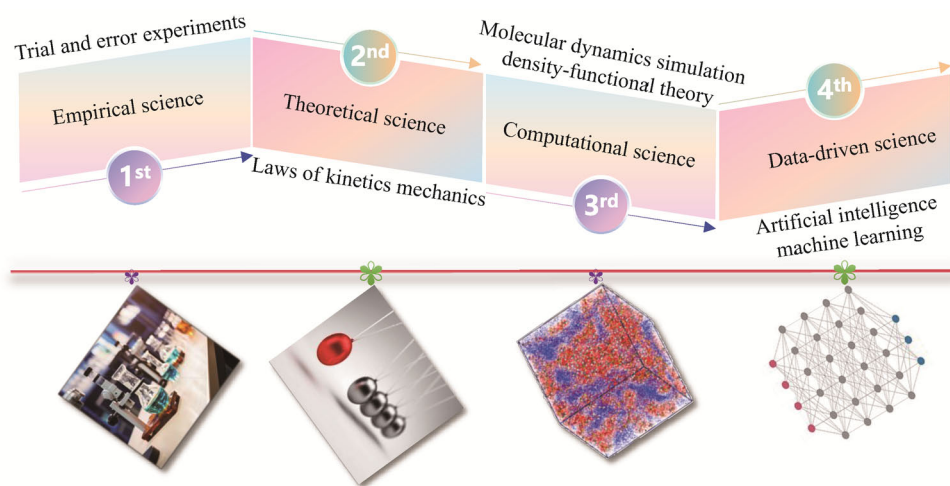


图 3 科学的 4 个范式: 经验型、理论型、计算型和数据驱动型

Fig. 3 Four paradigms of science: empirical, theoretical, computational and data-driven

近年来, 高通量计算与组合实验等方法迅速发展, 推动聚合物研究进入科学大数据阶段。以数据为核心的第四范式, 通过构建目标性能与高维特征之间的关联模型, 有效弥补了传统方法在效率和参数覆盖上的局限, 显著缩短了材料研发周期, 加速了新材料的筛选与研发。例如, 微软开发的 MatterGen 模型能够基于物理化学约束直接生成稳定新材料, 其结构稳定性较此前的 CDVAE 方法显著提升^[41]。同时, 机器学习技术也在自然语言处理、医学诊断、计算机视觉等领域得到广泛应用, 成为人工智能发展的重要方向^[42–46]。

聚合物材料具有多重可调参数, 包括聚合度、链长、化学组成、拓扑结构、立体构型、堆叠密度与化合价等, 为实现特定功能的聚合物设计提供了丰富的维度。历史上, 聚合物功能材料的设计主要依托于对重复单元的功能化调控, 并借助共聚、交联、嵌段等策略实现性能优化。随着技术进步, 受控/活性聚合、自组装及高通量合成等更为精准与系统的方法逐渐成为主流^[47–48]。然而, 聚合物

功能设计往往是非直观的过程, 其巨大的化学空间中仅有极小部分被实际探索, 仅依靠实验手段识别新材料面临显著挑战。因此, 该领域的未来发展必然依赖于高通量实验与数据驱动设计相结合的系统方法, 这正是机器学习与人工智能技术能够发挥颠覆性作用的领域。在材料的成分、结构及合成条件等数据中, 常蕴涵与性能相关的潜在模式与关联规律。通过机器学习挖掘这些数据, 可以识别隐藏趋势、构建 QSPR 模型, 并预测未知材料的性能, 从而为理性设计具备目标功能的聚合物材料奠定基础。

聚合物 AI 研究范式分为: 1) 传统分子指纹与机器学习范式: 通过摩根指纹(Morgan fingerprint)、扩展连接指纹(ECFP)等传统指纹, 将聚合物分子拆解为固定长度的子结构片段, 以二进制或计数向量记录子结构存在/不存在特征, 结合随机森林(RF)、支持向量机(SVM)等机器学习模型构建指纹–性能的预测关系。该范式是聚合物 AI 研究的基础框架, 特征提取速度快, 适用于大规模聚合物库的初步筛

选。2) 分子图和图神经网络(GNN)范式: 以聚合物的原子或重复单元为节点, 化学键为边构建分子图, 通过图卷积网络(GCN)等模型自动学习原子/单元间的拓扑关联与电子相互作用, 实现结构-性能关系建模。目前已开发出基于分子图的多任务预测模型, 可同时预测多种关联性能。然而存在模型训练依赖大规模高质量数据集以及可解释性不足等问题。3) 生成式 AI 范式: 基于 SMILES、BigSMILES 等分子序列表示, 通过变分自动编码器(VAE)、生成对抗网络(GAN)等生成式算法, 在化学规则约束下生成全新聚合物结构, 结合预测模型筛选高性能候选分子。4) 功能基元范式: 通过功能靶向筛选-不可替代性验证-尺度边界界定的准则, 精准锚定与目标性能直接相关的结构单元, 剔除惰性结构干扰, 这是其他范式难以实现的核心优势。通过将功能基元与聚合基元指纹(PUFp)和聚合基元图(Polymer-unit Graph)结合, 解决黑箱模型预测性差及可解释性不足等问题。

当前, 机器学习在聚合物材料研究与设计领域发展迅速, 已涌现出多篇综述文章, 从不同视角总结了相关进展^[49-53]。然而, 这些综述大多集中

于计算方法或算法框架, 对聚合物专用数据库的构建、数据驱动的功能基元表征以及模型可解释性等关键环节论述不足, 未能充分体现机器学习在各类聚合物功能材料研究任务中的广泛适用性。在此背景下, 亟须建立以人工智能为核心的新研究框架, 系统整合数据、算法与领域知识, 从而在探索聚合物领域广阔的特征与性能空间方面取得突破性进展。

2 有机聚合物功能材料中人工智能应用流程

人工智能具备从数据中学习规律、预测趋势、并挖掘新知识的能力, 为应对聚合物领域日益增长的数据量与体系复杂性提供了新机遇。它不仅有助于评估并弥补传统研究中的短板——如数据稀缺、描述符不足、聚合物编码方法不完善及算法复杂度高问题, 更能通过一套结构化的系统流程, 为材料的发现与设计提供系统性支持。如图 4 所示, 该流程通常涵盖 4 个核心且紧密关联的阶段: 数据生成与数据库构建、特征工程、模型构建与评估, 以及模型解释与知识挖掘。

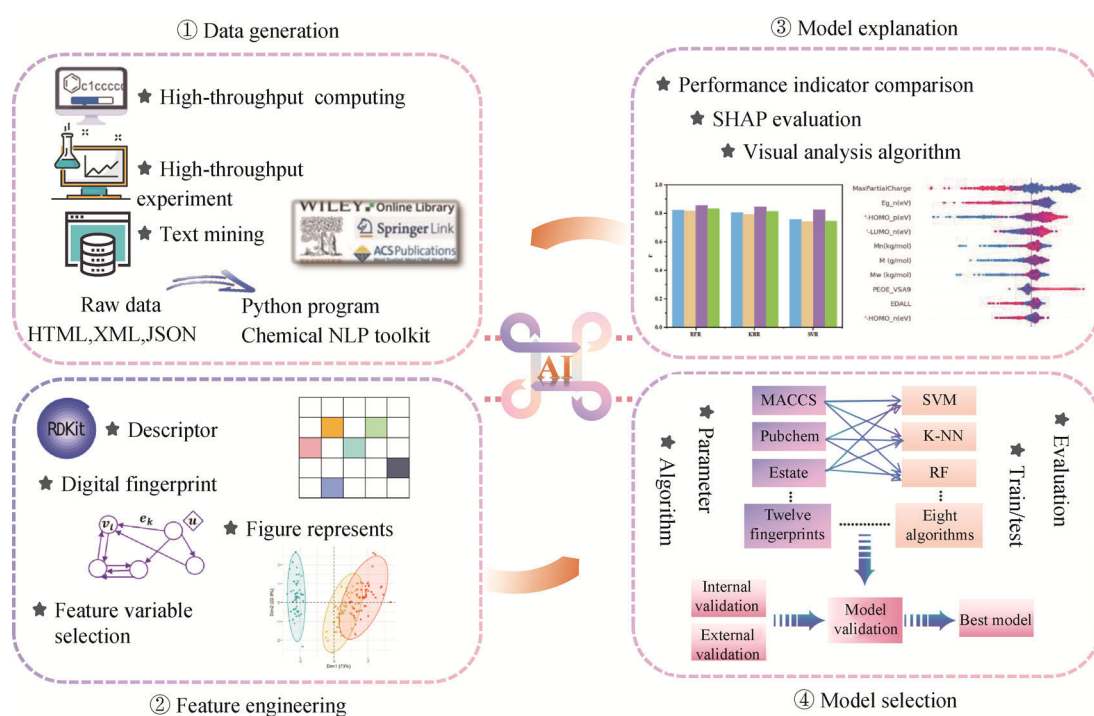


图 4 材料领域机器学习方法应用的一般流程示意图

Fig. 4 General workflow of machine learning methods applied in materials science

2.1 数据生成与数据库构建

材料数据是实现材料研发第四范式的基础, 数据库的构建是数据驱动研究的关键环节。依托材料

基因工程的理念, 实验数据库(如 CSD、COD、ICSD)与理论数据库(如 OQMD、AFLOW)近年来快速发展, 涵盖了无机晶体、有机分子及多种工程材料体

系,为科学研究提供了宝贵的数据资源^[54–56]。然而,数据质量与规模依然是当前数据驱动方法中最主要的瓶颈。实验数据常因仪器误差、样本污染等问题导致准确性受损,进而影响模型预测的可靠性。此外,由于实验本身难度大、成本高,加上聚合物体系固有的多尺度复杂特性,目前面向有机聚合物功能材料(如半导体、光伏、电介质材料)的可用数据集仍然较少,且化学多样性有限。为应对上述挑战,本团队系统整合了理论计算与实验数据,构建了 OPECM 数据库(有机聚合物能源材料数据库)^[57]。该数据库包含 3 大类共 3225 种聚合物,并在以下 3 方面取得关键进展:1) 建立了覆盖 3 类材料的实验性能基准,兼具稳定的分子结构与较高的化学多样性;2) 引入“功能基元”作为数据组织的基本构建单元,将材料功能映射至可设计的结构单元(如半导体/光伏材料的 π 共轭核、电介质材料的极性基团);3) 为每种聚合物补充了基于高通量 DFT 计算的量子化学描述符,从而使 OPECM 成为支持有机聚合物功能材料 QSPR 建模的综合数据资源。

OPECM 这类高质量、结构化的数据库,其构建依赖于高效、可靠的数据获取方法。目前,规模化数据生成主要依托 3 个相辅相成的途径:高通量实验、高通量计算与文本挖掘。这些方法协同作用,共同构成了数据驱动研究的核心基础设施。

2.1.1 高通量实验 高通量实验借助自动化、并行化与微型化技术,实现材料的批量合成、表征与性能测试,能够快速积累结构–性能关联数据,直接推动材料筛选与应用。其核心在于实验设计、样品平行制备与快速表征三者的协同闭环,有效解决了传统实验周期长、数据零散的问题,已成为数据驱动材料科学研究的重要支柱^[50, 58]:

1) 实验设计:旨在系统、高效地规划实验方案,通过合理选择与组合实验变量,建立工艺–结构–性能之间的定量关系,从而指导新材料合成与优化。在聚合物功能材料研究中,常用的实验设计方法(如全因子设计、正交试验设计及响应面设计等)能够通过数学方法精心安排的实验方案,在大幅减少所需实验次数的同时,系统且高效地分析与评估多个因素及其交互作用对材料性能的影响,从而快速导向优化目标。近年来,机器学习辅助的实验设计逐渐兴起,可基于已有数据建立代理模型,预测不同合成路径或配方的性能趋势,并推荐潜在的高性能候选体系,实现更具导向性的实验规划。例如,在聚合物光伏材料开发中,可通过设计不同给体单元的

化学结构、侧链类型与聚合条件,系统研究其对吸收光谱、能级与最终器件效率的影响,从而高效定位优势材料体系^[59]。

2) 样品平行制备:追求在精准控制合成/制备条件(如温度、时间、试剂用量)的前提下,仅改变目标参数,实现数百个样品的高通量制备。该方法将样品用量从传统毫克级降至微克级,显著降低了试剂成本与实验风险,同时减少了人为操作误差,大幅提升了实验效率。目前,薄膜气相沉积法是应用最为广泛的平行制备技术之一,该技术在真空、低压或等离子体环境下,将材料转化为气相后在基片表面沉积成膜^[60]。

3) 样品快速表征:作为高通量实验的数据输出环节,要求检测快速高效、能够多维度覆盖材料特性并实现自动化集成。分子结构表征(如高通量 FT-IR)、聚集态结构表征[如高通量 X 射线衍射 (XRD)]以及宏观性能表征(如 Newport Oriel Sol3A)等主流技术的应用,进一步提升了高通量实验的整体研发效率^[61–62]。

2.1.2 高通量计算 高通量计算融合了第三范式(计算模拟)与第四范式(数据驱动),基于量子力学方法批量获取材料的结构、能量及电子特性参数,为材料筛选、性能预测与逆向设计提供关键数据基础。计算精度与成本在很大程度上取决于泛函和参数的选择^[63]。高通量计算通常包括 4 个主要步骤:1) 初始结构获取;2) 自动化任务管理;3) 数据提取与标准化;4) 数据质控与筛选。

聚合物初始结构可通过查询实验数据库(如 OpenPoly、PolyInfo)获取。对于已知聚合物的结构信息,通过 SMILES/BigSMILES 编码统一表征。例如利用 SMILES 字符串生成聚合物分子链,或从聚合物数据库中筛选特定构型,经过去除不合理及重复结构、几何优化与能量最小化等预处理后,得到合理的计算初始模型^[64]。此外,生成式模型(如生成对抗网络、变分自动编码器)也可用于产生大量可靠结构作为计算输入。基于高性能计算(HPC)集群搭建自动化任务管理框架,将初始结构库中的每个聚合物体系封装为独立计算任务。自动化任务管理依托脚本驱动、工作流工具与高性能计算集群调度系统实现。根据性能需求分级配置计算参数,通过 Python 脚本读取聚合物初始结构与计算参数,批量生成 VASP、GROMACS 等软件的输入文件,并借助 SLURM、PBS 等作业调度系统提交任务,完成队列管理与资源分配,实现成百上千个计算任务的

并行执行。同时嵌入断点续算、异常重试机制,通过软件应用程序接口(API)实时监控任务运行状态,自动跳过计算失败的体系,确保数万至数十万量级的聚合物计算任务高效运行,大幅降低人工干预成本。数据提取与标准化是指计算完成后通过程序化脚本(Python/Shell)自动提取各聚合物体系的计算结果,包括聚合物的结构参数、性能数据等,统一其格式、单位与命名,并与初始结构及计算参数关联,构建结构化的聚合物数据库,避免计算数据碎片化。精准可靠的数据集有助于揭示目标变量与自变量之间的潜在关系,为确保数据质量需进行严格的数据质控与筛选,常用方法包括归一化、阈值筛选等。同时采用一致性校验、重复性验证等方式评估数据可靠性,从质控后的有效数据中快速筛选出满足性能要求的聚合物候选体系,为后续建模与预测提供高质量的数据基础^[65-67]。

在构建 OPECM 数据库的过程中,本团队系统整理了有机半导体、光伏及电介质聚合物材料的实验性能数据及其 SMILES 表示,并对所有 SMILES 字符串进行了标准化处理以确保一致性。随后,基于 Python 脚本生成三维原子坐标,并采用 Gaussian 16 软件包^[68]进行初始几何优化以获得初步结构。所有结构均经过分子可视化软件校验,以排除几何上不合理的构型,从而形成最终的初始计算模型。所有计算均在 B3LYP/6-31G(d)理论水平下完成。最终,利用 pymatgen、cclib 等工具脚本自动提取目标参数,形成了兼具唯一性、准确性与一致性的 OPECM 聚合物数据库。

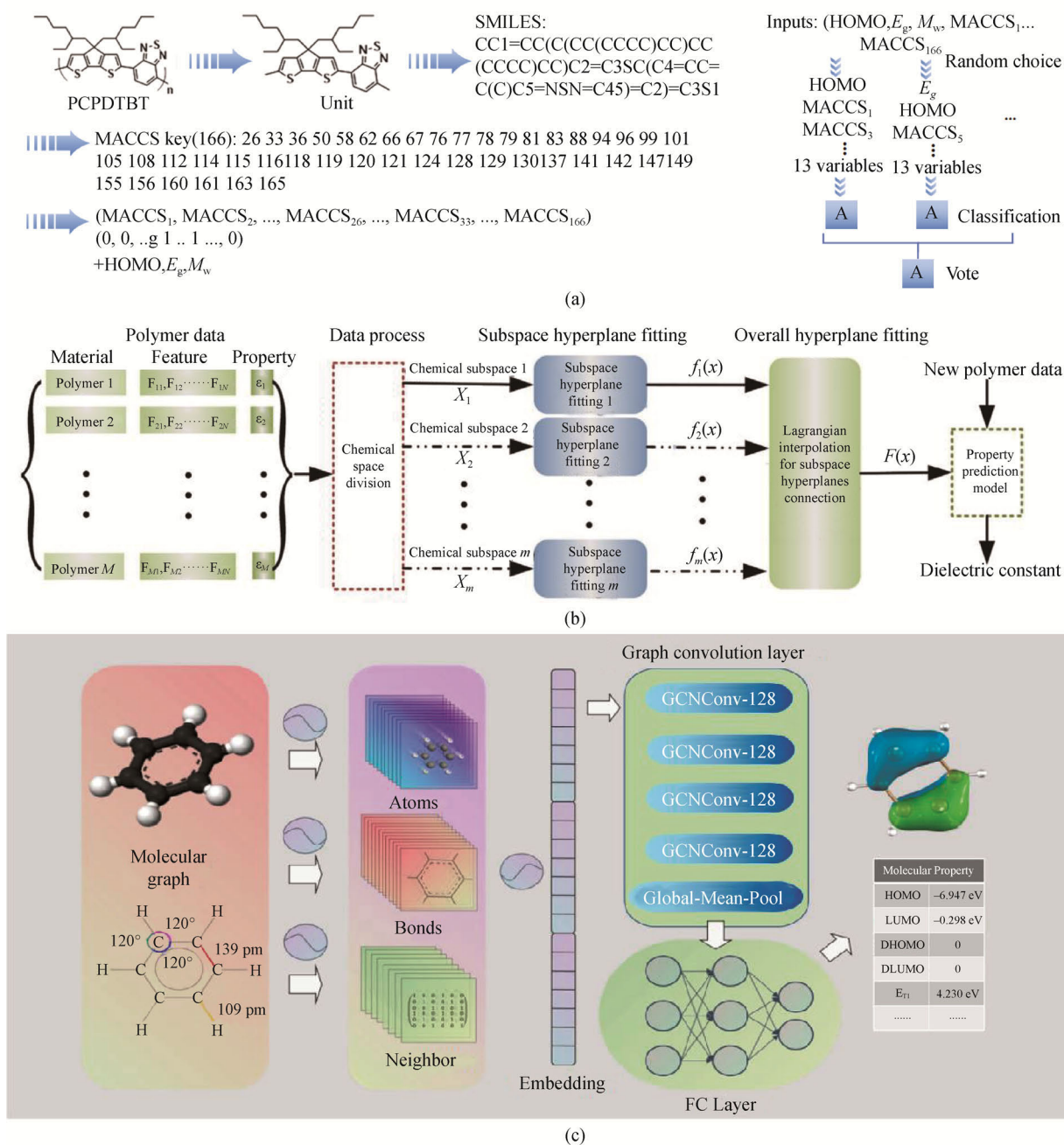
2.1.3 文本挖掘 随着互联网与科学出版的快速发展,海量的科学文献已成为获取材料信息的重要载体。然而,这些文献通常以 HTML、JSON 或 XML 等非结构化格式存储,难以直接用于数据分析。文本挖掘技术借助自然语言处理(NLP)、信息抽取与机器学习方法,能够从海量文本中自动识别并提取“结构-性能-参数”等关键关联信息,从而将其转化为结构化数据^[69-71]。随着大数据与机器学习方法的进步,文本挖掘在化学与材料科学中的应用日益广泛,显著加速了新催化剂、药物及光伏材料等领域的发现进程。例如, Huang 等^[72]利用文本挖掘技术,从超过 22 万篇学术论文中自动提取并构建了一个包含 292 313 条数据记录的电池材料数据库。该数据库的生成涵盖了文献检索、数据抽取、清洗、后处理及评估等多个步骤,为基于机器学习、数据挖掘和统计方法的电池材料设计与预测提供了宝贵的资

源。据报道,这是首个从大规模科学文献中全自动构建的电池材料数据库。研究者还提供了图形用户界面(GUI)以方便使用,对推动电池材料的智能化研发具有重要价值。Dong 等^[73]采用化学感知软件包 ChemDataExtractor 2.0(融合 NLP 与机器学习技术),基于 128 776 篇期刊文章,自动生成含 100 236 条有机半导体材料带隙记录及相关温度信息的数据库。对数据库的评估显示,加权精确率为 84%,加权召回率为 65%。这是迄今为止最大的开源非计算带隙数据库。数据库记录以 CSV 格式提供,同时提供机器可读的 JSON 和 MongoDB 格式,可以帮助数据挖掘和半导体材料发现。

2.2 特征工程: 从结构到可解释描述符

特征工程旨在从材料的物理、化学及结构特性中提取低维、可解释的关键描述符,从而在数据驱动设计与微观机理之间建立可计算的桥梁。该过程主要包括特征构建与特征选择两部分。在宏观层面,工艺条件、材料成分等变量可与目标性能参数直接关联,构建高精度预测模型,以指导材料性能优化。例如, Altarazi 等^[74]以碳酸钙的粒径和质量分数为输入,以抗拉强度、密度和延展性为目标,利用人工神经网络建立了 PVC 复合材料力学性能的预测模型,有效缩短了实验周期。在微观层面,常使用 SMILES 字符串、分子指纹或分子图等表示方法作为输入,绕过复杂的物理机制分析,直接预测材料性能。Nagasawa 等^[75]较早将监督学习引入有机光伏共轭聚合物的设计中,基于约 1000 个样本的实验参数(包括 HOMO 能级、分子量及 MACCS 指纹编码的分子结构)构建了人工神经网络和随机森林模型,其中随机森林表现出更高的预测精度,为聚合物主链与侧链设计提供了数据支持(图 5a)。Yi 等^[76]提出 OGA-ISVR 模型,用于预测聚合物弹性体在不同频率下的介电常数(ϵ_r),并揭示了特征与性能之间的映射关系,其预测效果优于高斯过程回归和核岭回归等方法(图 5b)。Wang 等^[77]将分子结构表示为图(原子为节点,化学键为边),借助图卷积神经网络与 LightGBM 模型,建立了从分子结构到光伏器件效率的定量预测框架,实现了有机光伏分子的高效筛选(图 5c)。这些方法通过特定的结构编码,将复杂的聚合物信息转化为机器可读的表示,在保留本征信息的同时简化了模型输入,从而实现了性能的快速、准确预测。

然而,在 AI 驱动的有机聚合物功能材料研究中,传统分子指纹或分子图表征往往可解释性较弱,其片段划分方式也存在一定随机性,难以准确反映

图 5 不同数据表示方法^[75-77]Fig. 5 Different data representation methods^[75-77]

聚合物骨架的区域化学差异，因而在识别对性能具有关键影响的子结构时面临局限。由于缺乏明确的功能基元导向，这些表示方法在构建 QSPR 模型时仍存在挑战。功能基元的划分并非主观割裂分子结构，而是先初步锁定与功能相关的分子片段；随后通过控制变量实验验证候选片段的功能不可替代性，即替换该片段为其他结构后，若材料的核心性能出现显著衰减，则判定为功能基元；最后，在尺度边界方面，以最小功能单元为边界，避免基元过度碎片化或过度冗余。为此，本团队提出了面向有

机聚合物功能材料的“功能基元”概念，并发展了 2 种基于基元的结构表示方法：PUFp 与 Polymer-unit Graph(图 6)^[78-79]。PUFp 采用独热(one-hot)向量对基元类型及数量进行编码，不涉及连接方式；Polymer-unit Graph 则以节点表示功能基元、边表示连接关系，将结构、属性及拓扑信息嵌入节点与边矩阵中，为定量解析基元-性能关系提供了清晰而有效的工具。这些方法显著提升了模型的可解释性与物理意义，推动了有机聚合物功能材料的 AI 理性设计，并具备向其他材料体系拓展的潜力。

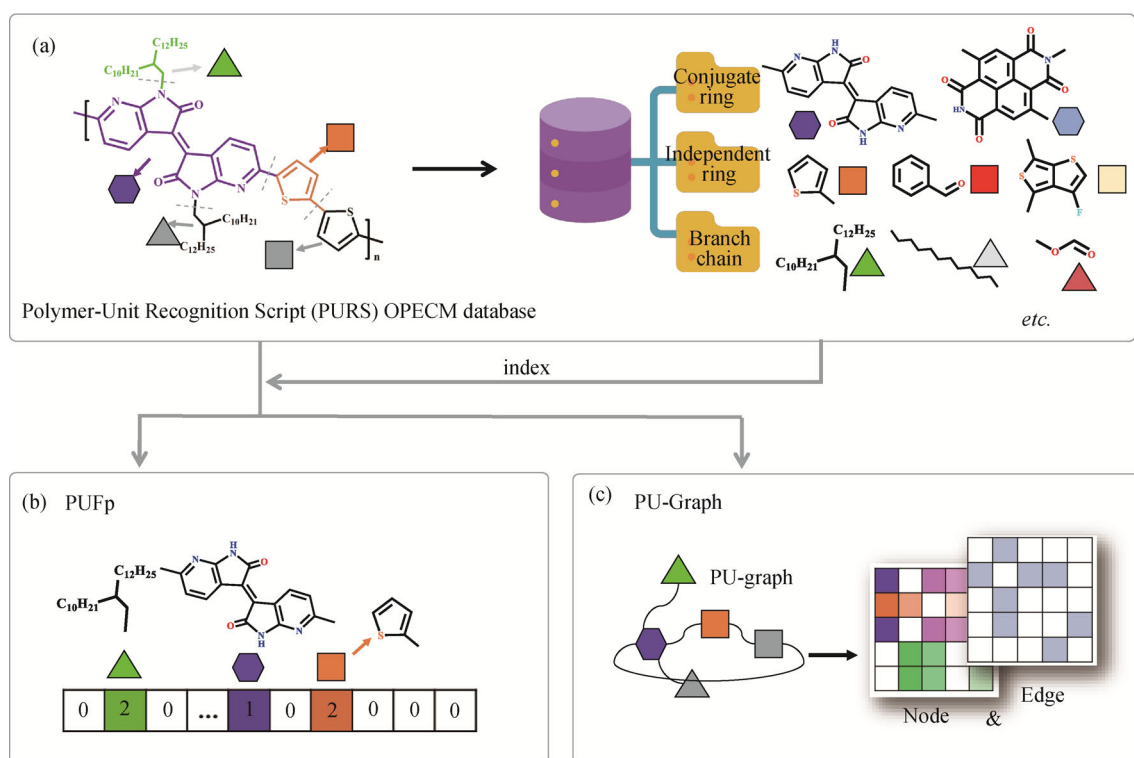


图 6 (a) 用于从有机聚合物功能材料中识别功能基元的 PURS 和功能基元数据库; (b) 基于功能基元的指纹 (PUFp) 和 (c) 基于功能基元的图 (PU-Graph), 用于机器学习的聚合物有机功能材料的可解释表达

Fig. 6 (a) PURS for identification of functional units from organic polymer functional materials and functional units database; (b) Functional-unit-based fingerprint (PUFp) and (c) functional-unit based graph (PU-Graph), accessible expression of polymer organic functional materials for machine learning

特征变量选择旨在剔除冗余、保留与目标性能强相关的变量, 从而降低模型复杂度和过拟合风险。通过合理的特征筛选(如采用降维或重要性评估方法), 可在减少计算开销的同时提高模型精度与可解释性。常用方法包括基于相关性分析的可视化(如散点图、热图)、线性或非线性的降维(如 PCA、*t*-SNE), 以及基于互信息、SHAP 值的重要性排序等, 这些手段共同指导特征集的优化与最终模型的构建。

2.3 模型构建、筛选与评估

模型开发包括算法选择、结构调参、模型训练与质量评估等关键步骤。算法通常根据学习模式分为监督学习、无监督学习和半监督学习。监督学习^[80]基于带标签样本进行训练, 适用于回归与分类任务; 无监督学习^[81] (如聚类) 则用于挖掘未标注数据中的潜在结构与模式。半监督学习^[82] 融合少量标注与大量未标注样本, 在标注成本高昂的场景中尤为适用。传统机器学习方法虽可解释性较强, 但依赖人工特征工程, 处理高维、非线性聚合物数据时存在局限。因此, 常需与深度学习相结合, 利用其自动特征提取能力, 以应对材料设计中的复杂挑战。

深度学习通过多层神经网络自动学习数据的层次化特征表示, 在图像、序列及图结构数据处理中展现出强大优势, 也已广泛应用于材料建模。其常用架构包括: GNN^[83], 可直接对分子或聚合物的图结构进行建模, 天然适用于具有明确拓扑关系的材料体系; Transformer 网络, 擅长捕捉序列或结构中的长程依赖关系, 已成为自然语言处理中的主导架构; 卷积神经网络(CNN)^[84], 多用于处理显微图像、光谱或空间网格数据。

模型调参是优化其性能的关键步骤, 需根据任务复杂度与数据规模, 调整网络深度、宽度、激活函数等结构参数, 以及学习率、正则化强度等训练超参数, 以在拟合能力与泛化性之间取得平衡^[85]。在训练阶段, 数据通常划分为训练集、验证集和测试集。通过在训练集上最小化损失函数来学习“结构-性能”映射, 并借助验证集监控过拟合、筛选最优参数。最终, 使用完全独立的测试集, 以均方根误差(RMSE)、决定系数(R^2)或分类准确率等指标, 客观评估模型的泛化能力。常见的评估策略包括 *k*-fold 交叉验证、Holdout 验证、Bootstrapping 验证等, 其中 *k*-fold 交叉验证可通过多次划分训练-测试子集

并取性能均值, 提供更稳健的模型性能估计。

2.4 模型解释与知识挖掘

模型解释旨在揭示预测背后的逻辑, 将“黑箱”输出转化为可理解的化学或物理洞察。这些洞察能够直接反馈, 指导研究者优化或发明新的特征描述符(如功能基元)及新材料设计策略, 从而形成“设计–预测–解释–再设计”的材料研发闭环。常用的解释方法包括: 1) 性能指标分析: 通过 RMSE、MAE、 R^2 等回归指标, 或准确率、精确率、召回率等分类指标, 从宏观上评估模型的预测质量与可靠性, 确立解释的信任基础。2) 特征重要性分析: 这是连接特征描述符(如功能基元)与宏观性能的桥梁。例如, SHAP 等方法能够量化每个特征(如特定功能基元、描述符)对单个预测或整体模型的贡献度, 从而在全局和局部层面揭示模型决策的化学依据^[86], 直接帮助研究者识别出对目标性能(如高迁移率、高介电常数)起关键促进或抑制作用的基元, 为定向分子工程提供明确指南。3) 深度学习可视化: 针对神经网络, 可通过可视化分析算法(如 Sensitivity Analysis、Guided Backpropagation、Layer-wise Relevance Propagation), 将隐藏层的激活或权重映射回输入空间, 以可视化的形式展示模型所关注的结构特征, 适用于解释分子图像、结构序列等非结构化数据^[87–88]。

综上, 人工智能技术以数据生成–特征构建–模型训练–机理挖掘为技术链路, 凭借其强大的数据学习、趋势预测及新知识发现能力, 为解决聚合物领域数据体量激增与体系复杂性攀升带来的技术挑战提供了创新性研究范式。该结构化应用流程不仅有效弥补了传统研究在数据稀缺性、描述符构建、聚合物分子编码体系设计及算法复杂度调控等方面的短板, 更为功能聚合物的高效发现与精准设计提供了系统性技术支撑。为进一步具象化人工智能技术在聚合物领域的应用场景和实施效能, 下文将聚焦有机半导体、有机光伏材料及聚合物电介质材料三类具有代表性的功能聚合物体系, 通过典型应用案例的系统解析, 深入阐述上述 AI 应用流程在具体材料研发中的创新突破, 以期对相关领域的研究者提供可借鉴的技术参考与研究思路。

3 人工智能在有机聚合物功能材料中的应用案例

人工智能凭借其强大的泛化与特征提取能力, 为有机聚合物功能材料的设计、性能预测与工艺优

化建立了新范式。它能够有效关联“组成–结构–性能”之间的高维复杂关系, 在从分子到聚集态的多尺度层面进行协同设计, 从而显著提升材料研发的效率与成功率。本节将围绕有机半导体、光伏和电介质三类代表性功能材料, 系统阐述人工智能在该领域的最新应用进展, 并结合本团队在相关方向的研究工作进行讨论。

3.1 半导体材料

有机半导体材料因具备质轻、柔性、可溶液加工及成本较低等优势, 在场效应晶体管、柔性显示与热电转换等领域具有广阔应用前景^[89–91]。其中, 载流子迁移率是衡量其性能的关键指标之一, 因此开发高迁移率有机半导体(OSCs)始终是研究重点。高性能 OSCs 的设计核心在于识别决定其电荷传输特性的关键结构单元(功能基元), 并理解其在宏观性能中的作用机制。

本团队提出了一种基于 PUFp 的机器学习框架, 用于高效设计与筛选高迁移率 OSCs^[78]。该框架的核心是一个基于 Python 开发的聚合物基元识别程序(PURS, 程序链接: <http://github.com/yecaichao/PURS-PUFp-PUGraph>), 其能够自动识别聚合物重复单元中的结构基元, 并生成对应的 PUFp 编码(图 7a)。PUFp 以一维向量的形式表征聚合物中各类基元的出现频次, 为机器学习提供了可计算的结构输入。基于收集的 678 个 OSCs 样本数据, 以 PUFp 作为结构描述符, 构建了迁移率分类预测模型, 实现了高达 85.2% 的预测准确率。利用 PURS 程序, 我们系统识别并筛选出超过 400 个结构基元, 构建了“聚合物单元数据库”, 并进一步分析识别出对迁移率具有显著影响的关键基元。该方法将高分子材料的结构解析至“基元”层面, 为研究“功能基元–性能”关系奠定了数据基础, 并实现了从被动性能预测向主动结构设计的范式转变。所提出的方案证明了通过机器学习预筛候选材料的能力, 为高迁移率 OSCs 的发现提供了一条高效路径。

为进一步增强模型的可解释性并更精准地表征聚合物拓扑结构, 本团队发展了一种基于 GNN 的 Polymer-unit Graph 表示方法(图 7b)^[79]。在该图中, 每个功能基元被视为一个节点, 基元之间的化学连接则构成边。通过图嵌入技术提取节点与边的特征, 从而将复杂的聚合物结构转化为适于 GNN 处理的图数据。这种表示方法不仅增强了 GNN 模型对聚合物结构信息的捕获能力, 也使得模型决策过程更为可视化与可解释。

基于此,本团队开发的 PU-gn-exp 模型能够从基元视角系统分析结构修饰(如支链工程、氟化取代、给体-受体组合)对 OSCs 性能的影响机制。在监督学习任务中,所构建的 PU-MPNN 模型在保持与基线模型相当预测精度的同时,将训练时间

缩短至后者的 2%,显著提升了计算效率。总而言之, Polymer-unit Graph 成功地将“功能基元”概念深度融入了 GNN 框架,为从基元层面精准分析和理解有机半导体材料的结构-性能关系提供了有力工具。

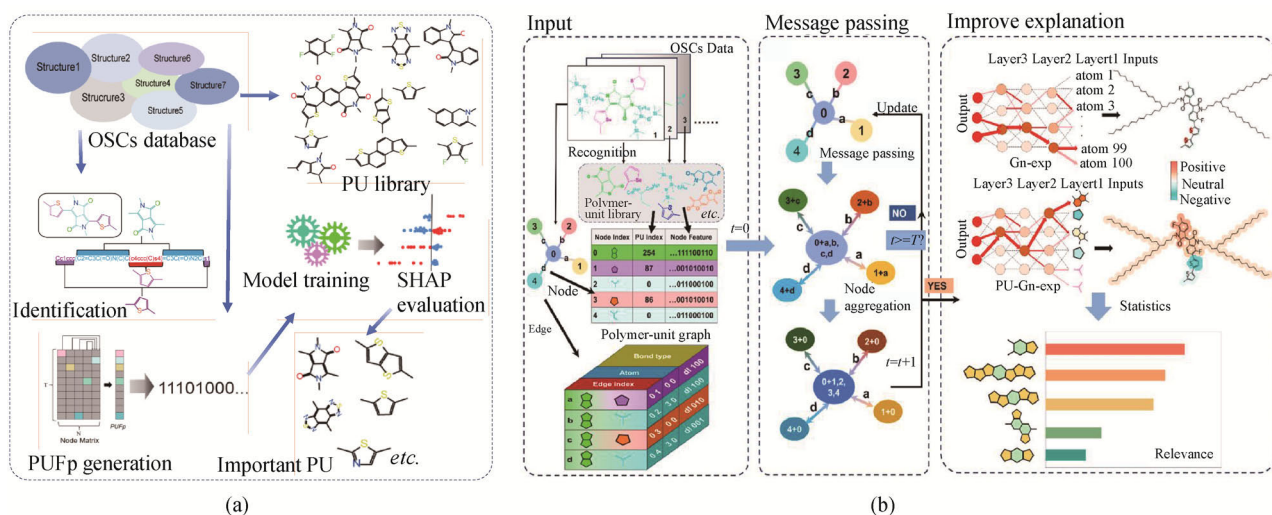


图 7 (a)PUFP 生成工作流程及其在 ML 开发高迁移率 OSC 材料中的应用; (b)聚合基元图的应用概述^[78-79]

Fig. 7 (a) Scheme of the PUFP generation workflow and its application in developing high-mobility OSC materials with ML; (b) Overview of the application of the Polymer-unit Graph^[78-79]

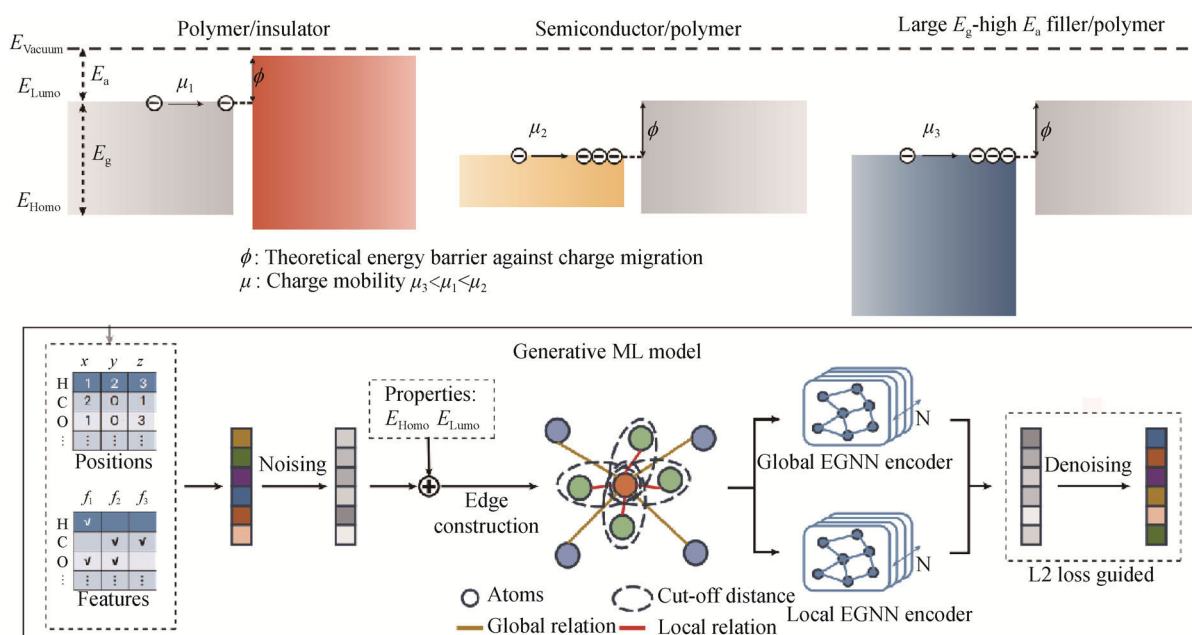
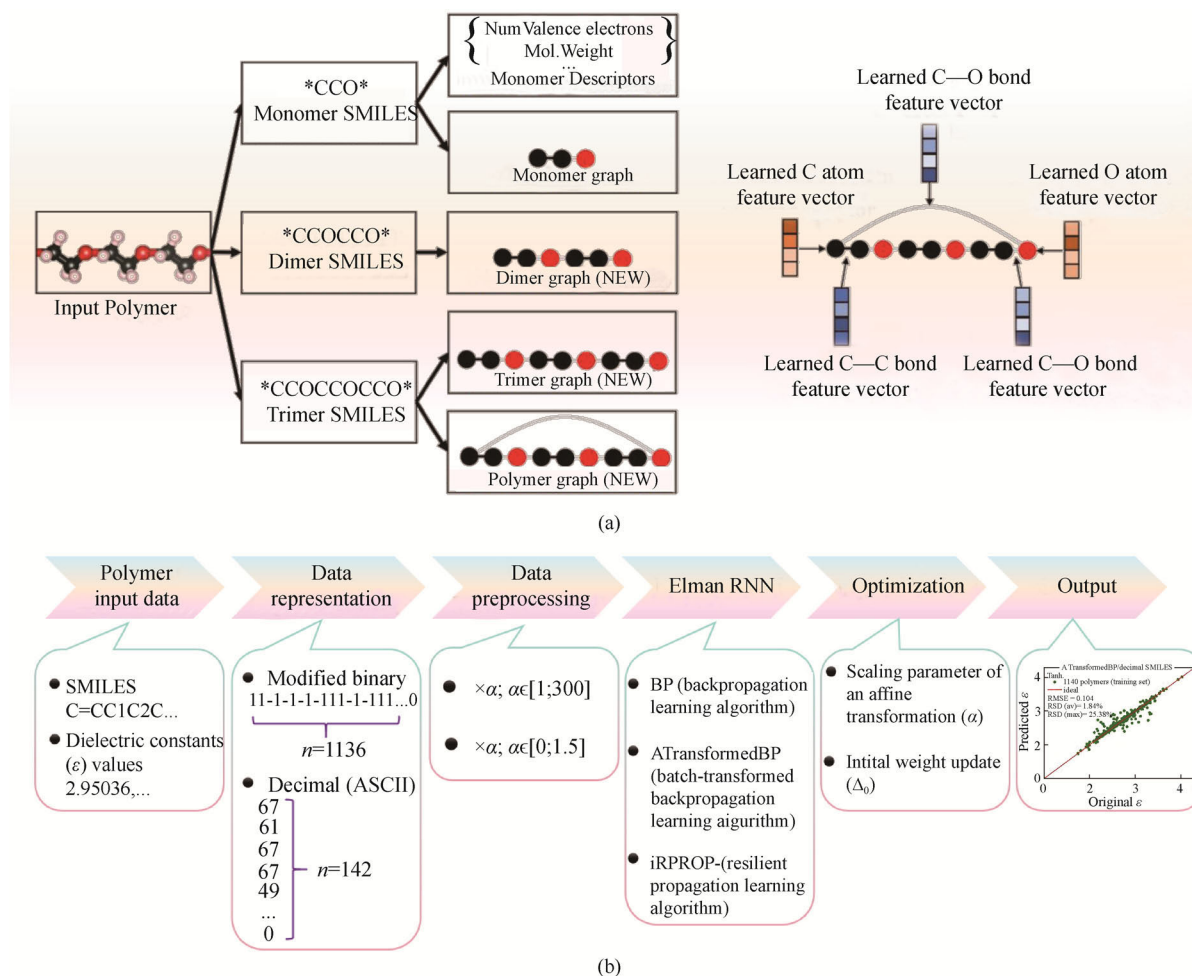
3.2 光伏材料

随着全球环保意识提升及对可再生能源需求的持续增长,有机光伏(OPV)技术凭借其成本低廉、重量轻、柔性好及可大面积溶液加工等突出优势,已成为新能源领域的研究热点。其中,非富勒烯受体(NFA)材料因其具有宽光谱吸收、易于调控的分子能级以及更高的光电转换效率潜力,受到了广泛关注^[92-94]。人工智能技术通过挖掘海量实验与计算数据,正加速高性能 OPV 材料,特别是 NFA 体系的发现与优化进程。

已有多个研究团队利用机器学习有效预测了 OPV 器件的关键性能指标。例如, Wu 等^[95]基于 565 个供体/受体组合的实验数据集,系统比较了多种机器学习算法对功率转换效率(PCE)的预测能力,发现提升回归树(BRT)和 RF 模型表现最优,并成功筛选出六种具有潜力的材料组合。Kranthiraja 等^[96]的工作则进一步扩展了数据规模,他们使用 RF 模型对包含聚合物给体与非富勒烯小分子受体的 OPV 体系进行预测,取得了高达 0.85 的相关系数,并对超过 20 万个虚拟共轭聚合物进行了高通量筛选,其合成验证结果与预测值相符,为 NFA-OPV 的理性设计提供了通用性范例(图 8a)。Sun 等^[97]提出了基于深度学习的 DeepAcceptor 框架,用于高效小分子受

体材料的设计与发现。该工作将图表示学习融入双向编码器表示(BERT),定制 abcBERT 模型(以受体分子中原子、键及连接信息为输入)预测 PCE,并结合 DFT 计算数据集与文献实验数据集进行预训练和微调。基于该模型为 PM6 设计受体材料。DeepAcceptor 提供的用户友好界面,显著提升了高性能受体材料设计与发现的便捷性和效率。

为在合成前高效鉴定高性能材料并深入理解其“结构-光伏性能”关系,本团队基于 1343 个经过实验验证的 OPV 受体材料数据,系统开发并比较了 5 种机器学习模型(图 8b)^[98]。结果表明,RF 模型预测性能最佳,并进一步验证了本团队提出的 PUFP 在预测精度上显著优于传统的 MACCS 指纹。PUFP 将分子结构分解为明确的功能基元进行编码,极大增强了模型的结构可解释性,使得准确预测由给体(D)与受体(A)基元组合所构成的共轭聚合物/大分子的 PCE 成为可能。利用构建的 PUFP-ML 模型,成功对大量受体材料进行了预评估与分类,并精准筛选出 2 种最具潜力的 NFA 候选材料。这项工作表明,将功能基元层面的特征表示与机器学习相结合,能够高效实现 PCE 预测与材料筛选,为下一代高性能 OPV 受体材料的设计提供了新思路与新工具。

图 9 用于输出大 E_g 和高 E_a 填充剂的生成式机器学习模型示意图^[104]Fig. 9 Schematic diagram of generative ML model for outputting fillers of large E_g and high E_a ^[104]图 10 (a)聚合物电介质材料表示方法；(b)基于 RNN 的聚合物介电常数预测流程图^[105–106]Fig. 10 (a) Polymer dielectric materials representation method; (b) Flowchart of our RNN-based polymeric dielectric constant prediction^[105–106]

割化学键,就能完整保持聚合物中原子的成键环境,进而与消息传递神经网络(MPNN)结合,在玻璃化转变温度、介电常数等 10 项性质的预测中,相较传统单体表示方法平均误差降低约 20%。这项工作表明,研发能更好反映聚合物复杂化学结构的表示方法,是提升机器学习预测准确性的关键;同时,将化学直觉融入表示构建,形成物理信息嵌入的机器学习框架,是提高模型泛化能力的有效策略。

另一方面,序列模型在处理线性聚合物表示方面也展现出优势。Nazarova 等^[106]采用递归神经网络(RNN)模型,建立了从 SMILES 字符串到介电常数的映射关系(图 10b)。他们通过 2 点优化策略提升了学习效率与精度:一是对比了 SMILES 的二进制与十进制表示,发现二进制格式能提供更有利的结构信息;二是引入了输入序列的仿射变换反向传播与弹性权重更新算法。优化后的 Elman 型 RNN 在预测 α -烯烃类聚合物介电常数时,显著优于未优化模型,验证了序列模型在捕捉聚合物非线性构效关系方面的能力。

上述研究表明,结合适配聚合物的结构表示方法(如图表示、序列表示)与合适的机器学习算法(如 MPNN、RNN),可显著提升对介电性能等关键性质的预测可靠性,从而为高性能聚合物电介质的高通量虚拟筛选与目标导向设计提供了强有力的计算工具。

针对聚合物电介质材料体系庞大、结构-性能关系复杂的问题,采用无监督学习方法实现材料的精准分类与化学空间可视化,为后续靶向设计提供指导。Li 等^[107]提出了一种机器学习驱动的策略,利用前馈神经网络来预测关键的代理参数,并从近 50 个库中向下选择聚合物候选物,借助高效和模块化的硫氟化物交换点击化学,成功合成和验证所选候选物。从 Polymer Genome 中编译了一个包括 6906 个具有报告 T_g 值的聚合物和 3381 个具有报告 E_g 值的聚合物数据集,通过主成分分析(PCA)完成高维特征降维,再利用 t -SNE 将特征映射至二维空间,结合层次聚类算法实现电介质材料耐热聚硫酸酯的无监督分类。该工作利用 ML 辅助方法发现和优化尖端介电聚合物的飞跃,为高温应用中聚合物介电薄膜的性能提供了新思路。

4 结论与展望

随着高通量计算、自动化实验以及标准化材料数据库的快速发展,材料科学已迈入以大数据为驱

动的新阶段。尽管目前高精度、多尺度的材料数据仍相对稀缺,但海量的结构与稳定性数据正为人工智能模型的训练与性能提升提供日益坚实的数据基础。在聚合物功能材料领域,人工智能的应用已取得一系列显著进展,尤其在材料逆向设计、性能精准预测及合成工艺优化等方面展现出强大潜力,持续推动该领域从依赖经验的传统模式,加速转向以数据和算法为核心的新范式。然而,要充分释放人工智能在聚合物研发中的全部潜力,仍需系统克服数据、算法及系统集成层面的多重挑战。本文系统展示了面向有机聚合物功能材料的专用数据库构建、基于功能基元的新型结构编码方法,以及可解释的构效关系解析框架,旨在弥合模型“黑箱”输出与人类可理解的化学知识之间的认知差距。我们相信,这一系列工作能够为促进数据驱动的人工智能创新与聚合物材料研究的深度融合,提供从数据基础、特征表示到模型理解的多维度支撑。

展望未来,为推动人工智能在有机聚合物功能材料研究中实现从辅助工具向创新核心的根本转变,应重点围绕以下方向系统开展研究,兼顾技术落地路径与核心挑战应对:1) 构建全面、开放且可互操作的标准数据库。以覆盖更广化学空间、集成多尺度性能参数为目标,整合高通量计算数据、实验表征数据及文献挖掘数据,统一数据编码与描述规范,打破跨平台数据壁垒,同时需应对数据质量参差不齐、敏感信息共享受限等问题,为稳健、可迁移的 AI 模型提供高质量数据支撑。2) 发展具有闭环反馈能力的生成与强化学习模型。嵌入聚合物化学先验知识(如聚合反应机理、构效关系)约束模型生成逻辑,使其能主动探索未知化学空间,提出新颖且合成可行的分子与材料设计方案,同时解决模型泛化能力不足、生成结构与合成技术脱节等难题,大幅提升新材料发现与性能优化效率。3) 推动实验与模型的动态耦合集成。通过标准化接口融合自动化聚合合成平台、智能原位表征设备与 AI 算法,打通数据采集、预测迭代、实验验证的数据流,实现“设计-预测-合成-表征”的自驱动研发循环,同步破解设备兼容性不足、实验与模型时序协同错位等瓶颈,促进实验系统与 AI 模型协同进化。4) 深化跨学科协作与平台化研发体系。打破化学、材料、数据科学与自动化工程领域的学科壁垒,构建“数据-模型-实验”一体化的协同工作流,实现从智能数据采集、特征工程到机器人实验验证的全链条贯通。唯有在这些方向上持续聚焦,精准突破技术瓶颈,

形成数据、算法与实验的高效协同,方能真正释放人工智能在聚合物功能材料领域的创新潜能,加速高性能有机聚合物材料从实验室研发向产业化应用的转化。

参考文献:

- [1] IQBAL A, SAEED A, UL-HAMID A. A review featuring the fundamentals and advancements of polymer/CNT nanocomposite application in aerospace industry[J]. *Polym Bull*, 2021, 78(1): 539–557.
- [2] TRZEPICINSKI T, NAJM S M, SBAYTI M, et al. New advances and future possibilities in forming technology of hybrid metal–polymer composites used in aerospace applications[J]. *J Compos Sci*, 2021, 5(8): 217.
- [3] LI Z Y, CHEN X, LIU D, et al. Recent advances in polymer-based composites for thermal management and electromagnetic wave absorption[J]. *Adv Compos Hybrid Mater*, 2025, 8(2): 210.
- [4] YANG S Q, DU S, ZHU J, et al. Closed-loop recyclable polymers: From monomer and polymer design to the polymerization–depolymerization cycle[J]. *Chem Soc Rev*, 2024, 53(19): 9609–9651.
- [5] KANEKO Y, WANG A H. Cross-coupling and dehalogenation reactions of novel π -conjugated polymers bearing high-soluble coumarin unit in the main chain and their acid-base responsive fluorescence[J]. *Eur Polym J*, 2024, 211: 113030.
- [6] HU Z D, LIU X Q, REN T L, et al. Research progress of low dielectric constant polymer materials[J]. *J Polym Eng*, 2022, 42(8): 677–687.
- [7] WANG K J, AMIN K, AN Z S, et al. Advanced functional polymer materials[J]. *Mater Chem Front*, 2020, 4(7): 1803–1915.
- [8] SHI J S, WANG J K, ZHAO Z H, et al. Fabrication of reduced graphene oxide doped poly(3, 4-ethylenedioxythiophene): Polystyrene (sulfonate-co-acrylic acid) conductive films with high conductivity and stability[J]. *Thin Solid Films*, 2022, 762: 139525.
- [9] NARAYAN M R, SINGH J. Exciton dissociation and design optimization in P3HT: PCBM bulk-heterojunction organic solar cell[J]. *Can J Phys*, 2014, 92(7/8): 853–856.
- [10] VARGHESE T, DUN C C, KEMPF N, et al. Flexible thermoelectric devices of ultrahigh power factor by scalable printing and interface engineering[J]. *Adv Funct Mater*, 2020, 30(5): 1905796.
- [11] AL-ITHAWI W, KHASANOV A, KOVALEV I, et al. TM-free and TM-catalyzed mechanosynthesis of functional polymers[J]. *Polymers*, 2023, 15(8): 1853.
- [12] LIPATOVA I M, MORYGANOV A P. Functionalization of synthetic fibrous materials using nanosized polymer carriers[J]. *Russ J Gen Chem*, 2017, 87(6): 1378–1385.
- [13] FORD M J, OHM Y, CHIN K, et al. Composites of functional polymers: Toward physical intelligence using flexible and soft materials[J]. *J Mater Res*, 2022, 37(1): 2–24.
- [14] UEHARA K, ISHITOBI M, ODA T, et al. First-principles molecular dynamics calculation of selenium clusters[J]. *Mol Simul*, 1997, 18(6): 385–394.
- [15] TANG Z, CHANG C F, BAO F, et al. Feasibility of predicting static dielectric constants of polymer materials: A density functional theory method[J]. *Polymers*, 2021, 13(2): 284.
- [16] SAHA L C, MIAN S A, JANG J K. Molecular dynamics simulation study on the carbon nanotube interacting with a polymer[J]. *Bull Korean Chem Soc*, 2012, 33(3): 893–896.
- [17] HIGUCHI Y. Fracture processes of crystalline polymers using coarse-grained molecular dynamics simulations[J]. *Polym J*, 2018, 50(8): 579–588.
- [18] ZOROMBA M S, ALHARBI F, AL-HOSSAINY A F, et al. Preparation of hybrid conducting polymers blend nanocomposite for energy conversion using experimental data and TD-DFT/DMO3 computations[J]. *J Mater Res Technol*, 2023, 23: 2852–2867.
- [19] LIU Y L, NIU C, WANG Z, et al. Machine learning in materials genome initiative: A review[J]. *J Mater Sci Technol*, 2020, 57: 113–122.
- [20] RAJAN K. Materials informatics: The materials “gene” and big data[J]. *Annu Rev Mater Res*, 2015, 45: 153–169.
- [21] MING Y Q, LI J L, WEN J L, et al. Machine learning in constructing structure–property relationships of polymers[J]. *Chem Phys Rev*, 2025, 6(2): 021305.
- [22] CHEN G, SHEN Z Q, IYER A, et al. Machine-learning-assisted *de novo* design of organic molecules and polymers: Opportunities and challenges[J]. *Polymers*, 2020, 12(1): 163.
- [23] CHEN J H, CHRISTAKOPOULOS P, IVANOV I N, et al. Artificial intelligence/machine learning of polymer materials and discovery chemistry[J]. *MRS Bull*, 2025, 50(11): 1364–1375.
- [24] MASHHADIMOSLEM H, ALI ABDOL M, ZANGANEH K, et al. Development of the CO₂ adsorption model on porous adsorbent materials using machine learning algorithms[J]. *ACS Appl Energy Mater*, 2024, 7(19): 8596–8609.
- [25] FENG H S, DING H, WANG S, et al. Machine-learning-assisted catalytic performance predictions of single-atom alloys for acetylene semihydrogenation[J]. *ACS Appl Mater Interfaces*, 2022, 14(22): 25288–25296.
- [26] XIAO B, AHMAD N, MAHMOOD A, et al. Data mining and machine learning analysis to find polymers for electronic and photovoltaics applications: A goal to achieve higher dielectric constant[J]. *Adv Theory Simul*, 2025, 8(9): 2500166.
- [27] KYHOIESH H A K, AWAD M A, SALEM K H, et al. Machine-learning-based design of organic chromophores with improved photovoltaic efficiency[J]. *ChemistrySelect*, 2025, 10(37): e03160.
- [28] WU M L, XIA P, HUANG H X, et al. π -Extension improves the photovoltaic performance: A helical perylene diimide oligomer based three-dimensional non-fullerene acceptor[J]. *Mater Chem Front*, 2019, 3(11): 2414–2420.
- [29] CHEN K X, LI L. Ordered structures with functional units as a paradigm of material design[J]. *Adv Mater*, 2019, 31(32): 1901115.
- [30] JIANG X M, DENG S Q, WHANGBO M H, et al. Material research from the viewpoint of functional motifs[J]. *Natl Sci Rev*, 2022, 9(7): nwac017.
- [31] 郭国聪, 姚元根, 吴克琛, 等. 结构敏感功能材料的基础研究[J]. *化学进展*, 2001, 13(2): 151–155.
GUO Guocong, YAO Yuangen, WU Kechen, et al. *Prog Chem*, 2001, 13(2): 151–155.
- [32] TEYSSIER J, SAENKO S V, VAN DER MAREL D, et al. Photonic crystals cause active colour change in chameleons[J]. *Nat Commun*, 2015, 6: 6368.
- [33] CHEN W S, GU J J, LIU Q L, et al. Two-dimensional quantum-sheet films with sub-1.2 nm channels for ultrahigh-rate electrochemical capacitance[J]. *Nat Nanotechnol*, 2022, 17(2): 153–158.
- [34] ZHAO R, YANG C, WANG H G, et al. Emergent multiferroism with magnetodielectric coupling in EuTiO₃ created by a negative pressure control of strong spin-phonon coupling[J]. *Nat Commun*, 2022, 13:

- 2364.
- [35] JEONG J, PARK S, PARK J, et al. Machine-learning-assisted process optimization for high-performance organic thermoelectrics[J]. *Adv Energy Mater*, 2025, 15(14): 2403431.
- [36] HUSSAIN E, TAHIR M H, DALAL A A, et al. Polymer design with enhanced crystallization tendency aided by machine learning[J]. *Phys B Condens Matter*, 2024, 694: 416437.
- [37] KOOKALANI S, CHENG B, TORRES J L C. Structural performance assessment of GFRP elastic gridshells by machine learning interpretability methods[J]. *Front Struct Civ Eng*, 2022, 16(10): 1249–1266.
- [38] CAO J Y, XU Z. Providing a photovoltaic performance enhancement relationship from binary to ternary polymer solar cells *via* machine learning[J]. *Polymers*, 2024, 16(11): 1496.
- [39] SHETTY P, RAJAN A C, KUENNETH C, et al. A general-purpose material property data extraction pipeline from large polymer corpora using natural language processing[J]. *npj Comput Mater*, 2023, 9: 52.
- [40] BARNETT J W, KUMAR S K. Modeling gas transport in polymer-grafted nanoparticle membranes[J]. *Soft Matter*, 2019, 15(3): 424–432.
- [41] ZENI C, PINSLER R, ZÜGNER D, et al. A generative model for inorganic materials design[J]. *Nature*, 2025, 639(8055): 624–632.
- [42] ZONG W W, ZHOU H M, HUANG G B, et al. Face recognition based on kernelized extreme learning machine[M]//*Autonomous and Intelligent Systems*. Berlin, Heidelberg: Springer 2011: 263–272.
- [43] DENG L, LI X. Machine learning paradigms for speech recognition: An overview[J]. *IEEE Trans Audio Speech Lang Process*, 2013, 21(5): 1060–1089.
- [44] BORCH C. Machine learning, knowledge risk, and principal-agent problems in automated trading[J]. *Technol Soc*, 2022, 68: 101852.
- [45] SHANG C, YOU F Q. Data analytics and machine learning for smart process manufacturing: Recent advances and perspectives in the big data era[J]. *Engineering*, 2019, 5(6): 1010–1016.
- [46] XING E P, HO Q, XIE P T, et al. Strategies and principles of distributed machine learning on big data[J]. *Engineering*, 2016, 2(2): 179–195.
- [47] GUO W J, CHAI S Y, ZHANG L, et al. Computer-aided design of crosslinked polymer membrane using machine learning and molecular dynamics[J]. *Chem Ing Tech*, 2023, 95(3): 447–457.
- [48] SIRISINHA K, MEKSAWAT D. Comparison in processability and mechanical and thermal properties of ethylene–octene copolymer crosslinked by different techniques[J]. *J Appl Polym Sci*, 2004, 93(3): 1179–1185.
- [49] RAZA Y, RAZA H, AHMED A, et al. Integration of response surface methodology (RSM), machine learning (ML), and artificial intelligence (AI) for enhancing properties of polymeric nanocomposites-a review[J]. *Polym Compos*, 2025, 46(15): 13591–13627.
- [50] TAN C, WU H J, YANG L, et al. Cutting edge high-throughput synthesis and characterization techniques in combinatorial materials science[J]. *Adv Mater Technol*, 2024, 9(10): 232038.
- [51] WANG K Y, SHI H Y, LI T J, et al. Computational and data-driven modelling of solid polymer electrolytes[J]. *Digit Discov*, 2023, 2(6): 1660–1682.
- [52] LI C H, BAO L Q, JI Y X, et al. Combining machine learning and metal–organic frameworks research: Novel modeling, performance prediction, and materials discovery[J]. *Coord Chem Rev*, 2024, 514: 215888.
- [53] YE H L, XIAN W K, LI Y. Machine learning of coarse-grained models for organic molecules and polymers: Progress, opportunities, and challenges[J]. *ACS Omega*, 2021, 6(3): 1758–1772.
- [54] KIRKLIN S, SAAL J E, MEREDIG B, et al. The Open Quantum Materials Database (OQMD): Assessing the accuracy of DFT formation energies[J]. *npj Comput Mater*, 2015, 1: 15010.
- [55] GOSSETT E, TOHER C, OSES C, et al. AFLOW-ML: A RESTful API for machine-learning predictions of materials properties[J]. *Comput Mater Sci*, 2018, 152: 134–145.
- [56] FLUCK E. Inorganic crystal structure database (ICSD) and standardized data and crystal chemical characterization of inorganic structure types (TYPIX) - Two tools for inorganic chemists[J]. *J Res Natl Inst Stand Technol*, 1996, 101(3): 217.
- [57] LIU X M, ZHANG X Y, WU Y B, et al. OPECM: A comprehensive dataset for organic polymer energy conversion materials enabling data-driven interpretative analysis[J]. *Phys Rev Materials*, 2025, 9(12): 125405.
- [58] IWASAKI Y, ISHIDA M, SHIRANE M. Predicting material properties by integrating high-throughput experiments, high-throughput *ab-initio* calculations, and machine learning[J]. *Sci Technol Adv Mater*, 2020, 21(1): 25–28.
- [59] O'BOYLE N M, CAMPBELL C M, HUTCHISON G R. Computational design and selection of optimal organic photovoltaic materials[J]. *J Phys Chem C*, 2011, 115(32): 16200–16210.
- [60] WANG J, HE F, ZHANG M. Innovative vapor transport deposition technique for selenium thin-film solar cell fabrication[J]. *Nanoenergy Adv*, 2025, 5(2): 8.
- [61] FUJIMOTO K, AIMI A, MARUYAMA S. Development of measurement tools for high-throughput experiments of synchrotron radiation XRD and XAFS on powder libraries[J]. *ACS Comb Sci*, 2020, 22(12): 734–737.
- [62] JOHNSON P M, STANSBURY J W, BOWMAN C N. FTIR microscopy for kinetic measurements in high-throughput photopolymerization: Experimental design and application[J]. *Macromol React Eng*, 2009, 3(9): 522–528.
- [63] JAIN A, HAUTIER G, MOORE C J, et al. A high-throughput infrastructure for density functional theory calculations[J]. *Comput Mater Sci*, 2011, 50(8): 2295–2310.
- [64] ZHENG X L, ZHENG P, ZHANG R Z. Machine learning material properties from the periodic table using convolutional neural networks[J]. *Chem Sci*, 2018, 9(44): 8426–8432.
- [65] HILGERS R, WORTMANN D, BLÜGEL S. Application of batch learning for boosting high-throughput *ab initio* success rates and reducing computational effort required using data-driven processes[J]. *Electron Struct*, 2025, 7(1): 015005.
- [66] YANG X Y, WANG Z G, ZHAO X S, et al. MatCloud: A high-throughput computational infrastructure for integrated management of materials simulation, data and resources[J]. *Comput Mater Sci*, 2018, 146: 319–333.
- [67] FOURCHES D, SASSANO M F, ROTH B L, et al. HTS navigator: Freely accessible cheminformatics software for analyzing high-throughput screening data[J]. *Bioinformatics*, 2014, 30(4): 588–589.
- [68] FRISCH M J, TRUCHS G W, SCHLEGEL H B, et al. Gaussian 16, Revision B.01[DB/OL]. Gaussian, Inc., Wallingford CT 2016.
- [69] GUPTA S, MAHMOOD A, SHETTY P, et al. Data extraction from polymer literature using large language models[J]. *Commun Mater*, 2024, 5: 269.
- [70] XU C W, WANG Y Y, BARATI FARIMANI A. TransPolymer: A Transformer-based language model for polymer property predictions[J].

- npj Comput Mater, 2023, 9: 64.
- [71] ISAZAWA T, COLE J M. Single model for organic and inorganic chemical named entity recognition in ChemDataExtractor[J]. J Chem Inf Model, 2022, 62(5): 1207–1213.
- [72] HUANG S, COLE J M. A database of battery materials auto-generated using ChemDataExtractor[J]. Sci Data, 2020, 7: 260.
- [73] DONG Q Y, COLE J M. Auto-generated database of semiconductor band gaps using ChemDataExtractor[J]. Sci Data, 2022, 9: 193.
- [74] ALTARAZI S, AMMOURI M, HIJAZI A. Artificial neural network modeling to evaluate polyvinylchloride composites' properties[J]. Comput Mater Sci, 2018, 153: 1–9.
- [75] NAGASAWA S, AL-NAAMANI E, SAEKI A. Computer-aided screening of conjugated polymers for organic solar cell: Classification by random forest[J]. J Phys Chem Lett, 2018, 9(10): 2639–2646.
- [76] YI Y, WANG L M, CHEN Z Y. Adaptive global kernel interval SVR-based machine learning for accelerated dielectric constant prediction of polymer-based dielectric energy storage[J]. Renew Energy, 2021, 176: 81–88.
- [77] WANG H S, FENG J, DONG Z H, et al. Efficient screening framework for organic solar cells with deep learning and ensemble learning[J]. npj Comput Mater, 2023, 9: 200.
- [78] ZHANG X Y, WEI G W, SHENG Y, et al. Polymer-unit fingerprint (PUFp): An accessible expression of polymer organic semiconductors for machine learning[J]. ACS Appl Mater Interfaces, 2023, 15(17): 21537–21548.
- [79] ZHANG X Y, SHENG Y, LIU X M, et al. Polymer-unit graph: Advancing interpretability in graph neural network machine learning for organic polymer semiconductor materials[J]. J Chem Theory Comput, 2024, 20(7): 2908–2920.
- [80] PONTIL M, VERRI A. Support vector machines for 3D object recognition[J]. IEEE Trans Pattern Anal Mach Intell, 1998, 20(6): 637–646.
- [81] WIEBE J. An NLP framework for interpreting implicit and explicit opinions in text and dialog[C]//Advances in Artificial Intelligence, 2015: 9091.
- [82] MA Y, CHEN D Y, WANG T, et al. Semi-supervised partial label learning algorithm *via* reliable label propagation[J]. Appl Intell, 2023, 53(10): 12859–12872.
- [83] TAMNANLOO J, MOKARIZADEH A H, TOISERKANI F, et al. Graph neural networks for surface tension prediction of polymers: A comparative analysis with descriptor-based models[J]. Macromolecules, 2025, 58(8): 3764–3773.
- [84] NGUYEN T, BAVARIAN M. A machine learning framework for predicting the glass transition temperature of homopolymers[J]. Ind Eng Chem Res, 2022, 61(34): 12690–12698.
- [85] SIPPER M. High per parameter: A large-scale study of hyperparameter tuning for machine learning algorithms[J]. Algorithms, 2022, 15(9): 315.
- [86] RODRÍGUEZ-PÉREZ R, BAJORATH J. Interpretation of compound activity predictions from complex machine learning models using local approximations and shapley values[J]. J Med Chem, 2020, 63(16): 8761–8777.
- [87] JUNG Y J, HAN S H, CHOI H J. Explaining CNN and RNN using selective layer-wise relevance propagation[J]. IEEE Access, 2021, 9: 18670–18681.
- [88] MONTAVON G, SAMEK W, MÜLLER K R. Methods for interpreting and understanding deep neural networks[J]. Digit Signal Process, 2018, 73: 1–15.
- [89] BHAT V, GANAPATHYSUBRAMANIAN B, RISCO C. Rapid estimation of the intermolecular electronic couplings and charge-carrier mobilities of crystalline molecular organic semiconductors through a machine learning pipeline[J]. J Phys Chem Lett, 2024, 15(28): 7206–7213.
- [90] AHMED D R, MUHAMMADSHARIF F F. A review of machine learning in organic solar cells[J]. Processes, 2025, 13(2): 393.
- [91] ALZHRANI F M A, MUHAMMAD ROUF A, NAJEEB J, et al. Computational design of new polymers having low exciton binding energy for organic solar cells fabrication: Chemical generation and visualization[J]. J Photochem Photobiol A Chem, 2024, 450: 115457.
- [92] ZHAO Z W, GENG Y, TROISI A, et al. Performance prediction and experimental optimization assisted by machine learning for organic photovoltaics[J]. Adv Intell Syst, 2022, 4(6): 2100261.
- [93] YANG B, ZHANG C R, WANG Y, et al. Open-circuit voltage loss and dielectric constants as new descriptors in machine learning study on organic photovoltaics[J]. Int J Quantum Chem, 2023, 123(5): e27039.
- [94] HACIEFENDIOGLU T, YILDIRIM E. Band gap and reorganization energy prediction of conducting polymers by the integration of machine learning and density functional theory[J]. J Chem Inf Model, 2025, 65(11): 5360–5369.
- [95] WU Y, GUO J, SUN R, et al. Machine learning for accelerating the discovery of high-performance donor/acceptor pairs in non-fullerene organic solar cells[J]. npj Comput Mater, 2020, 6: 120.
- [96] KRANTHIRAJA K, SAEKI A. Experiment-oriented machine learning of polymer: Non-fullerene organic solar cells[J]. Adv Funct Mater, 2021, 31(23): 2011168.
- [97] SUN J Y, LI D X, ZOU J, et al. Accelerating the discovery of acceptor materials for organic solar cells by deep learning[J]. npj Comput Mater, 2024, 10: 181.
- [98] LIU X M, ZHANG X Y, SHENG Y, et al. Advancing organic photovoltaic materials by machine learning-driven design with polymer-unit fingerprints[J]. npj Comput Mater, 2025, 11: 107.
- [99] SHENG Y, WU Y B, JIANG C, et al. Interpretable model of dielectric constant for rational design of microwave dielectric materials: A machine learning study[J]. J Mater Inf, 2025, 5(1): 7.
- [100] TAO Q L, XU P C, LI M J, et al. Machine learning for perovskite materials design and discovery[J]. npj Comput Mater, 2021, 7: 23.
- [101] QIN J C, LIU Z F, MA M S, et al. Machine learning-assisted materials design and discovery of low-melting-point inorganic oxides for low-temperature cofired ceramic applications[J]. ACS Sustainable Chem Eng, 2022, 10(4): 1554–1564.
- [102] WU C, CHEN L H, DESHMUKH A, et al. Dielectric polymers tolerant to electric field and temperature extremes: Integration of phenomenology, informatics, and experimental validation[J]. ACS Appl Mater Interfaces, 2021, 13(45): 53416–53424.
- [103] HE X J, WAN J C, ZHANG S Y, et al. Interpretable machine learning prediction of polyimide dielectric constants: A feature-engineered approach with experimental validation[J]. Polymers, 2025, 17(12): 1622.
- [104] YANG M Z, WAN C F, ZHOU L, et al. High-temperature polymer composite capacitors with high energy density designed *via* machine learning[J]. Nat Energy, 2025, 10(11): 1323–1333.
- [105] ANTONIUK E R, LI P, KAILKHURA B, et al. Representing polymers as periodic graphs with learned descriptors for accurate polymer property predictions[J]. J Chem Inf Model, 2022, 62(22): 5435–5445.

- [106] NAZAROVA A L, YANG L Q, LIU K, et al. Dielectric polymer property prediction using recurrent neural networks with optimizations[J]. *J Chem Inf Model*, 2021, 61(5): 2175–2186.
- [107] LI H, ZHENG H B, YUE T L, et al. Machine learning-accelerated discovery of heat-resistant polysulfates for electrostatic energy

storage[J]. *Nat Energy*, 2025, 10(1): 90–100.

作者贡献声明:

叶财超: 提出论文思路及构建框架、撰写修改论文

刘秀敏: 完善论文框架、绘制图表、论文撰写

Research and Design Progress on Organic Polymer Functional Materials under Artificial Intelligence Paradigm

LIU Xiumin^{1,2}, YE Caichao¹

(1. Department of Materials Science and Engineering, Southern University of Science and Technology, Shenzhen 518055, Guangdong, China; 2. School of Chemistry and Chemical Engineering, Nanjing University of Science and Technology, Nanjing 210094, China)

Extended Abstract

Polymer materials, as an important branch of materials science, are widely applied in some fields such as energy, electronics, and aerospace due to their adjustable structures, diverse properties, and low cost. Although polymer functional materials permeate into numerous technical fields, their multi-scale complex behaviors still pose significant challenges for precise design. In this context, the data-driven research paradigm advocated by the Materials Genome Initiative emerges, marking the entry of materials research into the “fourth paradigm” characterized by data-intensive approach. Machine learning as an important branch of artificial intelligence demonstrates outstanding performance prediction and design capabilities in various systems such as alloys, polymer dielectrics, and photovoltaic materials. Although machine learning makes progress in predicting and designing polymer properties, it still faces several challenges, *i.e.*, the scale and quality of data are limited by the costs of synthesis and characterization; the representation of polymer molecules and feature engineering still require the development of more effective encoding methods; and the experimental-computational loop has not yet been fully connected.

This review first, in conjunction with the changes in the research paradigm, systematically explores the progress of AI in the field of organic polymer functional materials. The fourth paradigm centered on data significantly shortens the material research and development cycle and accelerates the screening and design of new materials. Although several review articles summarize the relevant progress of machine learning in the field of polymer material research and design from different perspectives, there is still insufficient discussion on key aspects such as the construction of polymer-specific databases, data-driven characterization of functional units, and model interpretability. Establishing a new research framework centered on artificial intelligence is expected to achieve groundbreaking progress in exploring the vast feature and property space of polymers.

This review represents the general process of applying artificial intelligence to polymer materials, mainly consisting of four stages, *i.e.*, data generation, feature engineering, model selection, and model interpretation. Data quality and scale remain the weakest links in the existing data-driven approaches. The OPECM database, which contains a total of 3225 polymers, is constructed due to the scarcity of available datasets for organic polymer functional materials (such as semiconductors, photovoltaics, and dielectric materials), and the limited chemical diversity of these materials. To address the challenges faced by the polymer representation methods in quantitative structure-property relationship (QSPR) modeling, the “Polymer-Unit” concept is proposed and two novel structure-encoding schemes (*i.e.*, the Polymer Unit Fingerprint (PUFp) and the Polymer-unit Graph) are developed.

Through the lens of database construction, feature representation, and model interpretability, this review highlights machine learning applications in functional polymer design, with some case studies covering semiconductor, photovoltaic, and dielectric materials. PUFp-based framework enables pre-screening and classification of materials *via* machine learning, offering an efficient alternative route to discovering high-mobility organic semiconductors. The Polymer-unit Graph representation, implemented with graph neural networks (GNNs), provides a visually interpretable and topology-aware approach to polymer modeling.

Summary and Prospects AI achieves notable progress in polymer functional materials, particularly in inverse design, performance prediction, and process optimization, accelerating the shift from experience-driven to data-driven research. To fully unlock AI’s potential, some challenges in data, algorithms, and integration must be systematically addressed. This review represents a dedicated polymer database, a functional unit-based encoding method, and an interpretable structure–performance framework to bridge the gap between “black-box” model outputs and human-understandable chemical insight. Four key directions for future work should be concerned, *i.e.*, 1) building comprehensive, open, and interoperable polymer databases; 2) advancing generative and reinforcement learning for autonomous material discovery; 3) strengthening the dynamic coupling between experiments and models to form self-driving research loops; and 4) fostering interdisciplinary collaboration and integrated platform development. AI can evolve from a supportive tool into a core innovative engine for polymer functional materials, powered by the synergy of data, algorithms and experiments.

Keywords artificial intelligent paradigm; organic polymer functional materials; functional unit; data-driven; structure–property relationship