

## 核物理主题数据: 实验、理论与应用专题

## 核结构数据现代化: 提升评价效率与可用性\*

张一怒 罗皓天 袁岑溪<sup>†</sup> 李珂乐 李至奕 黄一鸣 蔡博帅

(中山大学中法核工程与技术学院, 珠海 519082)

(2025年11月10日收到; 2025年12月26日收到修改稿)

评价核结构数据库 (Evaluated Nuclear Structure Data File, ENSDF) 作为核物理领域的核心数据资源, 正面临数据量快速增长与传统格式僵化带来的挑战. 本文旨在系统探讨 ENSDF 的现代化进程, 深入分析其原始数据格式的历史局限性, 并综述国际核数据中心在数据结构现代化方面的主要努力. 文中重点介绍了 JSON、面向对象数据库及机器学习方法在核数据评价中的应用, 以及 NuDat 网络接口如何借助现代可视化技术实现数据的高效交互与传播. 此外, 本文还介绍了中国核数据团队在该领域的自主探索与实践, 特别是中山大学核数据团队在借鉴国际经验的基础上, 建立了本地化的原子核数据可视化查询系统. 基于这一框架, 本文进一步结合 JSON 化的核结构与衰变数据, 利用随机森林方法对超重核多种衰变模式的半衰期进行了系统建模与预测. 结果表明, 该模型在捕捉非线性关联、修正经验公式残差及识别主衰变模式方面具有显著优势, 与实验数据的一致率达 92.2%, 预测与实验半衰期均方根误差平均降低超过 50%. 研究结果显示, 现代化核结构数据库为人工智能在核物理领域的深入应用奠定了基础, 为我国核科学研究提供了坚实的数据支撑, 并体现了我国在核数据基础设施建设方面的自主创新能力. 本文数据集可在 <https://doi.org/10.57760/sciencedb.30258> 中访问获取.

**关键词:** 核数据, 可视化界面, 机器学习, 衰变**DOI:** 10.7498/aps.75.20251528**CSTR:** 32037.14.aps.75.20251528

## 1 引言: 核数据的基石

核物理学与核技术的持续发展包括核能、核医学、国防安全以及天体物理等领域, 对高质量、可追溯且可信的核数据提出了持续且不断增长的需求<sup>[1,2]</sup>. 评价核结构数据库 (Evaluated Nuclear Structure Data File, ENSDF)<sup>[3]</sup> 作为核物理领域的核心数据资源, 是一个收录经系统评价的核结构与衰变数据的综合性数据库, 其内容以标准化格式存储<sup>[4-6]</sup>. 该数据库由布鲁克海文国家实验室 (Brookhaven National Laboratory, BNL) 的国家核数据中心 (National Nuclear Data Center, NNDC) 负

责维护<sup>[7]</sup>, 并由国际核结构与衰变数据评价网络 (International Network of Nuclear Structure and Decay Data Evaluators) 持续贡献与更新, 以反映最新的实验结果与评价成果<sup>[8]</sup>.

ENSDF 涵盖所有已知核素的信息, 重点整合实验测量数据, 并作为核结构研究、核谱学应用、医学内辐射剂量计算以及《核数据表》(Nuclear Data Sheets) 等出版物的主要数据来源. 其关键数据类型包括核能级特性 (如激发能、半衰期、自旋、宇称及衰变模式)、 $\gamma$  射线特性 (如能量、强度、多极性), 以及  $\alpha$ ,  $\beta$  衰变特性等. 对于每个核素, ENSDF 均包含“采纳能级 (adopted levels)”数据集, 代表评价人员基于全部可用实验数据所确定的

\* 国家自然科学基金 (批准号: 12305129, 12475129)、广东省重点基础研究发展计划 (批准号: 2021B0301030006) 和核数据重点实验室基金 (批准号: JCKY2023201C152) 资助的课题.

<sup>†</sup> 通信作者. E-mail: [yuanxc@mail.sysu.edu.cn](mailto:yuanxc@mail.sysu.edu.cn)

最优推荐值.

作为 ENSDF 数据的主要用户接口, NuDat (Nuclear Data) 是由 NNDC 开发的网络化交互式应用程序, 可支持核结构、衰变及部分中子诱发反应数据的检索、查询、可视化与传播<sup>[9]</sup>. NuDat 的核心数据来源于 ENSDF 和核素卡片 (nuclear wallet cards), 并提供多种显示形式, 包括表格、能级示意图以及交互式核素图等.

随着基础核科学、核医学、反应堆设计与运行、地球物理、环境科学及辐射防护等领域对核数据的依赖不断增强, 加之实验技术的快速进步, 新数据源正以空前的速度积累. ENSDF 所依托的传统数据管理与传播体系正面临前所未有的挑战. 为了适应现代科学研究与信息化发展的需要, 对 ENSDF 及其关联系统 NuDat 的持续现代化改造已成为必然趋势. 由于原有的数据传播与访问机制沿用已逾数十年, 其功能与用户体验已难以满足现代网络技术与科研工作者的使用需求, 因此, 对这两个系统进行全面升级与重构势在必行<sup>[10]</sup>.

## 2 评价核结构数据库 (ENSDF): 简介与局限性

### 2.1 ENSDF 简介

ENSDF 是目前国际上唯一的综合性推荐核结构与衰变数据集. 其数据通过对所有可获得的实验结果进行严格审查, 并结合系统性趋势分析和理论模型约束而得到<sup>[5,6]</sup>. ENSDF 涵盖了核能级的主要特征信息, 如观测来源、激发能、半衰期、衰变模式、自旋与宇称; 以及  $\gamma$  射线的能量、强度、多极性、混合比和内转换系数等; 同时还包含特定辐射类型的能量与强度信息. 该数据库每年约有 220 个核素的数据得到全面评价, 此外还有若干核素的数据集被部分更新.

ENSDF 的维护工作由美国国家核数据中心 NNDC 代表国际核结构与衰变数据 (Nuclear Structure and Decay Data, NSDD) 网络负责执行, 该网络由国际原子能机构 (IAEA) 协调<sup>[11]</sup>. NSDD 网络目前汇集了来自 10 余个国家的 16 个数据中心及 20 多位国际公认的评价专家, 他们依照统一规范对全部已知同位素的核结构与衰变数据进行编制与评价. ENSDF 文件按质量数 ( $A$  链) 组织, 并在每个质量数内按照原子序数 ( $Z$  值) 递增排序.

对于每个核素, 相关信息以多个数据集的形式呈现, 通常每种实验类型对应一个独立的数据集. 此外, 每个核素均包含一个“采纳能级与  $\gamma$  射线”汇总数据集, 用以总结该核素的能级结构及  $\gamma$  射线特性.

中国核数据工作者长期活跃于 IAEA 组织与 NSDD 国际合作体系中, 承担了若干质量数链的 ENSDF 评价与更新任务, 并持续向 NuDat 等国际数据库提交高质量的结构与衰变数据条目. 其中, 中国原子能科学研究院负责  $A = 51, 62, 195-198$  质量链的评价工作, 吉林大学承担  $A = 52-56$  和  $A = 67$  质量链的评价任务. 与此同时, 国内团队依托放射性束装置与加速器实验平台 (如中国原子能科学研究院和中国科学院近代物理研究所等), 积极开展新实验结果的及时汇编、交叉核对及不确定度规范化处理, 推动国内测量成果快速纳入国际权威数据库.

此外, 中国团队还积极参与 NSDD 工作会议与技术讨论, 在评价流程标准化、参考文献溯源、同位素与同核异能态判据、 $\beta$  延迟发射数据及亚稳态数据一致性检查等方面提出了多项建设性方案. 通过工具链与培训资料的共享, 中国研究者显著提升了区域合作者在数据编制与质量控制方面的能力. 通过上述持续投入, 我国核数据群体不仅为国际核结构与衰变数据的完整性、可追溯性和时效性作出了重要贡献, 也为核物理研究、核能应用与核安全等领域提供了坚实的数据支撑<sup>[1,12,13]</sup>.

### 2.2 原始 ENSDF 格式的挑战与局限性

目前的 ENSDF 格式在数据管理与利用方面存在显著的固有局限, 这些局限直接制约了数据处理效率与质量.

如图 1 所示, 首先, 80 列 ASCII 固定宽度格式的限制带来了多重挑战. 该格式缺乏明确的分隔符与标签, 对非专业用户及评价人员而言均不易使用, 用户必须记忆各字段的固定位置和宽度, 而单位表达方式常不统一, 不对称误差亦缺乏规范支持. 这种空间受限的设计导致“大量数据被迫存储于注释中”, 且“缺乏进一步扩展空间”, 难以容纳新型核物理数据. ENSDF 的层级结构 (例如  $\gamma$  射线记录与能级记录的严格对应关系) 是通过文件中不同记录类型的顺序关系实现的, 任何数据转换操作都有可能破坏这种位置信息与层级连接, 造成数据完整性受损. 此外, 这种非标准化且僵化的 80

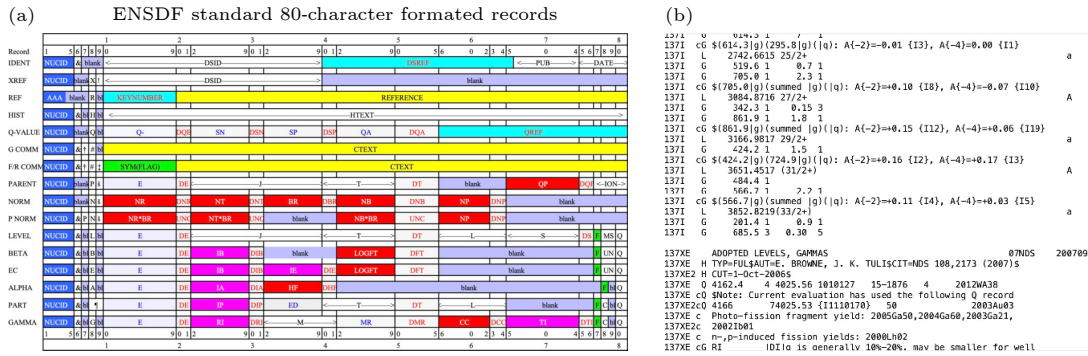


图 1 (a) ENSDF 标准 80 字符格式化记录说明图例; (b) ENSDF 数据实例节选

Fig. 1. (a) Illustration of the 80-character formatted record description according to ENSDF standard; (b) excerpt of ENSDF data instance.

列格式也严重限制了机器学习算法对核数据的高效解析与特征提取能力<sup>[14,15]</sup>.

其次, 评价吞吐量低是另一突出问题. 非机器可读文档及低效的手动数据提取是主要瓶颈. ENSDF 的原始数据, 尤其是来源于实验论文的表格信息, 通常以非结构化 PDF 文件形式存在. 现有基于规则的 PDF 表格解析工具效率有限且易出错, 导致大量提取误差, 需评价人员投入大量人力进行人工校正. 当处理规模达到数百至数百万文档时, 这些问题便成为“不可逾越的挑战”, 严重制约了评价流程的自动化与整体效率. 由于评价过程劳动密集且格式受限, ENSDF 的年度评价数量 (约 220—330 个核素) 明显低于实现“每个核素十年内至少更新一次”目标所需的工作量<sup>[14-17]</sup>.

第三, 数据访问与互操作性不足. 虽然 EXFOR, ENDF 与 ENSDF 等数据库在格式上已实现一定程度的标准化, 但其庞大的数据体量与复杂的层级结构仍使终端用户难以便捷获取特定应用所需信息. 部分特殊数据类型 (如  $\mu$  子诱发核反应数据) 虽包含于 ENSDF 中, 但其纳入方式尚未实现统一标准, 进一步削弱了系统的互操作性<sup>[15]</sup>.

此外, 数据完整性与准确性仍面临挑战. ENSDF 等数据库中仍可能存在信息缺失或不一致的问题, 例如部分核素 (如  $^{238}\text{U}$ ,  $^{232}\text{Th}$ ,  $^{233}\text{U}$ ,  $^{239}\text{Pu}$ ) 的能级方案中存在未观测能态, 或  $\gamma$  射线强度缺乏不确定度标注<sup>[18]</sup>. 这些问题往往并非源于评价团队的疏忽, 而是由于实验数据不足以支撑精确评价所致. 这表明, 即便数据格式与管理体制完全现代化, 评价结果的可靠性与完整性仍根本受限于实验数据的可获得性与质量, 反映出核物理实验数据获取的更广泛科学挑战.

### 3 ENSDF 的现代化

核结构与衰变数据是连接实验观测与理论模型的基础桥梁, 其系统性、完备性与可机读性直接决定了核物理研究的深度与可持续性. 在前两节中, 我们从历史沿革和现有局限角度论证了 ENSDF 等核结构数据库亟须现代化改造. 随着实验装置精度的提升和数据体量的指数级增长, 传统 ENSDF 数据体系在格式封闭、版本难溯源、误差难传递等方面的局限日益突出, 已无法充分支撑现代核物理的精细分析与数据驱动研究. 推动核数据的现代化, 不仅是技术层面的更新, 更是科学范式的转型. 为应对现代科研要求数据可发现性 (findable)、可访问性 (accessible)、可互操作性 (interoperable) 与可重用性 (reusable) 的 FAIR 原则需求, 对 ENSDF 进行系统的现代化改造已势在必行. 首先, JSON 格式的使用使每条数据都带有清晰的键值标签, 结合集中式的索引服务, 实现了数据的精确定位和可查找性, 研究者能够通过核素标识符、能级能量等多种条件检索到所需信息. 其次, 开放 API 确保了任何经过授权的用户或程序都可以实时访问数据库, 实现数据的可获取性. 无论是通过可视化界面还是脚本调用, 数据均可便捷获取, 不再被封闭在专业软件或孤立文件中. 第三, 遵循统一的 JSON Schema 和标准单位, 使 ENSDF 与其他数据库和工具能够互操作: 不同系统之间交换 ENSDF 数据时无需额外格式转换, 科研人员可以将 ENSDF 数据直接融入自己的分析流程或与天体物理、粒子物理等数据资源结合. 最后, 详细的原始数据和规范的格式保证了数据的上

下文和出处明晰,提升了数据的可复用性.未来的研究者即使在不同环境下,仍能够理解并重复利用当前发布的 ENSDF 数据,而不受限于过时的软件或人工解读.这种面向 FAIR 原则的现代数据结构使核数据真正变得“看得见、拿得到、用得了”,为开放科学和数据长期保存奠定了基础.通过引入现代化的核数据格式、面向对象数据库及自动化验证机制,可以实现实验数据与理论约束的一致性校验和不确定度的端到端传播,从而为研究核结构系统学规律提供高可信度的数据基础.更重要的是,现代化核数据结构为机器学习和统计推断方法能够在全库尺度上开展关联分析和外推预测开辟了新的途径.这一变革对核物理研究的促进作用体现在三个核心领域:在核结构研究中,JSON 格式使研究者可通过脚本或现成库直接调用能级、寿命、跃迁强度等数据,避免研究人员投入大量精力手动提取整理数据,显著提升理论模型与实验基准的比对效率与精度;在核天体物理领域,ENSDF 作为国际标准数据库,为恒星核合成、r-过程路径等关键问题提供远离稳定线原子核的寿命、反应率等基石数据.天体物理学者可以直接批量调用核素数据用于核网络计算中,大幅提高模拟预测的时效性与可信度;在核反应模拟方面,无论是 Hauser-Feshbach 等统计模型计算还是 Geant4 等蒙特卡罗仿真,均可从标准化的 JSON 数据库中自动提取复合核能级密度、衰变模式、射线强度等参数,确保理论计算与实验模拟始终基于最新权威数据,从而在核工程设计、医学剂量模拟等应用中实现更高精度.因此,核数据现代化不仅关乎数据管理效率,更承载着揭示原子核复杂多体结构与衰变规律的科学使命.本节从技术实现层面,系统讨论 ENSDF 现代化的具体路径与关键手段.需要强调的是,这些技术改进并非单纯的“格式升级”或“界面美化”,而是面向未来几十年核数据需求的基础设施重构:一方面,通过引入结构化、可机读的新格式和数据库体系,可以显著提升评价流程的自动化程度和吞吐能力,减少人工录入与校对环节中的系统性错误,支撑“每个核素定期更新”的长期目标;另一方面,现代化的数据结构天然符合 FAIR 原则,便于与反应数据库、天体核网络以及大型加速器装置的数据流对接,为跨库联合分析和多学科应用提供统一的数据底座.更重要的是,只有在底层数据结构充分标准化和接口化的前提下,机器学习、数据不确定度

分析等新方法才能在全库尺度上开展不确定度量化与外推预测,从而真正实现由“人工作表+经验判断”向“数据驱动+智能辅助评价”的范式转变.因此,ENSDF 的现代化既是提升当前评价效率的现实需求,也是保障未来核科学装置和前沿研究可持续发展的战略性投入.

### 3.1 新数据格式: JSON Schema

JSON(Java Script Object Notation)的引入是 ENSDF 现代化中的关键一步,其作为高度结构化、机器可读的数据交换格式,通过键值对结构实现了数据表示与数据本身的解耦,不仅能清晰表达对称、不对称等各类不确定性类型,还较传统格式更加直观灵活,可高效组织核数据并支持便捷扩展<sup>[19]</sup>.通过将 ENSDF 从传统纯文本迁移到基于 JSON 格式的数据库,数据项层次得以清晰化、含义自治化,彻底摆脱了列位限制和非标准注释的束缚.这一转变带来了多重优势:评价员无需再关注复杂格式更改,因专用编辑器可自动处理细节从而简化工作流程;JSON 在计算领域的广泛采用意味着解析、验证等工具可便捷借用或开发;其结构化与表达性支持内联文档和注释,显著提升了数据的可读性与可维护性.更重要的是,JSON Schema 作为强制数据一致性的强大验证机制,通过定义 JSON 文档的规则来规范允许的数据及其形式,其模块化设计支持组件重用与自动传播更改,确保全库一致性并最小化设计工作.在实践中,批量验证已借助该机制发现当前 ENSDF 中数百个错误,同时 Schema 还可从定义生成代码,进一步维护软件开发中的数据完整性.

### 3.2 新型面向对象数据库: CouchDB

新的面向对象数据库 CouchDB 的引入显著拓展了 ENSDF 的数据存储与管理能力. CouchDB 能够原生支持二进制对象的存储,从而可整合更加多样化的数据类型,包括类型化注释、图片、网页链接、音频文件以及扫描文档等.这一特性对于保存具有研究价值的历史评价员笔记、实验记录及其他非结构化文档至关重要,使得内部资料能够以原生格式进行存储与调用,无需繁琐的人工转换或编辑<sup>[15,20]</sup>.此外, CouchDB 通过“视图”机制提供了高效的查询与过滤功能.视图通过预计算索引实现快速访问,可显著提高特定数据的检索效率.例如,

用户能够在极短时间内获取包含所有核素的完整列表、具有已测半衰期的能级集合,或位于特定半衰期范围内的能级信息.通过灵活的数据存储结构与高性能的索引检索机制,CouchDB有望显著提升 ENSDF 的整体数据管理效率,加速核数据评价与传播的工作流程.

### 3.3 原子核结构数据现代化与机器学习的双向驱动

#### 3.3.1 ENSDF 现代化在机器学习应用中的驱动作用

现代化的核结构数据在推动机器学习方法应用于核物理研究中发挥了关键作用.随着评价核结构数据库 (ENSDF) 的持续更新与完善,研究人员能够获取更为系统、精确且覆盖广泛的实验与理论数据.这些数据不仅包含能级结构、衰变模式、寿命与分支比等基础信息,还在超重核、丰中子弱束缚核以及自发裂变等实验难以覆盖的区域提供了重要参考.机器学习在核质量与衰变预测、反应截面计算、核物质状态方程约束等理论问题,以及中微子实验事件重建、粒子识别等实验挑战中已被尝试使用,最近进展展示了机器学习在提高精度、处理高维数据、融合物理知识方面的优势<sup>[21]</sup>.机器学习算法的性能高度依赖于训练数据的质量和丰富度,高质量核数据一方面为模型训练提供了庞大而多维的输入特征空间,使算法能够捕捉核结构与衰变性质之间复杂的非线性关联;另一方面,通过交叉验证与模型泛化,研究者得以利用数据驱动的方式揭示潜在规律,从而弥补传统理论模型在预测未知核区时的不确定性.因此,现代化核结构数据不仅是机器学习模型有效运行的前提,更是推动核物理研究范式转型的重要基础<sup>[1,22]</sup>.

尽管核结构与衰变数据积累已久,但长期以来其存储形式(如 ENSDF 的 80 列 ASCII 格式)难以被机器学习算法直接利用.该格式结构僵化、变量分散、缺乏统一标准,导致机器学习模型在解析数据和特征提取阶段面临较大困难.研究者往往需要投入大量时间进行数据清洗、格式转换与变量标准化,才能构建适合算法训练的输入向量.这一问题成为制约机器学习在核物理中广泛应用的主要瓶颈之一.随着对超重核衰变性质等前沿问题的研究需求不断增强,如何将现代化核数据评价成果转

化为结构化、易访问、易处理的数据库格式,成为实现数据驱动核物理研究的关键环节.通过将传统 ENSDF 文件转换为标准化 JSON 格式,研究者可实现自动化读取与批量处理,显著降低数据预处理成本.结构化数据使得能级信息、衰变模式及其不确定度等可直接作为机器学习输入特征使用,从而加速模型构建与验证流程.借助这一改进,核物理学家能够快速建立大规模训练集,灵活设计特征工程,并在超重核衰变预测、核质量系统化研究等领域实现高效建模与数据验证.

ENSDF 作为评价数据库,不仅提供实验数据,还附有不确定度和数据质量的说明.在机器学习模型训练过程中,ENSDF 提供的不确定度信息可用于调整训练策略,从而提升模型对噪声数据的鲁棒性与预测质量<sup>[23,24]</sup>.常见方法包括:通过损失函数加权将不确定度越小的数据赋予更高权重,降低测量误差大的样本对模型训练的干扰;将不确定度作为输入特征,辅助模型识别样本可信度;以及依据不确定度进行数据增强,对目标值引入符合误差范围的扰动,提升模型的泛化能力.这些策略不仅使模型更关注高质量数据,还能帮助其学习数据中固有的误差结构,从而提高收敛效率与预测可信度.

#### 3.3.2 机器学习对 ENSDF 数据库维护和发展的驱动作用

前文指出,大量实验结果以非机器可读的 PDF 文档形式发布,核数据评价工作量巨大,新实验结果源源不断,评价人员需要耗费大量时间手动校正提取内容,严重影响了数据评价的效率与吞吐量.为克服这一瓶颈,NNDC 正在研发基于机器学习(machine learning, ML)与自然语言处理(natural language processing, NLP)的自动化表格解析方法,以实现表格的自动检测、内容提取与结构识别<sup>[15,20]</sup>.该方法将表格提取视为一个视觉识别问题,应用计算机视觉中的机器学习框架(例如基于 DETR 架构并使用约 100 万张表格预训练的 TATR 模型),实现对 PDF 文档中表格的自动检测、内容定位与顺序解析.该 ML 工具的性能表现优异:表格检测率达 93%—97%,内容识别率约 92%,结构预测准确率 88%,文本提取率超过 99%.相较于以往依赖手动对齐、易产生噪声的基于规则的提取工具,新系统在准确性与自动化程度上均有显著提升.NNDC 已基于此开发了面向 ENSDF 评价

人员的可视化网络工作界面. 用户可直接上传 PDF 文档, 系统将自动检测其中的表格区域, 并生成可交互的 HTML 或 CSV 输出, 便于后续核数据录入与验证. 该工具通过显著减少手动修正错误所需的时间, 显著提升了 ENSDF 数据评价与整理的整体工作效率.

此外, 机器学习在核数据评价领域的更广泛应用正成为国际研究的前沿方向. 传统核数据评价主要依赖物理模型 (如 TALYS 与 EMPIRE) 以及原子核结构数据库 (如 ENSDF). 然而, 在某些反应通道或能区缺乏实验数据、数据存在不一致性或可靠性不足的情况下, 评价过程往往依赖专家经验判断, 从而引入主观偏差并导致较大的不确定度<sup>[25-27]</sup>. 为此, 近年来研究者提出将机器学习与人工智能方法引入核数据评价流程, 以在不牺牲物理约束的前提下, 辅助自动化建模与不确定度量化, 从而减少人工偏差并显著加快数据生成过程. 例如, 通过对 ENSDF 数据进行交叉验证训练, 模型可能识别出某些数据点偏离整体规律, 这提示这些数据点要么实验误差偏大, 要么对应着新的物理现象. 对于前者, 评价人员可以重点复核文献, 检查是否存在数据误报; 对于后者, 可能需要进一步理论或实验研究来解释异常. 机器学习提供了一种数据驱动的质量控制手段, 由于其擅长发现高维数据中的异常模式, 可辅助人类评价人员发现传统方法难以及时察觉的问题数据. 例如, 一个训练预测核能级的模型若对某核素某能级产生了显著偏差, 就可能意味着该能级的实验数据或归属存在问题, 可提示评估者审查对应文献. 类似地, 分类算法可以根据已知核素的衰变模式训练, 从而对某些核素给出的衰变支路数据提出质疑. 具体来说, 文献 [16] 提出了端到端机器学习辅助核数据评价流程. 利用决策树 (decision tree), K-近邻 (KNN), Gradient Boosting 等方法, 从 EXFOR 数据学习反应截面, 然后对部分实验数据稀缺的核素进行预测, 结果比某些已评估库更接近新测数据. 文献 [25] 中, 研究者用机器学习 + 遗传算法来确定哪些实验组合能最有效地改进约 12200 个核数据值, 从而缩短核数据评价所需时间. 机器学习还被用于辅助评价中数据验证分析, 为评价人员提供除物理方法之外的判断工具<sup>[28]</sup>. 在工程应用中, 对反应堆设计中的核数据需求也指出机器学习与人工智能有助于提高效率、可重复性与不确定度量化<sup>[29]</sup>.

机器学习还可以协助 ENSDF 与其他数据库的数据整合. 比如, ENSDF 主要涉及核结构和衰变数据, 而核反应数据库 EXFOR 则包含大量截面测量值. 在很多情况下, 不同数据库的数据需要结合使用 (如从衰变数据推断反应产额或反之). 通过机器学习, 可以建立跨数据库的关联模型, 提高不同来源数据的一致性. 用机器学习研究 ENSDF 与 EXFOR 数据的相关性, 例如利用已知核素的衰变分支比和反应产物丰度的对应关系, 训练模型来预测未知核素在某些反应中的生成率. 这实际上相当于让机器学习在两个数据库间找到模式, 从而在缺乏直接实验数据时作出推断. 这种整合分析可以检测并提醒某些核素在不同数据库中的信息是否矛盾, 从而推动数据库之间的数据一致性审查.

### 3.4 NuDat: 核数据的高级可视化与传播

NuDat 作为 ENSDF 的核心网络接口, 为用户提供了一个集搜索、过滤与可视化于一体的综合性核结构与衰变数据平台, 其数据主要来源于 ENSDF<sup>[9]</sup>. NuDat 提供对 ENSDF 中经系统评价的核能级、衰变模式及相关实验数据的权威访问, 并通过定期更新以反映最新的 ENSDF 修订结果与新评价数据. 作为 NNDC 最受欢迎的服务之一, NuDat 占 NNDC 网站电子数据检索总量的近 70%, 每年处理超过 700 万次数据查询请求, 已成为全球核数据用户最常用的在线平台之一.

然而, NuDat 的技术架构最初设计于 15 年前, 已难以满足现代网络技术的发展需求. NNDC 认识到对其进行系统性现代化改造的必要性, 以提升用户体验并促进对核数据属性的深入探索. 关键改进包括引入 HTML Canvas 实现动态交互式可视化, 支持平滑的平移与缩放操作; 采用全站统一的导航菜单与视觉设计以增强可用性与数据库间的一致性; 并通过 Git 版本控制与 Gradle 构建系统实现网络开发流程的标准化, 从而保障系统质量与可持续扩展性<sup>[30]</sup>.

现代化后的 NuDat 不再是传统意义上的数据存储库, 而是一个功能完备的核数据探索与分析平台. 具有可定制的核素着色方案、一维与交叉变量绘图、多尺度与过滤控制功能, 以及多格式 (CSV, PNG, 可共享链接) 数据导出能力, 使用户能够以高度灵活的方式检索、交互与分析复杂的核属性. 这一转变极大提升了核数据的可访问性与可解释

性,使其更好地服务于核科学研究、教学培训及多领域应用,显著扩大了数据库的科学教育影响力.值得关注的是,NNDC正在研发基于Three.js框架的三维可视化版本.与传统的二维图表相比,三维界面能够以更直观的方式展现多维核数据(例如原子序数 $Z$ 、中子数 $N$ 与特定物理属性之间的相关性),有望揭示二维投影中难以辨识的规律与聚类结构.这一探索预示着核数据可视化领域的未来方向,即通过先进的交互与三维可视分析技术,深化对复杂核体系的理解,推动核数据认知与分析模式的创新.

除了NNDC的NuDat系统外,国际上亦在积极探索自助式核数据可视化工具.例如,法国国家科学研究中心(CNRS)开发的TkN平台<sup>[31]</sup>,同样致力于为用户提供可定制的核素图与能级可视化界面,进一步体现了全球核数据可视化技术向智能化与交互化方向演进的共同趋势.

#### 4 中国的自主探索:中山大学的实践

在全球核数据现代化的浪潮中,中国核物理研究团队正发挥着日益重要的作用.在中国核数据中心(China Nuclear Data Center, CNDC)的统筹与支持下,中山大学核数据团队积极探索将核数据评价与基础科研及工程应用相结合的新路径,致力于关键核数据的精确测量与系统化评价.依托我国在数据科学与计算物理领域的技术积累,并考虑到部分科研与教学环境下网络访问受限的实际情况,团队围绕ENSDF标准文件构建了从数据解析到可视化展示的完整离线流程.借鉴NNDC等国际

先进经验,我们在本地环境中搭建了符合ENSDF数据质量标准的离线可视化平台,形成了自主可控的原子核数据本地化查询与展示系统.该系统不仅支撑了国内在核能、核医学与国防等领域的关键数据应用需求,也为我国核数据基础设施的自主可持续发展奠定了坚实基础.该工作充分体现了CNDC推动“基础核数据研究与实际应用需求相结合”的战略方向在中山大学的落地与实践.

在此基础上,我们开发了ENSDFTransformer项目,用于读取并解析原始80列ASCII格式的ENSDF文件,将核素、能级、辐射及衰变信息结构化映射为JSON对象,并导入自主搭建的核数据管理数据库中.该项目是本地数据库的核心数据源,同时也是后续实现数据筛选、可视化分析以及引入机器学习等智能化方法的关键前置环节.借鉴国际ENSDF现代化经验,采用模块化JSON Schema设计思想,针对不同数据实体(包括核素、能级、衰变模式与辐射类型等)定义独立模式,以确保数据结构的层次化、可扩展性与一致性.

图2给出了核素对象的JSON Schema定义示例,用于展示该系统在结构化建模中的标准化设计理念.

#### 4.1 核结构数据转换流程

ENSDF的现代化转换流程可概括为“三步走”:解析、结构化与模式验证.首先,在解析阶段,程序按ENSDF既有行结构逐行读取文本文件,以记录为基本单元,根据第6列延续标识和第7列注释标识完成数据集与记录划分,并依据记录类型应用相应字段切分规则,将各字段解析为文本或数值

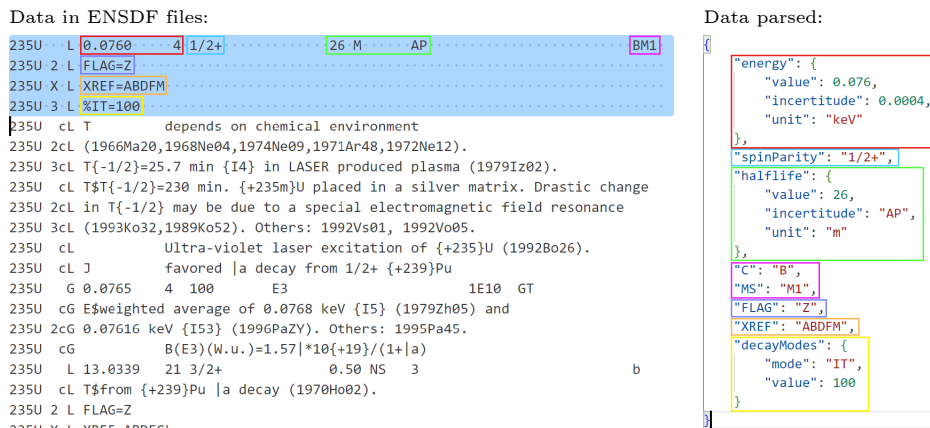


图2 左侧部分为ENSDF原始数据形式,右侧部分为对应的JSON格式数据

Fig. 2. The left part is in the original ENSDF data format, while the right part is the corresponding JSON format data.

并进行一致性检查, 最终生成初步 JSON 记录. 随后在数据结构化阶段, 这些线性记录被组织为层次化 JSON 对象: 以核素为顶层, 每个核素下包含数据集、历史信息、 $Q$  值与能级等子对象. 例如能级对象中, 能级能量、自旋宇称、半衰期及其不确定度均以键值对和嵌套字典形式存储, 从而兼顾物理含义与机器可读性. 最后在模式验证阶段, 通过预先定义的 JSON Schema 对生成文件进行结构与类型校验, 自动识别解析错误并实施如“半衰期不得为零”等强约束规则, 从而显著提升新格式在一致性与可靠性方面的优势, 为后续可视化展示与数据科学分析提供了稳固的数据基础.

### 4.2 核结构数据可视化平台

基于上述结构化数据, 本文构建了一个离线核数据可视化平台, 如图 3, 主要采用 Dash 与 Plotly 两个 Python 库实现. 平台的数据源为本地存储的 JSON 文件, 包括由 ENSDFTranformer 从 ENSDF 原始文件解析得到的数据.

Dash 用于搭建网页前端框架, 提供交互控件与页面布局功能, 用户无需掌握 JavaScript 即可快速生成交互式界面. Plotly 用于绘制图表, 与 Dash 高度兼容, 可实现鼠标滚轮缩放、拖拽平移、数据点悬停与点击事件监听等交互操作, 从而支持

动态的核数据探索与分析.

该可视化系统通过直观的图形交互与灵活的数据过滤, 实现了核结构与衰变数据的快速检索、比较与分析, 显著提升了数据的可用性与教学科研价值.

### 5 新的核结构数据文件的机器学习应用

在现代化核结构数据的基础上, 本文旨在充分利用现代化核结构数据的高可用性与机器学习方法的建模能力, 构建超重核衰变性质的系统预测框架. 通过将标准化、接口化的核数据直接输入机器学习平台, 可高效开展特征选择、模型训练与性能验证, 为后续理论研究与实验设计提供坚实的数据支撑.

研究超重核的衰变性质不仅是探索元素周期表边界的核心途径, 也是检验原子核结构理论模型的重要手段, 其核心科学目标是寻找理论预言的“稳定岛”区域 [32-36]. 半衰期是衡量超重核稳定性的关键指标, 决定了该核素能否在实验条件下合成并被探测到. 当半衰期过短时, 核素在生成后即迅速衰变, 难以被实验识别, 从而限制了新元素与新同位素的发现. 与此同时, 主衰变模式 (如  $\alpha$  衰变

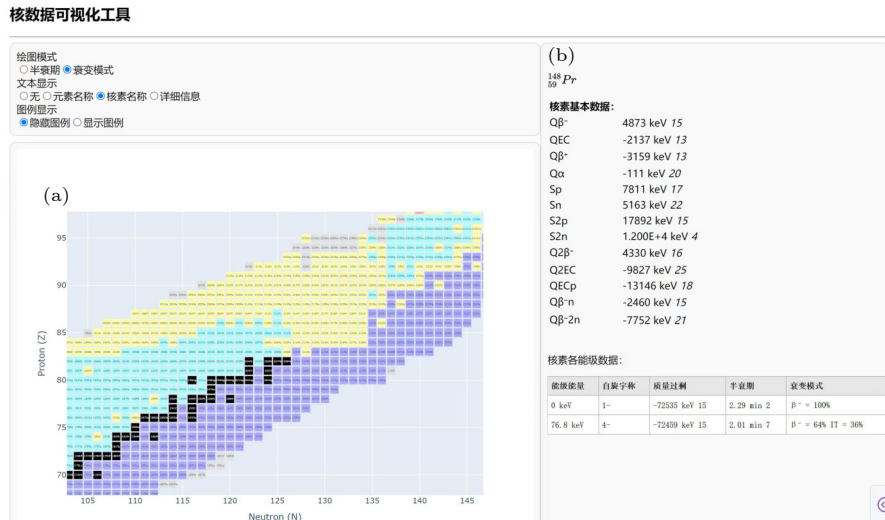


图 3 可交互核素图显示半衰期 (a) 该区域显示核素分布图, 并提供绘图选项控制. 用户可选择以半衰期或衰变模式为依据进行着色, 并可调节核素方块内文本显示与图例显隐; (b) 当用户在核素图中选中某个核素后, 其对应的全部 ENSDF 结构与衰变数据会自动加载并呈现, 支持筛选功能

Fig. 3. The interactive nuclide chart displays the half-life. (a) This area shows the nuclide distribution map and provides options for drawing control. Users can choose to color by half-life or decay mode, and can adjust the display of text within the nuclide blocks and the visibility of legends. (b) When a nuclide is selected in the nuclide plot, all the ENSDF structures and decay data corresponding to that nuclide will be automatically loaded and presented, and a filtering function is supported.

与自发裂变) 不仅反映了核子间的相互作用与壳效应, 还直接决定了超重核的寿命与可观测性. 系统研究超重核的半衰期与衰变模式, 有助于揭示超重质量区域的潜在稳定区间, 验证并约束核结构理论模型, 为实验合成与鉴定新核素提供重要参考.

此外, 由于超重核具有极大的核电荷数, 其核外电子运动速度接近光速, 相对论效应显著增强, 导致电子排布规律可能偏离已知周期性. 这一效应不仅影响元素的化学性质与周期表定位, 也使得超重核研究具有明显的跨学科意义, 连接核物理、原子物理与化学多个领域<sup>[32,33]</sup>. 理论预言的“稳定岛”上的同位素还可能具有潜在的应用价值. 基于这些原因, 《科学》(*Science*) 杂志早在 2005 年将“是否存在稳定的超重元素”列为 125 个最重要的未解科学问题之一<sup>[37]</sup>.

综上所述, 系统探索超重核的半衰期与衰变模式不仅是验证核结构理论与发现新元素的关键途径, 也为理解极端条件下的量子规律提供了独特窗口. 本文结合现代化核结构数据与机器学习方法, 聚焦超重核的衰变性质研究, 旨在为“超重稳定岛”的理论探索与实验发现提供数据驱动的支持.

在模型选择方面, 本文采用随机森林 (random forest, RF) 算法<sup>[38]</sup> 对超重核衰变性质进行系统建模与预测. 随机森林是一种基于多决策树集成的监督学习算法, 能够通过投票机制有效捕捉高维数据中的非线性关系并提高预测稳定性. 结合 JSON 格式 ENSDF 数据提供的多维输入特征 (如核素的质子数  $Z$ 、中子数  $N$ 、质量数  $A$ 、 $Z$  与  $N$  的奇偶性、衰变能等), 模型可同时实现对半衰期与主衰变模式的预测, 并通过特征重要性分析揭示关键核结构因素. 此种数据与算法的结合不仅为超重核衰变性质的探索提供了新的量化工具, 也展示了现代化数据格式与机器学习算法深度融合的应用潜力.

## 5.1 研究方法

本文聚焦于  $Z \geq 84$  和  $N \geq 128$  区域的原子核多种衰变模式的竞争. 先使用半经验公式计算此区域全部原子核基态的  $\alpha$  衰变、 $\beta^-$  衰变、 $\beta^+$  衰变、轨道电子俘获和自发裂变的分支半衰期, 然后使用随机森林算法对每个公式的残差和均方根误差进行训练. 通过训练前后残差的变化程度检验随机森林算法结合半经验公式计算分支半衰期的可行性. 原子核每一种衰变模式的强弱可以通过分支半衰

期的长短进行比较, 分支半衰期最短的衰变模式对应主要衰变模式. 所以, 通过比较所计算核素的不同衰变模式的分支半衰期, 可以预测核素的主衰变模式. 下面依次介绍本文采用的半经验公式、随机森林算法、数据构成和结果.

### 5.1.1 衰变半衰期经验公式

带电粒子统一衰变公式 (universal decay law, UDL)<sup>[39,40]</sup>

$$\log_{10} T_{1/2,\alpha} = aZ_{\alpha}(Z - Z_{\alpha})\sqrt{\mu'/Q_{\alpha}} + b\sqrt{\mu'Z_{\alpha}(Z - Z_{\alpha})(A_{\alpha}^{1/3} + (A - A_{\alpha})^{1/3})} + c \quad (1)$$

被用于拟合  $\alpha$  衰变半衰期  $T_{1/2,\alpha}$ , 其中  $Z_{\alpha}$ ,  $A_{\alpha}$ ,  $Q_{\alpha}$  和  $\mu' = A_{\alpha}(A - A_{\alpha})/A$  分别代表  $\alpha$  粒子的质子数、质量数、 $\alpha$  衰变能和约化质量数,  $Z$  和  $A$  分别代表母核的质子数和质量数. 衰变道假设是从母核基态衰变到子核基态.

对于自发裂变, 基于已存公式<sup>[41-46]</sup>, 提出了一个三参数公式 (记作 SF3):

$$\log_{10} T_{1/2,\text{SF}} = a \frac{(Z - \nu)^2}{(1 - \kappa I^2)A} + \frac{b}{A} + c, \quad (2)$$

用于自发裂变半衰期  $T_{1/2,\text{SF}}$  的计算<sup>[47]</sup>. 其中  $\nu$  表征不配对核子的阻塞效应 (对于偶偶核其值为 0, 对于其他核其值为 2)<sup>[43]</sup>;  $I = (N - Z)/A$  表征同位旋效应,  $N$  为母核的中子数;  $\kappa$  的值为 2.6<sup>[41,48]</sup>; 此外,  $a$ ,  $b$  和  $c$  是三个拟合系数. 由于系统性差异, (2) 式分别对  $Z < 104$  的核和其他核进行拟合, 如表 1 所列. 对于  $Z < 104$  的核, 随着  $Z$  减少, 库仑斥力减小,  $T_{1/2,\text{SF}}$  显著增加. 该区域内一些核素的自发裂变半衰期相对较长 ( $> 10^8$  s), 目前尚无法统一描述. 鉴于与其他衰变模式相比, 这些核素自发裂变的竞争显著较弱, 故未用于 (2) 式的拟合.

$\beta$  衰变半衰期  $T_{1/2,\beta}$  通过参考文献<sup>[49]</sup> 给出的公式进行估计. 假设基态的  $\beta$  衰变是有效加莫夫-泰勒 (Gamow-Teller, GT) 跃迁, 其分支半衰期可表示为

$$\log_{10} T_{1/2,\beta} = \log_{10} \kappa_1 - \log_{10} f_0 - \log_{10} B_{\text{GT}}, \quad (3)$$

其中  $\kappa_1 = \frac{2\pi^3 \hbar^7 \ln 2}{m_e^5 c^4 G_F^2} = 6147$  s,  $f_0$  是相空间因子,  $B_{\text{GT}}$  是 GT 约化跃迁概率. 对于轨道电子俘获, 相空间因子积分近似为

$$f_0^{\text{EC}} \approx 2\pi \left(\frac{Z}{137}\right)^3 \left[1 - \frac{1}{2} \left(\frac{Z}{137}\right)^2 + E_0\right]^2, \quad (4)$$

表 1 UDL, SF3 和 (3) 式的拟合参数和均方误差  
Table 1. The fitted parameters and mean square error of UDL, SF3 and Eq. (3).

模型	$a(\log_{10} B_{GT})$	$b$	$c$	RMSE	RMSE <sub>RF</sub>	降低程度
UDL	0.404	-0.360	-25.128	0.831	0.508	38.9%
SF3 ( $Z < 104$ )	-1.213	-1390.926	61.361	3.124	1.141	63.5%
SF3 ( $Z \geq 104$ )	-1.465	-12786.191	116.228	1.443	0.525	63.6%
公式(3) <sub><math>\beta^+</math></sub>	0.816	—	—	0.816	0.325	60.2%
公式(3) <sub><math>\beta^-</math></sub>	-1.967	—	—	1.675	0.705	57.9%
公式(3) <sub>EC</sub>	-1.653	—	—	2.196	0.846	61.5%

而对于  $\beta^\pm$  衰变, 相空间因子则积分近似为

$$f_0^{\beta^\pm} \approx \frac{\mp(E_0^5 - 10E_0^2 + 15E_0 - 6)2\pi(Z \mp 1)/137}{30 \{1 - \exp[\pm 2\pi(Z \mp 1)/137]\}}, \quad (5)$$

其中  $E_0$  是以电子质能为单位的  $\beta$  衰变能. 由于 ENSDF<sup>[5,6]</sup> 提供的  $\beta$  衰变能  $Q_\beta$  是原子质量差, 因此计算  $E_0$  时, 电子质量需要重新考虑并做如下处理:

$$E_{0,\beta^+} = \frac{Q_{\beta^+} + 2m_e c^2}{m_e c^2}, \quad E_{0,\beta^-} = \frac{Q_{\beta^-} + m_e c^2}{m_e c^2},$$

$$E_{0,EC} = \frac{Q_{EC} - m_e c^2}{m_e c^2}. \quad (6)$$

然后, 以  $\log_{10}(f_0 T_{1/2,\beta}/\kappa_1)$  的平均值作为  $\log_{10} B_{GT}$  的估计. 拟合结果列于表 1.

### 5.1.2 随机森林方法

随机森林算法是决策树算法和重取样算法的集成, 已被用于预测奇  $A$  核的基态自旋<sup>[50]</sup>. 决策树算法是一种非参数监督学习算法. 对于包含  $S$  个样本和  $I$  个特征 (变量) 的数据集  $\{(\theta_1, \dots, \theta_I)_s, s \in [1, S]\}$  以及目标值 (可观测量)  $\{y_s, s \in [1, S]\}$ , 它建立了一个二叉树结构, 以分段函数的形式根据特征值将数据集划分为  $L$  个子集, 每个子集称为一个叶节点. 通过为每个叶节点分配一个值构造函数  $f(\theta_1, \dots, \theta_I)$ , 最小化整个数据集的均方根误差 (root-mean-square error, RMSE) 为

$$RMSE = \sqrt{\frac{1}{S} \sum_{s=1}^S (y_s - f(\theta_1, \dots, \theta_I))^2}. \quad (7)$$

重取样是一种基于随机有放回抽样概念的统计方法. 通过这种方法, 数据集中数据的各种可能组合方式和权重自动被考虑<sup>[47,51,52]</sup>. 每次获得一个新数据集 (即 Bootstrap 样本) 时, 新的决策树会被训练出来, 并被用来计算整个数据集中每个样本的目标值. 通过重复这一过程  $M$  次, 便建造出了一个

由  $M$  棵树组成的森林. 数据集中每个样本的目标值的最终预测结果是森林中所有树计算结果的平均值, 即  $F(\theta_1, \dots, \theta_I) = \frac{1}{M} \sum_{m=1}^M f_m(\theta_1, \dots, \theta_I)$ . 模型的性能指标则采用随机森林的袋外评分进行估计, 即在构建每棵决策树时, 约有 1/3 的样本未被用于训练, 这部分“袋外”样本即可用于评估该树的预测误差; 最终对所有树的袋外误差进行平均, 即可得到整体模型泛化能力的一个有效估计, 从而降低了在计算未知核素时的偏差, 提高泛化能力. 本研究使用开源 Python 库 scikit-learn<sup>[53]</sup> 中的随机森林回归任务进行机器学习训练. 为了降低均方根误差的离散性, 此处假设森林由  $10^5$  棵树组成.

本文中  $\alpha$  衰变、 $\beta^-$  衰变、 $\beta^+$  衰变和轨道电子俘获的衰变半衰期经验公式的残差通过随机森林算法进行回归任务训练, 通过随机森林算法对经验公式的残差进行预测, 使所有训练核素经过随机森林修正的半衰期与实验半衰期的方均根最小. 所用特征包括  $Z, N, A$  和  $Z$  与  $N$  的奇偶性, 以及各自对应的衰变能. 由于自发裂变无法定义衰变能量, 这里使用文献<sup>[54]</sup> 中的裂变位垒 (fission barrier, FB) 来替代特征集中的衰变能, 并以此等价考虑裂变过程中的形变效应. 本文选择叶节点数为 11, 与我们之前研究中用于训练该区域结合能的叶节点数相同<sup>[49]</sup>. 如此, 对于给定衰变模式的给定经验公式所计算的分支半衰期  $\log_{10} T_{1/2,\text{beforeRF}}$ , 随机森林算法修正后的分支半衰期为  $\log_{10} T_{1/2,\text{afterRF}} = \log_{10} T_{1/2,\text{beforeRF}} + F(\theta_1, \dots, \theta_I)$ . 随后, 可通过残差

$$\log_{10} T_{1/2,\text{exp}} - \log_{10} T_{1/2,\text{beforeRF}}, \quad (8)$$

$$\log_{10} T_{1/2,\text{exp}} - \log_{10} T_{1/2,\text{afterRF}}, \quad (9)$$

以及均方根误差

$$RMSE_{\text{beforeRF}} = \sqrt{\frac{1}{S} \sum_{s=1}^S (\log_{10} T_{1/2,\text{exp},s} - \log_{10} T_{1/2,\text{beforeRF},s})^2}, \quad (10)$$

$$\text{RMSE}_{\text{afterRF}} = \sqrt{\frac{1}{S} \sum_{s=1}^S (\log_{10} T_{1/2, \text{exp}, s} - \log_{10} T_{1/2, \text{afterRF}, s})^2} \quad (11)$$

检验训练前后对分支半衰期的描述性能. 通过随机森林的拟合, 使所有训练核素经过随机森林修正后的半衰期预测值与其实验值之间的均方根误差最小化, 系统校正经验公式的系统偏差, 从而使修正后的理论预测值在整体上与实验数据更为吻合, 并依据不同衰变通道中分支半衰期最短者予以确定主衰变模式.

## 5.2 结果与讨论

图 4 展示了本文所用数据样本的分布. 我们从 ENSDF 数据库<sup>[3]</sup>中筛选出质子数  $Z \geq 84$  且中子数  $N \geq 128$  的核素, 共获得 472 个核素的 584 个衰变分支半衰期数据. 采用随机森林方法, 分别对五种衰变模式进行拟合, 各模式对应的样本数量为:  $\alpha$  衰变 211 个、自发裂变 (SF) 145 个、 $\beta^+$  衰变 88 个、 $\beta^-$  衰变 108 个, 以及电子俘获 (EC) 衰变 32 个. 此外, 还基于所得模型, 对该区域内所有已被实验确认存在、但半衰期或分支比尚不确定或在 ENSDF 文件中分支比未报告的核素进行了预测.

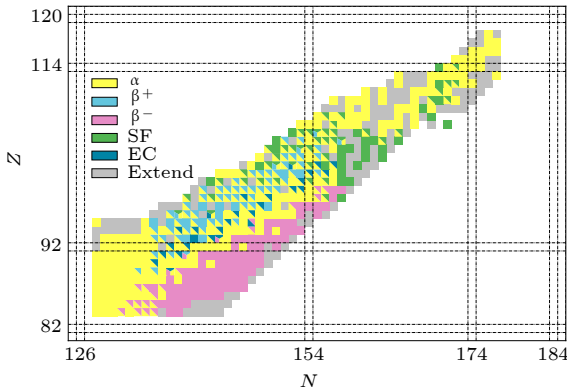


图 4 随机森林模型中训练集与外推集的样本点分布. 图中,  $\alpha$  衰变、 $\beta^-$  衰变、 $\beta^+$  衰变、自发裂变和电子俘获 (衰变的训练集数据点以图例对应颜色标示; 有色区域均为外推集, 即包含具有任意一种衰变能的核素. 已被实验确认存在半衰期或分支比尚不确定的核素.

Fig. 4. Distribution of sample points in the training set and extrapolation set of the random forest model. In the figure, training set data points for  $\alpha$  decay,  $\beta^-$  decay,  $\beta^+$  decay, spontaneous fission, and electron capture decays are represented by colors corresponding to the legend; all colored areas belong to the extrapolation set, i.e., they include nuclides with any type of decay energy; gray areas indicate nuclides that lack experimental half-life values or have excessively large uncertainties in the experimental data.

图 5 比较了这些衰变公式在随机森林算法训练前后的残差. 通过两个条件来确定离群点: 1) 计算值偏离实验值两倍均方根误差以上 (散点位于对应颜色的虚线之外); 2) 计算值与实验值的数量级相差 3 倍以上 ( $|\log_{10}(T_{1/2, \text{cal}}/T_{1/2, \text{exp}})|$  的值大于 3). 通过这种方式可以避免因均方根误差显著大而遗漏、显著小而增加离群点的情况. 与预期一致, 经过随机森林算法训练后, 这些衰变公式的离群点显著减少, 每种衰变半衰期的均方根误差也降低了 (见表 1). 需要注意的是, 因为此处的目标不是最大限度地减小均方根误差, 而是使其达到能够区分主要衰变模式的适当尺度, 所以为了避免过拟合和追求极小的均方根误差, 选择了相同的特征和叶节点数来训练随机森林学习不同衰变公式的残差.

已知实验数据主衰变模式与半衰期与 596 个核素主衰变模式与半衰期如图 6(a) 和图 6(b) 所示. 该图揭示了已知核素衰变模式的规律性分布: 从  $^{226}\text{Ra}$  到  $^{251}\text{Cf}$  存在一条典型的长寿命  $\alpha$  衰变谷, 其西南侧由  $\alpha$  衰变主导, 东南侧过渡为  $\beta^-$  衰变主

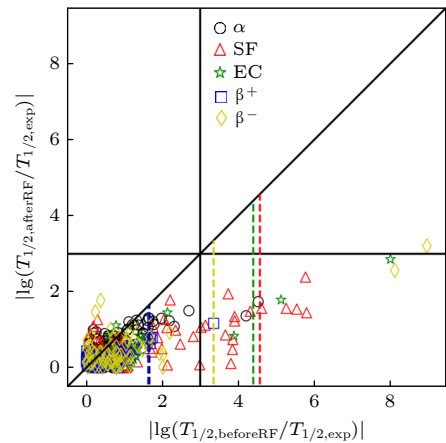


图 5 随机森林训练前后残差对比. 散点以不同颜色分别表示  $\alpha$  衰变、 $\beta^-$  衰变、 $\beta^+$  衰变, 自发裂变和电子俘获衰变的经验公式拟合残差与随机森林训练后的残差. 各颜色虚线标示了对应经验公式均方根误差 (RMSE) 的两倍位置. 随机森林显著提升了对各类衰变半衰期的描述能力

Fig. 5. Comparison of residuals before and after random forest training. Scatter points in different colors represent the residuals from the empirical formula fitting and the random forest training for  $\alpha$  decay,  $\beta^-$  decay,  $\beta^+$  decay, spontaneous fission, and electron capture decays. Dashed lines in corresponding colors indicate twice the root-mean-square error (RMSE) of each respective semi-empirical formula.

导, 而西北区则呈现  $\beta^+$  和 EC 衰变与  $\alpha$  衰变的竞争, 东北区则展现了  $\alpha$  衰变与自发裂变的激烈竞争. 尽管主衰变模式的分布有明确界限, 半衰期的分布却保持连续. 随机森林的预测结果则如图 6(c) 和图 6(d) 所示, 预测结果与实验结果高度一致, 其中主衰变模式与实验数据的一致率达到 92.2%. 对于主衰变模式预测不一致的情况, 往往具有两种竞争性衰变, 如  $^{255}\text{Rf}$ ,  $^{262}\text{Db}$  和  $^{286}\text{Fl}$  的  $\alpha$  衰变与自发裂变分支比都趋于 50%.

自发裂变对于超重核半衰期的研究非常重要. 如图 6(c) 和图 6(d) 所示, 未知核素的主衰变模式需要根据自发裂变、 $\alpha$  衰变和  $\beta^-$  衰变之间的竞争来确定. 对于丰中子核素, 主要的竞争在于自发裂变和  $\beta^-$  衰变之间; 而对于缺中子核素, 主要的竞争则在自发裂变和  $\alpha$  衰变之间. 现有的实验数据展示了一个从  $^{226}\text{Ra}_{138}$  到  $^{251}\text{Cf}_{153}$  的长寿命  $\alpha$  衰变区域, 其位于  $\beta^+$  和  $\beta^-$  衰变区域之间, 并以自发裂变终结. 模型正确地描述了这一现象. 当  $N$  超过 154

后, 存在一条半衰期为  $10^2 - 10^7$  s 的蓝色带向外蔓延, 并在  $Z = 114$  和  $N = 184$  的西南角存在一个淡蓝色岛形区域. 结合裂变位垒分析可以发现, 这是由于该岛形区域的裂变位垒较高, 从而导致了更长的自发裂变半衰期  $T_{1/2,\text{SF}}$ . 图 7 比较了自发裂变半衰期  $T_{1/2,\text{SF}}$  的实验值和裂变位垒随质量数的演变. 裂变位垒在  $A = 230$  之前随着质量数  $A$  的增加而下降, 随后表现为在 2—10 MeV 之间波动. 显然, 存在一个裂变位垒的阈值, 低于此阈值时自发裂变发生. 自发裂变半衰期较长的核素通常具有较小的自发裂变分支比. 此外, 自发裂变分支比小于 1% 的核素的裂变位垒普遍高于自发裂变分支比大于 1% 的核素, 这表明裂变位垒越高, 自发裂变越弱. 然而, 仅考虑自发裂变分支比小于 1% 或大于 1% 的核素时, 裂变位垒与自发裂变半衰期  $T_{1/2,\text{SF}}$  之间的对应关系变得更为复杂. 图 7 还演示了裂变位垒随  $Z$  的增大而增加, 这表明在超重核素中裂变位垒与库仑斥力之间存在竞争.

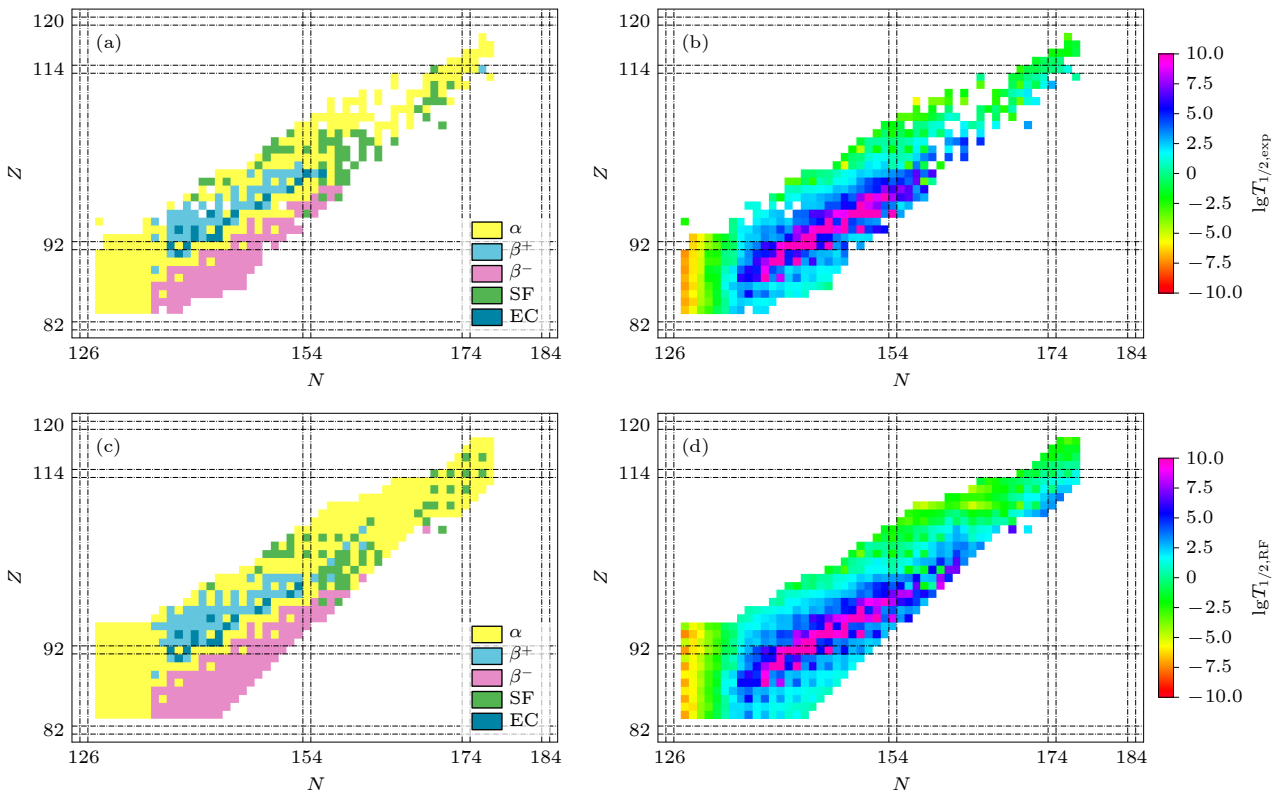


图 6 主衰变模式与总半衰期预测值与实验值对比 (a), (b) 取自美国国家核数据中心 (NNDC) 的实验数据; (c), (d) 为随机森林模型的预测结果, 该模型外推了那些仅有衰变能数据但缺乏半衰期实验测量值的核素

Fig. 6. Comparison between predicted and experimental dominant decay modes and total half-lives: (a), (b) Experimental data retrieved from the National Nuclear Data Center (NNDC); (c), (d) results predicted by the random forest model, which includes extrapolations for nuclides that have decay energy data but lack measured half-life values.

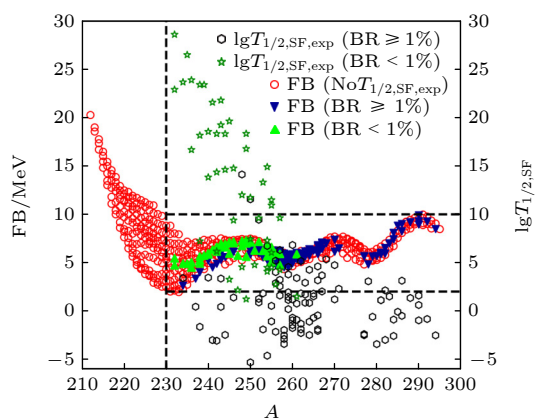


图 7  $T_{1/2,SF}$  (Y轴右侧) 和裂变势垒 (Y轴左侧) 随质量数变化的演变. 数据根据 SF 半衰期是否被实验测量到, 以及 SF 的分支比是否小于 1% 进行了分类

Fig. 7. Evolution of  $T_{1/2,SF}$  (on the right side of the Y-axis) and FB along the mass number (on the left side of the Y-axis). The datasets are divided according to whether the corresponding  $T_{1/2,SF}$  is measured and whether the branch ratio of SF is less than 1%.

## 6 结 论

本文系统梳理了评价核结构数据库 (ENSDF) 的发展历程与现存局限, 论证了其现代化改造的必要性与紧迫性. 通过引入 JSON Schema 结构化格式、CouchDB 面向对象数据库、机器学习自动抽取与验证技术以及 NuDat 高级交互式可视化界面, 核数据的采集、评价与传播流程实现了从静态汇编到智能管理的跨越. ENSDF 等核结构数据库的现代化建设不仅是一项面向大规模数据的技术革新, 更是推动核物理研究范式转变的重要契机. 通过实现数据的标准化、结构化与可追溯化, 现代化核数据体系为核结构系统学、核衰变机制以及极端核区可预测性研究奠定了统一而高精度的科学基石. 其科学意义体现在三个层面: 一是促进实验与理论数据的深度融合, 增强数据驱动的物理一致性与可验证性; 二是为机器学习等新型方法提供高质量特征空间, 使得核性质预测与规律发现具备可解释性与可量化不确定度; 三是推动开放、可重复与智能化的核数据生态构建, 使核数据真正成为连接核结构、反应与天体核物理的核心枢纽. 可以预见, 随着核数据现代化的持续推进, 未来核物理研究将从经验汇编走向数据智能驱动, 实现从“数据可用”到“知识可发展”的跃升.

在此框架下, 本文进一步结合 JSON 化后的核结构与衰变数据, 利用随机森林方法对超重核的多

种衰变模式半衰期进行了系统建模与预测. 结果表明, 机器学习在捕捉复杂的非线性关联、修正经验公式残差以及识别主导衰变模式方面具有显著优势. 与实验数据相比, 模型的主衰变模式预测一致率达到 92.2%, 预测与实验半衰期均方根误差 (RMSE) 平均降低 50% 以上. 这一结果证明了现代化核结构数据与机器学习算法结合的巨大潜力, 不仅可用于提高已知核素半衰期的预测精度, 也为未知超重核区域的外推提供了可靠工具.

以中山大学核数据团队为代表的中国研究力量, 在中国核数据中心的支持下, 已构建了自主可控的本地数据转换与可视化系统, 实现了原始数据的结构化解析、模式验证与交互展示, 这一实践表明, 中国在核数据现代化基础设施建设中已具备独立创新与持续贡献能力. 展望未来, 随着国际核数据体系的进一步互联与人工智能方法的深入融合, 现代化核结构数据库将在提升数据可用性、促进理论与实验协同以及支撑新核素发现、超重核稳定性研究、核天体物理建模以及核能与核安全等前沿应用中发挥愈发关键的作用.

## 数据可用性声明

本文的所有关联数据可以在科学数据银行 <https://doi.org/10.57760/sciencedb.30258> 中获取.

感谢中国核数据中心支持. 感谢中山大学、国家超级计算广州中心提供服务器以及计算资源支持.

## 参考文献

- [1] Ge Z, Chen Y 2020 *Nucl. Phys. Rev.* **37** 309 (in Chinese) [葛智刚, 陈永静 2020 *原子核物理评论* **37** 309]
- [2] Kolos K, Sobes V, Vogt R, Romano C E, Smith M S, Bernstein L A, Brown D A, Burkey M T, Danon Y, Elswawi M A, Goldblum B L, Heilbronn L H, Hogle S L, Hutchinson J, Loer B, McCutchan E A, Mumpower M R, O'Brien E M, Percher C, Peplowski P N, Ressler J J, Schunck N, Thompson N W, Voyles A S, Wieselquist W, Zerkle M 2022 *Phys. Rev. Res.* **4** 021001
- [3] ENSDF database (Evaluated Nuclear Structure Data File). <https://www.nndc.bnl.gov/ensdf/> [2025-12-29]
- [4] About ENSDF. <https://www.nndc.bnl.gov/ensdf/about.jsp> [2025-12-29]
- [5] Burrows T W 1990 *Nucl. Instrum. Methods Phys. Res. A* **286** 595
- [6] Tuli J K 1996 *Nucl. Instrum. Methods Phys. Res. A* **369** 506
- [7] National Nuclear Data Center. <https://www.nndc.bnl.gov> [2025]
- [8] Nuclear Data Working Group. <https://www.nndc.bnl.gov/>

- ndwg [2025-12-29]
- [9] NuDat 3. <https://www.nndc.bnl.gov/nudat3> [2025-12-29]
- [10] Dimitriou P, Basunia S, Bernstein L, Chen J, Elekes Z, Huang X, Hurst A, Iimura H, Jain A K, Kelley J, Kibédi T, Kondev F, Lalkovski S, McCutchan E, Mitropolsky I, Mukherjee G, Negret A, Nesaraja C, Nica N, Pascu S, Rodionov A, Singh B, Singh S, Smith M, Sonzogni A, Timár J, Tuli J, Verpelli M, Yang D, Zerkin V 2020 *EPJ Web Conf.* **239** 15004
- [11] International Network of Nuclear Structure and Decay Data (NSDD). <https://www-nds.iaea.org/nsdd/history.html> [2025]
- [12] Ge Z, Chen Y 2015 *Chin. Sci. Bull.* **3087** (in Chinese) [葛智刚, 陈永静 2015 科学通报 3087]
- [13] Ge Z G, Chen Y J 2019 *At. Energy Sci. Tech.* **53** 1742 (in Chinese) [葛智刚, 陈永静 2019 原子能科学技术 **53** 1742]
- [14] Hayes, E, McCutchan, S, Yoo, A, Mattera, S, McCorkle, B, Shu, A, Sonzogni, C, Soto, S, Zhu, F, Kondev, F 2021 *Bull. Am. Phys. Soc.* **66** DF.00003
- [15] Morse C 2022 Nuclear Data Week(s) 2022, New York, US, Oct.31–Nov.4 2022 E000003921
- [16] Vicente-Valdez P, Bernstein L, Fratoni M 2021 *Ann. Nucl. Energy* **163** 108596
- [17] Zerkin V, Pritychenko B, Totans J, Vrapcjenjak L, Rodionov A, Shulyak G 2022 *J. Instrum.* **17** P03012
- [18] Henning G, Kerveno M, Dessagne P, Claeys F, Dari Bako N, Dupuis M, Hilaire S, Romain P, de Saint Jean C, Capote R, Boromiza M, Olacel A, Negret A, Borcea C, Plompen A, Paradelo Dobarro C, Nyman M 2023 *EPJ Web Conf.* **284** 01022
- [19] Morse C. 2024 The 25th Technical Meeting of the NSDD network, Vienna, Austria, Apr.15–Apr.19, 2024 p48
- [20] Mason D, Ricard-McCutchan E 2022 *The 24th Meeting of the NSDD Network* Canberra, Australia, Oct. 24–28, 2022 p2
- [21] He W, Li Q, Ma Y, Niu Z, Pei J, Zhang Y 2023 *Sci. China Phys. Mech. Astron.* **66** 282001
- [22] Boehnlein A, Diefenthaler M, Sato N, Schram M, Ziegler V, Fanelli C, Hjorth-Jensen M, Horn T, Kuchera M P, Lee D, Nazarewicz W, Ostroumov P, Orginos K, Poon A, Wang X N, Scheinker A, Smith M S, Pang L G 2022 *Rev. Mod. Phys.* **94** 031003
- [23] Huang M, Wendt K A, Schunck N F, Holmbeck E M 2025 *Phys. Rev. C* **112** 034317
- [24] Wu X, Moloko L E, Bokov P M, Delipei G K, Kaizer J, Ivanov K N 2025 *Nucl. Sci. Eng.* **1**
- [25] Neudecker D, Cutler T E, Devlin M, Brain P, Gibson N, Grosskopf M J, Herman M W, Hutchinson J, Kawano T, Khatiwada A, Kleedtke N, Leal-Cidoncha E, Little R C, Lovell A E, Stamatopoulos A, Thompson E C, Vander Wiel S A, Williamson E 2025 *Phys. Rev. X* **15** 021086
- [26] Schnabel G, Capote R, Koning A, Brown D 2021 *arXiv preprint arXiv: 2110.10322 [nucl-th]*
- [27] Schnabel G, Sjöstrand H, Hansson J, Rochman D, Koning A, Capote R 2021 *Nucl. Data Sheets* **173** 239
- [28] Neudecker D, Grosskopf M, Herman M, Haeck W, Grechanuk P, Vander Wiel S, Rising M E, Kahler A, Sly N, Talou P 2020 *Nucl. Data Sheets* **167** 36
- [29] Huang Q, Peng S, Deng J, Zeng H, Zhang Z, Liu Y, Yuan P 2023 *Heliyon* **9** e13883
- [30] Shu B, Mason D 2023 *EPJ Web Conf.* **284** 18004
- [31] Dudouet J, Gruyer D 2024 *Eur. Phys. J. Plus* **139** 641
- [32] Zhou S G 2017 *Nucl. Phys. Rev.* **34** 318 (in Chinese) [周善贵 2017 原子核物理评论 **34** 318]
- [33] Zhou X H, Zhang Z Y, Gan Z G, Xu F R, Zhou S G 2020 *Sci. Sin. Phys. Mech. Astron.* **50** 2 (in Chinese) [周小红, 张志远, 甘再国, 许甫荣, 周善贵 2020 中国科学: 物理学 力学 天文学 **50** 2]
- [34] Hofmann S, Münzenberg G 2000 *Rev. Mod. Phys.* **72** 733
- [35] Giuliani S A, Matheson Z, Nazarewicz W, Olsen E, Reinhard P G, Sadhukhan J, Schuettrumpf B, Schunck N, Schwerdtfeger P 2019 *Rev. Mod. Phys.* **91** 011001
- [36] Smits O R, Düllmann C E, Indelicato P, Nazarewicz W, Schwerdtfeger P 2024 *Nat. Rev. Phys.* **6** 86
- [37] Science News Staff 2005 *Science* **309** 78
- [38] Breiman L 2001 *Mach. Learn.* **45** 5
- [39] Qi C, Xu F R, Liotta R J, Wyss R, Zhang M Y, Asawatangtrakuldee C, Hu D 2009 *Phys. Rev. C* **80** 044326
- [40] Qi C, Xu F R, Liotta R J, Wyss R 2009 *Phys. Rev. Lett.* **103** 072501
- [41] Bao X, Guo S, Zhang H, Xing Y, Dong J, Li J 2015 *J. Phys. G* **42** 085101
- [42] Soylu A 2019 *Chin. Phys. C* **43** 074102
- [43] Ren Z, Xu C 2005 *Nucl. Phys. A* **759** 64
- [44] Xu C, Ren Z, Guo Y 2008 *Phys. Rev. C* **78** 044329
- [45] Santhosh K P, Biju R, Sahadevan S 2010 *Nucl. Phys. A* **832** 220
- [46] Santhosh K P, Nithya C, Jose T A 2021 *Phys. Rev. C* **104** 024617
- [47] Cai B, Yuan C X 2023 *Nucl. Sci. Tech.* **34** 204
- [48] Royer G 2000 *J. Phys. G* **26** 1149
- [49] Cai B, Yu T, Lin X, Zhang J, Wang Z, Yuan C 2023 *At. Energy Sci. Tech.* **57** 704
- [50] Wen H F, Shang T S, Li J, Niu Z M, Yang D, Xue Y H, Li X, Huang X L 2023 *Acta Phys. Sin.* **72** 152101 (in Chinese) [温湖峰, 尚天帅, 李剑, 牛中明, 杨东, 薛永和, 李想, 黄小龙 2023 物理学报 **72** 152101]
- [51] Cai B, Chen G, Xu J, Yuan C, Qi C, Yao Y 2020 *Phys. Rev. C* **101** 054304
- [52] Cai B, Chen G S, Yuan C X, He J J 2022 *Chin. Phys. C* **46** 084104
- [53] Pedregosa F, Varoquaux G, Gramfort A, et al. 2011 *J. Mach. Learn. Res.* **12** 2825
- [54] Möller P, Sierk A J, Ichikawa T, Iwamoto A, Mumpower M 2015 *Phys. Rev. C* **91** 024310

SPECIAL TOPIC—Thematic data in nuclear physics: Experimental, theoretical and applied research

## Modernization of nuclear structure data: Promoting evaluation efficiency and usability\*

ZHANG Yinu LUO Haotian YUAN Cenxi<sup>†</sup> LI Kele LI Zhiyi  
HUANG Yiming CAI Boshuai

(*Sino-French Institute of Nuclear Engineering and Technology, Sun Yat-Sen University, Zhuhai 519082, China*)

( Received 10 November 2025; revised manuscript received 26 December 2025 )

### Abstract

The Evaluated Nuclear Structure Data File (ENSDF) serves as the cornerstone database for nuclear structure and decay data, underpinning research in nuclear physics, energy, medicine, and astrophysics. However, its legacy 80-column ASCII format, established decades ago, presents significant challenges in data accessibility, interoperability, and scalability amidst today's data-intensive research environment. This work systematically addresses the urgent need for modernizing ENSDF by analyzing its historical limitations, reviewing international modernization initiatives, and presenting a comprehensive framework that integrates advanced data formats, database technologies, machine learning (ML), and interactive visualization.

The primary objectives of this study are twofold: first, to articulate a pathway for transforming ENSDF into a FAIR (findable, accessible, interoperable, reusable)-compliant resource through structural and technological upgrades; and second, to demonstrate the scientific potential of modernized nuclear data by applying ML methods to predict decay properties of superheavy nuclei.

Methodologically, we propose and implement a multi-faceted modernization framework. This includes the adoption of JSON Schema to replace rigid column-based records with hierarchical, self-describing, and machine-readable data structures. We introduce CouchDB, an object-oriented database, to natively support diverse data types and enable efficient querying via precomputed views. Additionally, we discuss the integration of machine learning both as a tool for enhancing database maintenance—such as automated PDF table extraction and anomaly detection—and as a predictive engine for nuclear properties. Inspired by international efforts like the modernized NuDat interface, we also developed a localized visualization platform using Dash and Plotly, enabling interactive exploration of nuclide charts and level schemes based on locally parsed ENSDF data.

As a concrete application of modernized nuclear data, we focus on the decay properties of heavy nuclei. Using JSON-formatted structure and decay data derived from ENSDF, we train a random forest (RF) model to predict half-lives and dominant decay modes across  $\alpha$  decay,  $\beta^\pm$  decay, electron capture, and spontaneous fission. The model is trained to correct residuals from established semi-empirical formulas (e.g., the Universal Decay Law for  $\alpha$  decay and a new three-parameter formula for spontaneous fission). Input features include proton number  $Z$ , neutron number  $N$ , mass number  $A$ , parity, decay energies, and fission barriers.

Our results demonstrate a significant improvement in predictive accuracy. The RF model achieves 92.2% agreement with experimental data in identifying the dominant decay mode and reduces the root-mean-square error (RMSE) by an average of over 50% across all decay channels. Notably, the model successfully reproduces

\* Project supported by the National Natural Science Foundation of China (Grant Nos. 12305129, 12475129), the Guangdong Major Project of Basic and Applied Basic Research of Guangdong Province, China (Grant No. 2021B0301030006), and the Nuclear Data Key Laboratory Fund Project, China (Grant No. JCKY2023201C152).

<sup>†</sup> Corresponding author. E-mail: [yuanx@mail.sysu.edu.cn](mailto:yuanx@mail.sysu.edu.cn)

known systematic trends, such as the elongated  $\alpha$  decay valley and the competition between  $\alpha$  decay and spontaneous fission in the northeast region of the chart. It also predicts an island of enhanced stability southwest of  $Z = 114$ ,  $N = 184$ , correlated with higher fission barriers.

In conclusion, this work underscores that modernizing ENSDF is not merely a technical upgrade but a paradigm shift toward data-driven nuclear science. By implementing JSON-based structuring, object-oriented databases, ML-aided evaluation, and advanced visualization, we establish a foundation for scalable, interoperable, and intelligent nuclear data infrastructure. The successful application of RF to superheavy nuclear decay validates the synergy between modern data formats and ML, offering a powerful tool for exploring unknown nuclides and constraining theoretical models. These efforts, exemplified by the autonomous development at Sun Yat-sen University, reflect China's growing capacity for innovation in nuclear data infrastructure and its commitment to supporting future discoveries in nuclear physics, astrophysics, and applied nuclear technologies.

All the data presented in this work can be found at <https://doi.org/10.57760/sciencedb.30258>.

**Keywords:** nuclear data, visualization interface, machine learning, decay

**DOI:** [10.7498/aps.75.20251528](https://doi.org/10.7498/aps.75.20251528)

**CSTR:** [32037.14.aps.75.20251528](https://cstr.cn/32037.14.aps.75.20251528)



## 核结构数据现代化：提升评价效率与可用性

张一怒 罗皓天 袁岑溪 李珂乐 李至奕 黄一鸣 蔡博帅

## Modernization of nuclear structure data: Promoting evaluation efficiency and usability

ZHANG Yinu LUO Haotian YUAN Cenxi LI Kele LI Zhiyi HUANG Yiming CAI Boshuai

引用信息 Citation: *Acta Physica Sinica*, 75, 020119 (2026) DOI: 10.7498/aps.75.20251528

CSTR: 32037.14.aps.75.20251528

在线阅读 View online: <https://doi.org/10.7498/aps.75.20251528>

当期内容 View table of contents: <http://wulixb.iphy.ac.cn>

---

## 您可能感兴趣的其他文章

### Articles you may be interested in

基于神经网络方法研究  $\beta$  衰变释放粒子的平均能量数据

Average energy data of  $\beta$  decay nuclei based on neural networks

物理学报. 2025, 74(18): 182901 <https://doi.org/10.7498/aps.74.20250655>

基于波动与扩散物理系统的机器学习

Machine learning based on wave and diffusion physical systems

物理学报. 2021, 70(14): 144204 <https://doi.org/10.7498/aps.70.20210879>

机器学习在光电子能谱中的应用及展望

Application and prospect of machine learning in photoelectron spectroscopy

物理学报. 2024, 73(21): 210701 <https://doi.org/10.7498/aps.73.20240957>

生物分子模拟中的机器学习方法

Machine learning in molecular simulations of biomolecules

物理学报. 2023, 72(24): 248708 <https://doi.org/10.7498/aps.72.20231624>

机器学习在宇宙线粒子鉴别中的应用

Application of machine learning in cosmic ray particle identification

物理学报. 2023, 72(14): 140202 <https://doi.org/10.7498/aps.72.20230334>

蛋白质计算中的机器学习

Machine learning for *in silico* protein research

物理学报. 2024, 73(6): 069301 <https://doi.org/10.7498/aps.73.20231618>