文章编号：1006-9941（2024）06-0660-12

基于数据驱动的氮杂多环含能化合物的开发研究进展

刘友海[1,2]，黄　实[1]，张文全[1]，杨福胜[2]
（1. 中国工程物理研究院化工材料研究所，四川 绵阳 621999；2. 西安交通大学化学工程与技术学院，陕西 西安 710049）

摘　要： 含能材料的开发面临诸多挑战，传统“试错法”的研发模式会导致研发周期长，效率低。随着数据科学与人工智能技术的发展，基于数据驱动的研发模式为含能材料的发展开辟了新的路径。多环含能化合物是当前含能材料学科的研究热点，其中氮杂多环骨架由于存在π电子的离域共振和较多的可修饰位点，分子结构的稳定性得到提高，同时能量基团的存在保证了分子的能量水平，使得能量与稳定性之间的固有矛盾得到很好的平衡。研究简要介绍了数据驱动开发新型含能材料的工作流程，概述了数据驱动方法用于氮杂多环含能化合物开发的最新研究进展，最后对数据驱动的方法用于新型含能材料的开发提出展望。未来的发展方向应考虑通过数据增强、治理等手段补充数据量，以提高模型预测的准确性及泛化能力；可通过建立化学反应条件和合成路径筛选的机器学习模型预测分子的可合成性，从而加速新型氮杂多环含能化合物的开发。

关键词：含能材料；数据驱动；氮杂多环含能化合物；机器学习
中图分类号：TJ55；O64　　**文献标志码：**A　　**DOI：**10.11943/CJEM2024088

0　引言

含能材料是一类含有爆炸性基团或含有氧化剂及燃料的化合物或混合物，能够在外界特定能量刺激下发生剧烈的氧化还原反应，释放出大量的能量[1]。由于化学组成结构中蕴含着巨大能量，含能材料被广泛应用于武器装备、航天推进、工程建设及矿物开采等领域，对国防安全和经济建设有着重要意义[2-7]。然而，含能分子的开发具有极大的挑战性，需要在性能上兼顾能量、感度、力学强度等多个方面；在研制过程中，需要经历设计、合成、表征、应用多个阶段，仅仅依靠依赖科研人员经验的“试错方法”来探索新型含能分子将面临巨大的工作量。因此，有必要采用更科学的方法来发现、筛选并合成潜在的含能分子。

2011年美国政府宣布启动“材料基因工程计划”[8]，在各国政府的推动下，各类前沿科学与材料学深度交叉融合，出现了继经验驱动、理论驱动、计算驱动之后科学研究的第四范式——数据驱动材料设计[9]，其核心就是依靠人工智能技术对已有数据进行处理，分析并总结规律进而指导科学研究[10]。机器学习（machine learning，ML）作为一种重要的数据驱动方法在跨学科领域，如材料科学、化学、物理学和计算机科学等，展现出了巨大优势，并且已被逐步应用于大量新型材料的研究，如有机光电材料[11]、钙钛矿材料[12]、储能电池材料[13]、光伏材料[14-15]等。目前，采用机器学习方法进行含能材料领域的相关研究也取得了一定进展，建立了多种含能分子设计的机器学习模型，比如性能预测[16-17]、分子生成[18]等，极大地提升了含能分子的研发效率。

氮杂多环含能化合物具有较高的分子内能量，有利于提高含能材料的能量密度；多环结构的稳定性较高，具有较多的可修饰位点，可以通过合适的修饰方法满足不同的性能需求；氮杂多环含能化合物相比传统含能材料在燃烧过程中产生对环境更加友好的副产物，它在提高含能材料的能量和稳定性方面具有广阔的应用前景，因此成为目前含能材料领域的研究热点之一。氮杂多环含能化合物主要包括氮杂稠环和联环

收稿日期：2024-03-20；**修回日期：**2024-04-08
网络出版日期：2024-05-22
基金项目：国家自然科学基金（22375190）
作者简介：刘友海（1985-），男，博士研究生，主要从事含能材料计算与合成研究。e-mail：colour307@163.com
通信联系人：张文全（1986-），男，研究员，主要从事含能分子设计与合成研究。e-mail：zhangwq-cn@caep.cn

引用本文：刘友海，黄实，张文全，等. 基于数据驱动的氮杂多环含能化合物的开发研究进展[J]. 含能材料，2024，32(6):660-671.
LIU You-hai，HUANG Shi，ZHANG Wen-quan，et al. Research Progress of Nitrogen Heteropolycylic Energetic Materials Based on Data-driven[J]. *Chinese Journal of Energetic Materials（Hanneng Cailiao）*，2024，32(6):660-671.

化合物，其中氮杂稠环类含能化合物存在较大的π电子共轭结构，相较于单环化合物展现出更强的π-π相互作用，这对化合物的机械感度和热稳定性具有正面影响，同时拥有更多数量的N=N、N—N、N—O等化学键，使其相对于单环化合物拥有更高的生成焓和密度，从而表现出良好的能量性能[19-20]；而氮杂联环类含能化合物是通过化学键将2个及以上的环连接在一起，从而具有更多的可修饰位点，易于对分子进行性能调控，平衡含能分子能量和稳定性间的固有矛盾[21]。为此，本文简要介绍了数据驱动方法开发含能化合物的工作流程和基本步骤，重点介绍了其在氮杂多环含能材料开发上的最新研究进展，同时讨论其在含能材料领域面临的一些挑战，并展望了数据驱动在含能材料特别是氮杂多环含能化合物开发上的未来发展方向。

1 数据驱动开发含能材料的工作流程

在含能材料领域数据驱动的研究方式主要是通过学习已有的数据，利用机器学习算法建立一个准确的预测模型，模型可以有效地预测未知材料的性能，从而实现数据驱动的材料设计。图1展示了数据驱动开发含能材料的一般工作流程[22]，包含2个主要部分：一、数据挖掘及特征工程：从海量文献、数据库等各种途径中搜集相关信息并提取相关数据，用与目标属性强相关的特征或描述符表示数据；二、模型的构建及目标性能分子的筛选，利用机器学习算法对数据进行有效建模，用建立好的模型对分子的性能进行预测，找到所需性能的材料，并通过实验或理论计算验证，实现新材料设计和开发，下面将对这两个部分进行介绍。

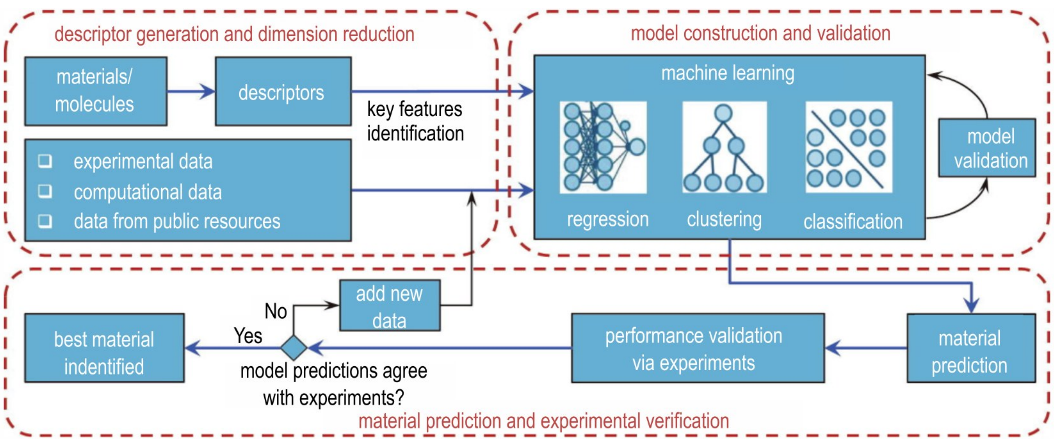


图1 数据驱动开发含能材料的工作流程[22]

Fig.1 Workflow for developing energetic materials by data-driven[22]

1.1 数据挖掘

数据是机器学习方法应用的前提，数据的大小、质量是决定模型预测精度的关键因素。数据的来源广泛，可从实验结果、软件（Gaussian、VASP等）计算以及文献中收集得到，如Song等[23-24]搜集了近千个样本数据，包括密度、爆速、爆压和热分解温度等多个维度，涵盖了不同类型的含能材料分子，包括脂肪族、芳烃、单环和多环化合物，这些数据被用来训练搭建的回归模型。Chandrasekaran等[25]利用实验结果获得了一个由104个样本组成的数据集，用来预测含能化合物的密度值。

除此之外，还可以从现有数据库（CCDC、PubChem等）中挖掘得到数据样本。Nguyen等[17]从剑桥晶体数据库中筛选出了10251个含有确定晶体密度值的含能分子，用于回归模型的训练并预测分子的晶体密度。Song等[23]也从剑桥晶体数据中心收集了365个非石墨层状堆积结构以及22个类石墨层状堆积结构的分子，用于分类模型的训练，以筛选出满足类石墨层状堆积这种稳定性较高的分子结构。Casey等[26]以氧平衡为标准，从GDB-17数据库中筛选出了26265个具有含能特性的分子，他们使用了电子结构计算程序（Gaussian）和热化学计算程序（Cheetah）的组合流程来计算分子的性能，并将这些计算数据作为训练数据用于卷积神经网络的训练。

数据的数量和质量对模型的预测精度至关重要[27]，然而，由于含能材料合成方法、性能测试标准、环境条件、国家安全等因素，目前尚无法建立一个标准统一、数据丰富且对外开放的含能材料数据库。特别是对于深度学习模型而言，数据量过少会导致模型的拟合效果较差。为解决含能材料数据过少的问题，可

以考虑通过片段对接、简化分子线性输入规范（simplified molecular-input line entry system，SMILES）枚举等方式生成新的训练样本，对已有数据进行扩充；还可以考虑利用药物分子研究领域预训练好的模型，通过微调等方式将其应用于构建含能分子机器学习模型；或是利用高通量计算方法，通过计算机生成大量虚拟数据来补充实验数据，从而丰富数据集。学习其他材料设计研究领域数据增强或数据扩充的经验，对完善含能材料数据集具有重要的启发意义，有望进一步提升含能分子性能预测模型的精度。例如，Li等[28]将303个高能分子通过完全枚举法的数据增强方式构建了一个新的数据集D_s，作者将由预训练、生成、预测3个循环神经网络（RNN）模型组成的深度学习框架分别对数据集D_s进行迁移学习及爆轰性能的预测，经过量子化学计算验证后筛选出了35个性能优异的新分子，数据增强及迁移学习方式如图2所示。

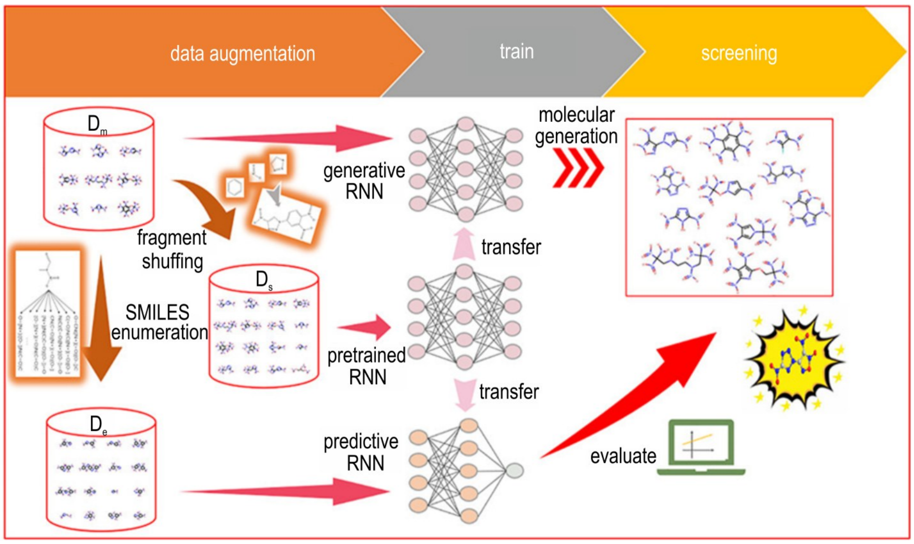


图2 基于RNN模型的含能分子数据增强及迁移学习[28]

Fig.2 Enhancement and transfer learning of energetic molecular data based on RNN model[28]

1.2 特征工程

分子结构的构建方式是机器学习实现的关键，不同构建方式的选取对模型的精度会产生重要影响。目前分子结构的构建方式主要分为三类：一维（1D）、二维（2D）和三维（3D）。1D分子结构是通过线性排列描述分子结构，表现形式包括SMILES以及指纹，对于指纹常用的是扩展连接指纹（ECFP）[29]和分子接入系统（MACCS）[30]；2D分子结构是将分子图结构映射到平面上，以便更直观地展示原子之间的连接关系和空间结构，其中原子属性对应节点信息，键属性对应边信息，键连接方式对应邻接矩阵以及特征矩阵，对于2D结构数据需要采用专门的图神经网络进行模型训练；3D分子结构则是描述分子的立体空间结构，包括原子的立体位置、键角度和手性中心等信息，通常被用于药物分子的药效学研究和分子对接。

而考虑到含能材料特殊的性质及结构特征，可以通过添加基于领域内知识和经验的自定义特征来补充描述分子结构，以提高模型的性能和预测精度，加入自定义特征后可以更全面地描述和区分不同结构的含能材料，为含能材料的研究和设计提供更有针对性的信息。如马里兰大学的Ealton等[31]系统研究了机器学习在含能材料性能预测中的应用，他们以一个含有300多个含能分子的数据库作为训练集，分别尝试了包括键加和（sum over bond）、库仑矩阵（Coulomb matrices）、键袋（bag of bonds）、多种分子指纹谱以及自定义描述符集（custom descriptor set）在内的多种特征（描述符）提取方法，结果表明，通过组合式的特征提取方法，模型精度会得到较大提升。Song等[23]构建了复合描述符集来表示分子，它由两部分组成，一部分为包含碳、氢、氧、氮和卤素的电拓扑状态指数（electrotopological state indice，E-state）指纹谱，另一部分是基于分子结构和组成的29种自定义描述符。Chen等[32]提出了空间矩阵描述符的概念，在此概念下构造了基于体积占用空间矩阵和热贡献空间矩阵的描述符作为模型的输入特征，其从原子结构的角度来表征高能分子质量和能量的空间分布，从而准确预测晶体密度和固相生成热。Deng等[33]以撞击感度为目标属性构建了17种自定义描述符作为输入特征，结果表明影

响撞击感度的主要因素为氧原子间的几何距离、含氧双键数以及亲水性。钱博文等[34]利用Dragon2.1软件计算了1481种分子描述符，并采用遗传算法对描述符进行筛选，最终确定了6个描述符，采用多元线性回归和人工神经网络在含有149个含能分子的数据集上构建了预测模型。Fathollahi等[35]从32种含能共晶的分子结构出发，提取出1600多种分子描述符，分别采用人工神经网络和多线性回归构建模型对含能共晶密度进行预测，在该研究中，他们提出了一种基于单分子描述符计算共晶描述符的方法，并且根据相关系数对特征描述符集进行降维，找到了3个与密度最为相关的系数。

1.3 模型构建及分子筛选

准备好数据并确定数据表示方式后，下一步就是选择合适的算法来训练模型。机器学习使用的基础算法包括朴素贝叶斯，K近邻（KNN），决策树（decision-tree），支持向量机（SVM），核岭回归（KRR），神经网络（NN）。除朴素贝叶斯仅能用于分类任务，核岭回归仅能用于回归任务外，其他几种算法既适用于回归也适用于分类，其具体介绍如下：

（1）朴素贝叶斯以给定数据作为先验知识，根据贝叶斯定理确定最可能的假设，算法的逻辑性简单且较为稳定，但是对输入数据的表达方式较为敏感，如果选择不合适的特征表述可能会导致分类效果不佳。

（2）KNN会对新样本与训练数据在样本假设空间内的距离进行计算，然后根据距离确定离新样本最近的K个点，并根据这K个点的值做出最终预测。KNN算法简单易懂，易于实现，不需要对数据分布做出假设，因此适用于非线性数据；同时训练复杂度较低，适用于大型数据集，然而当数据集较大时计算复杂度高，需要人为确定K值，选择不当可能导致较大误差。此外，KNN对异常值敏感，异常值可能会对预测结果产生较大影响。

（3）决策树具有类似流程图的结构，用于确定行动流程及其对应结果。完整的决策树包含根节点、叶节点及分支；根节点和叶节点中包含有每一步的问题或决策准则。决策树能够快速适应数据集，效率高，但是容易出现过拟合，数据的细微变化可能会导致树结构的巨大变化，使得结果不稳定。

（4）支持向量机、核岭回归具有一定的相似性，它们均利用核函数（核函数能够对输入量的相似度进行计算）将原始输入数据映射到更高维空间以简化计算，最终实现预测准确性的提升，可以解决高维数据的问题，泛化能力比较强，但是对非线性问题解决能力不强，对缺失数据敏感。

（5）人工神经网络及深度神经网络的基本单元是人工神经元，通常由输入层、隐藏层和输出层组成，神经元之间互相连接，学习的过程就是连接权重不断调整以尽可能地缩小预测值与实际值之间差距的过程。人工神经网络可以解决非线性问题，具有较高的鲁棒性及泛化能力，但是需要大量数据来提高准确性，且可能受到不充分数据的影响。

模型训练好后需要对预测模型进行不断优化改进，最终实现模型精度最优。含能分子设计的核心是根据性能预测结果对虚拟筛选空间中的分子结构进行排序，排名靠前的分子结构将被推荐给合成化学家进行进一步的合成探索[36-37]。因此，为保证含能分子设计的"成功率"，首要任务是构建高效的虚拟筛选空间。近年来，基于机器学习的高通量筛选技术被广泛应用在含能分子设计中，大大提高了含能分子的设计成功率。比如，Jiang等[38]利用神经网络获得了分子晶体的共晶可行性判据模型，并成功制备出一种新型含能共晶。Huang等[39]设计了晶体结构描述符，并将其作为梯度提升（XGBoost）等5种算法的输入特征，进而开发了以爆热、爆速、爆压、热分解温度和晶格能为目标性能的机器学习模型。王润文等[40]在预筛选出高密度骨架的基础上构建了潜在高能化合物的搜索空间，并将高通量计算与深度学习相结合预测了密度、生成焓、爆速、爆压和X—NO_2键解离能，最终筛选出了6个性能优异的候选分子。Xie等[41]提出了含能材料属性导向的自适应设计框架，首先通过枚举88个母环和13个取代基团所有可能组合生成了84083个样本构成搜索空间，并从中选择了88个数据作为初始训练数据集，以生成热和爆炸热作为目标属性，使用键加和（SOB）、扩展连通性指纹（ECFP），E-state指纹谱（E-state）、自定义描述符集（CDS）4种描述符来表征分子结构，接着作者利用线性回归、LASSO回归、核岭回归、支持向量机线性核（SVR.lin）、支持向量机径向核（SVR.rbf）、高斯回归（GPR）6种机器学习算法并结合"开发"、"探索"、知识梯度算法、随机算法等5种优化器来不断迭代筛选目标分子，最后通过高精度量子化学计算验证分子性能，虚拟筛选空间构建及分子筛选流程如图3所示。

Liu等[42]开发了含能材料高通量计算系统EM-Studio，其中的生成模块通过基于深度学习的RNN生成模型与片段对接相结合的方式，实现了分子

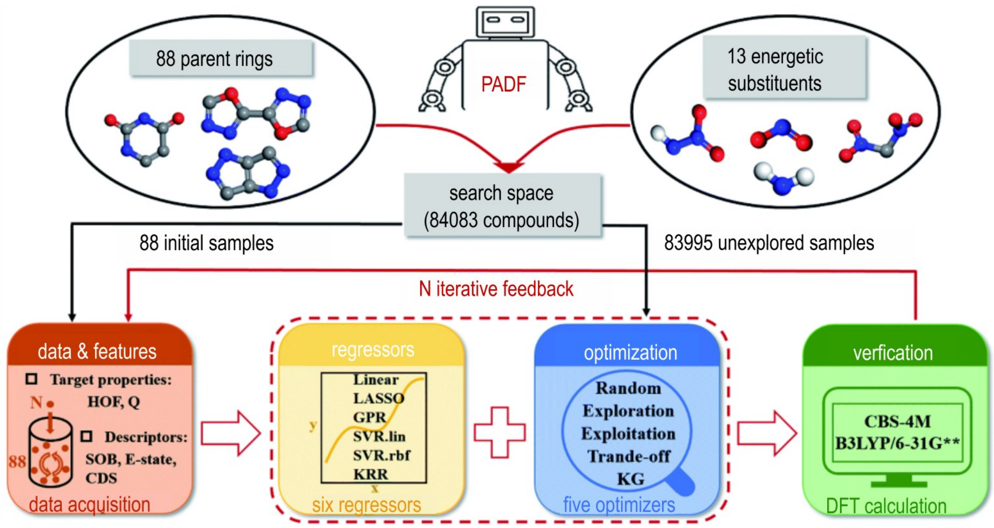


图3 基于启发式枚举法的虚拟筛选空间构建及目标分子筛选[41]

Fig.3 Construction of virtual screening space and screening of target molecules based on heuristic enumeration[41]

虚拟筛选空间的构建，如图4所示。然而，该方法对数据的需求量很大，需要针对不同的生成任务适应性地调整和优化参数，以提高分子生成的有效性。

这些研究工作将机器学习和高通量筛选技术应用于含能分子设计，为含能分子的高效筛选提供了解决方案。然而，在构建虚拟筛选空间和性能预测模型方面仍有改进的空间。对于虚拟空间的构建，需要进一步优化构建方法，将更多的分子结构与性质特征纳入考虑以提高筛选的准确性，而不仅仅依靠知识和经验选定目标结构单元和官能团。对于性能预测模型，已有许多方法可以实现与爆轰性能密切相关的爆速、爆压、密度等性质的准确预测，然而对于化学稳定性相关的性质，比如热分解温度、感度等的预测还有较大的提升空间。

2 氮杂多环含能化合物的开发

2.1 氮杂稠环含能化合物

氮杂稠环类含能化合物在平衡能量和稳定性上具有独特的优势，通过将机器学习和高通量筛选相结合，研究人员能够更高效地设计和筛选具有高能量密度和稳定性的氮杂稠环含能材料，这种数据驱动的方法为稠环含能材料的研发提供了新的途径。

2022年西南科技大学王润文联合中物院化工材料研究所刘健等[40]提出了预筛选分子骨架提升虚拟筛选空间整体性能的方案，该研究首先在CCDC及GDB-17两个数据库中预筛选出高密度分子骨架，通过片段组装构建了由潜在高密度含能分子组成的虚拟

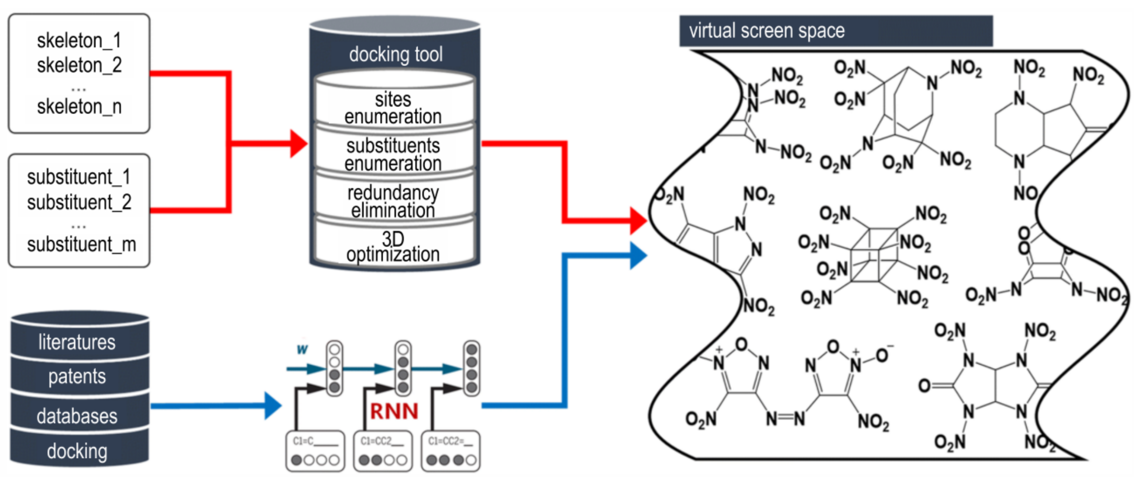


图4 深度学习生成模型与片段对接结合构建虚拟筛选空间[42]

Fig.4 Combining deep learning generative models with fragment docking to construct a virtual screening space[42]

筛选空间，然后在此基础上，利用深度学习模型、量子化学计算和爆轰产物状态方程等方法预测了分子的晶体密度、生成焓、爆轰性能、X—NO_2键解离能和撞击感度等指标，通过这些性能指标完成了含能分子的排序及筛选，最后得到了能量水平优于RDX、稳定性优于TNT的6个新型稠环含能分子，实现了含能分子的高效设计，开发过程如图5所示。虽然作者并没有通过实验验证这6个候选分子性能的预测准确性，但是作者提出的开发流程对提升高能分子的筛选效率具有借鉴意义。

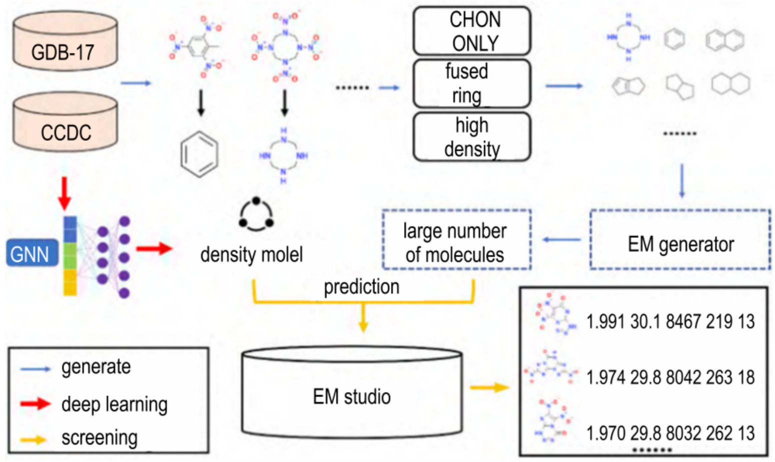


图5 高通量计算与深度学习模型相结合的稠环含能分子设计原理[40]

Fig.5 Principle of designing fused energetic molecules using high-throughput computing combined with deep learning model[40]

2022年，Song等[23]将高通量筛选与机器学习相结合，构建了机器学习辅助的含能分子高通量筛选系统，该系统将基于SMILES表示的分子结构作为输入，从已报道的文献中收集了上千条由碳氢氧氮及卤素组成的中性含能化合物的性能数据，包括密度、爆速、爆压、熔点、热分解温度及撞击感度6个维度的数据集，用于KRR预测模型的训练；同时，通过启发枚举式生成方法生成了包含25112个[5,6]稠环分子的筛选空间，通过引入密度、爆速、热分解温度的性能阈值后筛选得到99个[5,6]氮杂稠环含能分子；紧接着作者从CCDC中收集了365个非石墨层状堆积结构和22个类石墨层状堆积结构的数据，经过数据增强后构建了基于卷积神经网络（CNN）和长短期记忆网络（LSTM）的分类模型，从99个分子中筛选出具有类石墨层状堆积结构的候选分子7,8-二硝基吡唑并[1,5-*a*][1,3,5]三嗪-2,4-二胺（ICM-104），并实现了ICM-104的实验室合成和性能测试，研究的工作流程如图6所示。作者最后评估了含能化合物ICM-104的爆轰性能及热稳定性，并对实验结果和模型预测结果进行了比较，从结果上看密度、爆轰性能的预测误差较小，而分解温度则存在较大的误差，说明作者所选用的复合描述符不能准确地描述分子间相互作用。对于构建热分解温度模型，可考虑增加与热分解温度相关的描述特征，比如电子亲和能、离子化能、分子轨道能级等，并在数据收集时统一实验测试仪器的型号和测试条件，尽量减少不必要的噪声影响。

2022年Wen等[43]基于领域内的相关知识构建了一套筛选标准（例如：骨架由C、H、O、N四种元素组成；碳原子数/氮原子数≥0.33；杂环的总数≤4并且至少包含一个恶二唑环；没有三元环或四元环结构），从ZIN20数据库的可购买子集中筛选出171个类恶二唑稠环骨架，通过组合设计将骨架与氨基、硝基、叠氮基、甲基等基团组合生成含有约10^7个分子的筛选空间，以氧平衡、密度、合成可行性、爆轰性能以及静电平衡参数的阈值来筛选分子。作者首先利用氧平衡和合成可行性评分筛选出7500个分子，然后使用晶体密度拟合方程快速筛选出密度值排名前1%的分子，并使用密度泛函理论（DFT）结合EXPLO5进行精确计算。最后，利用静电平衡常数分析分子的静电感度，得到了兼顾能量与稳定性的6个分子，具体筛选过程如图7所示。遗憾的是，虽然通过理论计算验证了分子的性能，但作者并未判断分子实际合成的可行性，也未从实验上加以验证。

数据驱动方法在稠环含能分子的开发方面已取得一定的进展。首先，通过结合大规模的分子数据库和机器学习算法，研究人员能够更好地理解稠环结构与性能之间的关系，从而设计出更具潜力的稠环含能化

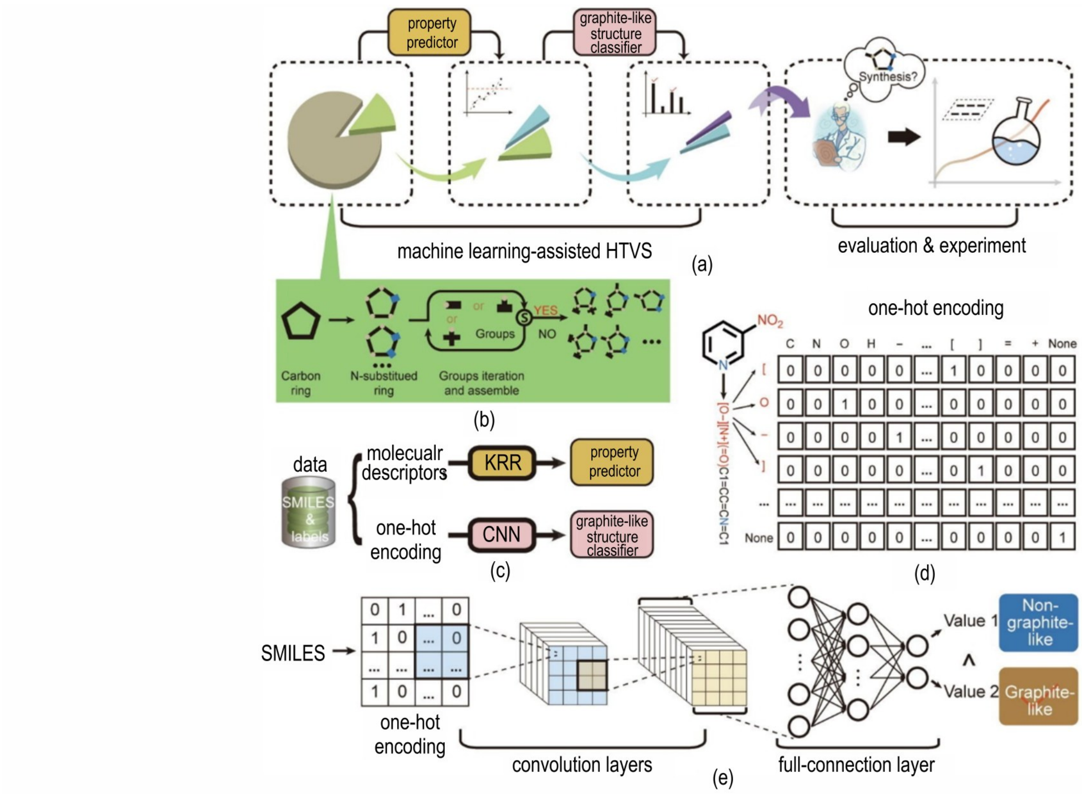


图6 机器学习辅助的高通量筛选及实验研究流程图[23]

Fig.6 Machine learning-assisted high-throughput screening and experimental research process diagram[23]

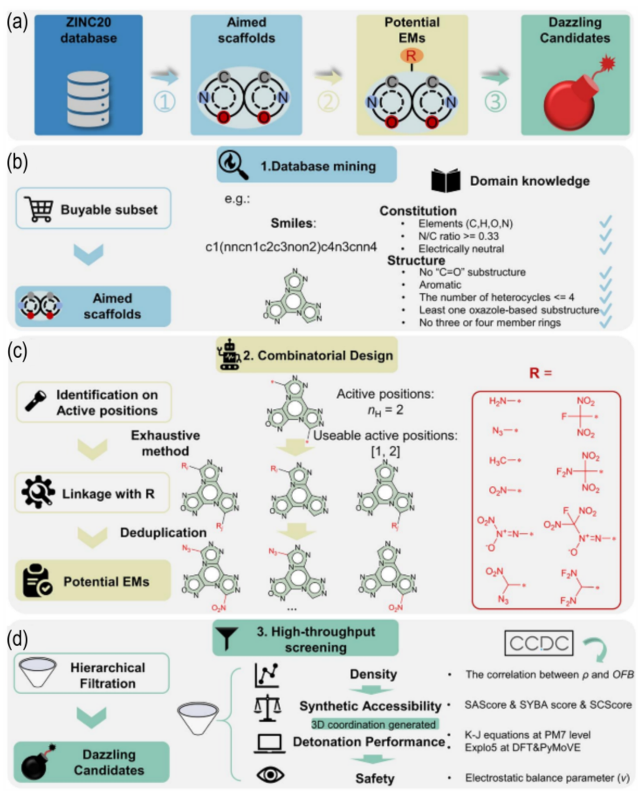


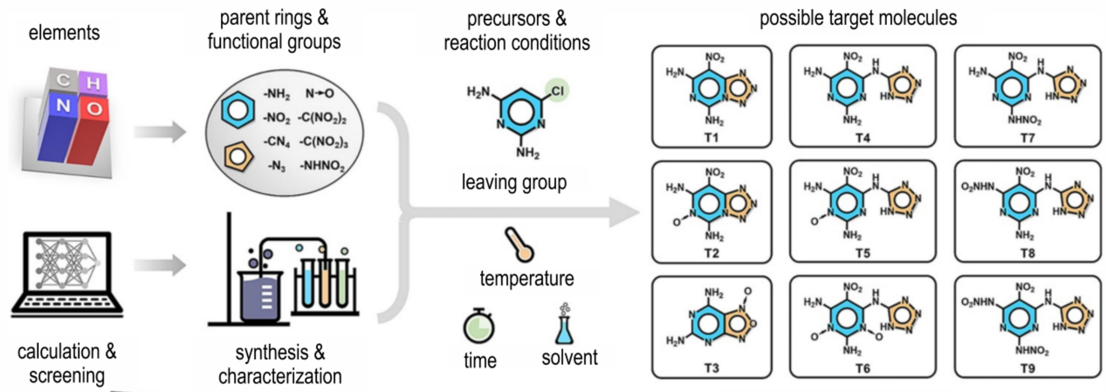
图7 筛选空间构建及稠环分子筛选过程[43]

Fig.7 Search space construction and the screening process of fused cyclic molecules[43]

合物，提高研发效率。其次，建立结构与性能之间的定量关系模型，使得研究人员能够在分子设计阶段准确预测其性能，指导后续的实验工作。总的来说，数据驱动方法为稠环含能分子的开发提供了新的思路和工具，然而，目前仍面临一系列问题和挑战。虚拟筛选空间中构建的分子虽然符合化合价规则，但现有研究工作未从实验合成角度考虑筛选出分子的可合成性，可能导致筛选得到的分子结构不合理和合成可行性不高，对于含能分子设计与合成的实际指导意义有限。因此，未来的工作需要考虑选择合适的合成路径、可控的反应条件，以提高数据驱动方法的准确性和实用性。

2.2 氮杂联环含能化合物

氮杂联环化合物的分子骨架具有较大范围的分子内共轭效应，在提高分子的热稳定性方面具有一定的优势。此外，氮杂联环含能化合物还具有更多的可修饰位点，可以通过分子修饰调节分子的性能，从而平衡分子的能量和稳定性。为加速氮杂联环含能化合物的开发，研究人员也开始将数据驱动方法作为一种重要的开发工具应用到其中。2023年Cao等[44]利用核岭回归机器学习模型对图8中T1~T9这9种经过不同官



No	ρ (g·cm⁻³)	ΔH_f (kJ·mol⁻¹)	D_v (m·s⁻¹)	P (GPa)	Nitro charge	No	ρ (g·cm⁻³)	ΔH_f (kJ·mol⁻¹)	D_v (m·s⁻¹)	P (GPa)	Nitro charge
T1	1.81	408.69	8536	28.8	-0.315	T6	1.85	449.14	8854	30.8	-0.328
T2	1.87	438.48	8901	33.2	-0.292	T7	1.84	532.55	8865	31.2	-0.340, -0.103
T3	1.74	305.22	8160	26.3	/	T8	1.84	586.77	8713	30.7	-0.339, -0.068
T4	1.77	472.84	8398	26.3	-0.371	T9	1.90	637.73	9170	35.4	-0.310, -0.050, -0.085
T5	1.82	450.72	8692	30.0	-0.348						

图8 基于嘧啶的含能材料的分子设计、数据筛选和性能预测[44]

Fig.8 Molecular design, data screening and performance prediction of pyrimidine-based energetic materials[44]

能团（氨基、硝基、硝铵基、氮氧化基）修饰的嘧啶衍生化合物的主要性能进行了预测，其中T4由于其较低的硝基电荷数和较高的能量，因此兼具良好的热稳定性及能量水平，接着作者通过实验合成了该分子并验证了模型预测的准确性，结果证明了该化合物具有优异的热稳定性、较低的机械感度和良好的爆轰性能，可与传统的耐热炸药相媲美。

2023年Lu等[45]利用高通量筛选系统来开发亚氨基桥联嘧啶环与五元杂环的含能材料，开发流程如图9所示。作者将嘧啶环与8种五元杂环骨架（吡唑、咪唑、1，2，3-三唑、1，2，4-三唑、1，2，4-恶二唑、1，2，5-恶二唑、1，3，4-恶二唑、四唑）通过亚氨基相连，构建了33个骨架，随后将氨基、硝基、羟基3种取代基团在骨架上循环添加，生成了包含12816个分子的初始筛选空间；接着根据能量特性、平面性、爆速、键解离能从搜索空间中筛选出了5个能量和稳定性都较好的分子，通过实验合成了其中3个分子，实验结果表明编号为“K19-21”的分子分解温度超过320 ℃，爆速接近8300 $m\cdot s^{-1}$，有望成为新型耐热炸药。

2002年Ma等[46]考虑到三唑并三嗪分子骨架在稠环分子中具有较好的氮含量以及较多的易修饰位点，利用SciFinder数据库筛选三唑并三嗪分子骨架，首先在SciFinder数据库中搜索到4617个相似度指数在0%~100%的三唑并三嗪稠环骨架，随后通过限制相似度指数小于60%后得到了708个骨架，在引入硝

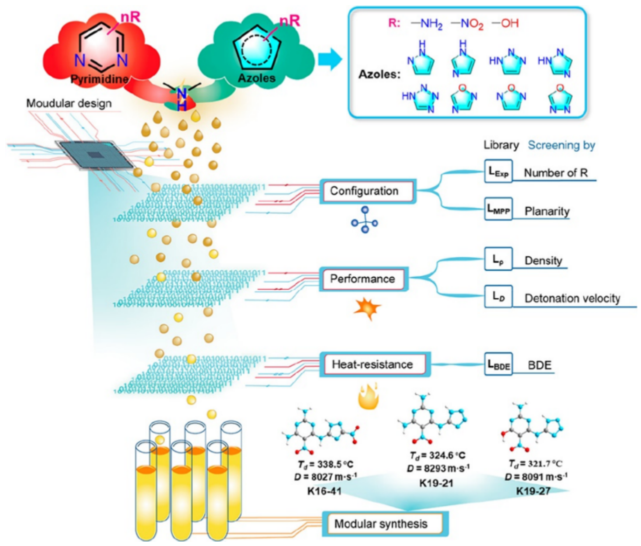


图9 虚拟空间构建及筛选流程[45]

Fig.9 Virtual screening construction and screening steps[45]

基基团后将这一数字减少至451个，最后使用常见官能团（叠氮基、硝基、双硝基、四唑基、氮氧化基、氨基、硝氨基）来取代骨架上的取代位点，在至少保留一个分子内氢键作用后得到了16个候选分子，如图10所示。作者最后合成了7-硝基-3-（1*H*-四唑-5-基）-[1，2，4]三唑[5，1-*c*][1，2，4]三嗪-4-胺-2-氧化（NTTO）这一合成性较高的分子，实验结果表明，NTTO具有1.811 $g\cdot cm^{-3}$的高密度，其能量性能优于RDX，并且具有较高的热稳定性。

2024年，Wen等[47]受到干细胞工作机制的启发

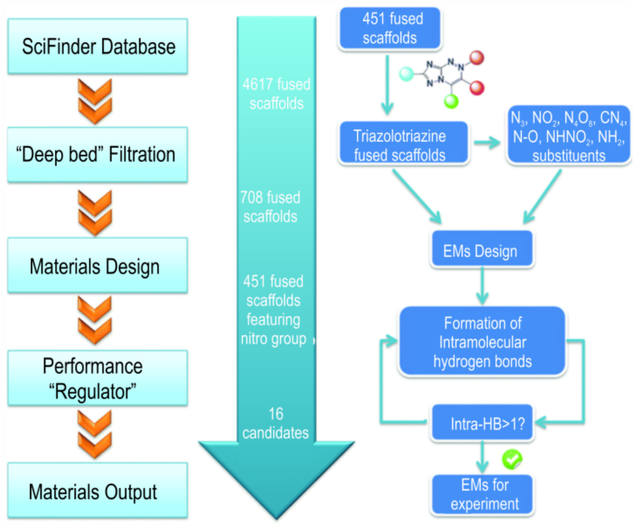


图10　基于三唑并三嗪骨架的开发流程[46]

Fig. 10　Development workflow based on triazolotriazine scaffold[46]

提出了具有多功能模块化的含能材料（MMEMs）的概念，可以根据需要通过不同的加工形式，实现具有不同应用功能的含能材料制备，比如晶体炸药、熔铸炸药、高能粘合剂、增塑剂等。作者通过组合设计将环氧丁烷、氮杂五元环、桥联键及含能基团相连接生成了含有近900万个分子的筛选空间，使用机器学习预测模型，根据合成可行性指数SAScore和SCScore、密度、落锤高度（h_{50}）和热分解温度筛选出了267个候选分子，作者最后对其中3个分子进行了实验合成及性能分析，结果表明所有分子都满足MMEMs预期的性能要求，即具有良好的热稳定性和能量性能，以及优异的聚合性能，分子生成及筛选过程如图11所示。

数据驱动应用于氮杂联环含能化合物的开发也面临着一些问题。首先，联环含能化合物的结构更加复杂多样，其设计需要考虑到多个环之间的相互作用，这增加了设计的复杂性。其次，联环含能化合物的合成通常更加困难，需要考虑到多步反应的选择和控制，以及可能涉及的多环连接方法，这对于数据驱动方法的应用也提出了更高的要求。此外，联环含能化合物性能变化也更加复杂，因为多个环之间的相互作用可能会导致分子性质的非线性改变，需要进行更加深入和全面的实验测试。因此，如何有效地利用数据驱动方法解决这些问题，提高联环含能化合物设计、合成和性能预测的效率和准确性，是当前需要解决的挑战之一。

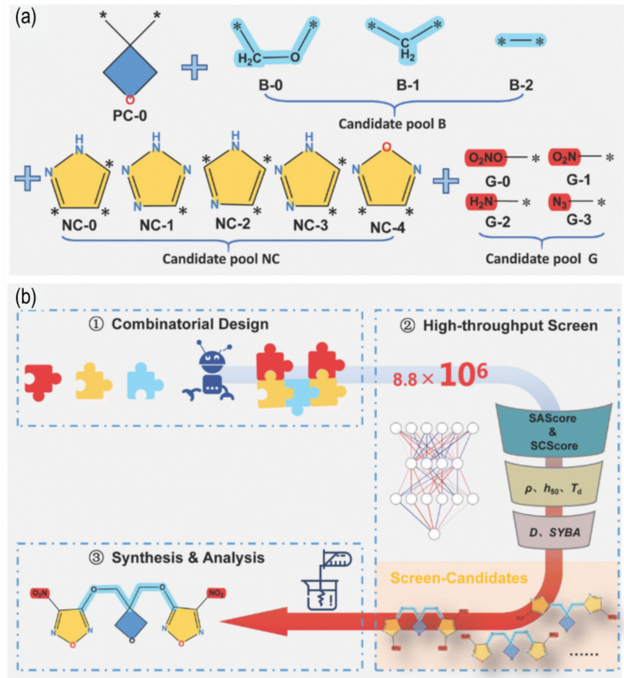


图11　分子生成及筛选工作流程[47]

Fig.11　Molecular generation and screening workflow[47]

3　结论与展望

DGPW简要介绍了数据驱动用于含能材料开发的流程，概述了数据驱动氮杂稠环和氮杂联环两类多环含能化合物研究的进展，同时讨论了当前数据驱动辅助多环含能分子开发过程中存在的问题，得到如下结论：

氮杂多环含能化合物因其较高的分子内能量和稳定的结构，在平衡能量和稳定性上具有独特的优势，是目前含能材料研究的热点。随着人工智能技术的快速发展和广泛应用，以机器学习为代表的数据驱动方法结合高通量筛选技术被越来越多地应用于多环含能材料的开发上，然而研究领域还存在诸多困难，各个环节仍有提升的空间。一方面，样本量不足、数据集质量不高、特征工程不充分等因素制约着模型的精度和泛化能力；另一方面，含能分子的合成涉及多种类型和机理的反应，高通量筛选出的分子也可能存在结构不合理、无法合成的情况。总体而言，与药物分子和其他材料相比，数据驱动辅助多环含能材料的开发还存在明显的差距，未来的发展可以从以下两个方面考虑：

（1）提高模型的预测精度和泛化能力。通过高通量计算生成虚拟数据补充样本量，或借鉴其他材

料设计中数据增强或数据扩充的经验有望解决含能分子样本容量较少的问题；通过数据治理，比如定义材料数据质量维度（数据完整性、数据测试条件一致性、可验证性），解决数据质量不高的问题；利用药物分子等其他领域预训练好的模型及算法，通过微调、迁移学习等方式来设计适用于含能分子机器学习的模型，针对不同的含能分子增加与其相关的描述特征符，有望提高模型的准确性及泛化能力。

（2）分子合成路线的智能设计。生成的大规模候选分子空间中，具有实际应用与合成价值的目标化合物通常只占很小比例，且含能分子的合成涉及多种反应类型、机理及条件，这也为分子的合成带来巨大挑战。目前在药物分子领域，机器学习可以加速化学反应条件的筛选，确定化学原料、操作条件（压力、温度、时间等），将其应用于氮杂多环含能化合物的反应条件、合成路径的设计将有助于进一步提升研发效率。

参考文献：

[1] 孙承纬，卫玉章，周之奎. 应用爆轰物理[M]. 北京：国防工业出版社，2000：216-239.
SUN Cheng-wei，WEI Yu-zhang，ZHOU Zhi-kui. Applied detonation physics[M]. Beijing：National Defense Industry Press，2000：216-239.

[2] ZHANG J，SHREEVE J N M. 3，3′-Dinitroamino-4，4′-azoxyfurazan and its derivatives：An assembly of diverse N-O building blocks for high-performance energetic materials[J]. *Journal of the American Chemical Society*，2014，136(11)：4437-4445.

[3] CHAVEZ D E，HISKEY M A，GILARDI R D. 3，3′-azobis(6-amino-1，2，4，5-tetrazine)：a novel high-nitrogen energetic material[J]. *Angewandte Chemie-International Edition*，2000，39(10)：1791-1793.

[4] FISCHER D，GOTTFRIED J L，KLAPOETKE T M，et al. Synthesis and investigation of advanced energetic materials based on bispyrazolylmethanes [J]. *Angewandte Chemie-International Edition*，2016，55(52)：16132-16135.

[5] WANG B，QI X，ZHANG W，et al. Synthesis of 1-(2*H*-tetrazol-5-yl)-5-nitraminotetrazole and its derivatives from 5-aminotetrazole and cyanogen azide：a promising strategy towards the development of C-N linked bistetrazolate energetic materials[J]. *Journal of Materials Chemistry A*，2017，5(39)：20867-20873.

[6] KLAPOETKE T M，PETERMAYER C，PIERCEY D G，et al. 1，3-Bis(nitroimido)-1，2，3-triazolate anion，the N-nitroimide moiety，and the strategy of alternating positive and negative charges in the design of energetic materials[J]. *Journal of the American Chemical Society*，2012，134(51)：20827-20836.

[7] 董海山. 高能量密度材料的发展及对策[J]. 含能材料，2004，12(Z1)：216-239.
DONG Hai-shan. Development and countermeasures of high-energy-density materials[J]. *Chinese Journal of Energetic Materials*（*Hanneng Cailiao*），2004，12(Z1)：216-239.

[8] 邓仲华，李志芳. 科学研究范式的演化——大数据时代的科学研究第四范式[J]. 情报资料工作，2013，4：19-23.
DENG Zhong-hua，LI Zhi-fang. Evolution of scientific research paradigms：The fourth paradigm of scientific research in the era of big data [J]. *Intelligence Information work*，2023，4：19-23.

[9] 陈明. 数据密集型科研第四范式[J]. 计算机教育，2013，9：107-110.
CHEN Ming. The fourth paradigm of data-intensive research [J]. *Computer education*，2013，9：107-110.

[10] MJOLSNESS E，DECOSTE D. Machine learning for science：state of the art and future prospects[J]. *Science*，2001，293(5537)：2051.

[11] REN F，WAED L，WILLIAMS T，et al. Accelerated discovery of metallic glasses through iteration of machine learning and high-throughput experiments [J]. *Science Advance*，2018，4：1566.

[12] LU S，ZHOU Q，OUYANG Y，et al. Accelerated discovery of stable lead-free hybrid organicinorganic perovskites via machine learning[J]. *Nature Communication*，2018，9：1-8.

[13] GÓMEZ B R，AGUILERA I J，HIRZEL T D，et al. Design of efficient molecular organic light-emitting diodes by a high-throughput virtual screening and experimental approach [J]. *Nature Materials*，2016，15：1120-1127.

[14] SAHU H，YANG F，YE X，et al. Designing promising molecules for organic solar cells via machine learning assisted virtual screening[J]. *Journal of Materials Chemistry A*，2018，6：4948-4954.

[15] ZHUO Y，TEHRANI A M，OLIYNYK A O，et al. Identifying an efficient，thermally robustinorganic phosphor host via machine learning[J]. *Nature Communication*，2018，9：1-10.

[16] KANG P，LIU Z，ABOU-RACHID H，et al. Machine-learning assisted screening of energetic materials [J]. *The Journal of Physical Chemistry A*，2020，124(26)：5341-5351.

[17] NGUYEN P，LOVELAND D，KIM J T，et al. Predicting energetics materials' crystalline density from chemical structure by machine learning[J]. *Journal of Chemical Information and Modeling*，*American Chemical Society*，2021，61(5)：2147-2158.

[18] SANCHEZ-LENGELING B，ASPURU-GUZIK A. Inverse molecular design using machine learning：generative models for matter engineering[J]. *Science*，*American Association for the Advancement of Science*，2018，361(6400)：360-365.

[19] KUMAR D，IMLER G H，PARRISH D A，et al. A highly stable and insensitive fused triazolo-triazine explosive（TTX）[J]. *Chemistry-a European Journal*，2017，23(8)：1743-1747.

[20] 霍欢，王伯周，廉鹏，等. 三种稠环硝胺化合物的爆炸性能估算及其硝化母体化合物的合成[J]. 火炸药学报，2014，37(1)：21-30.
HUO Huan，WANG Bo-zhou，LIAN Peng，et al. Estimation of explosive performances for three fused ring nitramine com pounds and synthesis of their nitrification parent ring compounds [J]. *Chinese Journal of Explosives and Propellants*，2014，37(1)：21-30.

［21］ SINGH J，STAPLES R J，HOOPER J P，et al. Pyrazole bridges ensure highly stable and insensitive bistetrazoles［J］. *Chemical Engineering Journal*，2022，431：133282.

［22］ TENG Z，ZHENG S，KAI S. Big data creates new opportunities for materials research：A review on methods and applications of machine learning for materials design［J］. *Engineering*，2019，5：1017-1026.

［23］ SONG S，WANG Y，CHEN F，et al. Machine learning-assisted high-throughput virtual screening for on-demand customization of advanced energetic materials［J］. *Engineering*，2022，10：99-109.

［24］ SONG S，CHEN F，WANG Y，et al. Accelerating the discovery of energetic melt-castable materials by a high-throughput virtual screening and experimental approach［J］. *Journal of Materials Chemistry A*，2021，9(38)：21723-21731.

［25］ CHANDRASEKARAN N，OOMMEN C，KUMAR V R S，et al. Prediction of detonation velocity and N-O composition of high energy C—H—N—O explosives by means of artificial neural networks［J］. *Propellants Explosives Pyrotechnics*，2019，44(5)：579-587.

［26］ CASEY A D，SON S F，BILIONIS I，et al. Prediction of energetic material properties from electronic structure using 3D convolutional neural networks［J］. *Journal of Chemical Information and Modeling*，2020，60(10)：4457-4473.

［27］ YANG C，CHEN J，WANG R，et al. Density prediction models for energetic compounds merely using molecular topology［J］. *Journal of Chemical Information and Modeling*，2021，61(6)：2582-2593.

［28］ LI C，WANG C，SUN M，et al. Correlated RNN framework to quickly generate molecules with desired properties for energetic materials in the low data regime［J］. *Journal of Chemical Information and Modeling*，2022，62(20)：4873-4887.

［29］ ROGERS D，HAHN M. Extended-connectivity fingerprints［J］ *Journal of Chemical Information and Modeling*，2010，50(5)：742-754.

［30］ DURANT J L，LELANDB A，HENRY D R，et al. Reoptimization of MDL keys for use in drug discovery［J］. *Journal of Chemical Information and Computer Science*，2002，42：1273-1280.

［31］ ELTON D C，BOUKOUVALAS Z，BUTRICO M S，et al. Applying machine learning techniques to predict the properties of energetic materials［J］. *Scientific Reports*，2018，8：1-12.

［32］ CHEN C，LIU D，DENG S，et al. Accurate machine learning models based on small dataset of energetic materials through spatial matrix featurization methods［J］. *Journal of Energy Chemistry*，2021，63：364-375.

［33］ DENG Q，HU J，WANG L，et al. Probing impact of molecular structure on bulk modulus and impact sensitivity of energetic materials by machine learning methods［J］. *Chemometrics and Intelligent Laboratory Systems*，2021，215：104331.

［34］ 钱博文，陈利平，陈网桦. 基于遗传算法的人工神经网络预测多硝基化合物撞击感度［J］. 含能材料，2016，24(7)：644-650.
QIAN Bo-wen，CHEN Li-ping，CHEN Wang-hua. The prediction of the impact sensitivity of polynitro compounds based on a genetic algorithm and artificial neural networks［J］. *Chinese Journal of Energetic Materials*（*Hanneng Cailiao*），2016，24(7)：644-650.

［35］ FATHOLLAHI M，SAIADY H. Prediction of density of energetic cocrystals based on QSPR modeling using artificial neural network［J］. *Structural Chemistry*，2018，29：1119-1128.

［36］ 刘锐，刘健，唐岳川，等. 人工智能辅助含能分子设计的应用与展望［J］. 含能材料，2024，32(4)：408-421.
LIU Rui，LIU Jian，TANG Yue-chuan，et al. Problems and thoughts of AI-assisted design of energetic molecules［J］. *Chinese Journal of Energetic Materials*（*Hanneng Cailiao*），2024，32(4)：408-421.

［37］ LIU R，TANG Y，TIAN J，et al. QSPR models for sublimation enthalpy of energetic compounds［J］. *Chemical Engineering Journal*，2023，474：145725.

［38］ JIANG Y，GUO J，LIU Y，et al. Coupling complementary strategy to flexible graph neural network for quick discovery of coformer in diverse co-crystal materials［J］. *Nature Communication*，2021，12(1)：1-14.

［39］ HUANG X，LI C，TAN K，et al. Applying machine learning to balance performance and stability of high energy density materials［J］. *iScience*，2021，24(3)：102240.

［40］ 王润文，杨春明，刘健. 高通量计算与深度学习相结合的稠环含能化合物设计［J］. 含能材料，2022，30(12)：1226-1236.
WANG Run-wen，YANG Chun-ming，LIU Jian. Exploring novel fused-ring energetic compounds via high-throughput computing and deep learning［J］. *Chinese Journal of Energetic Materials*（*Hanneng Cailiao*），2022，30(12)：1226-1236.

［41］ XIE Y，LIU Y，HU R，et al. A property-oriented adaptive design framework for rapid discovery of energetic molecules based on small-scale labeled datasets［J］. *RSC Advances*，2021，11(41)：25764-25776.

［42］ LIU J，ZHAO S，DUAN B，et al. High-throughput design of energetic molecules［J］. *Journal of Materials Chemistry A*，2023，11(45)：25031-25044.

［43］ WEN L，YU T，LAI W，et al. Transferring the available fused cyclic scaffolds for high-throughput combinatorial design of highly energetic materials via database mining［J］. *Fuel*，2022，324：124591.

［44］ CAO Y，SONG S，SHI J，et al. Synthesis and characterization of energetic molecules based on pyrimidine rings：Selection and verification of computational-assisted synthesis pathways［J］. *Chemical Engineering Science*，2023，282：119281.

［45］ LU Z J，HU Y，DONG W S，et al. From concept to synthesis：developing heat-resistant high explosives through automated high-throughput virtual screening［J］. *Journal of Physical Chemistry C*，2023，127(38)：18832-18842.

［46］ MA Q，CHENG Z，YANG L，et al. Accelerated discovery of thermostable high-energy materials with intramolecular donor-acceptor building blocks［J］. *Chemical Communications*，2022，58(28)：4460-4463.

［47］ WEN Y，WEN L，TAN B，et al. Bionic inspired multifunctional modular energetic materials：An exploration of new generation of application oriented energetic materials［J］. *Journal of Materials Chemistry A*，2021，9(38)：21723-21731.

Research Progress of Nitrogen Heteropolyclic Energetic Materials Based on Data-driven

LIU You-hai[1,2], **HUANG Shi**[1], **ZHANG Wen-quan**[1], **YANG Fu-sheng**[2]

（1. *Institute of Chemical Materials*, *CAEP*, *Mianyang* 621999, *China*; 2. *School of Chemical Engineering and Tecnology*, *Xi'an jiaotong University*, *Xi'an* 710049, *China*）

Abstract: The development of energetic materials faces many challenges, and the traditional trial-and-error research model often results in long development cycles and low efficiency. With the advancement of data science and artificial intelligence (AI) technologies, a data-driven research model has emerged as a new path for the development of energetic materials. Polycyclic energetic compounds are currently a hot topic in the field of energetic materials, among which nitrogen-containing polycyclic frameworks, due to the presence of π electrons for delocalized resonance and multiple modifiable sites, exhibit enhanced molecular structural stability. At the same time, the presence of energy groups ensures the energy level of the molecules, achieving a good balance between energy and stability, overcoming the inherent contradiction between them. This study briefly introduces the workflow of data-driven development of novel energetic materials, outlines the latest research progress of data-driven methods for the development of nitrogen-containing polycyclic energetic compounds, and finally proposes prospects for the application of data-driven methods in the development of novel energetic materials. Future directions should consider supplementing data volume through means such as data augmentation and governance to improve the accuracy and generalization ability of model predictions. Machine learning models can be used to predict the molecular synthetic feasibility by establishing chemical reaction conditions and synthetic pathways, thereby accelerating the development of novel nitrogen-containing polycyclic energetic compounds.

Key words: energetic materials; data-driven, nitrogen heteropolyclic energetic compounds; machine learning

CLC number: TJ55; O64 **Document code**: A **DOI**: 10.11943/CJEM2024088

Grant support: National Natural Science Foundation of China (No. 22375190)

（责编：卢学敏）