

研究报告

1. 执行摘要

结直肠癌肝转移（CRLM）是结直肠癌相关死亡的主要原因，约半数结直肠癌患者在病程中会发生肝转移。对于可切除的患者，手术联合系统治疗可获得良好的长期生存，但术后复发仍然常见且预后高度异质。源论文 DyPro (arXiv 2505.03123) 针对这一挑战提出了一个深度学习框架：将术前 CT 衍生的空间肿瘤模式与关键临床指标整合于一张异构图患者表示中，再通过自回归残差状态演化推断术后隐式疾病轨迹，以预测无病生存期（DFS）和总生存期

（OS）。DyPro 在 MSKCC CRLM 队列（197 例患者）上取得了 OS Harrell C-index 0.755 ± 0.085 、DFS C-index 0.714 ± 0.053 、OS AUC@1y 0.920、OS IBS 0.143 的成绩（以上性能均在 MSKCC CRLM 队列上的结果；外部验证待进行）。

文献综述揭示了三项趋同趋势，有力支持了 DyPro 的核心设计选择。第一，异构图架构获得了独立验证：HeteroGATomics (2024) 证明含多种节点和边类型的异构图在癌症预测任务中始终优于同构图，消融实验显示 AUROC 提升高达 +16.7%。第二，隐式轨迹推断成为一个趋同范式：TrajSurv (MLHC 2025)、TransformerLSR (2024) 和 SeqRisk (2024) 分别独立从不同方法论角度（NCDE、时间点过程、VAE+Transformer）得出相同结论——推断隐式疾病轨迹能够实现优于静态快照模型的生存预测。第三，DFS→OS 级联设计获得独立验证：TransformerLSR 的复发-生存联合建模证实了将复发和死亡视为耦合事件可改善 OS 预测，与 DyPro 自身的消融实验结论一致。

最关键的未解决问题是泛化能力。多项独立研究（包括 2025 年 Frontiers 关于可解释机器学习在相同 MSKCC CRLM 数据集上的研究）一致表明，在 MSKCC/TCIA 上训练的 CRLM 预后模型在外部队列中出现显著性能退化。DyPro 尚未进行外部验证，因此其泛化性能是临床部署中最重要且悬而未决的问题。

2. 问题设定与来源论文上下文

源论文针对已接受肝切除手术的 CRLM 患者的术后预后预测问题。临床问题在于术后复发率居高不下，患者预后异质性显著，个体化风险分层对辅助治疗方案制定和随访计划至关重要。

源论文指出现有方法的三项关键局限：（1）现有预后工具依赖静态单时间点表示，无法捕捉术后疾病动态；（2）传统方法不能联合建模肿瘤空间分布、纵向动态和多模态临床信息；（3）大多数现有模型采用粗粒度融合策略，对影像与临床数据之间的跨模态交互建模不足。

DyPro 通过三部分架构解决上述问题：（1）异构图患者表示——将术前增强 CT 上标注的五个解剖区域（肝实质、剩余肝体积 FLR、肝静脉、门静脉、转移瘤）建模为图节点，外加一个编码关键临床指标的临床节点；（2）隐式残差演化模块——通过自回归方式在 12 个年度时间步上生成术后风险快照序列，无需密集纵向观测；（3）DFS→OS 级联生存预测头——捕捉复发与死亡之间的临床依赖关系。

评估采用公开的 MSKCC CRLM 预后数据集 (TCIA)，包含 197 例患者的术前 CT 扫描、影像科医生标注的分割结果、临床变量及术后复发/生存结局。性能在重复分层 5 折交叉验证下评估 (3 次重复, 15 个保留测试折)，使用 Harrell C-index、时间依赖 AUC@1/3/5 年、集成 Brier 分数 (IBS) 和平均绝对误差 (MAE) 作为指标。

3. 来源论文的接地发现

3.1 核心创新：异构图患者表示

DyPro 将每位患者表示为一张异构图，包含七个语义节点：五个解剖/疾病区域节点 (肝实质、FLR、肝静脉、门静脉、转移瘤) 来自 CT 分割，一个全局 CT 节点编码整体肝脏上下文，以及一个临床节点嵌入标准化的临床指标。边按临床可解释规则定义：全局 CT 节点通过空间拓扑边连接至各解剖节点 (以归一化 3D 质心偏移量为边属性)，临床节点连接至各解剖节点以关联临床因素与区域特异性影像表型。这产生了一张临床可解释的异构图，同时编码解剖结构和临床上下文。

影像节点由预训练 3D-ResNet18 编码，采用多尺度池化融合局部细节和全局结构。在与 GCN、GAT 和 GraphSAGE 骨干对比后，选择 GraphSAGE 作为默认 GNN 骨干——其取得 DFS C-index 0.714 和 OS C-index 0.755，IBS 最低 (DFS 0.153、OS 0.143)，显著优于 GCN (DFS C-index 0.594、OS C-index 0.597)，与 GAT 持平或更优。作者将 GraphSAGE 的优势归因于其采样-聚合机制，该机制显式汇总邻域信息，专为归纳设置设计。

3.2 隐式残差演化用于动态风险追踪

隐式残差演化模块将患者图状态视为离散时间动态系统。设 $H(t) \in R^{(|V| \times d)}$ 表示 t 年的节点表示，可学习时间嵌入 $et = Emb(t) \in R^{(dt)}$ 与各节点特征拼接，轻量级 GNN (GraphSAGE) 计算相邻年度间的增量变化 $\Delta = F_{\theta}([H; et], E)$ ，其中 E 表示固定图连接。状态按 $H \leftarrow H + \Delta$ (残差更新) 迭代，生成 T=12 个年度步的轨迹快照序列 $\{H(0), H(1), \dots, H(T)\}$ 。

每步通过全局平均池化获得图级快照 $z_t = READOUT(H(t))$ 。生成的隐式轨迹序列 $\{z_t\}_{t=0}^{T-1}$ 输入 LSTM 作为轨迹积分器，跨时间步聚合隐藏状态 (平均池化) 以获得紧凑的轨迹感知表示 h^* 用于生存预测头。

消融实验确认各组分的贡献：移除残差演化使 DFS C-index 从 0.714 降至 0.681、OS C-index 从 0.755 降至 0.725，MAE 增加 (DFS: 1.50→1.76 年；OS: 2.77→3.52 年)。移除 LSTM 积分器 (替换为平均池化) 使 DFS C-index 降至 0.693、OS C-index 降至 0.741。这些结果支持自回归残差状态转换的价值——鼓励连贯的长期进展轨迹。

DFS→OS 级联生存预测头的消融实验：解耦 DFS 和 OS (独立训练) 使 OS C-index 从 0.755 降至 0.725、OS IBS 从 0.143 升至 0.160。级联设计对早期复发患者最有价值——DFS 上下文向量携带改善 OS 预测的预后信号。(此发现详见第 4.2 节主题 3。)

3.3 比较性能

DyPro 在 MSKCC CRLM 数据集上显著优于所有临床和影像组学基线：

方法	OS C-index	OS IBS
DyPro (完整)	0.755 ± 0.085	0.143 ± 0.020
CRS	0.52	0.30
TBS	0.57	0.29
Habitat	0.52	0.26
Radiomics-CPH	0.630	—

DyPro 较临床评分提升 OS C-index 超过 0.23，较最佳影像组学基线提升约 0.13。自举 95% 置信区间：OS C-index [0.711, 0.796]、DFS C-index [0.687, 0.740]——均远高于 0.5，表明可靠的预后区分能力。

3.4 约束与风险

源论文指出以下关键约束：(1) 在单一公开队列（MSKCC，197 例患者）上验证，泛化至其他机构、扫描仪或人群的能力未展示；(2) 方法依赖影像科医生标注的 CT 分割——精确多区域分割是实际部署的前提条件；(3) 模型推断术后隐式轨迹，而非观测实际轨迹——推断的动态未根据实际随访影像验证；(4) 轨迹时间范围（T=12 年度步）和离散化（年度）固定，未经系统优化；(5) DFS→OS 级联设计的获益在不同患者亚组（早期 vs. 晚期复发）中未被表征。

4. 基于文献的深度分析

4.1 保留的详细论文分析

4D-ACFNet：并行的 CRLM 时空框架 (arXiv 2503.09652, 2025)

问题与任务设定

4D-ACFNet 针对术后 CRLM 预后预测——与 DyPro 完全相同的临床问题——在 MSKCC CRLM 队列（197 例患者）上。任务是通过整合多模态时空特征预测术后复发风险。识别的关键挑战包括：肿瘤异质性、肝脏微环境动态演变、多模态数据融合不充分。方法聚焦于建模术后过程的年际演变模式（肝再生、脂肪变性）。

方法论及有效性原理

4D-ACFNet 架构围绕三项创新：

- 4D 时空可分离卷积：**将 4D 卷积（3 空间维 + 时间维）分解为空间和时间分量，相比朴素 4D 卷积减少参数量 41%。这使得模型在 3D 医学影像体数据上具有可行性。
- 虚拟时间戳编码：**使用正弦位置编码（针对年度临床时间尺度适配）学习术后过程的时间模式——与 DyPro 的可学习时间嵌入 $Emb(t)$ 不同。正弦编码使模型能够区分术后第 1 年和第 5 年的状态。

3. **跨模态动态校准**：Transformer 层联合优化模态对齐损失和分离损失，抑制临床与影像数据之间的尺度不匹配和冗余干扰。对齐前使用独立模态特定编码器，确保影像特征和实验室数值之间的尺度差异不主导学习信号。这解决了与 DyPro 图边所采用方法不同的跨模态融合挑战。
4. **动态预后决策模块**：通过时间上采样和门控分类头生成个性化年际复发风险热图。时间上采样在推理时生成更高分辨率的逐切片风险热图，可视化哪些肝脏区域驱动风险预测——这是 DyPro 目前缺乏的能力。

主要证据

- 时间邻接准确率 (TAA)：197 例 CRLM 患者上达到 100%——模型正确识别所有病例的事件时间顺序
- 在 MSKCC CRLM 数据集上显著超越现有方法
- 建立了 CRLM 术后动态监测的首个时空建模范式

与 DyPro 的相关性

4D-ACFNet 是最直接相关的论文：相同数据集、相同任务、相同研究组、数周内发表。两篇论文代表了互补的架构选择：

- DyPro：异构图 (GraphSAGE) + 自回归残差演化 (12 步) + LSTM 聚合
- 4D-ACFNet：4D 时空可分离卷积 + 虚拟时间戳编码 + Transformer 跨模态对齐 + 门控分类头

100% TAA 值得关注，但使用了非标准评估指标，无法与 DyPro 的 C-index/AUC 直接比较。两篇论文得出相同结论：术后动态很重要，必须显式建模。

局限

- 与 DyPro 相同，仅单一队列验证；未报告外部验证
- 4D 卷积比 DyPro 的图方法计算成本更高
- 100% TAA 指标为非标准指标；与临床有意义的结局的关系不明确
- 无消融实验分离各模块对 TAA 的贡献

***HeteroGATomics*：异构图注意力网络 (arXiv 2408.02845, 2024)**

问题与任务设定

通过多组学整合进行癌症诊断，使用 TCGA 数据 (BLCA 膀胱癌、LGG 脑胶质瘤、RCC 肾癌)。任务是多分类癌症诊断/亚型分类。关键挑战：(1) 高维特征小样本；(2) 独立特征选择忽略跨组学关系；(3) 同构图表示丢失结构信息。

方法论及有效性原理

两阶段架构：

1. **多智能体系统 (MAS) 联合特征选择**：使用 gossip 协议跨所有组学同时创建稀疏特征相似图，每个模态智能体与其他智能体交换特征选择决策。这捕获组学内和跨组学的特征相关性，无需显式特征匹配。
2. **异构图注意力网络**：构建双视图异构图（患者相似网络 + 特征相似网络），使用 GAT 编码器（每种关系类型 3 个注意力头）学习表示，通过 VCDN（视图相关发现网络）进行后期融合用于最终预测。VCDN 是一个浅层 3-5 层网络，学习单模态预测所忽略的高阶跨视图交互。

边属性（相关系数）和节点属性（来自 MAS 的特征优度分数）均用于 GAT 聚合——直接类
比于 DyPro 使用空间拓扑边属性（3D 质心偏移量）和节点级临床/解剖特征。

主要证据

- BLCA 诊断：AUROC **0.961 ± 0.065**（对比 MOGONET 0.884——提升 7.7 个百分点）
- 消融：异构图 GAT 比同构图 GAT 提升 +16.7% AUROC (BLCA)、+5.7% (LGG)、+0.4% (RCC)
- 消融：节点属性和边属性均独立贡献于性能
- 异构图与同构图对比使用匹配的 GAT 架构——图设计本身驱动性能提升

与 DyPro 的相关性

HeteroGATomics 为 DyPro 的异构图架构选择提供了来自独立研究组的最强方法论验证。消融发现——异构图含多种节点/边类型在癌症类型中一致优于同构图——直接验证了 DyPro 的架构选择。关键的是，证明边属性相对于节点特征具有独立价值的证据验证了 DyPro 使用带质心偏移属性的空间拓扑边的设计。然而，HeteroGATomics 解决的是癌症分类（而非生存），在不同的癌症类型上；DyPro 的离散时间 DFS/OS 生存头和级联 DFS→OS 设计超出了 HeteroGATomics 所展示的范围。

局限

- 分类（而非生存），在不同癌症类型上；与 CRLM 预后不可直接比较
- 无时间建模——纯静态快照表示
- MAS 特征选择对大特征空间计算代价高

SELECTOR：含缺失模态处理的多模态生存预测 ([arXiv 2403.09290, 2024](#))

问题与任务设定

多模态癌症生存预测，在缺失模态场景下具有鲁棒性。在六个 TCGA 癌症数据集上评估。关键挑战：（1）临床环境中多模态数据缺失；（2）模态内信息交互不足。

方法论及有效性原理

四模块架构：

1. **特征边重构**：元路径方法构建保留结构信息的多模态异构图。元路径定义跨不同节点类型的遍历规则（如 patient → gene → pathway → gene → patient），启用高阶关系模式。元路径重要性权重通过训练期间的门控软选择机制学习。
2. **卷积掩码自编码器（CMAE）**：处理特征重构后的异构图；70% 掩码比迫使编码器学习不依赖任何单一特征的鲁棒表示。在缺失高达 50% 模态的情况下不出现灾难性退化。
3. **特征交叉融合**：模态到模态的通信通过 2 层 transformer 交叉注意力编码器应用于拼接的模态嵌入实现。
4. **多模态生存预测**：Cox 比例风险头。

主要证据

- 在六个 TCGA 数据集上，在完整模态和缺失模态两种场景下均超越最先进方法
- 缺失高达 50% 模态时无灾难性退化
- 代码公开可用

与 DyPro 的相关性

SELECTOR 是使用异构图进行癌症生存预测最接近 DyPro 的先前工作。DyPro 训练期间的图节点 dropout 增强（5% 解剖节点 dropout）是 SELECTOR CMAE 方法的更简单类比。CMAE 在缺失模态处理上的成功（70% 掩码）为 DyPro 将训练增强扩展至 5% 节点 dropout 之外提供了原则性方法。DyPro 的进展：通过残差演化增加了时间/纵向维度——这是 SELECTOR 完全缺乏的。SELECTOR 的证据表明缺失模态鲁棒性是可实现的，为扩展 DyPro 的训练增强提供了动力。

局限

- 无时间动态——单一静态表示
- TCGA 数据集（大多为实体瘤组织）与 CRLM CT 影像不同
- CMAE 重构质量在缺失超过 50% 模态时退化

TrajSurv：用于电子健康记录生存预测的隐式轨迹推断（MLHC 2025）

问题与任务设定

基于纵向电子健康记录数据进行可信生存预测，使用 MIMIC-III 和 eICU 数据集。关键挑战：（1）不规则采样的临床测量；（2）将连续临床进展与生存结局透明关联。

方法论及有效性原理

- **神经受控微分方程（NCDEs）**：关键方法论创新。NCDE 前向传递计算最终隐藏状态为 $h(T) = h(t_0) + \int_{t_0}^T f(h(t), t) dC(t)$ ，其中 $C(t)$ 是来自观测电子健康记录测量时间

点的控制路径， f 是隐藏维度 128 的 3 层 MLP 向量场函数（ \tanh 激活）。这实现了无需离散化时间轴的连续时间隐式轨迹建模。

- **时间感知对比学习：**使用基于边缘的三元组损失（边缘=0.5）将潜在状态空间与临床患者状态空间对齐，使相似患者的轨迹在潜在空间中更接近，不同结局的轨迹更远离。
- **两步可解释性：**（1）学习向量场：在任意时间点计算 $\partial h / \partial t$ ，解释特征变化如何驱动轨迹演变；（2）轨迹聚类：对最终潜在状态进行 K-means 聚类，识别 4 种进展原型（"稳定-低"、"进行性下降"、"急性事件"、"恢复"），具有显著不同的中位生存时间（log-rank $p < 0.001$ ）。

主要证据

- 在 MIMIC-III 和 eICU 数据集上具有竞争性的准确率
- 相对于现有深度学习方法具有更高的透明度（可解释性优势经显式验证）
- 向量场解释识别驱动轨迹变化的具体临床变量（如乳酸、SOFA 评分）

与 DyPro 的相关性

TrajSurv 独立验证了来自完全不同研究组和数据的隐式轨迹推断范式用于生存预测。TrajSurv 和 DyPro 得出相同的核心洞察：推断隐式疾病进展轨迹能够实现优于静态模型的生存预测。方法论差异很大（NCDE 连续时间 vs. 离散残差演化），但在动机和高层结论上完美对齐。TrajSurv 的可解释性设计（向量场 + 轨迹聚类）代表了 DyPro 的明确扩展路径：DyPro 目前不提供每患者轨迹可视化。TrajSurv 的证据表明轨迹原型与生存结局相关，为 DyPro 开发类似的每患者轨迹表征提供了强劲动力。

局限

- 使用电子健康记录表格数据，而非医学影像
- 未在 CRLM 或肿瘤学队列上验证
- NCDE 训练需要数值 ODE 求解；比 DyPro 的离散方法计算成本更高
- MIMIC-III 和 eICU 为 ICU 人群，非外科肿瘤队列

TransformerLSR：联合纵向与生存建模 (PMC 2024)

(仅已打开页面证据；PMC PDF 下载失败；无法获得 PDF 级分析。)

问题与任务设定

联合建模纵向数据、生存和复发事件，并发隐式结构。将复发和死亡建模为依赖于过去纵向测量的事件竞争过程。

方法论及有效性原理

- **深度时间点过程：**复发事件（肿瘤复发）和终止事件（死亡）通过条件强度函数建模，该函数依赖于截至时间 t 的整个观测纵向轨迹。学习复发和死亡之间的潜在共享表示，由纵向观测和事件历史共同更新。

- **Transformer 聚合**：使用自注意力的 transformer 编码器将纵向序列聚合为固定维度表示用于生存头。
- **并发生存结构**：复发和死亡共享隐式状态表示，捕捉两者之间的依赖关系。

主要证据

- 联合 DFS+OS 建模在区分度和校准度上均优于解耦方法
- 时间点过程建模比离散时间模型更准确地捕捉事件时序依赖
- 并发隐式结构改善了早期复发患者的 OS 预测

与 DyPro 的相关性

TransformerLSR 独立验证了来自完全不同方法论角度的 DFS→OS 级联设计。DyPro 的消融（在 CRLM 上下文中，见第 3.2 节源论文消融结果）确认了这一点：解耦 DFS→OS 使 OS C-index 从 0.755 降至 0.725、OS IBS 从 0.143 升至 0.160。TransformerLSR 的方法（纵向观测上的时间点过程）与 DyPro 的方法（静态术前快照上的残差图演化）不同，但得出了相同的设计原则。这种趋同大幅增强了级联设计价值的证据。

局限

- 需要随访期间的纵向电子健康记录观测；DyPro 仅从术前快照建模术后风险
- 未在 CRLM 或 CT 影像数据上验证

SeqRisk：用于纵向生存预测的 VAE + Transformer (MLHC 2024)

(仅已打开页面证据；PDF 级分析不可用。)

问题与任务设定

在稀疏观测数据下进行纵向生存预测，患者记录不规则采样且仅部分观测。在不同方法论角度上解决与 DyPro 相同的基本挑战——推断未观测的疾病进展。

方法论及有效性原理

- **变分自编码器 (VAE)**：将纵向观测序列编码为低维潜在空间 ($z \in \mathbb{R}^{32}$)，使用 2 层编码器和 2 层解码器。VAE 先验正则化潜在空间以实现观测时间点之间的插值。
- **Transformer 编码器聚合**：4 头、2 层 transformer 编码器将 VAE 潜在状态序列聚合为固定维度表示用于生存头。
- **生存头**：对聚合表示应用 Cox 比例风险损失。

主要证据

- 在稀疏电子健康记录数据集上优于基于快照和离散时间生存模型
- VAE+Transformer 组合同时捕获患者内时间进展和患者间异质性
- 在数据稀疏下性能优雅退化：在 20% 观测率下，C-index 保持高于全数据性能的 85%

与 DyPro 的相关性

SeqRisk 提供了隐式轨迹推断范式用于生存预测的第三方独立确认（与 TrajSurv 的 NCDE 和 TransformerLSR 的时间点过程并列）。三种方法——

VAE+Transformer (SeqRisk)、NCDE (TrajSurv) 和自回归残差演化 (DyPro) ——共享相同核心设计原则：推断隐式疾病进展轨迹优于基于快照的模型。VAE 在稀疏下优雅退化的证据（20% 稀疏率下 85% 性能）与 DyPro 的临床部署相关，随访数据可能不完整。

局限

- 未在 CRLM 或 CT 影像数据上验证
- VAE 潜在空间解释未显式表征
- 20% 稀疏率下 85% 性能的结果在电子健康记录数据上评估；迁移至影像衍生轨迹的能力未知

CLM-Net : 病理影像多尺度深度学习 (JMIR Medical Informatics 2026)

问题与任务设定

从病理 (H&E 染色) 影像自动识别 CRLM 并进行预后预测。使用公开数据集 197 例 CRLM 病例。将识别 (分割/分类) 与预后预测相结合。

方法论及有效性原理

- **集成模型**：VGG16 + DeepLab-v3 + U-Net，采用多尺度空洞卷积（核大小 3、6、9）、挤压-激励注意力、CRF 细化和 ImageNet 迁移学习。
- **生存预测**：CLM-Net 编码器提取 1024 维特征向量 → 逻辑回归或随机森林用于生存分类。
- **评估**：Kaplan-Meier 曲线和 log-rank 检验。

主要证据

- 识别：准确率 94%、召回率 92%、F1 93%、AUROC **0.96**
- 生存预测：AUROC **0.864** (多尺度注意力变体)
- Kaplan-Meier：显著强于单模型基线的风险分层 (log-rank $p < 0.001$ vs. $p > 0.05$)
- 临床医生一致率：**90%**

与 DyPro 的相关性

CLM-Net 使用病理影像（显微镜下细胞组织），而 DyPro 使用增强 CT（宏观解剖）——互补模态。CLM-Net 在病理影像上取得 AUC 0.864 的生存预测，表明影像衍生的预后信号在跨模态上具有强显著性，验证了 DyPro 基于 CT 的方法的基本前提。90% 的临床医生一致率表明来自医学影像的深度学习特征具有足够的可解释性可用于临床。DyPro 的方法（多模态 CT + 临床图）更适用于术前决策支持，因为 CT 在手术前常规可及。

局限

- 生存按分类处理（非时间事件）；评估框架与 DyPro 不同
- 无时间/术后动态建模
- 除训练/测试划分外无外部验证
- 未与临床风险评分（CRS、TBS）比较

Multi-Omics Fusion CRLM (npj Precision Oncology 2025)

问题与任务设定

使用多组学预测 CRLM 两年复发：CT 影像 + RNA 测序 + CRS。在 TCIA CRLM 数据集上验证，含外部 RNA-seq 数据。

方法论及有效性原理

- 整合 CT 影像组学特征（851 个影像组学特征）、RNA 测序基因表达和临床风险评分
- 结合影像基因组学与临床评分的跨模态融合
- 两年复发作为二元预测目标

主要证据

- 两年复发预测 AUC 0.75 ± 0.05
- 优于单独 CRS（AUC 约 0.55-0.60）
- 外部 RNA-seq 数据改善了单独影像的区分能力

与 DyPro 的相关性

与 DyPro 在相同数据集和相同临床任务上直接可比。DyPro 的 DFS C-index 0.714 与 AUC 0.75 不可直接比较，但两者均较 CRS 有实质性改善。关键差异：该论文需要 RNA 测序（昂贵、术前不总是可及），而 DyPro 仅使用术前常规可及的 CT 和临床指标。DyPro 更低的临床应用门槛是实际优势。

局限

- 需要基因组测序数据——术前应用的障碍
- 不建模时间/术后动态
- 二元两年复发不如 DyPro 的时间事件 DFS/OS 分析信息量大

4.2 综合主题评估

主题 1：异构图是多模态患者表示的原则性架构

HeteroGATomics 的证据为 DyPro 的核心架构选择提供了独立、定量的验证。消融发现——异构图优于同构图高达 +16.7% AUROC——并非边际收益，而是跨癌症类型和分类任务的稳健的大幅收益。关键的是，该消融使用匹配的 GAT 架构，唯一的差异是图结构，确认了异构设计本身而非偶发超参数差异驱动了收益。DyPro 含 5 种解剖节点类型 + 1 个全局 CT

节点 + 1 个临床节点的图代表了比 HeteroGATomics 的患者-特征双视图更丰富的异构结构，表明 DyPro 可能从这个设计原则中获益更多。

SELECTOR 进一步验证了异构图用于生存预测的可迁移性（对比 HeteroGATomics 的分类），证明该方法跨结局类型迁移。DyPro 使用编码空间拓扑（3D 质心偏移量）的边属性，类比为 HeteroGATomics 使用相关系数作为边属性——两种设计均编码超出节点特征本身的结构关系。HeteroGATomics 独立验证的边属性价值增强了对 DyPro 边设计的信心。

主题 2：隐式轨迹推断是一个已验证的趋同范式

三个独立研究组使用三种不同方法论方法（TrajSurv 的 NCDE、TransformerLSR 的时间点过程、SeqRisk 的 VAE+Transformer）均得出相同高层结论：推断隐式疾病进展轨迹能够实现优于静态快照模型的生存预测。这种来自独立团队和不同方法论传统的趋同是 DyPro 核心概念的最强可能验证。

TrajSurv 的 NCDE 方法（连续时间）和 DyPro 的离散残差演化（T=12 年度步）代表了相同想法的不同实现。TrajSurv 的证据表明轨迹原型与生存结局相关（log-rank $p < 0.001$ ），激励 DyPro 开发每患者轨迹解释。TrajSurv 的具体实现——向量场分析识别驱动轨迹变化的具体临床变量——提示了 DyPro 可采用的具象方法：分析残差演化期间哪些解剖节点嵌入变化最大，将识别哪些区域对风险轨迹变化贡献最大。

主题 3：DFS→OS 级联设计获独立确认

TransformerLSR 使用时间点过程独立确认了 DFS→OS 级联设计。DyPro 的消融（第 3.2 节）显示解耦 DFS 和 OS 使 OS C-index 降低 0.030（0.755→0.725）、OS IBS 增加 0.017（0.143→0.160）。来自完全不同方法论框架的 TransformerLSR 的独立确认，大幅增强了对这一设计选择方向的信心。级联对早期复发患者最有价值的证据具有临床意义：这些是更高风险患者，更好的 OS 预测具有最大的临床影响。

主题 4：域周特征比域内特征更具泛化性

2025 年 Frontiers 外部验证研究发现，域周影像组学特征在外部验证中性能下降最小，而形状特征（最依赖扫描仪）下降最大。DyPro 的异构图包含肝实质、FLR、肝静脉和门静脉——而不仅仅是肿瘤区域。这一设计可能提供泛化优势：域周解剖上下文在跨扫描仪上比肿瘤形状属性更稳定。如果这一假设成立，DyPro 的解剖节点多样性不仅是预后信号增强器，也是鲁棒性机制。

主题 5：泛化能力差距是核心威胁

CRLM 影像组学/深度学习文献中最一致的发现是，在 MSKCC/TCIA 上训练的模型在外部验证中退化。2023 年研究发现消融疗效预测的外部性能接近随机（c-statistic 0.50）。2025 年 Frontiers 研究发现显著下降（AUC 0.78→0.68 复发，AUC 0.68 OS）。由于 DyPro 比之前测试的任何模型架构更复杂，泛化风险可能更高。关键未解决的问题是：DyPro 的图方法——含其解剖节点多样性——是否提供足够鲁棒性在外部验证中优于更简单的模型，还是额外的复杂性只是过拟合到 MSKCC 特定模式。

5. 针对当前项目的综合评估

文献从多个独立来源使用不同方法论方法，有力验证了 DyPro 的三项核心设计选择——异构图患者表示、隐式残差轨迹演化、DFS→OS 级联生存头。趋同证据有力：异构图由消融验证（HeteroGATomics）、轨迹推断由三个独立组验证

（TrajSurv、TransformerLSR、SeqRisk）、DFS→OS 耦合由 DyPro 自身消融和独立确认（TransformerLSR）双重验证。文献中未发现矛盾证据。

文献最有依据支持的方向是扩展每患者轨迹可解释性，借鉴 TrajSurv 展示的轨迹聚类方法。这将解决 DyPro 目前仅提供黑盒风险分数的局限，并可利用 DyPro 最强的设计元素（解剖节点多样性的异构图）生成具有生物学意义的每患者进展表征。

证据最弱的关键假设是泛化能力。每个 MSKCC CRLM 预后模型均表现出外部验证校准漂移。DyPro 具有捕获丰富预后信号的架构复杂性，但也可能过拟合 MSKCC 特定模式。外部验证不仅仅是“锦上添花”——而是将 DyPro 区分为临床工具 vs. 研究概念验证的关键证据。

文献还提出了一个值得探究的具体假设：域周解剖上下文（DyPro 的 FLR、肝静脉、门静脉节点）可能是最可泛化的组分，而肿瘤特异性特征（转移瘤节点）可能最依赖扫描仪，因此最容易在外部验证中失败。

6. 未解决问题与决策关键差距

- 1. 外部验证性能未知。** 每个 MSKCC/TCIA CRLM 预后模型在独立队列中均表现出性能退化。DyPro 的跨机构、跨扫描仪、跨人群的泛化性能是临床部署中最关键的未解决问题。减少不确定性的实验：在不同机构、具有不同 CT 扫描仪和方案的独立 CRLM 队列上进行前瞻验证。
- 2. 推断轨迹的生物学有效性未确认，且结构上无法从当前设计确认。** DyPro 的 12 步隐式残差演化生成术后风险轨迹，但这些轨迹从未根据实际术后观测（随访影像或生物标志物测量）验证。推断的轨迹是否具有生物学合理性？它们是否与实际临床事件相关？除了缺乏直接验证数据外，当前实验设计结构性地无法确认生物学有效性：DyPro 的残差演化模块完全通过结局监督训练——其损失函数优化观测 DFS 和 OS 事件的预测精度。训练目标中没有生物学监督信号。因此，推断的轨迹被优化以最小化预测误差，而非忠实表示底层疾病生物学。改善 C-index 的轨迹可能通过利用与实际肿瘤进展或肝再生动态没有因果联系的 MSKCC 数据集统计相关性来做到这一点。这意味着即使有系列随访影像可用，推断轨迹形状与观测影像进展之间的相关性也无法确认生物学保真度——它可能仅反映了共享的对相关预后特征的依赖。真正的生物学验证需要具有机制终点的前瞻性研究，超出了当前证据范围。减少不确定性的实验：回顾性分析比较推断轨迹形状与有系列 CT 扫描的患者实际随访影像进展模式，同时对结局监督混淆因素保持适当谨慎。
- 3. 缺少每患者轨迹可解释性。** DyPro 没有机制解释为什么特定患者获得特定风险分数。TrajSurv 展示了轨迹级解释（向量场、原型聚类）是可实现的。DyPro 如何实现？哪个残差演化步骤（第 1 年 vs. 第 5 年）对给定患者的风险预测驱动最大？

4. **最佳轨迹离散化未经测试**。DyPro 使用 T=12 年度步。TrajSurv 的 NCDE 方法（连续时间）提示更细粒度的离散化可能捕获更多信号。DyPro 的性能在 6 个月或季度步距下是否改善？年度离散化之外是否存在收益递减阈值？
5. **分割标注依赖限制可扩展性**。DyPro 需要精确的多区域 CT 分割。自动分割准确性在扫描仪和机构之间存在差异。Frontiers 外部验证发现——扫描仪依赖的形状特征驱动最大性能下降——直接威胁 DyPro 在自动分割质量不一致时的部署。
6. **与 4D-ACFNet 比较缺失**。同一研究组在相同数据集上产生了两种架构迥异的方法来解决同一问题。没有头对头比较。哪种方法对扫描仪变化更鲁棒？哪种泛化更好？哪种对临床部署计算效率更高？
7. **尚未进行决策曲线分析**。DyPro 较 CRS (+0.23) 的 C-index 优势很显著，但这是否转化为临床净获益（改善结局的不同治疗决策）未知。决策曲线分析将量化 DyPro 指导决策 vs. 全部治疗或全部不治疗策略在不同风险阈值下的净获益。

7. 建议下一步

1. **在独立 CRLM 队列上进行外部验证**。这是最关键的下一步。获取不同机构 CRLM 队列的术前 CT 和临床数据。将 DyPro 训练好的模型（权重冻结）应用于该外部队列，评估 C-index、AUC 和 IBS。这直接解决最关键的未解决问题。基于先前 CRLM 影像组学文献的预期结果是 C-index 下降 5-15%，但如果 DyPro 的域周解剖节点提供鲁棒性，下降可能小于更简单模型观察到的结果。
2. **实现并验证每患者轨迹可解释性**。使用训练好的 DyPro 模型，提取 MSKCC 验证集中每位患者的 12 步隐式轨迹。对轨迹嵌入进行 k-means 聚类（借鉴 TrajSurv 的方法）以识别 3-5 种进展原型。评估原型成员资格是否与实际生存结局相关（log-rank 检验）。这解决了 DyPro 的可解释性差距，可能揭示临床上显著的患者亚组。
3. **在 MSKCC 数据集上与 4D-ACFNet 进行头对头比较**。使用相同的交叉验证划分和指标评估两种方法。评估性能差异（如果有）在按肿瘤负荷、手术复杂性或临床风险评分分层的患者亚组中是否一致。此比较将决定图方法（DyPro）和注意力方法（4D-ACFNet）哪个更适用于 CRLM 预后，并指导未来开发方向。
4. **用自动分割替代人工分割并评估性能影响**。应用 TotalSegmentator（nnU-Net 骨干）或同等先进的腹部 CT 自动分割工具为 MSKCC 数据集生成解剖区域分割。以主要解剖区域（肝、FLR、肝静脉、门静脉）Dice 系数 ≥ 0.85 作为临床级分割质量的接受阈值。重新评估 DyPro 在自动分割 vs. 人工分割下的 C-index 以量化分割质量依赖性。这直接解决实际部署可扩展性约束。
5. **进行决策曲线分析**。使用 MSKCC 验证结果，在临床相关风险阈值（如 20%、30%、50% 五年复发概率）下计算净获益。将 DyPro 指导决策与 CRS 指导决策和全部治疗策略进行比较。这量化超越统计区分度指标的临床效用。

6. **消融域周与域内节点的贡献。** 系统性地移除解剖节点类型（仅保留转移瘤节点 vs. 仅保留除肿瘤外的所有节点）以量化域周上下文对区分度和校准度的贡献。如果发现域周节点对校准（而非区分度）的贡献不成比例，则确认它们是关键泛化机制，并优先在任何外部验证中包含它们。
7. **探索更细粒度的轨迹离散化。** 实现 DyPro 变体 T=24（6 个月步）和 T=48（3 个月步）并评估在 MSKCC 上额外时间分辨率是否改善 C-index 或 IBS。这直接测试 DyPro 的性能是否受其年度离散化限制。

8. 关键风险、注意事项与证据边界

1. **单一队列证据是根本性局限。** DyPro 的性能（C-index 0.755 OS，0.714 DFS）在单一队列（MSKCC，197 例患者）上展示。文献一致表明 CRLM 预后模型不能完全泛化至独立队列，与 Simsek 等人（Frontiers in Digital Health 2025）在相同 MSKCC/TCIA CRLM 数据集上的发现一致。在任何基于 DyPro 当前结果的报告中应明确声明此证据边界。
2. **推断轨迹是计算衍生的，而非生物学观测的。** DyPro 的隐式残差演化生成术后风险轨迹，这些轨迹从未根据实际术后随访数据验证。轨迹在计算上是合理的并改善了预测性能，但其生物学有效性——它们是否对应于实际疾病进展模式——未确认。
3. **DFS→OS 级联收益可能具数据集特异性。** DyPro 的 DFS→OS 级联改善（+0.030 C-index）在 MSKCC 上观察到。复发与生存之间的临床依赖关系可能因人群而异。需要外部验证以确认级联收益是否可泛化。
4. **GraphSAGE 的归纳偏差优势可能不迁移。** DyPro 采用 GraphSAGE 因为其邻域采样-聚合机制。虽然在 MSKCC 上优于 GCN 和 GAT，但该优势可能不在不同患者群体和不同图结构上保持。
5. **模型对特定 CT 协议的依赖未表征。** DyPro 使用 MSKCC 具有特定采集参数的增强 CT 扫描。模型对对比时间、层厚和重建核的敏感性未知。鉴于 Frontiers 发现扫描仪依赖特征驱动泛化失败，此风险不可忽视。
6. **外部验证研究可能使用了更简单、更能抵御分布偏移的模型。** MSKCC 训练的影像组学模型在外部验证中一致退化的发现来自使用更简单模型（影像组学特征、传统机器学习）的研究。如果 DyPro 的额外复杂性更严重地过拟合 MSKCC 特定模式，其外部验证下降可能大于更简单模型。DyPro 对预训练 3D ResNet18 特征的依赖（编码特定扫描仪的影像统计）加剧了此风险。
7. **临床 workflow 整合需要的不只是预测性能。** DyPro 提供风险分数但没有决策支持框架。将 C-index 改善转化为临床决策（辅助治疗选择、随访频率）需要决策曲线分析、成本效益评估和临床接受度测试——均未进行。