

## Medical Imaging-Pathology-Genomic Fusion and Its Applications in Clinical Diagnosis and Treatment

DUAN Jingxian<sup>1</sup>, ZHAO Yuanshen<sup>1</sup>, KANG Liangyuqi<sup>1</sup>, LIANG Dong<sup>1,2</sup>, ZHENG Hairong<sup>1,2,3</sup>, LI Zhicheng<sup>1,2\*</sup>

(1. Institute of Biomedical and Health Engineering, Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences, Shenzhen 518055, China; 2. State Key Laboratory of Medical Imaging Science and Technology Systems, Shenzhen 518055, China; 3. Institute of Brain Machine Interface, Nanjing University, Nanjing 210023, China)

**Abstract:** Medical imaging, pathology, and genomics respectively provide information on tumor spatial-morphological phenotypes, histopathological architecture, and molecular mechanisms. Single-modal approaches are constrained by scale discrepancies, sampling biases, and cross-center domain shifts, limiting their support for clinical decision-making. For precision oncology, imaging-pathology-genomics integration aims at semantical alignment and consistency validation among macroscopic imaging, microscopic histological, and mechanistic molecular evidence. This review systematically examines the field via fusion methodologies and clinical applications. We discuss the clinical advantages of multimodal integration, summarize key fusion paradigms, and emphasize its clinical necessity. In applications, focusing on completing the clinical evidence chain, we summarize its advantages in differential diagnosis, molecular subtyping, surgical planning, treatment response stratification, and systematic decision output, highlighting how integration turns predictive results into verifiable, actionable clinical decisions via cross-modal validation. Finally, we discuss emerging trends: spatial omics with multi-region sampling, longitudinal tumor evolution modeling, multimodal foundation models, and multi-center collaborative validation. We propose clinical translation recommendations and a utility evaluation system, offering a roadmap for next-generation intelligent multimodal systems in precision oncology.

### Highlights

1. Imaging-pathology-genomics integration completes the clinical evidence chain by fusing macroscopic, microscopic, and molecular evidence. It transforms fragmented observations into cohesive tumor biology understanding and enables verifiable, actionable decisions, beyond simple feature aggregation.
2. Fusion paradigms integrate into oncology workflows: differential diagnosis, molecular subtyping, surgical planning, treatment response stratification, and standardized reporting integrating multimodal predictions into decisions.
3. Emerging advances shape next-generation systems: spatial omics with multi-region sampling, longitudinal modeling, multimodal foundation models, and multi-center validation. Combined with translation guidelines and evaluation systems, they provide a deployment roadmap.

**Key words:** multimodal fusion; radiopathomics; radiogenomics; precision oncology

---

**Foundation items:** Shenzhen Medical Research Fund (No.A2303008); National Natural Science Foundation of China (Nos. T2525037, 62201557, 62571523); Guangdong Special Support Plan (No.2024TX08A213); Shenzhen Science and Technology Program (Nos.JCYJ20250604183020027, JCYJ20241202125014018); Strategic Priority Research Program of the Chinese Academy of Sciences (No.XDB0930302).

**Received:** 2026-01-09; **Revised:** 2026-03-11

\*Corresponding author, E-mail: zc.li@siat.ac.cn.

# 医学影像-病理-基因融合的智能分析和诊疗应用

段静娴<sup>1</sup>, 赵源深<sup>1</sup>, 康梁钰淇<sup>1</sup>, 梁栋<sup>1,2</sup>, 郑海荣<sup>1,2,3</sup>, 李志成<sup>1,2</sup>

(1. 中国科学院深圳先进技术研究院生物医学与健康工程研究所, 深圳 518055; 2. 医学成像科学与技术系统国家重点实验室, 深圳 518055; 3. 南京大学脑机接口研究院, 南京 210023)

**摘要:** 影像、病理与基因组学分别提供肿瘤的空间形态表型、组织学结构与分子机制信息。单一模态往往受限于尺度割裂、取样偏倚与跨中心域偏移, 难以支撑临床决策。针对肿瘤精准诊疗的临床需求, 影像-病理-基因融合的核心目标并非简单叠加特征或替代检测, 而是实现空间证据、组织学证据与机制证据的同域对齐与一致性校验。本文围绕融合方法学与临床应用两条主线展开综述: 在方法学部分, 系统讨论多模态融合的临床优势, 归纳关键融合范式, 强调影像-病理-基因融合在临床诊疗应用中的必要性; 在应用部分, 以多模态融合补全证据链为核心, 总结其在鉴别诊断、分子分型、手术规划、疗效分层与系统化输出中的独特优势, 突出融合如何将预测结果转化为可验证、可执行的临床决策。最后, 讨论空间组学与多区域采样、纵向演化建模、多模态基础模型与多中心协作验证等关键发展趋势, 并提出面向临床转化的建议要点与临床效用评价体系, 为下一代肿瘤多模态智能诊疗系统的构建与落地提供建议路线图。

**关键词:** 多模态融合; 影像病理组学; 影像基因组学; 精准肿瘤学

**中图分类号:** R3 **文献标志码:** A

**引用格式:** 段静娴, 赵源深, 康梁钰淇, 等. 医学影像-病理-基因融合的智能分析和诊疗应用[J]. 数据采集与处理, 2026, 41(2): 416-438. DUAN Jingxian, ZHAO Yuanshen, KANG Liangyuqi, et al. Medical imaging-pathology-genomic fusion and its applications in clinical diagnosis and treatment[J]. Journal of Data Acquisition and Processing, 2026, 41(2): 416-438.

## 引言

肿瘤的产生机制极其复杂, 通常认为是由遗传、化学或者物理等因素引起的基因突变, 所导致细胞增殖、分化和凋亡等调控机制失常而产生<sup>[1]</sup>。恶性肿瘤是严重威胁人类生命健康的难治疾病<sup>[2]</sup>。当前, 肿瘤治疗的核心方向是个性化的精准诊疗, 即基于肿瘤的个体差异, 实现“千人千策”的全流程治疗与随访管理, 最终改善患者预后、提升生存质量<sup>[3]</sup>。然而, 肿瘤异质性对其精准诊疗造成了巨大挑战。从肿瘤的微环境角度来看, 肿瘤内部存在多区域、多克隆及多微环境生态位的显著差异, 不同区域的肿瘤细胞在基因型、表型及所处微环境上均有所不同, 导致单次穿刺活检或病理切片仅能获取局部肿瘤信息, 无法全面代表肿瘤的全局特征, 易造成诊断偏差与治疗方案误判, 使患者难以获得良好的治疗效果<sup>[4-6]</sup>。与此同时, 肿瘤的动态演化特性使得其在治疗过程中会发生克隆替代现象, 即优势克隆体得以存活并增殖, 使新生成的肿瘤细胞展现出耐药性, 从而导致治疗效果逐渐衰减甚至失效<sup>[7-9]</sup>。这两大难

**基金项目:** 深圳市医学研究专项资金(A2303008); 国家自然科学基金(T2525037, 62201557, 62571523); 广东省“特支计划”(2024TX08A213); 深圳市科技计划项目(JCYJ20250604183020027, JCYJ20241202125014018); 中国科学院战略性先导科技专项(B类)(XDB0930302)。

**收稿日期:** 2026-01-09; **修订日期:** 2026-03-11

题相互关联,本质上均指向无法全面、精准、实时捕捉肿瘤多维度特征这一核心痛点,而破解该痛点的关键在于是否能够更全面的解析肿瘤。

近年来,随着人工智能(AI)技术的快速发展,有望改变生物医学研究和临床实践,对肿瘤的影像、病理和基因数据的智能分析研究带了新的机遇<sup>[10-11]</sup>。受早期的算力与算法限制,肿瘤智能诊疗主要利用单一模态数据。基于单模态的分析方法由于缺乏对肿瘤的全局认知,导致模型容易过拟合、泛化能力差、模型的可解释性差,以及对临床指导意义有限等<sup>[12-15]</sup>。不同单模态数据都蕴含其独特的技术优势和明显不足。影像技术可以做到无创定位全身病灶,可以轻松获取肿瘤的宏观分布与动态变化,但对肿瘤的发生机制及细胞组成的判断多为间接的经验判断,无法直接判断肿瘤的分型<sup>[16-17]</sup>;病理检查是肿瘤诊断的绝对标准,能够呈现肿瘤细胞形态、微环境结构,但由于存在跨中心偏移,病理切片的染色标准不一、扫描参数差异等,导致检测结果的一致性与代表性受限<sup>[18-20]</sup>;基因组学技术能够精准实现肿瘤分子分型、筛选靶向治疗靶点,明确肿瘤分子层面的核心异常,但缺乏肿瘤空间分布上下文信息,无法关联宏观形态与微观结构特征,且分子检测结果与临床诊疗动作之间的解释链条易断裂,难以直接转化为可落地的诊疗建议<sup>[21-22]</sup>。由此可见,仅从单一模态数据入手,很难实现对肿瘤的精准诊疗。

多模态数据整合已成为突破单模态诊疗局限、实现肿瘤精准治疗目标的核心路径。多模态数据融合能够实现数据信息的互补,通过整合影像、病理、基因组及临床数据,从不同视角对肿瘤展开全景式的认知<sup>[23]</sup>。同时,多模态整合有助于提升肿瘤诊疗精度,提升预后与疗效预测的准确度,还能够增强模型的泛化性与可解释性,促进肿瘤智能诊疗模型的临床应用<sup>[24-25]</sup>。需要强调的是,这种多模态整合并非简单的各个模态特征拼接,而是构建整合宏观影像、微观病理以及分子基因的跨尺度分析框架,不仅可推动肿瘤的精准诊疗,而且还能为探究肿瘤演化机制提供有力的工具。多模态融合在肿瘤医疗领域已取得显著进展,核心围绕影像、组学、病理和临床数据等模态整合。主流融合范式包括前融合、中期融合和后融合,其中图融合属中期融合子类。AI技术如卷积神经网络(Convolutional neural networks, CNNs)、生成对抗网络(Generative adversarial networks, GANs)、基础模型等赋能数据处理与特征提取,可应用于肿瘤筛查、诊断、预后预测及治疗响应评估<sup>[26-29]</sup>。当前研究已验证了其在提升诊断精度、优化风险分层等方面的优势,但仍面临数据异质性、模型可解释性不足、伦理合规及数据稀缺等挑战,未来需向轻量化、可解释化及多源数据深度整合方向推进<sup>[30-34]</sup>。我国在肿瘤多模态诊疗方面也出台了一系列政策,如《健康中国行动——癌症防治行动实施方案(2023—2030年)》和《“十四五”国民健康规划》等一系列重大战略文件均明确提出支持AI、大数据等新技术在肿瘤诊断和治疗中的应用,为多维度肿瘤多模态分析提供了坚实的政策支撑<sup>[35-36]</sup>。

本文围绕肿瘤影像、病理和基因融合的智能分析诊疗应用这一主题,通过系统性回顾这一领域的文献,明确当前领域研究的热点和存在的不足,为后续研究工作指明方法。本文的第一部分主要从肿瘤影像-病理-基因融合的方法学进行归纳,第二部分以肿瘤多影像-病理-基因融合的临床应用为主线介绍国内外研究前沿,最后对未来肿瘤多模态智能诊疗进行了展望。

## 1 肿瘤跨尺度证据链驱动的融合框架

精准医疗是肿瘤诊疗的核心诉求,需要根据患者的个体特征制定最优的治疗方案。因此需要对肿瘤具有全面的认知。临床实践与前沿研究日益证实,单模态数据无法完整表达个体肿瘤的差异化问题。因此,整合多维度数据,构建一条从宏观影像表型到微观病理组织结构、再到深层基因分子机制的完整认知链条,已成为推动医学决策智能化的关键<sup>[37]</sup>。

### 1.1 肿瘤影像、病理、基因的融合特点

不同维度的数据在不同生物学尺度上提供了不可替代的、互补的信息。它们的融合能够构建一个

从宏观生理结构到微观组织状态,再到分子机制的完整生物学叙事链条。磁共振成像(Magnetic resonance imaging, MRI)作为一种非侵入式成像技术,能够无创地描绘整个肿瘤及其周围环境的三维空间结构和生理功能,为医生提供明确的肿瘤大小、位置及边界等重要信息<sup>[38]</sup>。全玻片数字扫描技术(Whole slide imaging, WSI)是对目标肿瘤的病理学呈现,是界定肿瘤细胞和组织形态的金标准,为MRI观察到的宏观肿瘤提供了微观解释<sup>[39]</sup>。基因组学、转录组学和蛋白质组学等技术(统称为组学, Omics)从分子层面解释了在WSI和MRI上观察到的表型背后的分子驱动力<sup>[40-41]</sup>。融合这3个模态,就像是用不同焦距的镜头观察同一物体: MRI给出哪里有问题, WSI解释是什么的问题,而Omics则回答为什么会有问题。这种多维度信息的整合,使得我们能够构建一个从基因型到表型的完整通路,从而更深刻地理解肿瘤的生物学行为。

通过强制模型学习跨维度的样本信息,可以有效抑制由数据偏移导致的伪相关,从而提升模型的泛化性和可解释性<sup>[42]</sup>。传统的研究方法通常是通过影像特征直接预测基因信息<sup>[43-44]</sup>。但这种简单的端到端输出,易使模型学到由机器性能不足产生的噪点信息,而非真正的肿瘤生物学特征,从而导致模型的泛化性严重不足。因此通过建立成由影像到病理,再到基因的完整证据链路,可提高模型的鲁棒性与生物学上的可解释性。模型不再仅仅预测有无基因突变,而是被强制要求学习这三者之间的一致性关系。这种方法本质上是为深度学习模型引入了生物学先验知识的约束,使其从一个单纯的“黑匣子”关联学习器,转变为一个寻求机制性解释的推断引擎<sup>[45]</sup>。

融合模型的最终目标不应仅仅是输出一个抽象的概率值或分类标签,而生生成能够直接嵌入临床 workflow、指导医生决策的、可解释且可操作的信息<sup>[46-47]</sup>,如图1所示。在临床中,这种智能化决策可贯穿于诊疗全流程。术前规划阶段,融合模型通过解读影像的空间异质性,生成可预测分子表型的肿瘤“栖息地(Habitat)”图谱<sup>[48-49]</sup>。这不仅能指导靶向活检,有效克服肿瘤内部取样偏差以确保诊断准确性,还能在术中为切除边界的决策提供关键参考<sup>[46]</sup>。术后分析则可通过对微观与分子层面的分析,使模型能够预测肿瘤的发展情况,从而帮助医生进行治疗决策<sup>[50]</sup>。

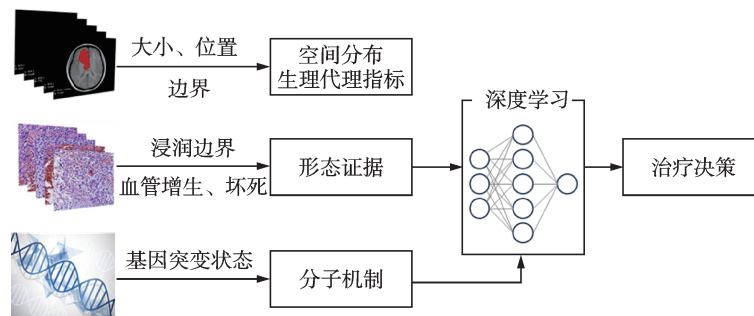


图1 患者级跨尺度证据链

Fig.1 Patient-level cross-scale evidence chain

## 1.2 肿瘤影像、病理、基因的融合方法

由于肿瘤影像、病理、基因数据在获取难易程度、分析方法等方面存在差异,因此导致不同模态之间出现了各种组合分析方法,此部分重点介绍常见的一些融合分析方法。

### 1.2.1 影像基因组学(Radiogenomics)

影像基因组学是一个新兴的交叉学科领域,旨在利用临床影像图片提取数字特征,并使之建立起与基因组学之间的关系<sup>[51]</sup>,其核心思想是肿瘤的宏观影像表征,如肿瘤大小、形状、内部纹理异质性等,在一定程度上反映了其潜在的分子生物学特征和基因调控网络<sup>[52]</sup>。通过将这两个维度的信息进行融

合,研究人员希望能够实现“透过现象看本质”,即从非侵入性的影像数据中推断出肿瘤的分型、基因突变状态、信号通路激活情况以及对治疗的反应。这对于指导个体化治疗、预测预后具有巨大潜力,尤其是在无法获取组织活检或活检样本不足的情况下<sup>[51-52]</sup>。

对影像基因组学最常见的研究方法是特征提取与关联分析。通过医学影像提取包含肿瘤的形态学、内部密度分布及纹理特征等影像组学信息,结合患者的基因组学数据,利用机器学习算法获取影像组学信息与基因的关联信息。例如, Ma等<sup>[53]</sup>以双能CT拍摄的影像为基础,预测非侵入性肺腺癌的EGFR突变状态。而随着深度学习的不断发展,近年来越来越多的研究倾向于直接从原始影像中学习基因组学特征相关的表达。如 Padmaja等<sup>[54]</sup>通过结合乳腺X线分析的卷积神经网络和用于基因组分析的LSTM+Transformer模型,实现了对乳腺癌96%的精准识别; Abdelaziz等<sup>[55]</sup>使用于U-Net对肿瘤进行分割,并结合EfficientNetB0与ResNet50的混合深度学习框架预测胶质瘤患者MGMT启动子甲基化状态,获得了95%的分类准确率; Yan和Sun等<sup>[56-57]</sup>通过建模,利用术前医学影像的深度学习模型预测患者的整体生存率,证明了某些深层特征与独特的信号通路相关,这些通路对患者具有预后意义。但在深度学习模型中,如何解释影像特征与基因组学事件之间的内在联系仍是一个开放问题。同时,研究结果的重复性和泛化能力也有待进一步验证,实现临床常规应用仍需大量前瞻性研究。

### 1.2.2 影像病理融合分析

医学影像与病理图像的融合,旨在结合宏观与微观的优势,实现互补增强,在做到快速定位病灶的同时,给予肿瘤的微观生物学解释。这种融合有助于更精确的肿瘤分级、分型、预后预测以及治疗反应评估<sup>[58-59]</sup>。影像与病理的融合可以应用于多个医学场景。首先是预后方面, Lin等<sup>[60]</sup>开发了一种影像与病理的融合模型,用于预测肺腺癌患者的预后情况和辅助化疗获益; Xie等<sup>[61]</sup>利用影像组学与病理学特征预测了肝细胞性肝癌的复发风险。此外,还有研究利用此方法预测生物标志物的表达。例如, Mao等<sup>[62]</sup>基于深度学习的影像病理学模型,预测了胰胆管发育不良儿童胆管中促动子环氧合酶-2的致癌反应。虽然影像与病理融合具有相当大的应用前景,但这两种模态在分辨率、视角及样本形变上的巨大差异,以及肿瘤内部的影像学和病理学异质性如何有效地在融合模型中体现,是其进一步发展的技术瓶颈。

### 1.2.3 病理基因融合分析

病理基因融合分析旨在直接建立病理图像的视觉表型与基因组学分子分型之间的桥梁。这种融合的意义是实现肿瘤表型特征与分子机制的联合解析,提升肿瘤分型、预后评估与疗效预测精度,为揭示肿瘤异质性及微环境机制提供新途径,是推动精准病理诊断与个体化肿瘤治疗的关键技术方向<sup>[63-64]</sup>。当前病理图像与基因组学融合的研究多以深度学习为核心,聚焦肿瘤分子分型、预后预测与基因突变关联建模。融合后的模型在多癌种中显著优于单模态方法。Chen等<sup>[65]</sup>提出了一种多模共注意力变换器(Multimodal co-attention transformer, MCAT)架构,通过注意力机制学习图像与基因之间的长期关系,并通过视觉展示多模态相互作用的可解释性; Yan等<sup>[66]</sup>基于病理图像和基因表达数据的多模态生存分析,建立起通路到斑块交叉融合模块,赋予了模型良好的可解释性,帮助病理学家筛选哪条通路起关键作用,定位通路在WSI中受影响的部位,并挖掘与预后相关的病理图像模式。虽然这种融合模式有良好的性能表现,但病理图像的跨中心采集差异和WSI的巨大数据量都是影响算法落地的重要因素。

### 1.2.4 影像-病理-基因融合分析

虽然两两模态的融合已经显示出显著潜力,但每种模态都有其固有的局限性<sup>[67]</sup>。医学影像提供了宏观的肿瘤解剖和功能信息,但缺乏细胞和分子细节;病理图像提供了微观的细胞形态和组织结构,但缺乏整体空间背景和非侵入性;基因组学提供了分子层面的驱动信息,但缺乏空间异质性和形态学表现。因此,将影像、病理和基因组学三模态数据进行综合融合分析,是实现肿瘤精准医疗的终极目标,

也是最具挑战性的方向。这种三模态融合旨在创建一个最为全面的肿瘤表征,通过互补优势来克服单一或两两模态的局限,从而有望在肿瘤诊断、分级、预后预测和治疗反应评估方面取得突破性的进展,为个体化治疗提供前所未有的深度洞察<sup>[68]</sup>。相较于两种模态数据的融合,三模态融合的研究对配对数据要求较高。Liu等<sup>[69]</sup>通过多模态融合亚型划分(Multimodal fusion subtyping, MOFS)框架,整合多模态数据,研究识别出3种独特的胶质瘤亚型(MOFS1、MOFS2和MOFS3),它们在预后、分子特征和治疗反应方面存在显著差异;Braman等<sup>[70]</sup>利用深度正交融合(Deep orthogonal fusion, DOF)模型,根据多模态数据预测胶质瘤患者的总体生存期。

### 1.3 肿瘤影像、病理、基因的融合策略

在实践中,如何将来自不同模态的数据进行有效的集成,是多模态学习的核心问题。广义上,不同模态的医学数据集维度和类型差异极大,包括数值数据、图像数据、文本数据和序列数据。此外,根据其来源,这些数据集可分为测序数据、临床数据和实验数据<sup>[71]</sup>。这些模态在数据结构、语义粒度、信息密度以及测量尺度上存在显著差异。简单的拼接或粗暴的整合难以捕捉模态间深层次的互补性,甚至可能因异质性过高而引入噪声或冗余信息,从而限制模型的性能<sup>[72]</sup>。研究人员根据融合发生的时间点和方式,发展出多种如图2所示的融合策略,其中每种策略都有其适用场景、优势和局限性。

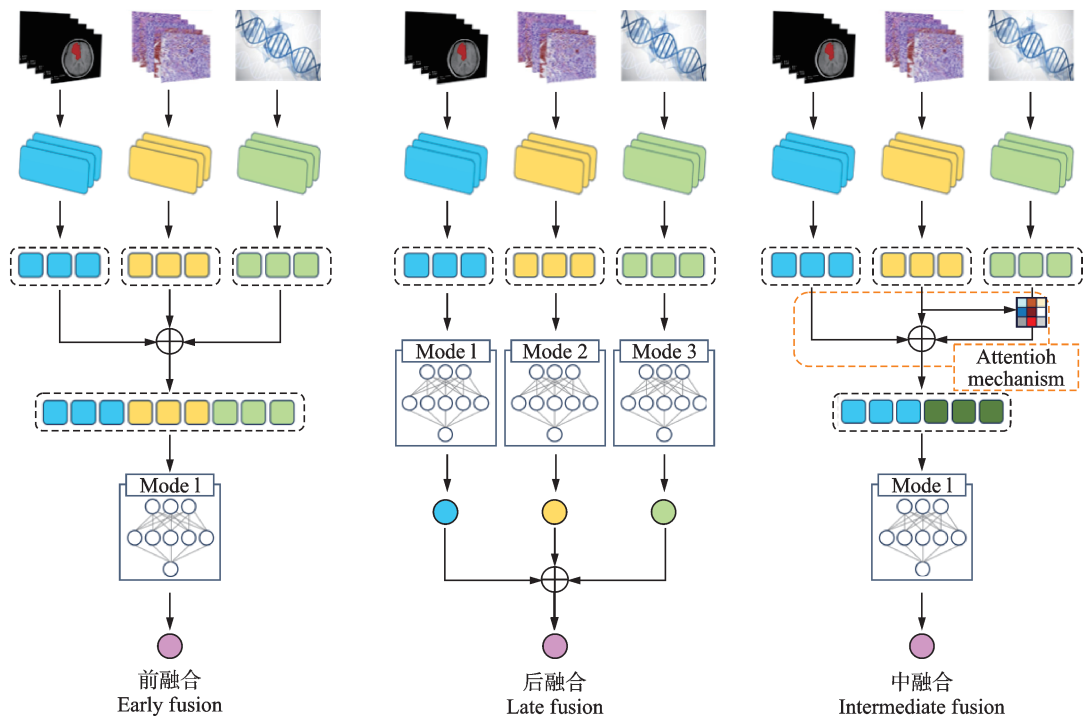


图2 多模态数据融合策略分类图

Fig.2 Taxonomy diagram of fusion strategies for multimodal data

#### 1.3.1 前融合

前融合(Early fusion),也称为数据级融合(Data-level fusion)或特征级融合(Feature-level fusion),是指在模型学习之前,将来自不同模态的原始数据或特征直接合并成一个统一的集合。简单来说,就是将不同模态的特征向量直接拼接起来,形成一个更长的特征向量,然后将这个合并后的特征向量作为一个单独的输入进行后续处理。前融合简单明了的特点,直观体现在可以相对快速地实现模型架构设

计,只需训练一个模型来处理融合后的数据,使模型学习到不同模态之间更深层次、更细粒度的交互模式。

前融合会视所有模态对当前任务同样重要且相关,但实际情况中并非总是这样<sup>[73]</sup>。首先,数据质量常会因各中心的设备与操作流程不同而不尽相同,从而产生数据缺失、噪声干扰和各模态数据不平衡等问题,直接融合容易放大噪声,使模型学习到并不存在的规律。研究人员常会使用数据清洗、缺失值填补、噪声抑制及模态平衡等方法,提升数据可靠性。其次,数据间异构性显著的问题也会影响数据的直接融合。不同模态医学数据的格式、尺度、采集标准存在本质差异,如影像数据为像素矩阵、临床数据为结构化数值、基因组数据为序列信息,难以直接进行拼接与整合。Attallah等<sup>[74]</sup>利用自编码器(Autoencoder, AE)的无监督学习能力,将高维异构的深度学习特征与手工特征低维共享隐空间,实现特征的统一编码与融合,消除特征维度、分布的异构性。另外,未经处理的原始数据常存在维度灾难的问题,若使用原始数据建模,易导致模型计算复杂度激增,引发过拟合风险。因此在建模前,常会对数据进行降维,利用特征筛选技术,仅保留关键特征信息。如Schulte-Sasse等<sup>[75]</sup>通过图卷积网络(Graph convolutional networks, GCNs)隐空间映射与组学数据预处理,实现对多组学高维数据的降维,既保留了核心生物学信息,又降低了计算复杂度,提升了模型泛化能力。最后,部分场景下多模态数据采集时间存在差异,如术前影像与术后临床指标的不同步,难以实现时序层面的精准对齐。研究人员会引入时序对齐算法,实现多模态数据的时序同步匹配。如Khorasani等<sup>[76]</sup>提出CTAF(Cross-temporal attention fusion)方法,通过双向跨时序注意力与软对齐损失,实现异步脑电图(Electroencephalography, EEG)与外周生理信号的时序对齐。

前融合对原始数据质量要求极高,当原始数据存在严重噪声、缺失或异构性较强时,算法也难以弥补融合效果的下降;同时,面对高维度医学数据(如高分辨率影像、全基因组数据)时,前融合会不可避免地出现计算效率低下、模型训练困难等问题;此外,该方法缺乏对数据语义信息的挖掘,难以充分利用多模态数据的互补性,在复杂临床场景(如多器官病变、罕见病诊断)中的适应性较差。因此,前融合更适用于各模态特征维度相对均衡、数据类型较为相似,且数据完整性较高的情况。

### 1.3.2 后融合

后融合(Late fusion),也称为决策级融合(Decision-level fusion),是指对来自不同模态的数据分别训练独立的模型,然后将这些独立模型生成的预测结果(如分类概率、回归分数、决策标签)在最终阶段进行集成,以得出最终的综合预测。与前融合相比,后融合可以处理简单的数据异构性。同时,后融合允许不同模态的数据使用最适合分析其特征的模型,如音频使用隐马尔可夫模型(Hidden Markov model, HMM)、图像使用支持向量机(Support vector machines, SVM)等<sup>[77]</sup>。各模态的独立建模也使得模型避免了多模态融合的维度灾难。即使某一模态缺失,其余模态也能独立做出预测,增强了系统的鲁棒性。

但后融合在决策融合阶段依然会受到来自原始数据和模型设计的影响。首先是各模态输出结果的异质性。不同模态输出的结果格式、置信度存在差异,难以直接进行融合。Tortora等<sup>[78]</sup>将不同模态的原始数据通过独立分类器生成统一格式的输出结果,消除软硬标签及维度差异,以保证决策融合阶段的有效性。其次,融合算法难以确定,不同模态在不同临床任务具有不同的贡献度,固化的融合规则难以适应复杂的临床场景需求。Hassan等<sup>[79]</sup>利用全局特征与局部特征的互补性设计融合逻辑,并采用能更好适配融合特征非线性分布特性的二次判别分析(Quadratic discriminant analysis, QDA)作为最终决策模型,以达到优化融合的效果。另外,单个模态输出的决策存在偏差累积效应,即当单一模态模型的输出结果存在偏差时,融合后会放大偏差,进而影响最终输出结果的准确性。Tripathi等<sup>[80]</sup>利用不同深度的残差网络的互补性,并通过Dempster-shafer Theory相互强化互补的有效输出,让冲突的偏差

输出相互抵消。最后,输出结果的置信度校准困难,不同模态的输出结果置信度缺乏统一的评价标准,难以实现合理的权重分配与融合。研究人员常采用置信度评估方法量化各模态决策的可靠性,为权重分配提供依据。

后融合难以充分利用多模态数据的底层细节信息与语义关联,存在早期信息丢失问题,在需要精准捕捉细微病变特征的场景中性能有限;其融合效果高度依赖于单模态模型的性能,若单模态模型性能较差,难以通过融合规则弥补。此外,融合规则的高度复杂性导致可解释性不足,难以明确各模态决策在最终结果中的贡献的占比,不符合临床决策的可追溯性需求。同时,该方法缺乏对多模态数据时序关联的挖掘,在时序性临床场景中的适应性较差<sup>[81]</sup>。因此,后融合适用于各模态数据差异大、难以进行早期对齐和拼接,或者对系统鲁棒性要求较高的场景。

### 1.3.3 中间融合

中间融合(Intermediate fusion)是介于前融合和后融合之间的一种特征融合策略,更多地被应用于深度学习的模型中。中间融合在特征提取阶段使用不同的模型对各模态的原始数据提取特征向量,并在隐藏层中对数据进行融合,再输入后续的决策层网络。与前融合不同的是,中间融合会通过损失反向传播,同时优化各个模态的特征提取与融合模块<sup>[82-83]</sup>。

近年来,注意力机制(Attention mechanism)在中间融合中扮演了越来越重要的角色,其核心机制是基于QKV(Query-key-value)三元组框架。在注意力机制中,经提取后的特征,会计算特征相似度得分并生成动态权重,加权聚合关键特征、抑制冗余噪声后,将融合特征传递至决策层完成预测<sup>[84-85]</sup>。注意力机制能够捕捉不同模态之间深层次的非线性关联,同时动态加权允许模型自主寻找重要特征,从而优化信息利用效率,提高模型可解释性。但注意力机制具有较高的计算复杂度,当序列长度 $n$ 增加时,计算量平方级增长,导致长序列处理效率低下、难以实时部署。Zhao等<sup>[86]</sup>提出SPAttention,引入原则结构稀疏(Principled structural sparsity)策略,提升了计算效率和模型性能。其次,注意力机制也有长距离依赖捕捉不足的特点,长序列中注意力权重易稀释,无法有效建模全局信息关联。Li等<sup>[87]</sup>提出MHANet(Multi-scale hybrid attention network),通过整合多尺度时序注意力、多尺度全局注意力与通道注意力的混合注意力机制,解决注意力机制长距离依赖捕捉不足的问题。另外,注意力权重可解释性弱且易受噪声干扰,权重分配呈黑箱特性,噪声易导致权重误分配。Jha等<sup>[88]</sup>提出头级熵正则化策略,搭配可学习的头级强度参数与Softmax温度系数,仅惩罚过度熵值,但又不限制模型表达能力,解决了注意力权重分布弥散、可解释性差等问题。

当前,注意力机制是多模态深度学习研究的热点。它在处理图像、文本、序列等多种异构数据融合时表现出色,是目前认为最有前景的融合策略之一,尤其适用于影像-病理-基因融合这种需要高度语义理解和关联的任务中。

### 1.3.4 图融合

图融合(Graph fusion)是注意力机制与多模态数据融合、图神经网络(Graph neural networks, GNNs)交叉领域中相当热门的研究方向,是利用图论(Graph theory)和图神经网络来建模和融合多模态数据的方法,其核心原理是先将各模态特征映射为图节点并构建关联边,再通过节点级/模态级注意力机制计算动态权重,经加权聚合更新节点表示,最后通过全局池化生成融合特征并传递至决策层<sup>[89-90]</sup>。

在图融合中,模态与结构的异质性显著,不同来源图的节点定义、边语义及特征空间差异大,易引发特征冲突与信息丢失。其次,图融合的计算复杂度高,大规模图数据融合易导致内存溢出、训练效率低下。Chen等<sup>[91]</sup>设计模态特定轻量级Adapter编码器,将不同任务图的异质特征映射至统一空间,并提出拓扑感知Wasserstein距离(Topology-aware Wasserstein distance, TWD)损失,实现图结构对齐,结

合无参数 MoE (Mixture-of-experts) 路由机制动态调节模态贡献,缓解特征冲突并优化参数冗余、内存占用高等问题。同时,图融合的语义一致性与可解释性较低,缺乏先验指导易生成模糊表示,难以完成决策过程的解释。Wu 等<sup>[92]</sup>提出 SAGE 框架,通过注入 SAM 语义先验、设计语义持久注意力模块及双级蒸馏三重损失约束优化,解决多模态图融合语义一致性低、可解释性弱及依赖大模型的问题。

在肿瘤多模态数据融合中,可以根据数据内部和不同模态间的复杂关系构建图结构,将多模态数据作为抽象节点,寻找节点间空间关联及生物相关性等信息。图融合完美适配非欧几里得数据“无固定网格结构、依赖节点关联”的本质特征,能捕捉长距离依赖与局部关联。因此,其在生存预测、空间域识别等任务中,能融合多模态互补信息,显著提升预测精度与聚类一致性,但图结构和训练模型对参数极其敏感,容易出现过拟合等情况。另外,图融合的计算复杂程度与训练样本的规模要求同样在限制其产业化落地<sup>[93-95]</sup>。在肿瘤研究中,当需要关注患者群体间的相似性、肿瘤微环境的细胞相互作用或基因调控网络与表型之间的关联时,图融合提供了一种强大的框架。

## 2 应用:以融合补全证据链为核心的肿瘤临床问题图谱

肿瘤临床决策往往同时依赖三类证据的相互印证。第1类是空间证据,主要来自医学影像,为病灶的全局范围、亚区异质性、多灶分布及纵向变化提供观察窗口;第2类是组织学证据,来自病理切片与细胞层面的结构解析,能够呈现肿瘤细胞形态、坏死与血管增生等组织结构特征,以及肿瘤微环境中免疫细胞和基质细胞的空间组织;第3类是机制证据,来自基因组学与多组学,刻画驱动事件、通路状态与潜在可干预靶点。肿瘤内异质性与分支进化不仅是普遍现象,而且直接限制了单次局部取样对整体生物学状态的代表性,从而削弱单模态证据对治疗选择与风险分层的支撑力度<sup>[96]</sup>。在这一前提下,多模态融合的应用价值在于临床应用性(见图3),多模态融合的核心贡献在于补全证据链并实现闭环验证(见表1)。

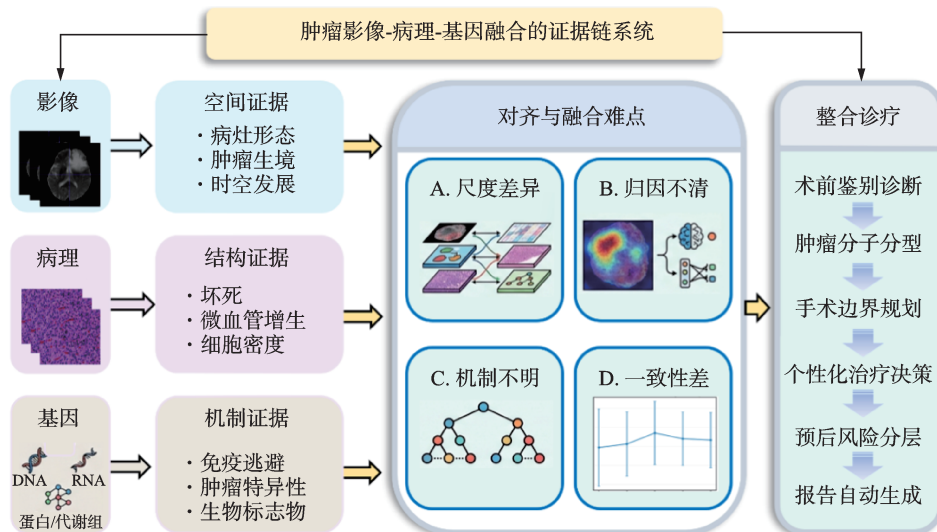


图3 病理-影像-基因融合难点与临床应用

Fig.3 Challenges and clinical applications of pathology-imaging-genomics fusion

影像将肿瘤表达为可定位的空间单元,例如病灶层级、亚区层级或肿瘤栖息地层级,从而为点状的病理与组学采样提供空间锚定。病理作为结构证据能够对影像表型所隐含的生物学过程进行验证与解释,并以弱监督学习等范式在大规模切片数据上获得稳定的区域级证据输出。组学作为机制证据进

一步将表型与可干预的分子过程连接起来,使得影像所见不仅是相关性提示,而且能够被通路和网络层面的解释框架支撑。与单模态相比,融合模型的产出更接近临床需要的结论形式,即对关键空间热点、对应组织学依据与机制线索的联合呈现,并伴随可信度评估与复核建议,从而降低过度推断的风险<sup>[97]</sup>。

表 1 肿瘤临床问题图谱:证据链要素与融合优势

Table 1 Oncology clinical problem atlas: Elements of evidence chain and integration Advantages

临床问题	关键空间证据 (影像)	关键组织学证据 (病理)	关键机制证据 (组学/通路)	融合优势
术前诊断与鉴别	病灶分区,强化核心、坏死、非强化浸润前沿、水肿边界;多灶分布	坏死、微血管增生、细胞密度梯度、浸润结构;MIL 关键区域	驱动事件方向;缺氧、血管生成、增殖、侵袭相关通路	空间补全与一致性闭环
分子分型与标志物提示	生境引导采样导航;浸润前沿与核心异质性	结构与细胞图谱;关键区域复核	通路活性与网络枢纽优先;免疫抑制、修复、缺氧等机制分层	将无创预测转成采样动作
手术规划	边界不确定区与高风险核心定位	浸润梯度、血管形态、细胞密度;切缘与区域证据	侵袭、缺氧、血管生成、修复相关通路;枢纽方向性	边界证据链化与区域级归因
系统治疗与疗效分层	反应/耐受的病灶	免疫浸润结构、肿瘤细胞密度、血管形态、坏死与炎症反应区分	修复、缺氧、血管生成、免疫抑制、抗原呈递等机制分层	从疗效概率到机制分层与策略提示证据框架
预后风险分层	复发/死亡风险	结构性高风险表型与细胞生态位	高风险通路与枢纽;空间语境下的机制解释	风险空间化与可干预线索
报告自动生成与系统化输出	纵向变化与表型描述	结构与细胞图谱	通路/枢纽方向性解释,机制分层解释	报告从结论变证据链

## 2.1 术前诊断与鉴别:将影像主导的判断推进为跨模态证据闭环

术前诊断与鉴别的任务是确定肿瘤类型、评估侵袭性与高级别风险,并据此制定取样强度、切除范围与后续治疗路径。该阶段的临床限制在于信息获取顺序具有固定性,影像通常最先获得并决定初步策略,而病理与组学证据多在有限取样后才能形成。因此,术前决策需要一种能够把影像的全局信息与后续组织学、机制证据有效衔接的推理框架<sup>[98]</sup>,以减少由单模态不完备信息导致的路径偏差。

仅依赖影像时,浸润、坏死与血管生成等关键过程往往通过强化模式、弥散和灌注等代理指标间接反映,其生物学特异性有限且易受扫描仪器、重建参数与设备差异影响,跨中心稳定性不足会削弱术前诊断结论的可迁移性<sup>[96]</sup>。仅依赖病理或组学时,证据强度较高,但空间覆盖受限,穿刺或有限切片可能遗漏最具代表性的高侵袭亚区,单点活检所带来的代表性不足在肿瘤内异质性与分支进化背景下尤为突出,可能导致分型、分级与风险判断出现系统性偏差。

多模态融合的首要作用是空间补全。影像能够将病灶表达为可定位的空间单元,典型包括强化区、坏死区、非强化浸润区及水肿边界等亚区,并进一步形成病灶层级与亚区层级的表型描述。病理与组学证据一旦锚定到这些空间单元,就能够将点状取样的结论转译为空间分布层面的解释,从而显式对抗取样偏倚。在这一思路下,注意力机制与多实例学习可用于建立空间单元与病理证据之间的对应关系,并在个体层面给出证据贡献的权重分配。Vanguri 等<sup>[99]</sup>提出的 DyAM 框架将影像、病理与基因组

信息联合建模,通过动态注意力实现跨模态自适应加权,其输出不仅提升了风险分层稳定性,也使不同模态在特定患者上的贡献可被量化,从而支持术前阶段的证据整合与优先级判断。

多模态融合的第2个作用是证据闭环。病理结构特征如坏死与微血管增生,以及细胞密度梯度和浸润前沿,能够对影像亚区的生物学含义提供可重复验证的证据支撑。组学层面的驱动事件与通路活动进一步提供对影像表型形成机制的解释,使空间表型、组织结构与分子机制之间形成一致性检查框架。以跨模态交互为核心的融合模型在此具有代表性。如Chen等<sup>[100]</sup>提出的Pathomic fusion将全切片病理表征与多组学特征进行门控与交互建模,能够同时输出关键病理区域与重要分子特征,并将两者通过交互项连接为可解释链路。该类方法对于术前鉴别的意义在于,它将病理可复核的结构证据与可干预的机制证据纳入同一推断过程,从而减少了影像或病理单点结论在机制层面的悬空,提高了临床沟通与复核的可操作性。

术前阶段在多癌种中常伴随多时间点影像或治疗前评估信息,单时点表型容易受到炎症反应、扫描差异与观察者偏差影响。引入纵向融合能够以时间一致性增强对侵袭性表型与风险趋势的识别稳健性。Gao等<sup>[101]</sup>提出的多模态反应预测(Multi-modal response prediction, MRP)模型将治疗前后或多阶段影像与临床、病理评估联合建模,强调跨模态知识挖掘与可解释输出,使得模型更贴近临床实际信息流。该类纵向多模态框架提示,证据闭环不仅发生在模态之间,也应延伸至时间维度,从而把术前判断从单次观测提升为动态证据整。

### 2.1.1 典型的多模态融合输出

面向术前应用,融合模型的输出宜采用证据链结构而非单一概率值。输出通常包括疑似类别或级别的风险评分,用于提供方向性决策支持;关键空间热点的定位,用于明确最可疑的病灶或亚区并辅助制定取样与切除优先级;对应组织学证据的摘要,例如基于弱监督学习定位与诊断相关的切片区域并形成区域级证据条目;机制提示则以通路活动或驱动事件方向性解释影像与病理所见,并在跨模态不一致或证据不足时降低置信度,提示复核或补充取样。DyAM的动态注意力权重与Pathomic fusion的关键区域与关键分子特征输出均体现了这一证据链导向的输出原则,并为术前阶段将模型结论转译为临床可读信息提供了范式参考<sup>[99]</sup>。

### 2.1.2 临床获益

融合在术前诊断与鉴别中的直接优势主要体现在证据覆盖与决策稳健性两方面:其一,融合通过空间补全与区域级归因,能够更早识别高侵袭性空间单元并将其转化为取样与切除策略的优先级安排,从而提高有限采样对关键生物学信息的覆盖效率,降低因单点取样遗漏导致的分型或风险误判;其二,融合以跨模态一致性实现证据闭环,减少单纯依赖影像表型或局部病理所引发的路径偏差,并在多时间点信息可用时,通过纵向融合增强对风险趋势的判读稳定性。MRP等纵向多模态框架对临床信息流的拟合进一步表明,融合模型若能够同时处理空间异质性与时间演化,将更接近真实世界术前决策所需的证据结构<sup>[101]</sup>。

## 2.2 分子分型与标志物提示:从预测突变走向指导取样与机制分层

肿瘤分子分型与标志物评估在临床路径中具有基础性地位,但其结论的可靠性强烈依赖取样的空间代表性。多区域测序研究已系统性揭示实体瘤存在广泛的肿瘤内异质性与分支进化,单点活检可能遗漏关键克隆或关键微环境状态,从而导致分型与靶点判断出现偏差<sup>[96]</sup>。在这一背景下,多模态融合的目标不应被定义为替代检测,而应聚焦于优化3个更贴近临床的问题:取样位置如何选择、取样是否充分和机制分层是否可信。融合将影像的空间覆盖优势与病理的结构证据、组学的机制证据贯通,使分子分型从单点结论转化为带有空间语境与一致性校验的综合判断框架。

术前与术中阶段,临床需要确定哪些区域必须采样,是否需要多区域取样以覆盖异质性,尤其在病

灶内部存在强化核心与非强化浸润前沿等差异表型时更为突出。术后阶段,临床需要评估既有分子分型结果对全肿瘤的代表性,明确其是否足以支撑后续治疗选择与试验入组分层。多区域测序在肾癌与肺癌等癌种中显示,不同空间区域的突变谱与克隆结构可显著不同,支持将多区域取样作为提高分型可靠性的策略基础。

影像到分子的无创推断存在两类结构性风险:其一是模型容易学捷径特征,导致跨中心泛化不稳,并将非生物学差异误解为分子差异;其二是同一分子事件或通路状态在肿瘤内部呈空间不均,影像表型与分子状态的对应关系可能仅在特定亚区成立,从而限制患者级单一预测结论的可用性。针对这一点,影像基因组学领域大量工作仍集中于单基因或少数标志物预测,且受样本规模与外部验证不足的制约,显示出从预测走向可行动决策仍存在方法鸿沟<sup>[102]</sup>。

组学检测的短板则主要来自取样代表性与空间语境缺失。单点测序可能测不到关键亚克隆,或无法反映浸润前沿与核心区在微环境组成、免疫状态与代谢通路上的差异,从而影响靶点判定与耐药风险评估。多区域测序研究强调,肿瘤内异质性会系统性限制基于单一活检样本的个体化治疗策略,这一结论对所有实体瘤具有普遍警示意义<sup>[96]</sup>。

多模态融合的首要优势是基于空间单元的取样导航。影像能够将病灶分解为可定位的亚区或肿瘤栖息地,并以此作为锚点组织病理与组学证据,使模型输出从无创预测转化为取样动作建议。该策略在概念上等价于将分子分型问题由患者级分类转译为亚区级证据获取问题,进而服务于术前与术中决策。多区域测序在多个癌种中证明空间异质性的广泛性,为基于亚区的取样优化提供了生物学依据<sup>[96]</sup>。第2个优势是通路级稳定性优先。与直接预测单一突变相比,通路活性、网络枢纽与细胞互作状态在跨平台、跨中心的可迁移性通常更强,也更容易与病理形态和微环境结构建立可复核对应关系。影像与病理更多提供表型与结构证据,组学在通路层面提供机制证据,三者协同可将模型输出组织为机制分层而非孤立的突变提示,从而提高临床解释性与外部验证可行性。第3个优势是一致性校验机制。融合框架允许对影像亚区、病理结构与通路活动之间的一致性进行显式检查。当三者支持同一方向的分型或机制推断时,结论置信度提高;当三者出现不一致时,模型应输出较低置信度并提示补充取样或复核,避免将局部证据泛化为全局结论。这类一致性约束与不确定性表达是融合区别于单模态预测的关键临床安全属性。

在方法实现层面,已有多项多模态影像基因组学研究提供了可借鉴的算法范式。Moreno等<sup>[103]</sup>提出的放射基因组学集成方法以影像特征预测EGFR与KRAS等突变状态,代表了单基因预测在工程实现上的常见路径,同时也反映出此类任务对数据规模与外部验证的依赖,提示需要进一步引入跨模态一致性与空间语境来提升临床可用性。Hoivik等<sup>[104]</sup>将MRI放射组学风险分层与基因表达签名关联,展示了以基因特征集与分子亚型对影像分群进行机制解释的思路,为以通路与表达程序而非单突变为融合目标提供了证据。此外,跨模态交互建模框架如Pathomic fusion在配对病理与组学数据上验证了病理结构证据与分子机制证据的联合推断能力,为通路级机制分层与证据链输出提供了方法学支撑<sup>[105]</sup>。

### 2.2.1 针对分子分型与标志物的融合模型典型输出

面向这类临床流程,多模态融合输出宜采用可行动的结构化形式。建议输出包括3部分:第1部分为建议采样亚区的排序列表,明确优先覆盖的空间单元,并给出与分型或机制相关的理由条目;第2部分为预计信息增益,刻画多区域取样对覆盖异质性的提升程度,可用亚区差异度或克隆覆盖的指标表示;第3部分为分型或通路层面的置信度评估,并在跨模态不一致时触发复核与补采提示。该输出形式将模型的无创预测能力转化为取样策略优化工具,使分子分型由术后被动结果变为术前术中可规划过程。

### 2.2.2 临床获益

融合在分子分型与标志物提示中的临床获益主要体现在分型准确性与可重复性提升,以及将预测转化为取样质量控制工具两个方面。通过影像驱动的空间导航与跨模态一致性校验,关键分子信息被捕捉的概率提高,减少因单点取样遗漏关键克隆或关键微环境状态导致的分型偏差,这一点与多区域测序关于肿瘤内异质性的结论一致<sup>[96]</sup>。同时,融合输出以证据链形式显式表达不确定性与复核需求,使模型在真实世界场景中更符合临床安全与可解释性要求,推动分子分型从结果导向走向过程导向与质量导向。

### 2.3 手术规划:最大安全切除的融合逻辑,将边界写成证据链

手术规划的核心在于在肿瘤控制与功能保护之间取得可解释的平衡。其一是切除范围的确定,需要识别最可能承载高侵袭性与耐受风险的核心区域,同时区分功能风险较高但侵袭证据不足的区域,避免不必要的扩大切除;其二是多区域取样策略的制定,需要在有限时间与组织量约束下,覆盖浸润前沿与核心区的异质性,以提高术后病理与分子分型对全肿瘤的代表性。

仅依赖影像时,影像学边界并不等同于细胞浸润边界,水肿区与非强化区的生物学含义随癌种、治疗状态而显著波动。影像能够提供空间范围,但对浸润梯度、微血管结构与细胞密度等关键组织学信息的指向性有限。仅依赖病理或组学时,证据强度高但空间覆盖受限,无法提供全局浸润图谱,常常只能对局部切缘或少数取样点作出解释,难以直接支撑术中边界调整与取样优先级排序。实体瘤中普遍存在的肿瘤内异质性进一步放大了这一矛盾<sup>[106]</sup>,使单点结论在术后分型与靶点判断上的稳定性受到系统性限制。

#### 2.3.1 算法范式

在不同癌种的手术场景中,融合策略已形成若干具有代表性的临床与算法范式,体现了将边界写成证据链的可行路径。

第1类范式以术中成像增强空间证据,典型代表为荧光引导与术中磁共振。Stummer等<sup>[106]</sup>开展的5-氨基乙酰丙酸荧光引导切除随机对照多中心试验显示,荧光引导能够提高影像学完全切除率,并改善无进展生存相关结局,其本质是将肿瘤代谢相关的荧光信号与术前影像边界融合,从而在术中提供更敏感的空间提示。Senft等<sup>[107]</sup>关于术中MRI的随机对照研究进一步证明,术中MRI作为实时空间复核手段能够提高肿瘤切除范围的达成率,使术前影像边界与术中残留评估形成闭环。这类案例说明,空间证据的迭代更新是实现最大安全切除的重要基础,但其临床可用性需要与组织学与机制证据协同,避免将单一空间信号直接等同于细胞浸润边界。

第2类范式以影像与病理对齐构建区域级真值,用于学习可解释的肿瘤范围映射。在前列腺癌场景中,面向根治术标本的全切片病理与术前多参数MRI的配准被广泛用于建立空间对应关系,从而为影像侧学习肿瘤范围与边界提供高质量监督信号。相关配准与标注映射工作为构建生物学图谱与区域级预测提供了方法基础<sup>[108]</sup>。Priester等<sup>[109]</sup>进一步提出将多模态影像与活检信息联合输入的深度学习模型,生成三维肿瘤范围估计地图与边界信息,用于辅助术前或介入规划,其核心思路是以多源证据共同约束空间范围推断,并以三维地图形式服务临床决策。

第3类范式以术中快速病理或局部成像增强组织学证据,并与影像空间信息闭环。乳腺癌保乳手术中,术中切缘评估直接决定再切除率与局部控制。Veluponnar等<sup>[110]</sup>提出基于术中超声图像的计算机辅助肿瘤边界描绘方法,以病理切缘结果作为金标准评估,实现对阳性或近切缘的预测,并展示了将术中影像与病理结果闭环用于边界决策支持的路径。该类工作强调,边界评估应当形成可反馈的学习系统,使影像侧的边界提示能够被病理结果持续校准,从而提高跨操作者与跨中心的稳定性。

上述3类范式覆盖了空间证据强化、影像病理对齐学习与术中边界闭环校准3条主要路径(见表2),能够在泛癌外科场景中形成可迁移的融合方法学框架<sup>[106]</sup>。

表2 手术边界的区域定义与特异性融合算法设计

Table 2 Regional definition of surgical boundaries and design of a specific fusion algorithm

空间单元	定义方式	常用融合算法	可复核证据	临床应用
病灶级	分割/检测	跨注意力	病理总体负荷	靶病灶随访
栖息地级	参数聚类/分区	门控/MIL注意力	坏死、MVP、密度	取样排序/切除优先
边界前沿级	距离环带	反事实遮挡	浸润梯度	切缘复核/补采
细胞邻域级	细胞图/图划分	图注意力	免疫抑制结构	机制分层讨论
纵向轨迹级	多时点对齐	时序融合	反应异质性	耐受监测/调整

### 2.3.2 临床获益

手术规划阶段的多模态融合不应停留在边界可视化层面,而应形成由空间定位、组织学校准与机制解释共同支撑的证据链。荧光引导与术中MRI展示了空间证据迭代更新对切除范围达成的增益,影像与全切片病理配准及三维范围估计地图展示了区域级真值构建与模型归因的可行性,术中影像与病理结果闭环展示了边界评估的可持续校准路径。未来更具临床可用性的方向在于,将上述路径统一到同一输出范式中,使高价值切除与采样区域具备可复核的证据条目与置信度表达,从而在最大切除与功能保护之间提供可追溯的决策依据。

融合将风险空间化并证据化,有助于更合理地平衡最大切除与功能保护。风险定位明确后,术者可以在高风险亚区优先实现更充分切除,同时对功能风险高但侵袭证据弱的区域采取更审慎策略。融合驱动的系统化多区域取样能够降低遗漏关键亚克隆或关键微环境结构的概率,提高术后分型与靶点判断的可重复性,并为后续放疗靶区与系统治疗分层提供更可靠的组织学与机制基础<sup>[106]</sup>。

### 2.4 系统治疗与疗效分层:融合将疗效预测推进为机制分层与策略提示

系统治疗阶段的核心矛盾在于,疗效并非单一肿瘤负荷变化所能完全刻画。不同病灶、不同亚区在治疗压力下呈现差异性反应与耐受演化,影像提供空间与纵向窗口,病理提供细胞与微环境结构证据,组学提供驱动事件与通路层面的机制解释。多模态融合的应用价值在于将疗效评估从相关性预测提升为证据链推理,使临床能够理解反应差异来自何处、由何种组织学结构支撑、对应哪些可干预机制环节,从而为分层与组合策略讨论提供可复核的证据框架。

多模态融合的优势可在空间、组织学与机制3个层面展开。空间层的关键在于定位主导疗效与耐受的病灶或亚区。影像能够将肿瘤表达为病灶层级与亚区层级的空间单元,并在纵向随访中捕捉各单元的反应轨迹。融合模型通过区域级归因将疗效贡献分解到具体空间单元,用于解释同一患者内部反应不一致的现象,并支持对高风险亚区进行重点监测或补充评估。在免疫治疗场景中,Vanguri等<sup>[99]</sup>提出的DyAM将CT影像、PD-L1免疫组化切片与基因组特征联合建模,显示多模态整合对免疫治疗反应预测具有增益,并提供了跨模态自适应加权机制,有助于把患者级结论拆解为多源证据对同一决策的支持力度。

组织学层可以解释影像变化的生物学含义。病理能够提供肿瘤细胞密度、免疫浸润结构、血管形态与坏死等结构证据,从而区分影像变化中哪些更接近真实肿瘤负荷变化,哪些更可能反映治疗相关反应或微环境重塑。对于新辅助治疗等存在明确病理终点的场景,多模态模型往往以病理完全缓解或残余肿瘤负荷作为监督信号,进而将影像与病理证据对齐,形成解释链。Mao等<sup>[111]</sup>在《*Science Advances*》上报道的乳腺癌新辅助治疗研究构建了多模态自动化系统用于预测病理完全缓解,强调多源信息整合对预测能力的提升,其思路具有代表性。Guo等<sup>[112]</sup>将提出的跨模态深度学习模型用于预测乳腺癌新辅助治疗的病理完全缓解,体现了将不同模态表征对齐并进行跨模态推断的范式。

机制层的关键在于将不敏感或耐受风险指向可干预环节。组学与通路信息能够将疗效差异解释为DNA修复、缺氧与血管生成、免疫抑制与抗原呈递、代谢重编程等方向的机制差异。融合模型若以通路或网络枢纽为输出目标,通常比预测单一基因事件更稳健,也更容易与病理结构证据对齐,从而形成机制分层而非黑箱概率。Wang等<sup>[113]</sup>报道的DeepAFM代表了多模态深度学习用于免疫治疗反应预测并强调生物标志物发现的路径,说明融合不仅提升预测性能,还可用于机制线索挖掘与分层解释。在更通用的多模态融合研究中,Pathomic fusion等框架通过病理与组学交互建模输出关键区域与关键分子特征,为机制分层提供了方法学支撑<sup>[100]</sup>。

#### 2.4.1 临床终点与算法设计

系统治疗场景中,可按临床终点与融合深度分为3类,分别对应空间归因、组织学校准与机制分层的不同侧重。第1类以免疫治疗反应为终点,强调跨模态自适应加权与证据条目化输出。DyAM在非小细胞肺癌队列中整合CT、PD-L1免疫组化切片与基因组信息,展示了多模态融合对免疫治疗反应预测的增益,并提供动态注意力机制以反映不同模态在个体层面的贡献差异,适合用于解释型分层与证据权重呈现<sup>[99]</sup>。DeepAFM同样聚焦免疫治疗反应预测,并强调个体化治疗策略与生物标志物发现,代表了融合用于机制线索挖掘的路径<sup>[113]</sup>。

第2类以新辅助治疗病理完全缓解为终点,强调跨模态对齐与纵向信息整合。Mao等<sup>[111]</sup>将提出的多模态自动化系统用于预测病理完全缓解,体现了将多源信息整合为可执行预测模型的工程化方向。Guo等<sup>[112]</sup>提出的跨模态深度学习模型进一步展示了跨模态学习用于新辅助治疗反应预测的可行性,为空间层与组织学层的闭环验证提供了范式参考。

第3类以放疗或放化疗响应为终点,强调多尺度表征与机制可解释性。Jin等<sup>[114]</sup>报道的多尺度融合模型将不同尺度的特征用于预测放疗响应,体现了将组织学表征与机制线索融入疗效评估的方向,为耐受机制定位与分层提供了方法学路径。

#### 2.4.2 临床获益

系统治疗阶段的多模态融合应以证据链闭环为目标,完成从疗效预测到机制分层与策略提示的迁移<sup>[99]</sup>。多模态融合为组合策略讨论提供线索,其价值不在于直接给出用药建议,而在于形成策略生成的证据框架,例如当主导耐受亚区呈现缺氧与血管生成相关线索且与病理血管结构证据一致时,后续策略讨论可围绕微环境重塑方向展开;当免疫抑制相关结构与通路信号一致时,可围绕免疫微环境调节方向展开。DyAM等研究在免疫治疗反应预测中展示了多源证据整合的增益<sup>[99]</sup>,为这类框架提供了可参考的实现路径。

### 2.5 预后风险分层:融合的优势在于把风险落到空间并干预

预后风险分层在肿瘤全流程中承担枢纽作用,其价值不止于预测生存或复发概率,更在于指导随访与辅助治疗强度,并支撑临床试验的入组与分层。单模态模型往往只能输出1个风险值,难以回答风险来自何处、由何种组织学结构支撑、对应何种机制环节。多模态融合的应用优势在于将风险空间化并证据化,使风险分层从统计预测推进为可复核的证据链推理,并为后续干预讨论提供更具体的靶向语境。

病理-基因-影像融合的第1项优势是风险可定位。融合模型不仅输出风险值或风险组,还应指出风险主要由哪个病灶或哪个亚区主导,并给出与之对应的组织学结构证据与机制证据条目。此类输出将风险从患者级抽象量转化为空间级证据链,便于在术后复盘与随访规划中落地。融合的第2项优势是降低取样偏倚并重建组学信号的空间语境。组学信号在影像空间中被重新解释后,其预后含义更加具体。例如同一条高风险通路如果主要出现在浸润前沿对应的亚区,其临床含义更倾向于边界控制与局部复发风险;若主要出现在坏死或缺氧相关亚区,则更可能提示耐受生态位与治疗抵抗风险。融合由

此把分子信号从孤立向量转化为具有空间定位的机制线索,提高跨个体与跨中心解释的一致性。融合的第3项优势是可行动分层。风险分层结果能够被转译为行动框架,包括重点监测的空间单元、需要重点复核的组织学证据类型,以及需要优先讨论的机制方向。该框架强调证据可复核性与一致性校验,避免将模型输出直接等同于具体治疗决策,而是用于支撑多学科讨论与后续策略选择的证据结构。

### 2.5.1 预后分层的融合算法策略

预后分层的算法案例可按融合深度分为3类,分别为强化风险定位、机制解释与临床可用性。第1类以病理与组学的交互融合为核心,强调机制可解释与跨模态归因。将全切片病理表征与基因组学与转录组学特征进行门控与双线性交互建模,在TCGA的胶质瘤与肾透明细胞癌数据上验证其对生存预测与分级任务的增益,并能够输出关键病理区域与重要分子特征的联合解释,适合将风险拆解为结构证据与机制证据的组合条目<sup>[100]</sup>。第2类以跨注意力为核心,强调对不同信息源的自适应加权与可迁移整合。SurvPGC将病理图像、基因组数据与临床记录联合建模,通过跨注意力模块融合多源表征,并在TCGA数据集验证其预后预测能力,同时提供注意力可视化以揭示不同模式的贡献差异。该类模型对于术后分层的启示在于,临床信息与分子机制并非与病理影像竞争,而是共同构成风险语境,融合应当显式表征其相对贡献并形成可解释证据链<sup>[115]</sup>。第3类面向泛癌生存预测,强调跨癌种一致的风险表征与域偏移鲁棒性。UMPSNet提出统一的多模态泛癌生存网络,针对多癌种数据异质与域偏移问题设计融合框架,并报告了可复现的预后基因特征发现与生存预测增益,代表了从单癌种走向泛癌可迁移风险分层的方向<sup>[116]</sup>。以跨注意力为主线的CATfusion同样聚焦泛癌生存风险,融合病理特征与多组学特征以提升生存预测表现,体现了双向注意力在多组学与病理之间同步表征的可行路径<sup>[117]</sup>。

综合来看,高质量的预后分层不应止于更高的C-index,而应输出可定位的风险来源、可复核的结构证据与可干预的机制线索,并在评价时同时报告区分度、校准与净获益,以形成面向临床决策的完整证据闭环<sup>[118]</sup>。

### 2.5.2 临床获益

多模态融合使患者预后分层更可信,跨模态一致性为风险结论提供结构与机制的双重支撑,降低单模态捷径与单点取样偏倚导致的系统性误差。融合也使分层更可用,空间归因与机制归因将风险转译为可沟通、可复核的证据条目,为随访重点、复核策略与试验分层提供可落地依据,并为后续干预讨论提供明确语境。

## 2.6 结构化报告生成:融合让报告从结论走向证据链

多模态融合在临床落地的关键瓶颈之一,是模型输出与临床沟通语言之间存在显著鸿沟。单模态模型常给出一个分数或一句结论,临床医生需要额外完成解释、定位与复核,才能把模型结果纳入多学科讨论与后续决策。融合模型的独特优势在于天然具备多源证据条目,影像提供可定位的空间单元,病理提供可复核的组织学结构与细胞层证据,组学提供可干预的机制线索。若将这些条目以结构化方式组织,报告生成可从结果陈述升级为证据链呈现,降低误解与过度外推风险,并显著减少临床侧将模型翻译成医学语言的成本。多模态医学人工智能综述指出,跨模态整合的临床价值不仅在于性能提升,更在于形成可解释、可行动的综合输出以支撑真实世界决策流程<sup>[119]</sup>。

### 2.6.1 报告自动生成与系统化输出

从算法发展看,报告生成已从单纯的影像到文本迁移,逐步走向以结构化中间表示或多智能体流程提升可控性与可解释性。数据集与信息抽取范式为放射学报告内容结构化提供了通用框架,使模型可先生成实体与关系,再生成自然语言报告,从而降低幻觉与表达漂移风险。在此基础上,多智能体框架开始将视觉分析、检索、草稿生成、复核与一致性检查拆解为可审计步骤,增强报告的结构化程度与可解释性。病理侧同样出现面向临床报告生成的端到端模型,HistoGPT等工作证明从多张全分辨率

病理图像生成临床级病理报告具有可行性,并指出报告生成可用于降低劳动负担与提升报告一致性<sup>[120]</sup>。这些进展共同支持在融合诊疗系统中采用证据链驱动的结构化报告模板。

面向泛癌融合诊疗应用,结构化报告建议采用统一的证据链框架,核心是将结论、空间证据、组织学证据、机制证据与不确定性提示组织成可复核条目,便于在多学科讨论中逐条核查。

### 2.6.2 临床获益

从应用获益角度看,结构化报告生成将融合模型输出转化为临床可沟通产物,减少医生将模型输出重新组织成医学语言的负担,并提升多学科讨论的效率与一致性。结构化表示与两阶段报告生成范式为放射学报告的可控生成提供了可复用路径,病理侧HistoGPT等工作展示了以临床报告为目标的生成式模型能够输出接近真实报告结构的内容,这为融合系统在放射、病理与组学证据条目化输出上的统一提供了方法学基础。

## 3 总结与展望

肿瘤影像、病理与基因组学的融合正在把智能诊疗从单一模态的预测工具推进为可落地的证据链系统。真正具有临床价值的融合模型不应止步于更高的区分度指标,而应能够输出可验证、可干预的推理结构。所谓可验证,是指模型给出的空间热点、组织学结构与机制线索能够被后续病理、实验或随访结果检验并闭环校准。所谓可干预,是指输出能够直接对接临床动作,例如取样布局、手术与放疗规划、随访策略与试验分层,而不是停留在患者级风险分数。未来的评价标准将从单纯的性能排名转向临床效用导向,重点考察模型能否改善净获益,能否减少不必要的侵入性操作,能否在真实世界流程中降低不确定性并提升决策一致性。

解决取样偏倚将成为下一代融合诊疗系统的核心问题之一。多区域测序与空间组学的发展使得肿瘤的分子状态不再是无空间语境的向量,而是可定位、可分层、可追踪的空间过程。空间转录组、多区域测序、三维病理与数字病理细胞图谱的结合,将推动模态对齐从患者级相关迈向区域级证据一致性,并进一步逼近区域级因果线索。融合系统需要把取样位置、亚区归属与组织学结构显式纳入建模与报告,使模型输出能够说明何处的风险更高、何处的机制更强、何处需要补充证据,从而把无创推断转化为取样与复核策略的优化工具。

纵向建模将成为从静态分型走向动态管理的关键转折点。肿瘤在治疗压力下发生表型与机制漂移,单次手术或单时点组学难以长期代表患者状态。未来的融合模型应当把治疗过程与演化纳入学习目标,围绕不同病灶与不同亚区的反应轨迹建立稳定表征,并将漂移监测作为系统能力的一部分。纵向证据链能够解释同一患者内部反应不一致的来源,区分短期炎症或水肿所致影像变化与真实肿瘤负荷变化,并在机制层面提示耐受方向的演化趋势,从而将疗效评估和风险分层从一次性结论升级为持续更新的临床工具。

基础模型与多中心协作将决定融合系统的可迁移与可维护性。现实世界的的数据稀缺、缺失模态与域偏移不可避免,依赖单中心高质量闭环数据训练的模型难以广泛部署。下一代系统需要以多模态基础表征为起点,通过少样本适配与跨中心验证形成稳定的可迁移能力,并把数据标准、对齐策略、模型更新与漂移监测纳入工程化体系。可泛化不仅要求跨设备、跨协议、跨人群的稳健表现,也要求模型在缺失模态情况下仍能给出合理的不确定性表达与复核建议,避免在证据不足时输出过度自信的结论。

面向临床转化,落脚点应回到一份可执行清单。数据层面需要明确对齐层级与对齐误差来源,报告缺失模态比例及其对性能与不确定性的影响,并采用严格的外部验证设置而非仅内部交叉验证。模型层面应优先支持区域级融合与证据条目化输出,配套不确定性估计与一致性校验,使输出能够被复核并触发补采或复检。评价层面应形成校准、净获益与临床可行动输出的联合标准,确保模型不仅能

预测,更能以结构化报告、分层建议与取样导航的形式进入多学科讨论与真实决策流程。上述要点共同指向同一结论,肿瘤多模态智能诊疗的下一步不是更复杂的网络结构,而是从模型走向证据链系统,从预测走向可验证与可干预的临床决策支持。

#### 参考文献:

- [1] AKHTAR M S, SWAMY M K. Anticancer plants: Clinical trials and nanotechnology[M]. Singapore: Springer Singapore, 2017: 1-3.
- [2] HANAHAN D, WEINBERG R A. Hallmarks of cancer: The next generation[J]. *Cell*, 2011, 144(5): 646-674.
- [3] 贾馨竹, 金铮, 朱波. 肿瘤治疗中的科学困境与精准治疗未来路径探讨[J]. *中国科学基金*, 2025, 39(1): 24-33.  
JIA Qingzhu, JIN Zheng, ZHU Bo. Dilemmas in cancer treatment and exploration of pathways for precision therapy[J]. *Bulletin of National Natural Science Foundation of China*, 2025, 39(1): 24-33.
- [4] SUN P, BUSH S J, WANG S, et al. STMiner: Gene-centric spatial transcriptomics for deciphering tumor tissues[J]. *Cell Genomics*, 2025, 5(2): 100771.
- [5] MO C K, LIU J, CHEN S, et al. Tumour evolution and microenvironment interactions in 2D and 3D space[J]. *Nature*, 2024, 634(8036): 1178-1186.
- [6] 石峥. 辐射调节非小细胞肺癌异质性和肿瘤微环境的研究[D]. 兰州: 中国科学院大学(中国科学院近代物理研究所), 2026.  
SHI Zheng. Study on the radiation-regulated heterogeneity and tumor microenvironment in non-small cell lung cancer[D]. Lanzhou: University of Chinese Academy of Sciences (Institute of Modern Physics, Chinese Academy of Sciences), 2026.
- [7] CHEN H, LI Y, LI H, et al. NBS1 lactylation is required for efficient DNA repair and chemotherapy resistance[J]. *Nature*, 2024, 631(8021): 663-669.
- [8] NICOLAS A M, PESIC M, ENGEL E, et al. Inflammatory fibroblasts mediate resistance to neoadjuvant therapy in rectal cancer[J]. *Cancer Cell*, 2022, 40(2): 168-184.
- [9] 宋国利, 陈杰. 病理图像分析的深度学习研究方法研究综述[J]. *中国科学基金*, 2022, 36(2): 225-234.  
SONG Guoli, CHEN Jie. A review of deep learning methods for pathology image analysis[J]. *Bulletin of National Natural Science Foundation of China*, 2022, 36(2): 225-234.
- [10] LI Z C, QIN W, LIANG D, et al. Biomedical AI: Evolving from digital to physical and biological intelligence[J]. *The Innovation Informatics*, 2025, 1(1): 100010.
- [11] HUANG T, XU H, WANG H, et al. Artificial intelligence for medicine: Progress, challenges, and perspectives[J]. *The Innovation Medicine*, 2023, 1(2): 100030.
- [12] ARDILA D, KIRALY A P, BHARADWAJ S, et al. End-to-end lung cancer screening with three-dimensional deep learning on low-dose chest computed tomography[J]. *Nature Medicine*, 2019, 25(6): 954-961.
- [13] MCKINNEY S M, SIENIEK M, GODBOLE V, et al. International evaluation of an AI system for breast cancer screening[J]. *Nature*, 2020, 577(7788): 89-94.
- [14] LITJENS G, KOOI T, BEJNORDI B E, et al. A survey on deep learning in medical image analysis[J]. *Medical Image Analysis*, 2017, 42: 60-88.
- [15] ZHANG Z, WANG Z, LI R, et al. Machine learning model on multi-omics data enables risk stratification and identifies molecular heterogeneity and therapeutic targets in glioblastoma[J]. *Molecular Cancer*, 2026. <https://doi.org/10.1186/s12943-026-02637-2>.
- [16] LAI Q, VONG C M, CHEN C. Weakly supervised semantic segmentation via dual-stream contrastive learning of cross-image contextual information[J]. *IEEE Transactions on Industrial Informatics*, 2024, 20(10): 11635-11643.
- [17] 杨农. 基于影像组学和基因组学的非小细胞肺癌治疗疗效预测研究[D]. 长沙: 中南大学, 2025.  
YANG Nong. Prediction of therapeutic efficacy in non-small cell lung cancer treatment based on radiomics and genomics[D]. Changsha: Central South University, 2025.
- [18] 白玉. 标注受限场景下的病理图像分析算法研究[D]. 北京: 北京邮电大学, 2026.  
BAI Yu. Research on pathology image analysis algorithms under annotation-limited scenarios[D]. Beijing: Beijing University of

- Posts and Telecommunications, 2026.
- [19] GARG V, AGARWAL A, SACHDEVA K, et al. Comparison of intra- and interobserver reliability analysis of orientation proximal femur physis parameters on radiographs[J]. *Journal of Clinical Orthopaedics and Trauma*, 2025, 68: 103093.
- [20] WANG X, WANG Z, WANG W, et al. IDH-mutant glioma risk stratification via whole slide images: Identifying pathological feature associations[J]. *iScience*, 2025, 28(1): 111605.
- [21] 张滨宇. 癌症智能医疗辅助关键技术研究[D]. 北京: 北京邮电大学, 2026.  
ZHANG Binyu. Research on key technologies of intelligent medical assistance for cancer[D]. Beijing: Beijing University of Posts and Telecommunications, 2026.
- [22] WANG D, LIU B, ZHANG Z. Accelerating the understanding of cancer biology through the lens of genomics[J]. *Cell*, 2023, 186(8): 1755-1771.
- [23] TIAN R, HOU F, ZHANG H, et al. Multimodal fusion model for prognostic prediction and radiotherapy response assessment in head and neck squamous cell carcinoma[J]. *npj Digital Medicine*, 2025, 8(1): 302.
- [24] SIAM A, ALSAIFY A R, MOHAMMAD B, et al. Multimodal deep learning for liver cancer applications: A scoping review [J]. *Frontiers in Artificial Intelligence*, 2023, 6: 1247195.
- [25] KEYL J, KEYL P, MONTAVON G, et al. Decoding pan-cancer treatment outcomes using multimodal real-world data and explainable artificial intelligence[J]. *Nature Cancer*, 2025, 6(2): 307-322.
- [26] ZHAO Y, LI L, YU X, et al. SurvGraph: A hybrid-graph attention network for survival prediction using whole slide pathological images in gastric cancer[J]. *Neural Networks*, 2025, 189: 107607.
- [27] TANG Z, LI J, CHENG J, et al. Pre-operative overall survival prediction of diffuse glioma enhanced by longitudinal data[J]. *IEEE Journal of Biomedical and Health Informatics*, 2025, 29(7): 5158-5167.
- [28] WANG W, ZHAO Y, TENG L, et al. Neuropathologist-level integrated classification of adult-type diffuse gliomas using deep learning from whole-slide pathological images[J]. *Nature Communications*, 2023, 14(1): 6359.
- [29] DUAN J, ZHAO Y, SUN Q, et al. Imaging-proteomic analysis for prediction of neoadjuvant chemotherapy responses in patients with breast cancer[J]. *Cancer Medicine*, 2023, 12(23): 21256-21269.
- [30] DOAN L M T, SHAHHOSSEINI K, VERMA S, et al. Bridging modalities with AI: A review of AI advances in multimodal biomedical imaging[J]. *Communications Engineering*, 2026, 5(1): 30.
- [31] TRUHN D, ECKARDT J N, FERBER D, et al. Large language models and multimodal foundation models for precision oncology[J]. *npj Precision Oncology*, 2024, 8(1): 72.
- [32] DELLAMONICA D, RUAU D, GRIFFITHS B, et al. The AI revolution: How multimodal intelligence will reshape the oncology ecosystem[J]. *npj Artificial Intelligence*, 2025, 1(1): 40.
- [33] BOEHM K M, KHOSRAVI P, VANGURI R, et al. Harnessing multimodal data integration to advance precision oncology [J]. *Nature Reviews Cancer*, 2022, 22(2): 114-126.
- [34] STEYAERT S, PIZURICA M, NAGARAJ D, et al. Multimodal data fusion for cancer biomarker discovery with deep learning[J]. *Nature Machine Intelligence*, 2023, 5(4): 351-362.
- [35] 国家卫生健康委, 国家发展改革委, 教育部, 等. 关于印发健康中国行动——癌症防治行动实施方案(2023—2030年)的通知[EB/OL]. (2023-10-30)[2026-02-18]. [https://www.gov.cn/zhengce/zhengceku/202311/content\\_6915380.htm](https://www.gov.cn/zhengce/zhengceku/202311/content_6915380.htm).
- [36] 国务院办公厅. 国务院办公厅关于印发“十四五”国民健康规划的通知[EB/OL]. (2022-04-27)[2026-02-18]. [https://www.gov.cn/zhengce/content/2022-05/20/content\\_5691424.htm](https://www.gov.cn/zhengce/content/2022-05/20/content_5691424.htm).
- [37] XU X, LI J, ZHU Z, et al. A comprehensive review on synergy of multi-modal data and AI technologies in medical diagnosis [J]. *Bioengineering*, 2024, 11(3): 219.
- [38] KATTI G, ARA S A, SHIREEN A. Magnetic resonance imaging (MRI)—A review[J]. *International Journal of Dental Clinics*, 2011, 3(1): 65-70.
- [39] 付灵傲. 基于深度学习与多示例学习的病理图像细胞识别[D]. 成都: 电子科技大学, 2022.  
FU Ling'ao. Recognition of cell in pathological image based on deep learning and MIL[D]. Chengdu: University of Electronic Science and Technology of China, 2022.
- [40] AMETAJ B N. Periparturient diseases of dairy cows[M]. Cham: Springer International Publishing, 2017: 1-6.
- [41] CHAKRABORTY S, HOSEN MD I, AHMED M, et al. Onco-multi-OMICS approach: A new frontier in cancer research[J].

- BioMed Research International, 2018, 2018: 1-14.
- [42] ZECH J R, BADGELEY M A, LIU M, et al. Variable generalization performance of a deep learning model to detect pneumonia in chest radiographs: A cross-sectional study[J]. *PLOS Medicine*, 2018, 15(11): e1002683.
- [43] ZHANG M, BAO Y, RUI W, et al. Performance of 18F-FDG PET/CT radiomics for predicting EGFR mutation status in patients with non-small cell lung cancer[J]. *Frontiers in Oncology*, 2020, 10: 568857.
- [44] BI W L, HOSNY A, SCHABATH M B, et al. Artificial intelligence in cancer imaging: Clinical challenges and applications[J]. *CA: A Cancer Journal for Clinicians*, 2019, 69(2): 127-157.
- [45] WANG J W, MENG M, DAI M W, et al. Correlation does not equal causation: The imperative of causal inference in machine learning models for immunotherapy[J]. *Frontiers in Immunology*, 2025, 16: 1630781.
- [46] KARAMBELKAR V, BORKAR P, PAUL A, et al. AI-assisted robotic systems for brain tumor biopsy procedures[C]// *Proceedings of 2025 International Conference on Next Generation of Green Information and Emerging Technologies (GIET)*. Gunupur, India: IEEE, 2025: 1-5.
- [47] XIANG J, WANG X, ZHANG X, et al. A vision-language foundation model for precision oncology[J]. *Nature*, 2025, 638(8051): 769-778.
- [48] SUJIT S J, AMINU M, KARPINETS T V, et al. Enhancing NSCLC recurrence prediction with PET/CT habitat imaging, ctDNA, and integrative radiogenomics-blood insights[J]. *Nature Communications*, 2024, 15(1): 3152.
- [49] SHANG Y, ZENG Y, LUO S, et al. Habitat imaging with tumoral and peritumoral radiomics for prediction of lung adenocarcinoma invasiveness on preoperative chest CT: A multicenter study[J]. *American Journal of Roentgenology*, 2024, 223(4): e2431675.
- [50] MARRA A, MORGANTI S, PAREJA F, et al. Artificial intelligence entering the pathology arena in oncology: Current applications and future perspectives[J]. *Annals of Oncology*, 2025, 36(7): 712-725.
- [51] CEN X, ZOU J, TONG Y, et al. Redefining imaging genomics for the next decade[J]. *Science Bulletin*, 2025, 70(22): 3720-3723.
- [52] GUO Y, LI T, GONG B, et al. From images to genes: Radiogenomics based on artificial intelligence to achieve non-invasive precision medicine in cancer patients[J]. *Advanced Science*, 2025, 12(2): 2408069.
- [53] MA J W, JIANG X, WANG Y M, et al. Dual-energy CT-based radiomics in predicting EGFR mutation status non-invasively in lung adenocarcinoma[J]. *Heliyon*, 2024, 10(2): e24372.
- [54] PADMAJA C, RAMACHARAN S, BALA KRISHNAN S, et al. Breast cancer diagnostics: Integrating deep learning and radiogenomics[C]// *Proceedings of 2025 Third International Conference on Networks, Multimedia and Information Technology (NMITCON)*. Bengaluru, India: IEEE, 2025: 1-5.
- [55] ABDELAZIZ N M, DAWOOD E A A, TANTAWY A A. Integrating deep learning and radiogenomics: A novel approach to glioblastoma segmentation and MGMT methylation prediction[J]. *Journal of Imaging*, 2025, 11(11): 403.
- [56] YAN J, ZHAO Y, CHEN Y, et al. Deep learning features from diffusion tensor imaging improve glioma stratification and identify risk groups with distinct molecular pathway activities[J]. *eBioMedicine*, 2021, 72: 103583.
- [57] SUN Q, CHEN Y, LIANG C, et al. Biologic pathways underlying prognostic radiomics phenotypes from paired MRI and RNA sequencing in glioblastoma[J]. *Radiology*, 2021, 301(3): 654-663.
- [58] WANG J, DAI J, CHENG Y, et al. Advances in the use of radiomics and pathomics for predicting the efficacy of neoadjuvant therapy in tumors[J]. *Translational Oncology*, 2025, 58: 102435.
- [59] ZHAO Y, LIU F, ZHU C, et al. Integrating radiology and histology via co-attention deep learning for predicting progression-free survival in patients with metastatic prostate cancer[J]. *Chinese Medical Journal*, 2025, 138(22): 3013-3015.
- [60] LIN H, HUA J, GONG Z, et al. Multimodal radiopathological integration for prognosis and prediction of adjuvant chemotherapy benefit in resectable lung adenocarcinoma: A multicentre study[J]. *Cancer Letters*, 2025, 616: 217557.
- [61] XIE Q, ZHAO Z, YANG Y, et al. A clinical-radiomic-pathomic model for prognosis prediction in patients with hepatocellular carcinoma after radical resection[J]. *Cancer Medicine*, 2024, 13(11): e7374.
- [62] MAO H M, ZHANG J J, ZHU B, et al. A novel deep learning radiopathomics model for predicting carcinogenesis promotor cyclooxygenase-2 expression in common bile duct in children with pancreaticobiliary maljunction: A multicenter study[J]. *Insights into Imaging*, 2025, 16(1): 74.

- [63] 张雪芹, 李悦欣, 刘畅, 等. 融合病理图像和基因组学多模态的癌症生存预测[J]. 华东理工大学学报(自然科学版), 2025, 51(4): 505-513.  
ZHANG Xueqin, LI Yuexin, LIU Chang, et al. Cancer survival prediction by integrating multimodal pathological images and genomics[J]. Journal of East China University of Science and Technology, 2025, 51(4): 505-513.
- [64] ZHANG Z, ZHAO Y, DUAN J, et al. Drop and align: Fusing pathology and genomics via graph learning for cancer survival prediction[C]//Proceedings of 2025 IEEE 22nd International Symposium on Biomedical Imaging (ISBI). Houston, USA: IEEE, 2025: 1-5.
- [65] CHEN R J, LU M Y, WENG W H, et al. Multimodal co-attention transformer for survival prediction in gigapixel whole slide images[C]//Proceedings of 2021 IEEE/CVF International Conference on Computer Vision (ICCV). Montreal, Canada: IEEE, 2021: 3995-4005.
- [66] YAN R, ZHANG X, JIANG Z, et al. Pathway-aware multimodal transformer (PAMT): Integrating pathological image and gene expression for interpretable cancer survival analysis[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2026, 48(1): 896-913.
- [67] WAQAS A, TRIPATHI A, RAMACHANDRAN R P, et al. Multimodal data integration for oncology in the era of deep neural networks: A review[J]. Frontiers in Artificial Intelligence, 2024, 7: 1408843.
- [68] MTOOR S, RASHIDIAN N, MESSAOUDI N, et al. Integrating genomics, radiomics, and pathomics in oncology: A scoping review and a framework for AI-enabled surgomics[J]. Bioengineering, 2026, 13(1): 117.
- [69] LIU Z, WU Y, XU H, et al. Multimodal fusion of radio-pathology and proteogenomics identify integrated glioma subtypes with prognostic and therapeutic opportunities[J]. Nature Communications, 2025, 16(1): 3510.
- [70] BRAMAN N, GORDON J W H, GOOSSENS E T, et al. Deep orthogonal fusion: Multimodal prognostic biomarker discovery integrating radiology, pathology, genomic, and clinical data[C]//Proceedings of International Conference on Medical Image Computing and Computer-Assisted Intervention. Cham: Springer International Publishing, 2021: 667-677.
- [71] LIU J, CEN X, YI C, et al. Challenges in AI-driven biomedical multimodal data fusion and analysis[J]. Genomics, Proteomics & Bioinformatics, 2025, 23(1): qzaf011.
- [72] CHAABENE S, BOUDAYA A, BOUAZIZ B, et al. An overview of methods and techniques in multimodal data fusion with application to healthcare[J]. International Journal of Data Science and Analytics, 2025, 20(4): 3093-3117.
- [73] ZHONG Z, SCHNEIDER D, VOIT M, et al. Anticipative feature fusion transformer for multi-modal action anticipation[C]//Proceedings of 2023 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV). Waikoloa, USA: IEEE, 2023: 6057-6066.
- [74] ATTALLAH O, ANWAR F, GHANEM N M, et al. Histo-CADx: Duo cascaded fusion stages for breast cancer diagnosis from histopathological images[J]. PeerJ Computer Science, 2021, 7: e493.
- [75] SCHULTE-SASSE R, BUDACH S, HNISZ D, et al. Integration of multiomics data with graph convolutional networks to identify new cancer genes and their associated molecular mechanisms[J]. Nature Machine Intelligence, 2021, 3(6): 513-526.
- [76] KHORASANI A, DEMAZURE T. Cross-temporal attention fusion (CTAF) for multimodal physiological signals in self-supervised learning[EB/OL]. (2026-02-02). <https://doi.org/10.48550/arXiv.2602.02784>.
- [77] 任泽裕, 王振超, 柯尊旺, 等. 多模态数据融合综述[J]. 计算机工程与应用, 2021, 57(18): 49-64.  
REN Zeyu, WANG Zhenchao, KE Zunwang, et al. Survey of multimodal data fusion[J]. Computer Engineering and Applications, 2021, 57(18): 49-64.
- [78] TORTORA M, CORDELLI E, SICILIA R, et al. RadioPathomics: Multimodal learning in non-small cell lung cancer for adaptive radiotherapy[J]. IEEE Access, 2023, 11: 47563-47578.
- [79] HASSAN M, ALI S, ALQUHAYZ H, et al. Developing liver cancer drug response prediction system using late fusion of reduced deep features[J]. Journal of King Saud University—Computer and Information Sciences, 2022, 34(10): 8122-8135.
- [80] TRIPATHI P C, BAG S. A computer-aided grading of glioma tumor using deep residual networks fusion[J]. Computer Methods and Programs in Biomedicine, 2022, 215: 106597.
- [81] CUI C, YANG H, WANG Y, et al. Deep multimodal fusion of image and non-image data in disease diagnosis and prognosis: A review[J]. Progress in Biomedical Engineering, 2023, 5(2): 022001.
- [82] STAHLSCHEMIDT S R, ULFENBORG B, SYNNERGREN J. Multimodal deep learning for biomedical data fusion: A

- review[J]. *Briefings in Bioinformatics*, 2022, 23(2): bbab569.
- [83] GUARRASI V, AKSU F, CARUSO C M, et al. A systematic review of intermediate fusion in multimodal deep learning for biomedical applications[J]. *Image and Vision Computing*, 2025, 158: 105509.
- [84] NAGRANI A, YANG S, ARNAB A, et al. Attention bottlenecks for multimodal fusion[EB/OL]. (2022-11-30). <https://doi.org/10.48550/arXiv.2107.00135>.
- [85] CIERNIK L, MORIK M, THEDE L, et al. Beyond the final layer: Attentive multilayer fusion for vision transformers[EB/OL]. (2026-01-14). <https://doi.org/10.48550/arXiv.2026.09322>.
- [86] ZHAO M, HU W, WANG J, et al. Making every head count: Sparse attention without the speed-performance trade-off[EB/OL]. (2025-11-27). <https://doi.org/10.48550/arXiv.2511.09596>.
- [87] LI L, FAN C, ZHANG H, et al. MHANet: Multi-scale hybrid attention network for auditory attention detection[EB/OL]. (2025-05-21). <https://doi.org/10.48550/arXiv.2505.15364>.
- [88] JHA N K, REAGEN B. Regularizing the entropy landscape of self-attention: Towards a soft inductive bias in LLMs[C]// *Proceedings of OPT 2025: Optimization for Machine Learning*. [S.l.]: [s.n.], 2025.
- [89] VELIČKOVIĆ P, CUCURULL G, CASANOVA A, et al. Graph attention networks[EB/OL]. (2018-02-04). <https://doi.org/10.48550/arXiv.1710.10903>.
- [90] SUN C, LI C, LIN X, et al. Attention-based graph neural networks: A survey[J]. *Artificial Intelligence Review*, 2023, 56(S2): 2263-2310.
- [91] CHEN J, QIAO Z, ZHANG Q, et al. G-merging: Graph model merging for parameter-efficient multi-task knowledge consolidation[C]// *Proceedings of the Fourteenth International Conference on Learning Representations*. [S.l.]: [s.n.], 2026.
- [92] WU G, LIU H, FU H, et al. Every SAM drop counts: Embracing semantic priors for multi-modality image fusion and beyond [C]// *Proceedings of the Computer Vision and Pattern Recognition Conference*. Piscataway, NJ, USA: IEEE, 2025: 17882-17891.
- [93] ZHANG Z, ZHAO Y, DUAN J, et al. Pathology-genomic fusion via biologically informed cross-modality graph learning for survival analysis[EB/OL]. (2024-04-11). <https://arXiv.org/pdf/2404.08023>.
- [94] SUN W, ZHANG P, XU J, et al. Bridging attention fusion-based multi-view graph neural networks for spatial gene expression prediction[J]. *Pattern Recognition*, 2026, 171: 112207.
- [95] LUO B, TENG F, TANG G, et al. StereoMM: A graph fusion model for integrating spatial transcriptomic data and pathological images[J]. *Briefings in Bioinformatics*, 2025, 26(3): bbaf210.
- [96] GERLINGER M, ROWAN A J, HORSWELL S, et al. Intratumor heterogeneity and branched evolution revealed by multiregion sequencing[J]. *New England Journal of Medicine*, 2012, 366(10): 883-892.
- [97] ZWANENBURG A, VALLIÈRES M, ABDALAH M A, et al. The image biomarker standardization initiative: Standardized quantitative radiomics for high-throughput image-based phenotyping[J]. *Radiology*, 2020, 295(2): 328-338.
- [98] DAGOGO-JACK I, SHAW A T. Tumour heterogeneity and resistance to cancer therapies[J]. *Nature Reviews Clinical Oncology*, 2018, 15(2): 81-94.
- [99] VANGURI R S, LUO J, AUKERMAN A T, et al. Multimodal integration of radiology, pathology and genomics for prediction of response to PD-(L)1 blockade in patients with non-small cell lung cancer[J]. *Nature Cancer*, 2022, 3(10): 1151-1164.
- [100] CHEN R J, LU M Y, WANG J, et al. Pathomic fusion: An integrated framework for fusing histopathology and genomic features for cancer diagnosis and prognosis[J]. *IEEE Transactions on Medical Imaging*, 2022, 41(4): 757-770.
- [101] GAO Y, VENTURA-DIAZ S, WANG X, et al. An explainable longitudinal multi-modal fusion model for predicting neoadjuvant therapy response in women with breast cancer[J]. *Nature Communications*, 2024, 15(1): 9613.
- [102] MENDES SERRÃO E, KLUG M, MOLONEY B M, et al. Current status of cancer genomics and imaging phenotypes: What radiologists need to know[J]. *Radiology: Imaging Cancer*, 2023, 5(6): e220153.
- [103] MORENO S, BONFANTE M, ZUREK E, et al. A radiogenomics ensemble to predict EGFR and KRAS mutations in NSCLC[J]. *Tomography*, 2021, 7(2): 154-168.
- [104] HOIVIK E A, HODNELAND E, DYBVIK J A, et al. A radiogenomics application for prognostic profiling of endometrial cancer[J]. *Communications Biology*, 2021, 4(1): 1363.

- [105] ZHANG X, LAI Q, CAO J, et al. Radiogenomics: Current understandings and future perspectives[J]. MedComm, 2026, 7(2): e70583.
- [106] STUMMER W, PICHLMEIER U, MEINEL T, et al. Fluorescence-guided surgery with 5-aminolevulinic acid for resection of malignant glioma: A randomised controlled multicentre phase III trial[J]. The Lancet Oncology, 2006, 7(5): 392-401.
- [107] SENFT C, BINK A, FRANZ K, et al. Intraoperative MRI guidance and extent of resection in glioma surgery: A randomised, controlled trial[J]. The Lancet Oncology, 2011, 12(11): 997-1003.
- [108] LI L, SHIRADKAR R, GOTTLIEB N, et al. Multi-scale statistical deformation based co-registration of prostate MRI and post-surgical whole mount histopathology[J]. Medical Physics, 2024, 51(4): 2549-2562.
- [109] PRIESTER A, FAN R E, SHUBERT J, et al. Prediction and mapping of intraprostatic tumor extent with artificial intelligence [J]. European Urology Open Science, 2023, 54: 20-27.
- [110] VELUPONNAR D, DE BOER L L, GELDOF F, et al. Toward intraoperative margin assessment using a deep learning-based approach for automatic tumor segmentation in breast lumpectomy ultrasound images[J]. Cancers, 2023, 15(6): 1652.
- [111] MAO N, DAI Y, ZHOU H, et al. A multimodal and fully automated system for prediction of pathological complete response to neoadjuvant chemotherapy in breast cancer[J]. Science Advances, 2025(18): eadr1576.
- [112] GUO J, CHEN B, CAO H, et al. Cross-modal deep learning model for predicting pathologic complete response to neoadjuvant chemotherapy in breast cancer[J]. npj Precision Oncology, 2024, 8(1): 189.
- [113] WANG Z, LIU X, HAN K, et al. Multimodal deep learning for immunotherapy response prediction and biomarker discovery in non-small cell lung cancer[J]. Journal of the American Medical Informatics Association, 2025, 32(11): 1641-1653.
- [114] JIN Y, ARIMURA H, IWASAKI T, et al. Multiscale fusion models with genomic, topological, and pathomic features to predict response to radiation therapy for non-small cell lung cancer patients[J]. Laboratory Investigation, 2025, 105(10): 104204.
- [115] HOU J, ZHANG R, XIE Y, et al. Multimodal deep learning for cancer prognosis prediction with clinical information prompts integration[J]. npj Digital Medicine, 2025, 9(1): 76.
- [116] ZHANG B, LI S, JIAN J, et al. From single-cancer to pan-cancer prognosis[J]. The American Journal of Pathology, 2025, 195(10): 1869-1884.
- [117] HU Y, LI X, YI Y, et al. Deep learning-driven survival prediction in pan-cancer studies by integrating multimodal histology-genomic data[J]. Briefings in Bioinformatics, 2025, 26(2): bbaf121.
- [118] VICKERS A J, ELKIN E B. Decision curve analysis: A novel method for evaluating prediction models[J]. Medical Decision Making, 2006, 26(6): 565-574.
- [119] ACOSTA J N, FALCONE G J, RAJPURKAR P, et al. Multimodal biomedical AI[J]. Nature Medicine, 2022, 28(9): 1773-1784.
- [120] TRAN M, SCHMIDLE P, GUO R R, et al. Generating dermatopathology reports from gigapixel whole slide images with HistoGPT[J]. Nature Communications, 2025, 16(1): 4886.

#### 作者简介:



段静娴(1991-),女,副研究员,硕士生导师,研究方向:医学人工智能,E-mail: jx.duan@siat.ac.cn.



赵源深(1988-),男,副研究员,硕士生导师,研究方向:医学人工智能。



康梁钰淇(2000-),男,研究实习员,研究方向:医学人工智能。



梁栋(1977-),男,研究员,博士生导师,研究方向:生物医学成像的交叉研究。



郑海荣(1977-),男,中国科学院院士,研究员,博士生导师,研究方向:医学成像技术及先进医疗设备研发、磁共振和超声成像。



李志成(1982-),通信作者,男,研究员,博士生导师,研究方向:医学人工智能,E-mail: zc.li@siat.ac.cn.

(编辑:王静)