arXiv:2601.00685v1 [physics.med-ph] 2 Jan 2026

基于类人 AI 的自动野中野全脑放射治疗计划：结合对话式大语言模型反馈

Adnan Jafar[1], An Qin[1], Gavin Atkins[1], Xiaoyu Hu[1],Yin Gao[2], XunJia[1*]1. 约翰斯·霍普金斯大学放射肿瘤学与分子放射科学系，美国马里兰州巴尔的摩 2. 宾夕法尼亚大学放射肿瘤学系，美国宾夕法尼亚州费城

排版版本：2026年1月5日

* 通讯作者。电子邮件：xunji@jhu.edu

摘要

背景： 全脑放射治疗（WBRT）因其简便且有效而成为一种常见的治疗方式。尽管现代治疗计划系统中的自动野中野（Auto-FiF）功能有助于制定全脑放射治疗计划，但要生成最优计划仍需人工操作，包括定义患者特异性超参数，以及根据质量反馈优化计划。

目的： 本研究介绍了一种自动化全脑放射治疗计划流程，该流程整合了深度学习（DL）超参数预测模型，用于生成患者特异性参数，以及基于大语言模型（LLM）的对话式界面，用于交互式计划优化。

方法： 超参数预测模块使用55例全脑放射治疗病例进行训练，利用临床靶区体积（CTV）和危及器官（OARs）的几何特征，在RayStation治疗计划系统中确定最佳Auto-FiF设置。随后，在预测的超参数下生成治疗计划。对于生成计划欠佳的病例，通过对话模块采集语音输入形式的质量反馈，使用Whisper进行转录，并由GPT-4o进行解读，以调整计划设置。在15例独立病例中，采用临床指标和专家评审对计划质量进行评估，并通过特征重要性分析支持模型可解释性。

结果： 15个深度学习生成的计划中有14个在临床上可接受。在与临床计划的临床靶体积 D95%进行归一化后，深度学习生成的计划与临床计划在眼睛、晶状体的剂量，以及临床靶体积剂量指标 D1%和D99%方面均无统计学显著差异。基于深度学习的计划计算耗时不到1分钟，通过一次鼠标点击即可在约7分钟内完成整个工作流程，而手动计划需要15分钟。对于需要调整的病例，对话模块成功改善了剂量适形度并降低了热点。

结论： 所提出的系统提高了计划效率，同时保持了临床可接受的计划质量。该系统证明了将基于深度学习的超参数预测与大语言模型交互相结合，用于简化并生成高质量全脑放射治疗计划的可行性。

目录

I. 引言 1

II. 方法 3 II.A. 超参数预测模块.4 II.A.1. 数据集.4 II.A.2. 网络结构.5 II.A.3. 几何特征.5 II.A.4. 网络训练.7 II.B. 对话模块.8 II.C. 评估.9

III. 结果 10

III.A. 超参数预测模块 .10III.A.1. 深度学习增强的 Auto-FiF 计划 .10III.A.2. 可解释性分析 .12

III. B. 对话模块 14

IV. 讨论15

V. 结论 17

参考文献 18

I. 引言

约25%的癌症患者会发生脑转移瘤，最常见的原发部位是肺和乳腺[1]。随着癌症治疗效果提升、患者生存期延长，影像学技术进步使脑转移瘤得以更早发现，以及黑色素瘤等易发生转移的癌症患病率上升，脑转移瘤的发生率也随之增加[2]。脑转移瘤通常采用姑息治疗[3,4]，，治疗方案包括化疗、手术、放射治疗，或这些方法的联合应用。立体定向放射外科（SRS）和全脑放射治疗（WBRT）是治疗脑转移瘤的常见放射治疗方式。对于单发且可手术的脑转移瘤患者，手术后接受WBRT是标准治疗方案，已证实可改善生存和肿瘤学结局[5–7]。对于最多三处转移瘤，SRS较WBRT具有更好的局部控制效果和更少的副作用，但治疗区域外复发的风险较高[8–10]。尽管治疗方法多种多样[11]，，WBRT仍是治疗脑转移瘤的广泛应用方案[11,12]。

全脑放射治疗通常采用两束平行对向侧向射束，并让患者处于仰卧位[2,3]。全脑放射治疗计划的目标是制定治疗计划，使临床靶区体积（CTV）获得均匀剂量覆盖（通常包括全脑实质），同时尽可能降低晶状体、视神经/视交叉等危及器官所受的辐射剂量。在当前实践中，计划制定者通常与医师协作，根据医生反馈采用正向计划方法制定全脑放射治疗计划。由于野中野（FiF）技术简单、对摆位变化具有较强的稳健性，且可采用标准步进式照射方式进行交付，因此常被使用[13,14]。全脑放射治疗的野中野计划流程包括勾画器官和临床靶区体积的轮廓、设置等中心以及指定处方剂量。计划制定者首先确定由多叶准直器（MLC）为两束相对侧向射束形成的照射野，然后反复向每束射束添加子野以减少热点，并手动调整各子野的机器跳数（MUs），直至获得满意的计划。为简化这一流程，商用治疗计划系统（TPS）已实现 Auto-FiF算法，通过迭代方式为两束相对侧向射束添加射束子野，旨在利用射束权重优化屏蔽热点并尽量减少低剂量区域[15]。Auto-FiF算法已在多种疾病部位得到成功评估，包括

全脑、直肠、宫颈和乳腺[14, 16–18]。

尽管 Auto-FiF算法成功实现了计划流程自动化，但仍有两个主要因素会影响其效果。第一是 Auto-FiF算法中超参数的选择。这些超参数定义了用户对生成计划各方面的偏好和指导，因此会显著影响计划质量及其临床可接受性。以 RayStation治疗计划系统（RaySearch Laboratory，瑞典斯德哥尔摩）为例，这些参数包括靶区覆盖优先级、子野数量、每分次最小子野机器跳数（MU）、最小子野面积、最小开放叶片对数以及最小叶片末端间距。设置不当可能导致计划质量欠佳。第二个因素出现在使用 Auto-FiF算法生成初始计划之后。如果根据医师的意见还需要进一步优化，那么简化计划更新流程至关重要，以便治疗计划系统能够按照预期方向修订计划。

深度学习（DL）的最新进展显著改变了放射治疗计划[19–23]。已有研究提出基于深度学习的方法来自动制定全脑放射治疗计划[5,24]。这些方法要么预测对向侧野的射束孔径[5]，要么预测两个侧向射束的多叶准直器位置[24]。本研究将探索利用深度学习，通过为 RayStation治疗计划系统中的 Auto-FiF算法生成患者特异性计划超参数设置，来实现 Auto-FiF功能。与构建虚拟计划系统以操作治疗计划系统的研究方向[20,25]，类似，这种方法使我们能够使用商业治疗计划系统直接生成计划，同时由该系统处理计划中细致但至关重要的物理建模部分，例如剂量计算和多叶准直器建模，从而确保计划可交付性。

为简化计划优化流程，我们提出使用现代人工智能工具进行人机交互。近年来，大语言模型（LLM）在放射治疗计划中展现出巨大潜力[26,27]，，这得益于其通过在大规模、多样化语料库上进行自监督训练，并进一步与人类偏好对齐[28]而获得的零样本能力。它们能够理解、解释并将非结构化临床输入概括为结构化表示，因此在医疗保健领域的各种决策支持任务中具有重要价值[28,29]。因此，在本研究中，我们将开发一个人机对话模块，用于指导计划工具，从而无缝接收人类指令

并促进生成符合临床要求的计划。

II. 方法

商用治疗计划系统中的 Auto-FiF算法依赖一组直接影响计划质量的超参数。在常规实践中，设定这些超参数是一个需要反复迭代、有时还需多次调整的手动过程。依赖手动微调会降低效率，并导致不同计划制定者之间出现差异。为应对这些挑战，我们设计了一套基于 Auto-FiF 的全自动全脑放射治疗计划工作流，如图 1 所示。该工作流由两个紧密集成的模块组成：超参数预测模块和对话模块。

超参数预测模块采用监督式深度学习方法，预测治疗计划系统中最优的 Auto-FiF 超参数设置。输入为临床靶体积及邻近危及器官的几何描述符，并将其映射为超参数值，以生成可模拟经医师批准的临床全脑放射治疗计划的方案。我们选择监督学习框架，是因为该任务定义明确，适合采用监督回归/分类方法。我们特意选用几何特征而非完整 CT体积数据作为输入，因为全脑放射治疗涉及相对简单的解剖几何结构，靶区与器官之间的空间关系预计足以为计划制定提供指导。这一选择也增强了模型可解释性，并降低了计算负担。为进一步提高透明度和临床信任度，该模块纳入了一个可解释性层，用于突出对预测超参数影响最大的几何特征。随后，系统根据预测超参数启动 Auto-FiF功能以生成计划。

当自动生成的计划仍需进一步改进以满足临床预期时，对话模块便提供了一种新颖的医师与治疗计划系统交互渠道。该模块由大语言模型驱动，使医生能够以自然语言直接表达所需的修改，例如“减少顶叶区域的热点”或“改善左侧颞叶的均匀性”。系统会将这些指令转化为治疗计划系统中的可执行调整。这种直接途径不仅加快了迭代速度，还减少了沟通瓶颈，并加强了医师对最终计划

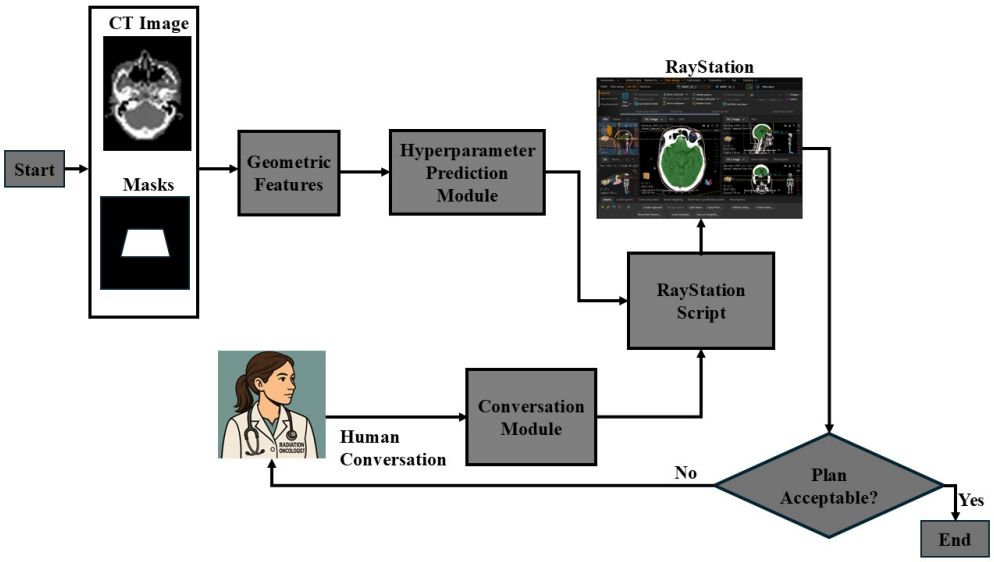

图 1：用于全脑放射治疗的 Auto-FiF 治疗计划工作流程，并结合了基于深度学习的可解释超参数预测模块和由大语言模型驱动的对话模块。

质量的监督。

后续各节将介绍这两个模块的技术实现，重点阐述其设计原则、在治疗计划系统中的集成方式，以及在临床全脑放射治疗场景中的评估。

II.A. 超参数预测模块

II.A.1. 数据集

我们回顾性选取了此前在本院接受过全脑放射治疗的70名患者。对于每个病例，我们从RayStation治疗计划系统导出了医学数字成像和通信格式的治疗计划数据，包括CT图像、RT结构和RT计划文件。由于超参数预测模块被定义为一项监督学习任务，因此构建的数据集以几何特征作为模块输入，并以相应的Auto-FiF超参数真实值设置作为输出。为获取这些作为输出的超参数真实值，我们编写了Python脚本，在RayStation治疗计划系统中针对一系列候选超参数设置系统地运行Auto-FiF算法，并采用网格搜索策略，找出最能复现医师批准的临床计划的超参数配置。通过这一过程，我们确定了四个关键参数：靶区覆盖优先级（25%至100%，步长为25%）、子野数量（1至3，步长为1）、

每次分次的最小子野机器跳数（2至8机器跳数，步长为2），以及最小子野面积（4、6、9和10 cm^2）；这些参数需要调整，才能忠实复现临床计划。其余参数，例如开放叶片对的最小数量（1）和最小叶片末端间距（0 cm），在所有病例中均保持默认值。

II.A.2. 网络结构

如图 2 所示，我们将学习最优 Auto-FiF 超参数的任务构建为一项由靶区及周围解剖结构的几何特征驱动的多类分类任务。为此，我们设计了三个采用全连接架构的深度学习模型。第一个模型预测靶区覆盖优先级，将其按 25% 的增量离散为从 25% 到 100% 的类别。第二个模型预测子野数量（1、2 或 3）。第三个模型与子野数量预测网络共享其骨干架构，预测一个联合标签，用于编码每次分次的最小子野 MU（2、4、6 或 8）与最小子野面积（4、6、9 或 10）的组合。对于第三个模型，我们将标签空间限制为四种具有临床相关性的组合：（2，4）、（4，6）、（6，9）和（8，10），因为这些参数组合始终足以复现经医师批准的临床计划。

II.A.3. 几何特征

我们选择脑、左眼和右眼，以及左晶状体和右晶状体作为感兴趣器官，这与医生为全脑放射治疗制定治疗计划时通常评估的结构一致。我们没有将这些结构的三维轮廓直接输入网络进行预测，而是从中提取相关几何特征，以避免在相关数据量较小时出现过拟合，并支持结果的可解释性。最初选取的几何特征包括各器官的体积、表面积和直径，器官间质心到质心的距离，等中心到各器官质心的距离，以及从等中心到各器官质心的极角和方位角。对于具有横向对称性的特征，例如等中心到左右眼的距离，我们对相应特征值取平均，以消除候选特征间的冗余。最终共得到 21 个特征，如图3 底部所示。

要移除 re　冗余信息，并聚焦于所提供的最相关信息

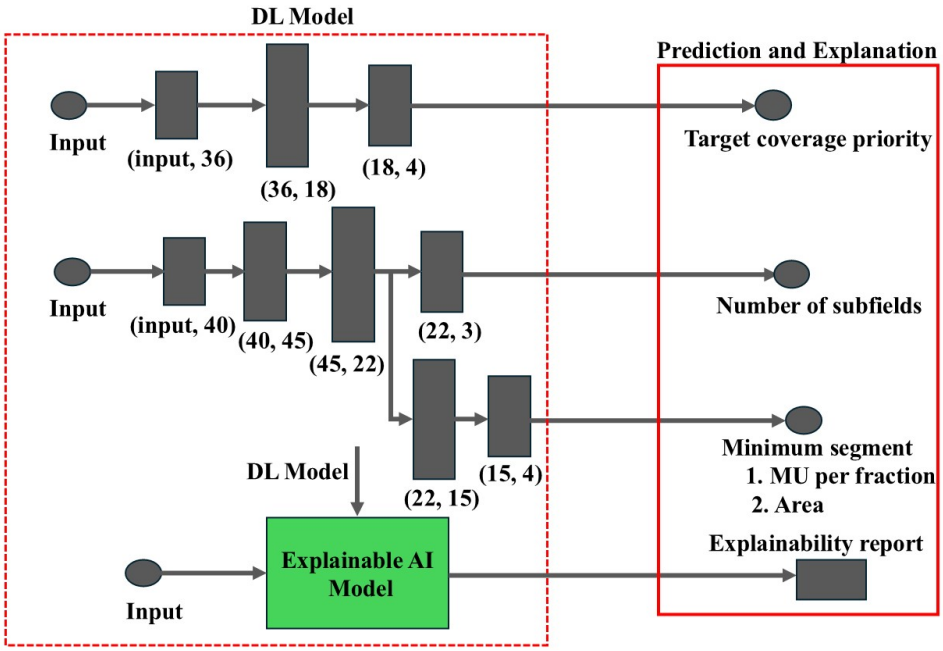

图2：用于 Auto-FiF 超参数预测的深度学习模型架构，包括一个用于靶区覆盖优先级的模型、一个用于子野数量的模型，以及第三个复用子野分类器主干并增加额外层的模型，用于预测每次分次的最小子野机器跳数和最小子野面积。采用可解释人工智能模块，将模型预测结果归因于几何特征。

几何特征集合，我们采用凝聚层次聚类[30] 对标准化输入特征进行聚类（图3），并使用欧几里得距离度量。凝聚层次聚类根据所选的连接准则和距离度量迭代合并簇，从而灵活地对特征进行分组[30]。通过在特定的水平阈值处递归切割树状图，我们迭代筛选出能够提升模型性能的特征。对于未能提高分类器准确率的特征，则予以剔除；这一过程持续进行，直至确定最优特征子集，从而提升模型性能。最终，选用于训练深度学习模型的几何特征包括脑、眼睛和晶状体的直径（脑直径、眼睛直径、晶状体直径）；脑、眼睛和晶状体的质心至等中心的距离（脑至等中心距离、眼睛至等中心距离、晶状体至等中心距离）；以及眼睛和晶状体的质心相对于等中心的极角和方位角（眼睛等中心极角、晶状体等方偏振、眼睛等中心方位角、晶状体等中心方位角）。这一结果表明，这些器官的大小及其相对于等中心的位置，提供了对治疗计划至关重要的独特信息。

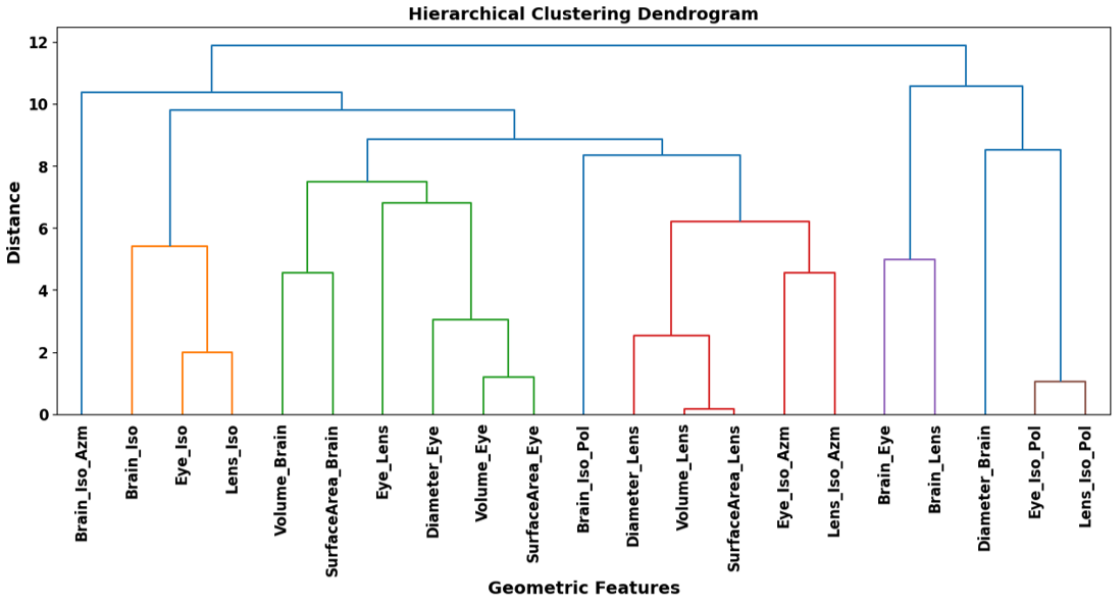

图3：树状图展示了几何特征的层次聚类，用于识别特征关系，并指导迭代式特征选择，以优化模型性能。

II.A.4. 网络训练

我们使用55例全脑放射治疗病例训练了三个深度学习模型，训练1500轮，批量大小为5，并以交叉熵损失作为优化目标。在训练队列上进行了五折交叉验证，以支持模型选择并减轻过拟合。通过监测验证损失实施早停。所有模型均使用Adam优化器进行训练，该优化器因具有自适应学习率且能稳健处理稀疏梯度而被选用。另使用一组独立的15例全脑放射治疗病例评估最终模型性能，这些病例未纳入训练或验证。

对于靶区覆盖优先级模型，鉴于数据集规模相对较小，我们在第一个全连接层之后采用了35%的丢弃率，以抑制过拟合。隐藏层采用 Leaky ReLU 和 ReLU 激活函数的组合，旨在兼顾前几层的梯度稳定性与后续层的计算效率。学习率和权重衰减均设为0.0002，这些取值在初步实验中实现了稳定收敛。

对于子野数模型和最小子野MU–子野面积模型，采用了更保守的正则化方案：在第一层后设置 10% 的丢弃率，在第二层后设置 15% 的丢弃率。这些模型还在

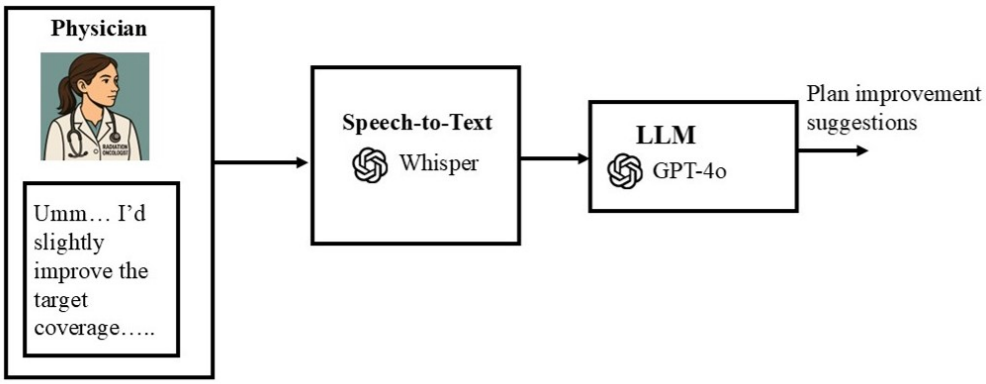

图 4：用于医生参与规划的对话模块。医师的口头反馈首先通过 Whisper 转换为文本，然后由 GPT-4o 进行解读，生成符合医师意图的计划改进建议。

第一层使用 Leaky ReLU 激活，后续各层则使用 ReLU 激活，以增强非线性，同时保持稳定性。学习率和权重衰减分别设为 0.001 和 0.00009，以加快收敛，同时防止对训练数据过拟合。

II.B. 对话模块

对于对话模块，医师输入经过采集、转录、解读，并转换为治疗计划系统中的优化目标。医师语音首先通过 *sounddevice* 库（用于录音的 Python 软件包）自动录制，然后使用 OpenAI Whisper（一种先进的语音转文本模型，以能够稳健处理口音和临床术语而闻名）转录为文本。随后，转录文本由 GPT-4o API（温度设置 $= 0.2$）进行处理，并根据预定义架构生成结构化 JSON 输出（图 4）。选择较低的温度设置是为了最大限度地减少随机性，并确保生成适用于临床决策的一致且可复现的输出。

一个关键的设计选择是使用 GPT-4o 进行少样本提示。在这一策略中，向模型提供了若干输入–输出对示例，以展示医生指令与结构化 JSON 之间所需的映射关系。例如，反馈“靶区覆盖较差”被映射为 `{ "target coverage": "bad", "OAR dose": "unchanged" }`。这些示例引导模型以结构化且机器可读的格式一致地解读临床反馈。这种方法至关重要，因为医生指令的措辞通常各不相同，而治疗计划系统要求

使用标准化命令来调整超参数。反馈可能包括提高靶区覆盖、降低特定危及器官的剂量，或修改初始控制点射束权重等请求。

生成后，系统会解析结构化 JSON，并将其直接转换为优化目标。例如，将改善靶区覆盖的要求具体化为：与初始计划相比，D_{95} 或 D_{99} 至少提高 5%。为实现这些目标，系统会围绕超参数预测模块预测的超参数值进行局部网格搜索。为提高计算效率，局部网格搜索仅针对两个参数进行：靶区覆盖优先级和子野数量，因为这两个参数可直接控制靶区覆盖和热点降低。搜索首先从靶区覆盖优先级开始；若仅调整该参数仍不够，则扩展至同时调整这两个参数，以实现医生偏好的权衡。若未找到满意的配置，则恢复为超参数预测模块得出的计划。

这一设计有效地闭合了反馈回路，使医生能够以自然的对话方式提供指导，并将其可靠地转化为可执行的优化策略，从而绕过传统的、反复进行的医师与剂量师沟通周期。

II.C. 评估

我们从三个互补层面评估超参数预测模块，以全面考察深度学习模型的技术准确性及整体工作流的临床效用。首先，我们在包含 15 例全脑放射治疗病例的独立测试集上，评估各多分类深度学习模块的预测表现。评估指标包括精确率、召回率和 F1 分数；针对每个类别分别计算这些指标，再对所有类别取宏平均，以避免偏重出现频率较高的类别。此项分析量化了模型预测自动化 FiF 计划所需离散超参数设置的可靠程度。其次，我们采用标准剂量学指标，对生成的治疗计划进行定量评估；评估前，针对每个病例，将深度学习生成的计划归一化至与临床计划相同的 D95%。对于临床靶体积（脑），评估 D1% 和 D99%；对于危及器官（眼睛和晶状体），评估平均剂量。通过比较深度学习增强的 Auto-FiF

计划与医师批准的临床参考计划，我们得以评估自动化工作流再现临床可接受剂量分布的程度，以及其是否维持了靶区覆盖与正常组织保护之间的微妙平衡。最后，由经验丰富的医学物理师进行定性专家评审，按照 5 分制评估深度学习增强的 Auto-FiF 计划的整体临床可接受性（5 =临床可接受，$4=$ 归一化后经小幅修改即可接受，$3=$ 比从头开始制定计划更好，$2=$ 需要大幅修改，$1=$ 不可用）。评审不仅关注覆盖和危及器官保护等客观标准，也考量敏感解剖区域是否存在热点、剂量分布是否符合临床预期，以及计划的总体质量等主观方面。

对话模块通过功能测试、计划调整准确性评估和专家评审相结合的方式进行评估。为展示其解读医师反馈的能力，我们设计了一组具有代表性的临床场景，用户在其中提供口头指令，并据此对计划进行优化。随后，将生成的 Auto-FiF 计划与基于预测超参数生成的基线计划进行比较，以确定是否成功纳入了所要求的修改。定量评估重点考察目标指标是否按照用户要求得到改善；定性评估则由医学物理师开展，以确认剂量分布变化具有临床意义，且未损害计划质量的其他方面。

III. 结果

III.A. 超参数预测模块

III.A.1. 深度学习增强的 Auto-FiF 计划

表1 展示了超参数预测模块中三个深度学习组件在靶区覆盖优先级、子野数量、每次分次的最小子野 MU 和最小子野面积方面的宏平均精确率、召回率和 F1 分数，表明各项任务的表现始终出色。图5展示了三个临床计划与对应的三个深度学习优化计划的比较。其中一个经物理评估人员审核后立即获批，一个经小幅修改后获批，还有一个被拒绝。

表1：基于五折交叉验证的三个深度学习模块的宏平均性能（平均值 ± 标准差）。

模块	精确率	召回率	F1分数
靶区覆盖优先级（4 类别）	0.89±0.03	0.87±0.04	0.88±0.03
子野数量（3 数量类别）	0.90±0.03	0.88±0.03	0.89±0.03
每分数和面积的MU（4 个类别）	0.92±0.02	0.90±0.03	0.91±0.02
总体宏平均	**0.90±0.03**	**0.88±0.03**	**0.89±0.03**

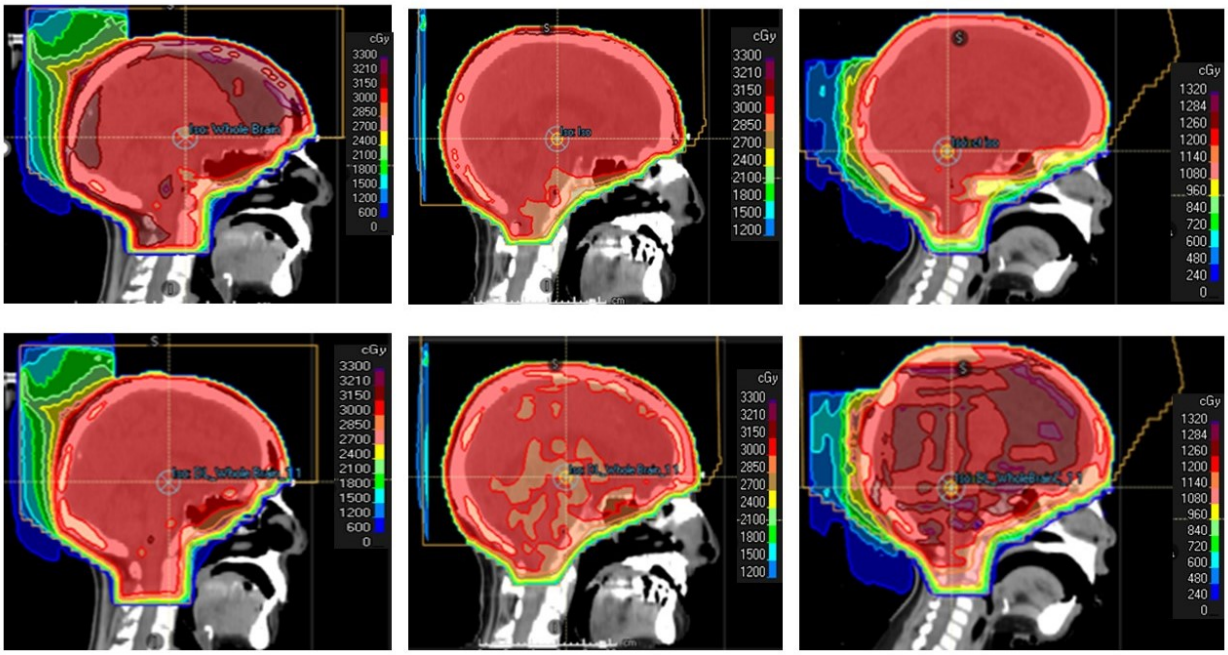

图5：临床批准的计划（上）与超参数预测模块生成的计划（下）的示例，分别对应批准（左）、经小幅修改后批准（中）和拒绝（右）的病例。中间的红色区域表示临床靶体积。

表2展示了超参数预测模块生成的计划与临床计划之间的剂量学比较。以$p = 0.05$为阈值水平，根据 t检验，在所有评估指标中均未观察到具有统计学意义的差异。我们还在表3中展示了医学物理师对计划的评估。15 个深度学习计划中，11 个（73.3%）无需修改即可接受，3 个（20%）需要小幅修改，1 个（6.7%）因热点过多且晶状体剂量相对较高而不可接受。

在计算时间方面，超参数预测模块的推理时间平均始终低于一分钟。该流程的总端到端运行时间

约为七分钟，其中包括从 RayStation 导入医学数字成像和通信数据、进行几何特征提取、深度学习模块推理以及 Auto-FiF计划生成。值得注意的是，整个过程完全自动化，只需单击一次鼠标即可触发，启动后无需人工干预。尽管对于经验丰富的计划制定者而言，全脑放射治疗的 Auto-FiF 计划制定被认为是相对简单的任务，但通常的手动计划时间约为 15 分钟，在此期间，计划制定者必须反复调整超参数并验证计划质量。相比之下，所开发的自动化工作流不仅缩短了计划制定时间，还通过消除手动步骤简化了计划制定流程。这一改进凸显了该系统在保持计划质量一致性的同时提升临床效率的潜力。

表2：超参数预测模块生成的计划（深度学习计划）与临床计划之间的剂量学指标比较。

结构	**剂量指标**	**临床 Plan (Gy)**	**DL Plan (Gy)**	**p值**
脑（临床靶体积）	D1%	31.7±4.4	31.4±4.0	0.59
脑（临床靶体积）	D99%	29.9±5.5	29.7±5.3	0.84
左眼	Mean	10.1±4.9	9.8±4.4	0.54
右眼	Mean	10.9±5.2	10.5±5.1	0.91
左眼晶状体	Mean	3.2±1.1	3.2±1.0	0.80
右眼晶状体	Mean	3.3±1.1	3.3±1.1	0.79

表3：医学物理师对计划的定性评估。

带评分的计划				**临床接受度**		
5	4	3	≤ 2	可接受	可接受，但需小幅修改	优于从头开始构建
11	3	1	0	11（73%）	3（20%）	1（7%）

注： 没有任何计划获得低于3分的评分。

III.A.2. 可解释性分析

我们模型的一项优势是可解释性，这使我们能够深入了解预测行为和结果。图 6（左）展示了与靶区覆盖优先级相关的模型预测中的特征贡献。贡献最大的前三个特征是晶状体等方偏振、脑至等中心距离和眼睛至等中心距离。这些发现表明，靶区覆盖优先级

受到等中心、脑和眼睛的相对位置关系影响，该位置关系决定了算法必须在确保足够的靶区覆盖与保护眼部结构之间实现的平衡。首先，研究发现，这些结构到等中心的距离十分重要。尽管晶状体至等中心距离 未列入贡献最大的前三个特征，但它与眼睛至等中心距离高度相关，因此其贡献已被隐含体现。其次，眼睛的极角晶状体等方偏振被确定为一个重要特征。该指标大致反映了眼睛相对于等中心的横向分布，从而影响沿射束方向的靶区覆盖。

图6（中）展示了影响深度学习模型对子野数量预测的特征贡献。贡献最大的三个特征是脑至等中心距离、晶状体至等中心距离和脑直径。从治疗计划的角度来看，脑的大小影响尤为显著，因为它决定了为实现临床靶体积足够均匀的剂量覆盖所需的子野数量。此外，脑与晶状体之间的距离，以及它们各自相对于等中心的偏移，在塑造剂量均匀性方面也发挥着重要的几何作用。与前一种情况类似，晶状体至等中心距离和眼睛至等中心距离预计呈现较强的相关性，因此后者的影响会通过前者得到隐式体现。

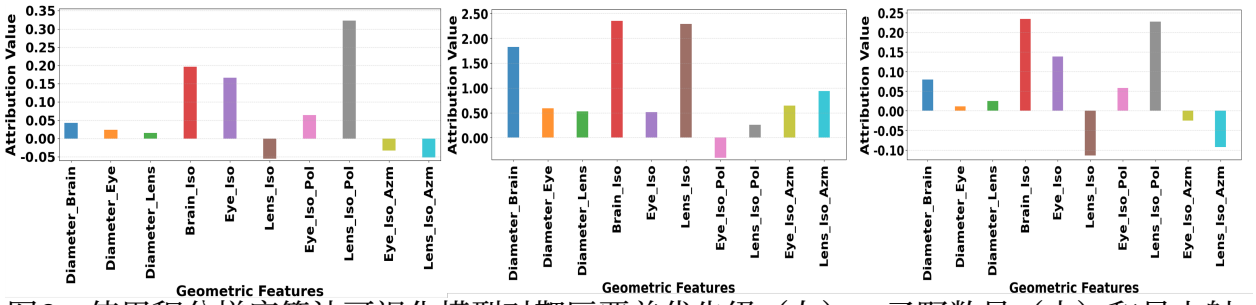

图6：使用积分梯度算法可视化模型对靶区覆盖优先级（左）、子野数量（中）和最小射野段设置（右）的预测所对应的特征贡献。

最后，图6（右）展示了所有最小射野段设置（每次分次的机器跳数、子野面积）下的特征贡献：(2, 4)、(4, 6)、(6, 9) 和 (8, 10)。排名前三的特征分别是脑至等中心距离、晶状体等方偏振和眼睛至等中心距离。这表明，分割设置可能取决于等中心、脑和眼部结构之间的空间关系；眼睛或晶状体距离越近，就需要越精细的射野段

形状，以控制这些敏感区域周围的剂量。

III.B. 对话模块

图 7 展示了两个案例：医学物理师通过对话模块以自然语言提出计划优化建议，并对比了原始的深度学习生成的计划（左）与根据反馈修订后的计划（右）。第一个案例（左上）中，初始深度学习计划的靶区覆盖不够理想，D95%为28.5 Gy，低于30 Gy的处方剂量；危及器官平均剂量分别为6.8 Gy（右眼）、5.9 Gy（左眼）、4.2 Gy（右侧晶状体）和4.0 Gy（左侧晶状体）。反馈内容为："靶区覆盖看起来不太理想——我肯定希望改善这部分。其余部分似乎还不错。"根据通过对话模块提供的反馈，靶区覆盖优先级设置从50%提高到75%，从而加强了优化器对靶区覆盖的重视。修订后，如图7右上子图所示，D95%提高至30.2 Gy，相应的危及器官平均剂量分别为7.0 Gy、6.1 Gy、4.3 Gy和4.1 Gy。

在第二个病例（左下）中，深度学习生成的计划具有充分的靶区覆盖，D95% $= 30.$ 9 Gy，但右眼（23.3 Gy）和左眼（19.6 Gy）的平均剂量偏高，而右侧晶状体和左侧晶状体的剂量分别为 6.1 Gy 和 8.4 Gy。对话模块提供的反馈为："右眼和左眼的平均剂量看起来有点高——建议降低。靶区覆盖似乎没问题。"作为回应，子野数量设置从 2 增加到 3，以提高调制灵活性，从而降低眼睛剂量。如图7右下方子图所示，修订后靶区覆盖保持稳定（D95% $= 30.$1 Gy），同时成功实现了所要求的眼睛剂量降低——右眼和左眼剂量分别降至 18.9 Gy 和 15.6 Gy，晶状体剂量也进一步降至 5.6 Gy（右侧晶状体）和 3.9 Gy（左侧晶状体）。

值得注意的是，这些改进并非简单归一化所能实现，因为归一化只会对所有剂量进行统一缩放；这些改进实际上源于对话模块指导下的针对性重新优化。这些示例共同表明，该模块能够将非正式的人类输入转化为可执行的计划优化。

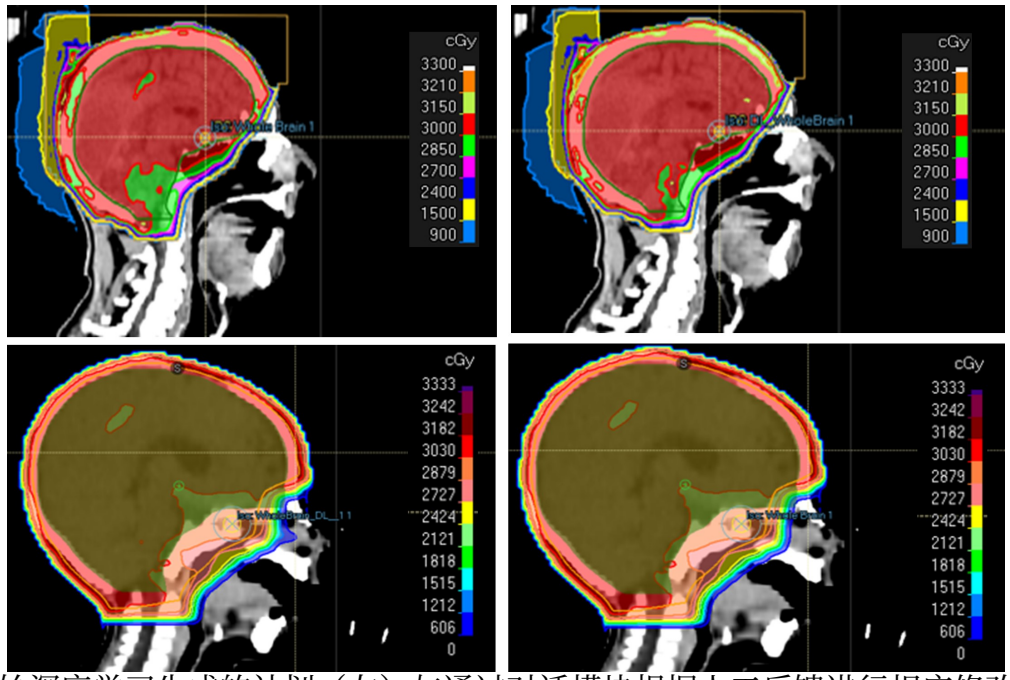

图7：原始深度学习生成的计划（左）与通过对话模块根据人工反馈进行相应修改后的计划（右）的比较。第一种情况下，初始深度学习计划的靶区覆盖不理想；第二种情况下，初始计划对左右眼施加了高剂量。

IV. 讨论

本研究提出了一种端到端深度学习框架，旨在自动确定 RayStation治疗计划系统中 Auto-FiF算法的超参数设置。该方法实现了全脑放射治疗计划的自动化，减少了耗时的手动调整和专家干预。由合格医学物理师进行的定性分析结果显示，在大多数测试病例中，深度学习优化计划均符合临床要求，仅需进行少量调整，即可使剂量分布符合临床标准。这些发现有力地支持了所提出的深度学习框架在全脑放射治疗中的临床可行性，彰显了其简化并改进计划流程的潜力。

由 Whisper 和 GPT-4o 驱动的对话模块，能够将非结构化人工反馈无缝转换为可直接指导治疗计划的结构化指令。该模块的主要目的是解决治疗计划流程中持续存在的人工干预需求。即使初始自动计划质量很高，也无法保证所有临床目标始终

得到满足。传统上，解决此类问题需要医生与剂量师反复沟通，这既耗时，也容易产生理解偏差。相比之下，对话模块允许医生直接提供口头指令。这项功能支持与治疗计划系统进行实时自然语言交互，从而有效绕过传统的医生与计划者反馈循环。该模块在临床专业知识与自动计划之间发挥了关键桥梁作用，确保治疗计划能够通过交互式、高效的方式进行优化，以满足个性化临床目标，极大提升了自动计划工作流程的灵活性和稳健性。在需要频繁、及时地优化计划并作出决策的场景中，例如在线自适应放射治疗，这项功能预计将更加重要[31,32]，；在涉及复杂治疗计划优化模型的情境中也是如此[33–35]。

使用积分梯度进行的可解释性分析进一步增强了所提出的基于深度学习的全脑放射治疗工作流程在临床环境中的透明度。在医疗保健领域的基于深度学习的模型中，通常很难理解模型如何做出具体的医疗决策，这可能会导致临床应用中的不确定性和犹豫[36]。通过应用积分梯度，我们能够确定哪些几何特征最显著地影响了模型在确定 Auto-FiF算法具体超参数设置时所做出的决策。这种透明的基于深度学习的全脑放射治疗工作流程有助于增强人们对系统应用和临床整合的信任。此外，可解释性分析还有助于检测模型中潜在的错误或偏差，从而实现持续改进，并确保治疗计划的安全性、可靠性和有效性。

脑转移瘤发病率不断上升，这归因于 MRI 成像技术的进步、对高风险患者筛查的改进以及全身治疗的增强，给应对这一情况的医疗资源带来了巨大压力[37]。所提出的端到端深度学习工作流程有望简化全脑放射治疗流程，使医疗资源能够投入到更重要的任务中。与此同时，我们认识到，全脑放射治疗计划是一项相对简单的任务。我们选择这项任务来展示该框架的整体可行性。据我们所知，这是首项将可解释人工智能与对话式反馈相结合并集成到治疗计划系统工作流程中的研究，使医生能够通过自然语言输入无缝地完善并最终确定计划。尽管这一概念验证聚焦于全脑放射治疗，但该框架具有通用性，可扩展至更复杂的疾病部位和治疗技术，在这些场景中，治疗计划更具挑战性——

——。我们预计，将当前框架扩展到其他场景也能带来类似优势，例如缩短计划制定时间、减少人为错误并确保治疗计划的一致性，从而进一步提升整个流程的效率和可靠性。

V. 结论

为简化全脑放射治疗计划，我们开发了一种自动化计划框架，整合了两个关键组件：超参数预测模块和对话模块。超参数预测模块利用监督式深度学习，根据患者特异性几何特征自动推断最优 Auto-FiF 参数设置，从而无需手动调整参数即可高效生成临床可接受的计划。与之互补的是，对话模块将人工反馈转化为结构化的计划目标，使计划制定者能够通过自然语言指令直接与治疗计划系统交互，以优化计划。在一组 15 例全脑放射治疗患者中进行测试后，超参数预测模块生成的计划有 93% 达到临床可接受标准（其中 73% 可直接使用，20% 经小幅修改后可用）；与临床计划相比，剂量学指标无统计学显著差异（$p > 0.05$）。对话模块则在需要调整时有效促进了交互式计划优化。对超参数预测模块进行的可解释性分析，为了解模型的决策过程提供了有价值的见解。上述组件共同实现了完全自动化的一键式计划工作流程，不仅缩短了计划制定时间，还保留了计划制定者输入意见并进行监督的灵活性。

作者贡献

A.J. 和 X.J. 提出了这一构想。A.J.、A.Q.、G.A. 和 X.H. 收集并整理了数据。A.J. 和 A.Q. 开发了计算算法及其实现。X.H. 和 Y.G 参与了评估方法的制定和结果解读。所有作者均核验了方法和结果。A.J. 和 X.J. 起草了初稿。所有作者均对最终稿进行了严格审阅并批准。

利益冲突

作者声明不存在利益冲突。

致谢

本研究部分由美国国立卫生研究院资助，资助项目编号为 R37CA214639、R01CA227289 和 R01CA254377。

参考文献

[1] P. Y. Wen、B. P. McLaren 和 J. S. Loeffler，《转移性癌症的治疗》，第2655–2670页，Lippincott Williams and Wilkins，宾夕法尼亚州费城，2001年。[2] E. McTyre、T. Emory、J. Scott 和 P. Chinnaiyan，《脑转移瘤的全脑放射治疗》，《国际神经外科学》**4**，S236（2013年）。[3] V. Sharma、P. Gaye、S. Wahab、N. Ndlovu、T. Ngoma、V. Vanderpuye、A. Sowunmi、J. Kigula-Mugambe 和 B. Jeremic，《非洲姑息性放射治疗实践模式（第一部分）：骨转移和脑转移》，《国际放射肿瘤学·生物学·物理学杂志》**70**，1195–1201（2008年）。[4] C. Nieder、A. Pawinski、E. Haukland、R. Dokmo、I. Phillipi 和 A. Dalhaug，《估算成年癌症患者对姑息性体外放射治疗的需求》，《国际放射肿瘤学·生物学·物理学杂志》**76**，207–211（2010年）。[5] J. Yu、Y. Goh、K. Song、J. Kwak、B. Cho、S. Kim、S. Lee 和 E. Choi，《使用深度学习进行全脑放射治疗自动计划的可行性》，《应用临床医学物理学杂志》**22**，184–190（2021年）。[6] N. Lamba 等，《颅内转移瘤切除术后立体定向放射外科与全脑放射治疗的比较：系统综述和荟萃分析》，《放射肿瘤学》**12**，106（2017年）。

[7] P. Brown 等，《单独放射外科与放射外科联合全脑放射治疗对1至3个脑转移瘤患者认知功能的影响：一项随机临床试验》，《美国医学会杂志》**316**，401–409（2016年）。[8] M. Kocher 等，《对1至3个脑转移瘤进行放射外科或手术切除后，辅助性全脑放射治疗与观察等待的比较：EORTC 22952–26001研究结果》，《临床肿瘤学杂志》**29**，134–141（2011年）。[9] H. Aoyama 等，《立体定向放射外科联合全脑放射治疗与单独立体定向放射外科治疗脑转移瘤的比较：一项随机对照试验》，《美国医学会杂志》**295**，2483–2491（2006年）。[10] M. Yamamoto 等，《多发脑转移瘤患者的立体定向放射外科治疗（JLGK0901）：一项多中心前瞻性观察研究》，《柳叶刀·肿瘤学》**15**，387–395（2014年）。[11] D. Andrews 等，《对1至3个脑转移瘤患者进行全脑放射治疗（可伴或不伴立体定向放射外科加量）：RTOG 9508随机试验的III期结果》，《柳叶刀》**363**，1665–1672（2004年）。[12] F. Franchino、R. Rud`a 和 R. Soffietti，《癌症脑转移的机制与治疗》，《肿瘤学前沿》**8**，161（2018年）。[13] H. Kim、J. Kwak、J. Jung、C. Jeong、K. Yoon、S. Lee、S. Ahn、E. Choi、S. Kim 和 B. Cho，《结合脚本应用程序接口与用户执行程序的乳腺正向调强放射治疗自动化野中野（FIF）计划框架》，《癌症研究与治疗技术》**17**，1533033818810391（2018年）。[14] K. Huang、S. Hernandez、C. Wang、C. Nguyen、T. Briere、C. Cardenas、L. Court 和 Y. Xiao，《自动化野中野全脑放射治疗计划》，《应用临床医学物理学杂志》**24**，e13819（2023年）。[15] R. Byrd、P. Lu、J. Nocedal 和 C. Zhu，《一种用于有界约束优化的有限记忆算法》，《SIAM科学计算杂志》**16**，1190–1208（1995年）。[16] K. Huang、P. Das、A. Olanrewaju、C. Cardenas、D. Fuentes、L. Zhang、D. Hancock、H. Simonds、D. Rhee、S. Beddar 和 T. Briere，《放射治疗自动化

直肠癌治疗计划，《应用临床医学物理学杂志》**23**，e13712（2022年）。[17] K. Huang 等，《可应用于多种疾病部位的全自动野中野算法》，Wiley，2021年。[18] C. Onal、A. Sonmez、G. Arslan、E. Oymak、A. Kotek、E. Efe、S. Sonmez 和 Y. Dolek，乳腺照射中野中野技术与切线楔形射束的剂量学比较，《日本放射学杂志》**30**，218–226（2012年）。[19] X. Li、J. Zhang、Y. Sheng、Y. Chang、F.-F. Yin、Y. Ge、Q. Wu 和 C. Wang，通过静态射野通量预测实现自动调强放疗计划（AIP-SFFP）：用于实时前列腺治疗计划的深度学习算法，《物理学与医学和生物学》**65**，175014（2020年）。[20] Y. Gao、Y. Park 和 X. Jia，头颈癌放射治疗的类人智能自动治疗计划，《物理学与医学和生物学》**69**，115049（2024年）。[21] X. Jia 等，NRG肿瘤学组织对人工智能用于放射治疗临床试验自动治疗计划的评估：现状与未来，《国际放射肿瘤学·生物学·物理学杂志》（2025年）。[22] C. Shen、D. Nguyen、Z. Zhou、S. Jiang、B. Dong 和 X. Jia，医学物理学中的深度学习导论：优势、潜力与挑战，《医学与生物学物理》**65**，05TR01（2020年）。[23] H. Shan、X. Jia、P. Yan、Y. Li、H. Paganetti 和 G. Wang，以深度学习协同推进医学成像与放射治疗，《机器学习：科学与技术》**1**，021001（2020年）。[24] E. Han、C. Cardenas、C. Nguyen、D. Hancock、Y. Xiao、R. Mumme、L. Court、D. Rhee、T. Netherton、J. Li 和 D. Yeboa，全脑放射治疗自动化治疗计划的临床实施，《应用临床医学物理学杂志》**22**，94–102（2021年）。[25] C. Shen、L. Chen 和 X. Jia，基于分层深度强化学习框架的前列腺癌调强放射治疗智能自动治疗计划，《医学与生物学物理》**66**，134002（2021年）。

[26] S. Liu、O. Pastor-Serrano、Y. Chen、M. Gopaulchan、W. Liang、M. Buyyounouski、E. Pollom、Q.-T. Le、M. Gensheimer、P. Dong、Y. Yang、J. Zou 和 L. Xing，由 GPT-4Vision 引导的自动化放射治疗计划，2024年。[27] Q. Wang 等，使用大语言模型智能体自动化放射治疗计划的可行性研究，《医学与生物学物理》**70**，075007（2025年）。[28] H. Naveed、A. Khan、S. Qiu、M. Saqib、S. Anwar、M. Usman、N. Akhtar、N. Barnes 和 A. Mian，大语言模型综合概述，arXiv **abs/2307.06435**（2023年）。[29] A. Thirunavukarasu、D. Ting、K. Elangovan、L. Gutierrez、T. Tan 和 D. Ting，医学领域的大语言模型，《自然·医学》**29**，1930–1940（2023年）。[30] D. M¨ullner，现代分层凝聚聚类算法，arXiv 预印本 **arXiv:1109.2378**（2011年）。[31] S. Lim-Reinders、B. M. Keller、S. Al-Ward、A. Sahgal 和 A. Kim，《在线自适应放射治疗》，《国际放射肿瘤学·生物学·物理学杂志》**99**，994–1003（2017年）。[32] N. Li 等，基于初始计划剂量体积直方图指导的自适应放射治疗自动治疗计划再优化，《医学与生物学物理》**58**（2013年）。[33] M. Zarepisheh、A. F. Uribe-Sanchez、N. Li、X. Jia 和 S. B. Jiang，一种用于放射治疗计划优化、具有体素依赖参数的多准则框架，《医学物理》**41**，041705（2014年）。[34] D. Nguyen、M.-H. Lin、D. Sher、W. Lu、X. Jia 和 S. Jiang，自动化治疗计划的进展，载于《放射肿瘤学研讨会》，第32卷，第343–350页，爱思唯尔，2022年。[35] J. A. M. Cunha、R. Flynn、C. B´elanger、C. Callaghan、Y. Kim、X. Jia、Z. Chen 和 L. Beaulieu，近距离放射治疗的未来方向，载于《放射肿瘤学研讨会》，第30卷，第94–106页，Elsevier，2020年。[36] X. Jia、L. Ren 和 J. Cai，人工智能技术的临床实施需要可解释人工智能模型，《医学物理》，1–4（2020年）。

[37] K. Singh、S. Saxena、A. Khosla、M. McDermott、R. Kotecha 和 M. Ahluwalia，《脑转移管理进展》，《神经治疗学》**19**，1772–1781（2022年）。