

AI 预测 NMR 化学位移的技术概览

1. 最新的AI模型与算法

近年来，多种基于机器学习的模型被用于化学位移预测。**图神经网络（GNN）**是最常见的方法之一，可直接利用分子图结构进行原子级预测。例如，Han 等提出一种可扩展的 GNN，仅将重原子作为节点，结合注意力机制和残差连接，并利用节点级和图级嵌入，显著提升了大分子（含 ^{13}C 和 ^1H ）的预测精度^{1 2}。**CASCADE** 工具（Paton 组）采用基于 SchNet 的 3D GNN，从原子坐标出发预测 ^1H 和 ^{13}C 位移，支持从 SMILES 自动生成构象并进行 Boltzmann 平均^{3 4}。更近期的工作如 **GT-NMR** 将 Transformer 架构引入分子图，将二维结构与节点级注意力相结合，实现了更高精度的位移预测（在标准 nmrshiftdb2 数据集上，GT-NMR 在 $^1\text{H}/^{13}\text{C}$ 位移预测中分别达到了 $\text{MAE}=0.158$ 和 1.189 ppm⁵）。

除了深度神经网络，还有基于描述符的机器学习回归模型。**SHIFTML**系列采用核岭回归（KRR）对局部原子环境（如 SOAP 或 LMBTR 描述符）进行学习，可预测固态分子的 ^1H 、 ^{13}C 等位移^{6 7}。最新版本的 ShiftML2 已扩展覆盖元素包括 H、C、N、O、F、P、Cl、Na、Ca、Mg、K 和 S，并保持与 DFT 近似一致的速度和精度^{8 7}。此外，一些工作采用弱监督学习技术，如 Park 等人提出的方法仅需分子级总位移标签，通过排列不变损失函数训练 GNN，从而无需手动标注原子级位移，也能获得与监督学习可比的预测精度^{9 10}。综上，多种 AI 模型（包括图神经网络、Transformer、核方法和深度神经网络回归等）已被开发，用于从分子结构预测 ^1H 、 ^{13}C （乃至 ^7Li 、 ^{15}N 等）核的化学位移。

2. 与传统DFT方法的对比

传统的量子化学方法（如 GIPAW-DFT）可以提供较高精度的化学位移计算，典型的计算误差可达到约 1.5 ppm (^{13}C) 和 0.2 ppm (^1H)¹¹。但这类方法计算成本高昂，一般只能处理较小体系或计算量庞大。AI 方法则以极低的计算成本实现近似于 DFT 的精度。例如 ShiftML 在固态分子数据集上预测 ^1H 位移的 RMSE 为 0.47 ppm，而对应的 DFT 计算为 0.35 ppm，预测速度提高了 4 个数量级⁶。Paton 组开发的 3D GNN 模型同样显示出数千倍于 DFT 的加速比¹²。另一方面，这些 ML 模型的绝对精度一般略逊于高水平 DFT，对于较复杂或训练外的体系精度有所下降。GT-NMR 在相对简单分子上（复杂度指标 $n\text{SPS} \leq 27.7$ ）可达到与最佳 DFT 计算相当的水平，但随着分子复杂度增加，误差显著增大¹³。综合来看，AI 方法在精度上已接近 DFT 级别，但更加快速高效；其适用性受限于训练集覆盖范围和模型结构，在复杂或未知领域仍需要谨慎对待^{13 6}。

3. 不同分子类型的预测表现

- **有机小分子**：针对常见有机分子（如药物或天然产物等），AI 模型已取得较好表现。利用 nmrshiftdb2 等实验数据库训练的模型（如 Jonas 和 Kuhn 2019 年提出的基于二维结构的 GNN）在小分子上可达 ^1H $\text{MAE} \approx 0.28$ ppm、 ^{13}C $\text{MAE} \approx 1.43$ ppm¹⁴，但这些模型往往未考虑分子的构象和立体化学。相比之下，支持三维构型的模型（如 CASCADE）能够区分立体异构体和构象异质性，从而提高了异构体识别的准确性。总体而言，若模型训练集中包含该类结构，预测误差可控制在 ppm 量级。
- **金属配合物及其它无机体系**：当前研究相对较少。固态机器学习模型 ShiftML 已扩展到含有 Na、Ca、Mg、K 等离子化合物¹⁵，但对过渡金属中心（如催化剂或配合物）的预测尚未充分覆盖。部分工作尝试在电池材料中应用机器学习 NMR。例如，Lin 等（2022）提出的 ML 流程用于预测含有过渡金属层的高压 Na 器件中 ^{23}Na 的 NMR 位移，揭示了不同堆垛序列对应的实验 NMR 峰¹⁶。这表明 ML 可用于固态电池中含金属离子的复杂体系中。
- **电解质溶液与离子**：锂离子电池电解质中， Li^+ 的局部溶剂化环境复杂且浓度依赖性强。最近有工作利用机器学习从 LiFSI/DME 电解液的 MD 模拟提取的溶剂团簇结构来训练 ^7Li 化学位移预测模型¹⁷。该

模型在不同浓度下成功再现了实验中 ^7Li 化学位移的非线性变化，并揭示了两种竞争溶剂化结构的切换¹⁷。以上研究表明，针对特定体系（如离子液或电解质）的 ML 模型需要结合动力学采样以捕捉溶剂化和浓度效应。目前通用模型对强烈依赖环境的化学位移预测仍存在挑战，需要更多样化的数据和专门建模。

总之，不同类别分子在 AI 预测中的表现差异主要取决于训练数据覆盖及所用模型的物理信息：常见有机分子（尤其数据丰富的领域）表现最好，而含金属中心或非规整结构的体系误差较大。此外，忽略溶剂、温度和构象变化的模型可能出现系统性偏差（如对灵活分子或立体化学敏感性不足）¹⁴¹³。

4. 开源工具与研究平台

当前已经涌现出若干开源工具和平台，用于 NMR 化学位移的机器学习预测：

- **ShiftML**：由 Ceriotti/Emsley 组开发，可通过 Materials Cloud 平台使用。ShiftML 输入晶体结构，输出 ^1H 、 ^{13}C 、 ^{15}N 、 ^{17}O 等核的各向同性位移，预测精度接近 DFT⁶⁷。最新版本 ShiftML2 支持多达 12 种元素并给出预测不确定度。
- **CASCADE**：由 Paton 组提供的在线预测工具（及 GitHub 开源代码），全称 Chemical Shift Calculation with DEep learning。用户可输入 SMILES 或分子编辑器构建分子，后台生成构象并利用 3D GNN 实时预测 ^1H / ^{13}C 位移³。CASCADE 对分子立体化学敏感，可辨别同分异构体，并支持批量预测和 Boltzmann 平均。
- **nmrshiftdb2**：一个开放的有机分子 NMR 数据库，包含数万条 ^1H 和 ^{13}C 谱数据及结构。它不仅可用于查询和分享实验数据，也是训练和验证 ML 模型的重要资源¹⁸¹⁹。常用数据集如 NMR8K/DFT8K/Exp5K 都来源于 nmrshiftdb2。
- **Chemprop**：由 Yang 等开发的图卷积神经网络框架（基于 D-MPNN 架构），可用于各种分子性质的 ML 预测。虽然原始 Chemprop 并非专门针对 NMR，但研究者可以利用其灵活的 Python 接口定制模型并训练 NMR 化学位移预测任务²⁰。

此外，还有一些学术项目和工具平台正在发展，如通过云计算集成 NMR 预测的 NMRCryst（学界工具）或商业软件插件。上述开源资源的出现降低了 ML 模型的开发和应用门槛，促进了社区共同进步。

5. 主要挑战与发展趋势

尽管 AI 在 NMR 化学位移预测方面取得了快速进展，但仍面临多项挑战：首先**数据多样性不足**。目前可用于训练的标注数据（尤其是实验测定的位移）集中在常见有机小分子，对于金属络合物、电解质、含稀有核的体系数据稀缺，限制了模型的泛化能力。其次，**环境与动力学效应难以建模**。大多数模型仅使用静态分子构型作为输入，缺乏对溶剂效应、温度涨落和构象分布的考虑，导致在真实条件下预测误差增大。再者，**模型解释性和泛化性**仍需提升。图神经网络虽然自动学习环境信息，但对模型内部如何捕捉化学环境物理本质认识有限；而对于超出训练域的“外推”体系（如更大的分子、复杂的立体化学、多核共振等），误差会显著增加¹³¹⁴。

未来研究趋势包括：采用更强大的模型架构（例如更复杂的图 Transformer、几何对称网络等）来提高表达能力；结合**多任务学习与迁移学习**策略，以利用不同核种或从理论计算向实验数据的知识转移（Paton 组已验证这种方法可降低实测误差）²¹¹⁹；引入**多物理尺度仿真**（如与分子动力学相结合）来模拟环境和动态效应¹⁷¹⁶；以及构建更大规模、高质量的数据库。总之，随着算法创新、数据积累和计算资源提升，AI 辅助的 NMR 化学位移预测有望向更高精度、更广应用领域发展，并助力结构解析和材料设计等应用。

参考文献： 本文内容引用了相关文献和公开资源²²⁹⁶⁷²³³¹⁷¹⁶等，以确保信息的准确性和及时性。

- 1 2 **Scalable graph neural network for NMR chemical shift prediction - Physical Chemistry Chemical Physics (RSC Publishing)**
<https://pubs.rsc.org/en/content/articlelanding/2022/cp/d2cp04542g>
- 3 **Real-time prediction of ^1H and ^{13}C chemical shifts with DFT accuracy using a 3D graph neural network - Chemical Science (RSC Publishing) DOI:10.1039/D1SC03343C**
4
12 <https://pubs.rsc.org/en/content/articlehtml/2021/sc/d1sc03343c>
19
21
- 5 13 **GT-NMR: a novel graph transformer-based approach for accurate prediction of NMR chemical shifts | Journal of Cheminformatics | Full Text**
18 23
<https://jcheminf.biomedcentral.com/articles/10.1186/s13321-024-00927-9>
- 6 8 11 **research-collection.ethz.ch**
15 <https://www.research-collection.ethz.ch/bitstream/handle/20.500.11850/575171/2/acs.jpcc.2c03854.pdf>
- 7 **ShiftML**
<https://tools.materialscloud.org/shiftml/>
- 9 **Predictive Modeling of NMR Chemical Shifts Without Using Atomic-Level Annotations | Request PDF**
10
https://www.researchgate.net/publication/343123728_Predictive_Modeling_of_NMR_Chemical_Shifts_Without_Using_Atomic-Level_Annotations
- 14 **Frontiers | Machine learning in computational NMR-aided structural elucidation**
<https://www.frontiersin.org/journals/natural-products/articles/10.3389/fntpr.2023.1122426/full>
- 16 **A machine learning protocol for revealing ion transport mechanisms from dynamic NMR shifts in paramagnetic battery materials - Chemical Science (RSC Publishing)**
<https://pubs.rsc.org/en/content/articlelanding/2022/sc/d2sc01306a>
- 17 **[2501.07321] Competing Effects of Local Solvation Structures on Chemical Shift Changes of Liquid Electrolyte**
<https://arxiv.org/pdf/2501.07321>
- 20 **Chemprop: A Machine Learning Package for Chemical Property ...**
<https://pubs.acs.org/doi/10.1021/acs.jcim.3c01250>