

医学 AI 领域的多模态医学数据融合与理解

摘要

医学人工智能（AI）领域正经历一场深刻的变革，其中多模态医学数据融合与理解已成为推动精准医疗、个性化治疗及新药研发的关键驱动力。本报告旨在全面调研当前医学 AI 领域中多模态数据融合与理解的相关问题，深入剖析其问题概要、背景与应用价值、研究现状与技术瓶颈、过往研究里程碑以及产业化应用案例。通过整合分析医学影像、电子健康记录、基因组学数据、临床文本等多源异构信息，多模态 AI 展现出超越单模态分析的巨大潜力。以图神经网络（GNN）、Transformer 等为代表的新兴模型在处理复杂医学数据、捕捉模态间深层关联方面取得了显著进展。然而，数据异构性、模态缺失、模型可解释性、数据隐私安全以及临床集成等挑战仍是该领域亟待突破的瓶颈。产业界已涌现出一批专注于多模态医疗 AI 解决方案的企业，并在肿瘤学、放射学等领域实现了初步应用，预示着广阔的市场前景。未来，该领域的发展将更侧重于可解释 AI、因果推断、医学基础模型以及临床转化研究，旨在构建更智能、更可靠、更易于临床应用的医学 AI 系统，最终惠及广大患者。

一、问题概要

1.1 定义多模态医学数据融合与理解

在当代医学实践与研究中，数据呈现出前所未有的多样性与复杂性。多模态医学数据融合（Multimodal Medical Data Fusion）是指系统性地整合来自不同来源、具有不同结构与特性的医学信息的过程。这些信息来源广泛，涵盖了医学影像（如 X 射线、计算机断层扫描（CT）、磁共振成像（MRI）、正电子发射断层扫描（PET））、电子健康记录（EHRs）、实验室检测结果、基因组与蛋白质组数据、生理信号（如心电图（ECG）、脑电图（EEG））、病理报告以及来自可穿戴设备的环境与行为数据等¹。这种多模态特性源于对患者健康状况进行全面评估的需求，因为单一类型的数据往往只能反映问题的某个侧面。例如，¹和²提及，“多模态的概念”源于“从图片、档案、通信系统、电子健康记录、在线文档、放射学报告和具有特定数值信息的不同风格的临床记录中获取数据”。这

种数据的多样性不仅带来了机遇，也对数据分析技术提出了新的要求。多模态机器学习（MMML）作为一门专注于分析多源数据的学科，旨在通过整合这些信息以增强算法的灵活性和分析深度³。

进一步而言，多模态医学数据的理解（Understanding）超越了简单的数据合并。它强调运用人工智能，特别是机器学习和深度学习模型，从融合后的、往往是高维且复杂的集成数据中提取有意义的模式、深层关联和临床知识。其最终目标是支持更为精准的临床决策，改进疾病诊断的准确性和效率，优化预后预测的可靠性，并为实现个性化治疗方案提供坚实的数据基础³。例如，⁴指出，多模态 AI 能够帮助医疗系统“共同理解各种类型的信息”，从而实现“更快的诊断、更准确的预测和一致的护理”。因此，融合是实现理解的手段，而理解则是将原始数据转化为可操作临床智能的关键。

1.2 该问题的核心挑战与意义

尽管多模态数据融合与理解带来了巨大的潜力，但在实际应用中面临着诸多核心挑战。首先，**数据异构性**是一个根本性难题。不同模态的数据在格式、尺度、维度、噪声水平、数据质量、完整性乃至语义层面都存在显著差异⁵，这使得直接整合极为困难。其次，有效捕捉和建模不同数据模态之间复杂的**非线性、高阶交互关系**对模型设计提出了高要求。再次，选择**最优的数据融合策略**——何时以及如何融合（例如，早期在特征层面融合，中期在模型内部融合，或晚期在决策层面融合）——是影响最终性能的关键，但尚无普适性答案²。此外，许多先进的深度学习模型（如 Transformer、GNN）虽然性能强大，但其“黑箱”特性导致**模型可解释性**较差，这在要求高可靠性和高风险性的临床决策中是一个主要障碍⁵。最后，**数据量与质量**问题，包括小样本学习的挑战、数据标注成本高昂、部分模态数据缺失或不完整等，也严重制约了模型的训练效果和泛化能力⁵。

这些挑战的复杂性在于，医学数据的多样性既是其价值的源泉，也是技术实现的主要障碍。各种数据类型（如影像、基因组、临床记录）本身具有独特的结构和信息含量¹。将这些本质上不同的数据流有效地结合起来，以形成一个统一且信息更丰富的表征，同时确保模型能够从中学习到有意义的跨模态关联，是当前研究的核心。例如，影像数据提供了

形态学信息，基因组数据揭示了分子层面的特征，而临床记录则包含了患者的宏观表型和病史。只有成功克服异构性带来的整合难题，才能真正释放多模态数据“1+1>2”的潜力，否则可能仅仅是噪声的叠加。

尽管挑战重重，但成功实现多模态医学数据融合与理解的核心意义极为深远。它有望**显著提升临床诊疗的精准度**。通过整合来自多个信息维度的互补证据，可以克服单一数据源的局限性，从而在疾病的早期诊断、精确分型、个性化风险评估和预后预测方面取得突破⁴。例如，MMHA 算法通过整合多组学数据，在 LGG 数据集上的分类准确率提升了 4.5%，为精准医疗提供了新工具⁸。其次，它是**实现个性化医疗的关键赋能技术**。通过构建患者全面的多维健康画像，可以为每位患者量身定制最优的治疗方案，最大化治疗效果，同时最小化潜在副作用⁷。再者，它能够**加速创新药物的研发进程**。通过整合药物特性、生物效应、临床试验等多源数据，有助于发现新的药物靶点、预测药物反应性，从而缩短研发周期，降低研发成本⁹。此外，多模态 AI 还有助于**优化医疗资源分配和提升服务效率**，例如通过辅助医生进行复杂决策，自动化部分诊断流程，从而减轻医生的工作负荷，尤其是在数据密集型科室如放射科¹。正如²所述，“在人工智能驱动下，医学领域多模态的内在益处对疾病诊断框架的性能产生重大影响”。

二、问题背景及应用价值

2.1 医学数据多模态化的趋势与驱动因素

近年来，医学领域的数据以前所未有的速度和规模持续增长，呈现出显著的多模态化趋势。这一趋势主要由以下几个关键因素驱动：

首先，**技术的飞速进步**是主要推手。高通量测序技术（如基因组学、转录组学、蛋白质组学）的成熟与成本下降，使得大规模获取分子层面数据成为可能⁹。医学影像技术的不断创新（如更高分辨率的 MRI、功能性 PET、三维超声等）产生了日益精细和多样化的影像数据²。同时，电子健康记录系统（EHRs）的广泛部署积累了海量的临床诊疗数据，包括结构化的实验室结果、用药记录以及非结构化的临床笔记和出院小结。此外，可穿戴健

康设备和移动健康应用的普及，也使得连续性的生理参数、行为模式和环境暴露数据得以被采集和利用⁵。²明确指出，医学大数据涵盖“患者病历或临床试验数据、医学遗传学和基因组数据、X 射线、超声、MRI 等医学影像，以及一些现代类型的大数据，如 3D 成像和生物传感器读数”。

其次，**日益增长的临床需求**，特别是对复杂疾病（如癌症、神经退行性疾病、慢性代谢性疾病等）更深层次理解和更精准诊疗策略的需求，极大地推动了对多维度信息的综合分析。临床医生在日常诊疗中，本身就会综合考虑患者的病史、体征、影像学检查、实验室结果等多方面信息。多模态 AI 的目标正是要系统化、智能化地模拟并增强这一过程，以应对疾病的复杂性和个体差异性⁵。例如，⁵提到“整合多种模态进行多模态数据分析，对于增进我们对复杂健康状况的理解和管理具有重大前景”。

再次，**计算能力的显著提升和 AI 算法的持续突破**为处理和分析大规模、复杂的多模态数据提供了坚实的基础。图形处理器（GPU）等高性能计算硬件的发展，以及深度学习等先进 AI 算法（特别是能够处理异构数据和捕捉复杂模式的架构，如注意力机制、图神经网络和 Transformer 模型）的成熟，使得曾经因计算限制而难以实现的多模态数据融合分析成为可能³。¹⁰的研究指出，向多模态方法的转变“与计算框架的快速进步、硬件能力的提高以及计算、内存和数据存储成本的降低相吻合”。

这种由技术进步、临床需求和计算能力共同驱动的多模态化趋势，正在从根本上改变医学数据的格局，为医学 AI 的发展开辟了全新的可能性。它反映了医学界从“一维”或“单点”视角向“多维”或“整体”视角理解疾病和健康的转变，是实现更深层次医学洞察和更高效能医疗服务的前提。

2.2 在精准医疗、个性化治疗及新药研发中的核心价值

多模态医学数据融合与理解在推动现代医学向更精准、更个性化方向发展的过程中，展现出不可替代的核心价值。

在**精准医疗**领域，其核心价值体现在对疾病和患者进行更精细的刻画与分层。通过融合

基因组学、影像组学、临床表型、生活方式等多维度数据，AI 模型能够识别出传统方法难以发现的疾病亚型，预测特定治疗手段的敏感性和抵抗性，以及评估患者的预后风险⁸。例如，⁸中提及的 MMHA 算法，通过整合多组学数据，在区分疾病亚型方面表现出色，显著提升了低级别胶质瘤（LGG）等癌症的分类准确率，为精准医疗提供了新的分析工具。SOPHiA GENETICS 公司也致力于通过其 SOPHiA DDM™平台整合多模态肿瘤数据，以加速精准医疗的临床应用¹²。这种精细分层是实现“在正确的时间，对正确的患者，施以正确的治疗”这一精准医疗核心理念的基础。

在**个性化治疗**方面，多模态数据分析使得“量体裁衣”式的治疗方案成为可能。AI 系统能够整合患者独特的生物标志物组合、影像特征、临床病史、乃至生活习惯和环境因素，构建全面的个体健康画像⁴。基于此画像，模型可以预测不同治疗方案在该患者身上的疗效和潜在副作用，从而辅助医生制定出最适合当前个体的治疗策略。⁷中提到，AI 可以通过分析可穿戴设备数据、饮食日志和血糖读数，为糖尿病等慢性病患者推荐个性化的生活方式和药物调整建议。这种高度个体化的干预有望显著提高治疗成功率，改善患者的生活质量。

在**新药研发**领域，多模态 AI 的应用正在重塑药物发现、开发和临床试验的流程。通过整合药物的化学结构、生物活性数据、基因表达谱、蛋白质相互作用网络以及已有的临床试验数据，AI 模型可以高效筛选候选药物、预测药物靶点、发现药物新适应症（即“老药新用”）⁹。例如，⁹中详细介绍了几种基于图神经网络和 Transformer 的模型（如 GCNGAT、AMDGT），它们通过融合药物和疾病的相似性数据及复杂的生化信息，来预测潜在的药物-疾病关联，从而支持药物重定位。此外，多模态数据分析还能优化临床试验设计，例如通过更精准地筛选入组患者，预测不同患者亚群对试验药物的反应，从而提高试验效率和成功率¹³。Tempus 公司的 Loop 平台正是通过整合真实世界数据（RWD）和生物模型进行新型靶点识别，加速临床前治疗开发¹⁴。

2.3 提升诊断准确性、治疗效果及医疗效率的潜力

多模态医学数据融合与理解不仅在上述前沿领域具有核心价值，其在提升现有医疗实践的

诊断准确性、治疗效果和整体医疗效率方面也蕴藏着巨大潜力。

首先，在**提升诊断准确性**方面，多模态 AI 能够整合来自不同检查手段的互补信息，从而减少漏诊和误诊，尤其对于早期、非典型或复杂的病例诊断具有重要意义⁴。例如，在肿瘤诊断中，将病理影像、基因检测结果与临床症状描述相结合，可以提供比单一信息源更全面的诊断依据。⁴明确指出，多模态 AI 能够“发现疾病迹象，监测病情变化，并帮助进行更快、更精确的诊断”，因为它能发现单一模态中不明显的关联。¹⁵中提出的多模态融合分类网络（MFCN）在诊断任务上的表现也优于单模态方法。

其次，在**改善治疗效果**方面，通过更精准的患者分层、更优化的治疗方案选择以及对治疗反应的动态监测和预测，多模态 AI 有助于提高整体疗效，延长患者生存期，并改善其生活质量⁷。例如，⁸中 MMHA 算法在治疗反应预测方面的潜力，以及⁷中 AI 在慢性病管理中通过个性化建议改善疾病控制，都体现了这一点。这种从数据到更优临床结果的转化，是多模态 AI 价值的核心体现。诊断准确性的提高直接为更有效的个性化治疗铺平了道路，而个性化治疗的成功又进一步验证了诊断的精确性，形成了一个正向循环，共同推动治疗效果的提升。

再次，在**提高医疗效率**方面，多模态 AI 系统可以作为医生的智能助手，辅助进行复杂的病例分析和决策，自动化部分重复性的数据解读任务（如影像初筛），从而优化临床工作流程，减轻医生的工作负担，尤其是在医疗资源相对紧张或特定专业领域（如放射科、病理科）医生短缺的情况下¹。¹提及的临床决策支持系统能够辅助临床医生和医院管理者做出更优决策。⁴也认为多模态 AI 可以“减轻工作负荷”。这种效率的提升不仅能让医生将更多精力投入到与患者的沟通和复杂疑难病例的处理上，还能缩短患者的等待时间，优化医疗资源的整体配置。

综上所述，多模态医学数据融合与理解的应用价值是多维度且相互促进的。它不仅推动着精准医疗、个性化治疗和新药研发等前沿领域的突破，也为提升日常医疗服务的质量和效率提供了强大的技术支撑，预示着未来医疗健康事业的深刻变革。

三、研究现状及瓶颈

3.1 主要融合技术与框架

多模态医学数据融合的核心在于如何有效地整合来自不同来源、具有不同特性的数据，以产生一个比任何单一模态都更具信息量和判别力的表示。根据信息融合发生的阶段，主要的技术框架可以分为早期融合、晚期融合、中期融合以及混合融合策略。

早期融合 (Early Fusion)，也称为特征级融合 (Feature-level Fusion)，是在特征提取之后、模型训练之前，将来自不同模态的特征进行直接连接 (concatenation) 或组合 (如加权平均)，形成一个统一的、高维的特征向量，然后将此向量输入到单个机器学习或深度学习模型中进行后续处理²。这种策略的优点在于能够较早地捕捉和学习模态间的低层次交互关系⁵。然而，它也面临显著的挑战：要求不同模态的数据在时间和空间上严格对齐，对数据维度和采样率的差异较为敏感，且当某一模态数据缺失时，整个融合过程可能受到严重影响⁵。在实际应用中，由于医学数据的异构性，早期融合的适用场景相对有限，例如¹⁰的系统综述中提到，在其分析的 97 项研究中仅有 5 项采用了早期融合。

晚期融合 (Late Fusion)，也称为决策级融合 (Decision-level Fusion)，则采取不同的策略。它首先为每个数据模态单独训练一个或多个模型，得到各个模态独立的预测结果 (如分类概率、回归值或决策标签)，然后在决策层面对这些来自不同模型的输出进行整合，例如通过投票机制、加权平均、堆叠 (stacking) 或更复杂的元学习器 (meta-learner) 来产生最终的预测结果²。晚期融合对模态间的异构性和数据缺失具有较好的鲁棒性，因为各个模态可以独立处理，模型设计也更为灵活⁵。然而，其主要缺点在于可能忽略了在特征学习阶段模态之间潜在的早期交互信息，因为各模态在很大程度上是独立分析的，直到最后才汇集结果⁵。尽管如此，由于其实现相对简单且对数据要求较为宽松，晚期融合是目前研究中应用最为广泛的策略之一，¹⁰的综述显示有 64 项研究采用了此方法。

中期融合 (Intermediate Fusion)，亦可称为联合融合 (Joint Fusion) 或共享表示融合 (Shared Representation Fusion)，试图在早期融合和晚期融合之间取得平衡。在这种

框架下，不同模态的数据首先通过各自独立的网络分支进行初步的特征提取，然后在模型的中间层（即隐藏层）将这些不同模态的特征进行融合，学习一个共享的、跨模态的特征表示，最后再基于这个共享表示进行后续的任务处理（如分类或回归）⁵。中期融合允许模型既能学习模态特异性的特征，也能在中间阶段捕捉模态间的交互信息。它在处理模态维度不平衡和部分模态缺失方面比早期融合更有弹性⁵。¹⁰ 将其归类为混合融合

（midfusion），并指出有 20 项研究采用。这种策略的设计更具灵活性，但也可能更为复杂。

****混合融合（Hybrid Fusion）****则更为灵活，它结合了早期、中期和晚期融合中两种或多种策略的元素，以适应特定任务和数据模态的复杂需求⁵。例如，一个模型可能首先对部分紧密相关的模态进行早期融合，然后将融合后的特征与其他模态的特征在中期层面进行整合，最后再与其他独立处理的模态结果进行晚期融合。这种策略提供了最大的定制化空间，但也带来了更高的设计和优化难度⁵。

在具体的融合方法层面，除了传统的机器学习技术（如支持向量机 SVM、随机森林等，常用于晚期融合或处理已融合的特征向量）外，深度学习已成为主流。多层感知机

（MLP）可用于融合不同来源的特征向量。卷积神经网络（CNNs）主要用于处理图像数据，但其提取的特征也可与其他模态融合；特定设计的 CNN 架构（如多通道 CNN）本身就可以处理多模态输入⁷。循环神经网络（RNNs）及其变体如长短期记忆网络

（LSTMs）则擅长处理序列数据，如电子健康记录中的时序事件或生理信号（如 ECG）⁷。自编码器（Autoencoders）及其变体（如变分自编码器 VAEs）常被用于学习模态间的共享潜在表示或进行跨模态数据生成。

近年来，****注意力机制（Attention Mechanisms）****在多模态融合中扮演了越来越重要的角色。它允许模型在融合过程中动态地为来自不同模态的信息或特征的不同部分分配不同的权重，从而突出重要信息，抑制无关或噪声信息，显著提升了融合的效率 and 效果⁸。例如，⁸中提出的多模态层次注意力（MMHA）算法便是利用自适应加权来聚合多模态特征。¹⁵中也设计了一个基于可学习自注意力机制的融合网络来整合影像和临床数据。¹⁸和

¹⁹ 更是详细阐述了基于自注意力的 Transformer 模型如何通过将不同模态特征投影到统一空间并应用注意力计算来实现高效融合。

下表总结了主要的医学多模态数据融合策略：

Table 1: 多模态数据融合策略对比 (Comparison of Multimodal Data Fusion Strategies)

融合策略 (Fusion Strategy)	描述 (Description)	优点 (Advantages)	缺点 (Disadvantages)	医学领域典型用例/文献参考 (Typical Use Cases in Medicine / Snippet References)
早期融合 (Early Fusion)	在特征提取后、模型训练前直接组合各模态特征。	能学习模态间低层交互。	要求数据对齐；对异构性、缺失数据和维度差异敏感。	紧密耦合且同步采集的数据融合，如融合同一时间点的多生理信号；相对较少见 ² 。
中期融合 (Intermediate Fusion)	在模型的中间隐藏层进行特征融合，学习共享表示。	平衡模态特异性学习与模态间交互；对维度不平衡和缺失数据有一定鲁棒性。	设计相对复杂；共享表示的学习可能具有挑战性。	融合图像特征和临床结构化数据进行疾病分级；癌症亚型分类 ⁵ 。
晚期融合 (Late Fusion)	为各模态单独训练模型，在决策层融合其输出结果。	对模态异构性和缺失数据鲁棒性好；模型设计灵活，易于实现。	可能忽略模态间的早期交互信息；若模态间关联强，效果可能不如早期/中期融合。	整合不同影像模态（如 CT、MRI）的诊断结果；结合基因表

			合。	达谱和临床数据预测预后；应用广泛 ² 。
混合融合 (Hybrid Fusion)	结合多种融合策略的特点，根据需求定制。	高度灵活性和适应性。	设计和优化难度大；理论基础相对不完善。	针对特定复杂任务，如结合早期融合处理图像内部多尺度特征，再与其他临床数据进行中期融合 ⁵ 。

3.2 新兴模型：图神经网络与 Transformer 的应用

随着人工智能技术的不断演进，图神经网络（Graph Neural Networks, GNNs）和 Transformer 模型已成为处理和融合复杂多模态医学数据的两大新兴支柱，它们各自凭借独特的优势，为解决传统方法难以应对的挑战提供了新的途径。

图神经网络 (GNNs) 因其卓越的捕捉实体间复杂关系和结构信息的能力，在医学领域的应用日益广泛。医学数据中天然存在大量的图结构，例如患者与药物、疾病、诊疗程序之间的关联构成了医疗异构图¹⁶；基因、蛋白质及其相互作用形成了生物分子网络；知识图谱则可以表示医学概念及其语义关系。GNNs 能够直接在这些图结构数据上进行学习，提取节点特征，学习边的关系，从而进行节点分类（如疾病诊断）、边预测（如药物-疾病关联预测）或图分类等任务⁹。例如，⁹中提及的 GCNGAT 模型，通过整合图卷积网络（GCN）和图注意力网络（GAT），在药物重定位任务中分析药物-疾病异构图，有效提取了关键的交互特征。Trans4DD 模型则进一步将 GNN 与 Transformer 结合，应用于基于医疗异构图（包含患者、药物、诊疗三种节点）的疾病诊断任务，旨在克服传统 GNN 消息传递机制的局限性¹⁶。

尽管 GNNs 表现出色，但也面临一些固有的挑战，其中最突出的是**过平滑（over-smoothing）**问题。随着 GNN 层数的增加，节点的表示向量会趋向于相似，导致模型

难以区分不同的节点，从而影响预测精度⁹。研究人员正在通过引入残差连接、更深层次的注意力机制或改进聚合函数等方法来缓解这一问题⁹。

Transformer 模型 最初是为自然语言处理（NLP）任务设计的，其核心是**自注意力机制（Self-Attention）**。该机制能够动态地评估输入序列（或特征集合）中不同元素之间的相互重要性，并据此加权聚合信息，从而有效捕捉长距离依赖关系和上下文信息¹⁵。由于其强大的表示学习能力和并行处理优势，**Transformer** 迅速被推广到计算机视觉、基因组学分析以及多模态数据融合等多个领域¹⁷。在医学 AI 中，**Transformer** 不仅用于处理临床文本、基因序列等序列数据，也被成功应用于医学影像分析（如 **Vision Transformer, ViT**）和多模态特征的融合¹⁵。例如，¹⁵ 中提出的基于可学习自注意力机制的多模态融合网络，能够有效整合医学影像特征和临床数据。¹⁸ 和 ¹⁹ 详细描述了如何利用 **Swin Transformer** 等模型，将 CT 影像、临床文本等不同模态的特征投影到统一的特征空间，然后通过窗口化自注意力和层级特征聚合来实现高效融合，捕捉模态内和模态间的全局关联。

然而，**Transformer** 模型在医学领域的应用也并非没有挑战。一个主要问题是其计算复杂度，特别是对于大规模图结构数据或非常长的序列，标准的自注意力机制可能导致二次方级别的计算开销¹⁶。此外，尽管 **Transformer** 在许多任务上取得了成功，但其在医学多模态领域的应用潜力仍有待进一步深入探索和验证，特别是在模型的可解释性和对医学领域知识的有效结合方面¹⁷。

值得注意的是，研究领域正出现一种将 **GNNs** 和 **Transformer** 的优势相结合的趋势。这种融合旨在利用 **GNNs** 显式建模关系结构的能力，同时借助 **Transformer** 强大的全局上下文感知和特征交互学习能力。例如，**Trans4DD** 模型¹⁶ 正是这样一个例子，它试图用 **Transformer** 的全局注意力机制来弥补 **GNN** 在处理大规模医疗异构图时消息传递机制的不足。这种融合思路反映了研究者们追求更全面、更深入理解复杂医学数据的努力：既要理解医疗实体（如患者、疾病、药物）之间的“网络关系”，也要理解每个实体或数据模态内部丰富的“内容信息”。这种结合有望产生比单一模型范式更强大的医学 AI 系统，能够

同时处理和推理显式的实体间关系（例如，患者服用特定药物）以及每个实体或数据模态内部丰富的上下文信息（例如，临床笔记的详细文本内容或影像中的像素模式），从而实现对患者状态更整体和细致的把握。

3.3 当前面临的挑战与技术瓶颈

尽管多模态医学数据融合与理解取得了显著进展，但其在理论研究和实际应用中仍面临诸多挑战和技术瓶颈，这些挑战广泛分布于数据、模型、系统及伦理等多个层面。

数据层面的挑战首当其冲：

- **异构性 (Heterogeneity):** 这是多模态融合的根本难题。不同来源的医学数据在数据类型（如数值、文本、图像、序列）、格式、尺度、分布、噪声水平、完整性、稀疏性以及语义解释上均存在巨大差异⁵。例如，⁵和⁶指出，“每种模态可能具有不同的数据类型、格式、尺度、噪声水平和可解释性级别”。这种异构性使得数据难以直接对齐、比较和融合，需要复杂的预处理和特征工程。
- **模态缺失 (Missing Modalities):** 在临床实践中，由于各种原因（如检查未进行、数据记录不完整、患者不配合等），某些患者可能缺失一种或多种模态的数据。这对模型的训练和预测的鲁棒性提出了严峻考验，因为模型需要能够在部分信息缺失的情况下仍能做出可靠判断⁵。
- **小样本问题 (Small Datasets):** 许多医学研究，特别是针对罕见病或特定亚群的研究，能够获取到的完整多模态数据集规模往往较小。而复杂的深度学习模型通常需要大量数据进行训练以避免过拟合，这导致在小样本情况下模型性能不佳或泛化能力差⁵。⁷提到 RNN 在小数据集上容易过拟合。
- **维度不平衡 (Dimensionality Imbalance):** 不同模态的特征维度可能存在巨大差异。例如，高分辨率的医学影像数据或基因组数据可能包含数百万甚至数十亿个特征，而临床实验室指标或人口统计学信息可能只有几十个特征⁵。如何在融合过程中平衡这些不同维度的信息，避免高维特征淹没低维但重要的特征，是一个技术难题。
- **数据标注与质量 (Data Annotation and Quality):** 高质量的医学数据标注（如病灶勾

画、病理分级) 通常需要经验丰富的临床专家, 成本高昂且耗时。同时, 原始医学数据中可能存在噪声、伪影、记录错误或不一致性, 这些都会直接影响模型的性能和可靠性。

模型层面的挑战同样不容忽视:

- **可解释性 (Interpretability):** 许多先进的深度学习模型, 尤其是端到端的融合模型, 其决策过程往往像一个“黑箱”, 难以向临床医生清晰地解释其判断依据⁵。这种透明度的缺乏严重限制了 AI 模型在关键临床决策中的应用和医生的信任度。¹³ 中 PROmAI 联盟也将提升 AI 在肿瘤学中的“可解释性”作为关键目标之一。
- **泛化能力 (Generalization):** 模型在训练数据集上可能表现优异, 但在应用于来自不同医院、不同设备、不同人群或不同时间段的新数据集时, 性能可能会显著下降。提升模型的跨域泛化能力是实现临床广泛应用的关键。
- **鲁棒性 (Robustness):** 模型对输入数据的微小扰动 (如图像噪声、文本中的错别字) 或对抗性攻击的敏感性, 也是衡量其临床适用性的重要指标。
- **最优融合技术选择 (Optimal Fusion Technique):** 如前所述, 目前尚无一种普适性的“最佳”融合策略。针对不同的临床问题、数据特性和可用的模态组合, 选择或设计最合适的融合架构本身就是一个需要深入研究的开放性问題⁵。

系统与伦理层面的挑战则关乎技术的实际落地和可持续发展:

- **数据隐私与安全 (Data Privacy and Security):** 医学数据包含高度敏感的个人健康信息, 其收集、存储、处理和共享必须严格遵守相关的法律法规 (如 GDPR、HIPAA) 和伦理准则。多模态数据的融合往往涉及多源数据的汇集, 这进一步加剧了隐私泄露和数据安全风险⁴。
- **伦理考量 (Ethical Considerations):** 包括数据偏倚 (bias) 可能导致的算法歧视 (例如, 模型在特定人群中表现不佳), AI 辅助决策下的责任归属问题, 以及如何确保技术的公平性和可及性, 避免加剧医疗资源分配不均等问题⁴。
- **计算资源需求 (Computational Resource Demands):** 训练复杂的多模态深度学习模型通常需要大量的计算资源 (如高性能 GPU 集群) 和较长的训练时间, 这对于一些

资源有限的医疗机构或研究单位可能构成障碍⁷。

- **临床集成与工作流程 (Clinical Integration and Workflow):** 如何将 AI 模型无缝、高效地集成到现有的临床工作流程中，使其真正为临床医生接受和有效使用，而不是增加额外负担，是技术能否成功转化的关键⁴。

这些挑战并非孤立存在，而是常常相互交织，形成一个复杂的“问题网络”。例如，为了解决小样本问题（⁵提及），研究者可能倾向于汇集更多来源的数据，但这无疑会加剧数据隐私和安全的顾虑（⁷提及）。同样，为了有效处理高度异构的数据（⁵提及），可能需要采用更为复杂的深度学习模型，但这又可能牺牲模型的可解释性（⁵提及）。选择融合策略时，也必须考虑到模态缺失的可能性（⁵提及）：早期融合对数据完整性要求高，而晚期融合虽更鲁棒但可能丢失模态间信息。此外，高性能模型（如大型 Transformer）对计算资源的高需求（⁷提及）可能限制其在资源不足环境下的应用，从而引发关于技术公平性的伦理讨论（⁷提及）。这种挑战间的联动效应意味着，多模态医学 AI 的发展需要系统性的解决方案和多方面的权衡，而非针对单一问题的孤立修补。

下表总结了多模态医学数据融合面临的主要挑战及其潜在的应对思路：

Table 2: 多模态医学数据融合的主要挑战与潜在解决方案 (Key Challenges in Multimodal Medical Data Fusion and Potential Solutions)

挑战类别 (Challenge Category)	具体挑战 (Specific Challenge)	描述 (Description)	潜在解决方案/缓解策略 (Potential Solutions/Mitigation Strategies)	文献参考 (Key Snippet References)
数据层面 (Data Level)	异构性 (Heterogeneity)	不同模态数据在格式、尺度、分布等方面差异巨大。	特征归一化/标准化；模态特异性编码器；学习共享潜在空间表	⁵



			示；基于 Transformer 的 对齐。	
	模态缺失 (Missing Modalities)	临床数据常有部 分模态不完整。	缺失数据插补技 术；设计对缺失 模态鲁棒的融合 机制（如晚期融 合、部分中期融 合）；生成对抗 网络（GANs） 补全数据。	5
	小样本问题 (Small Datasets)	特定疾病多模态 数据集规模有 限，易过拟合。	迁移学习；自监 督学习；数据增 强；元学习；正 则化技术。	5
	维度不平衡 (Dimensionality Imbalance)	不同模态特征维 度差异悬殊。	特征选择/降维； 加权融合策略； 注意力机制动态 调整特征重要 性。	5
模型层面 (Model Level)	可解释性 (Interpretability)	深度学习模型决 策过程不透明， 难以获得临床信 任。	LIME, SHAP 等 事后解释方法； 设计内生可解释 模型（如基于规 则、注意力可视 化）；因果推 断。	5

	最优融合技术选择 (Optimal Fusion Technique)	缺乏普适性的最佳融合策略。	基于任务和数据特性进行经验验证；自适应融合网络；神经架构搜索 (NAS)。	5
系统与伦理层面 (System & Ethical Level)	数据隐私与安全 (Data Privacy & Security)	医学数据高度敏感，融合共享面临法规和安全风险。	联邦学习；差分隐私；同态加密；安全多方计算；数据脱敏和匿名化。	4
	伦理考量 (Ethical Considerations)	数据偏倚导致算法歧视；AI 决策责任归属等。	公平性算法设计；偏倚检测与缓解；透明度原则；建立清晰的伦理审查和问责机制。	4
	计算资源需求 (Computational Resource Demands)	训练复杂模型需大量计算资源。	模型压缩与剪枝；知识蒸馏；轻量化网络设计；分布式训练。	7

四、过往研究的里程碑

医学人工智能领域从最初的单模态分析发展到当前复杂的多模态融合与理解，经历了一个逐步演进的过程，其间涌现了若干关键的理论突破、模型创新和标志性研究成果。

4.1 从单模态到多模态 AI 的演进

在早期阶段，机器学习在医学领域的应用主要集中于处理单一类型的数据。例如，计算机

辅助诊断系统利用图像处理和模式识别技术分析医学影像（如 X 光片、CT 扫描），生理信号处理算法分析心电图（ECG）或脑电图（EEG）数据，自然语言处理技术则用于从临床文本（如病历、文献）中提取信息³。¹⁰ 在其系统回顾中描述了传统机器学习最初如何依赖单模态数据，并在放射学、皮肤病学和病理学等特定领域取得了显著成就。这些早期的单模态模型为后续更复杂的研究奠定了基础，但也逐渐暴露出其局限性。

驱动这一领域从单模态向多模态转变的核心因素，是研究者们日益认识到单一数据源在描述复杂生物学现象时的不足，以及不同数据模态之间潜在的互补性与协同效应¹⁰。正如¹⁰所强调的，单模态模型在提供对患者状况的全面理解方面存在“固有的局限性”。例如，仅凭影像学表现可能难以区分某些肿瘤的良恶性或预测其侵袭性，而结合基因表达谱或临床病理特征则可能提供更准确的判断。尽管¹⁰指出“多模态方法相对于单模态模型的附加价值仍不完全清楚”本身是一个持续的研究问题，但也承认“整合异构数据类型带来了方法学和技术上的障碍”，这间接说明了尽管存在挑战，但整合的需求已被广泛认知。

近五到十年间，多模态医学 AI 研究经历了显著的增长和加速。这一方面得益于医学数据的爆炸式增长和可及性的提高（如大型生物样本库、电子健康记录数据库的建立），另一方面则归功于计算能力的飞跃（特别是 GPU 的普及）和深度学习算法的突破，尤其是那些能够有效处理异构数据、捕捉复杂依赖关系的先进模型架构，如注意力机制、图神经网络（GNNs）和 Transformer 模型的出现与广泛应用³。³中的图表（图 2A）直观地显示了从 2019 年到 2024 年，关于多模态 AI（特别是在医疗保健和放射学领域）的研究呈现出“显著的激增”。¹⁰也将向多模态的转变与“计算框架的快速进步、硬件能力的提高以及计算、内存和数据存储成本的降低”联系起来。这些因素共同为多模态医学 AI 的蓬勃发展创造了有利条件。

4.2 关键理论突破与模型创新

在多模态 AI 的演进过程中，一系列理论和模型的创新起到了里程碑式的推动作用：

- **深度学习的引入与普及:** 卷积神经网络（CNNs）在医学影像分析领域取得了巨大成功，为从图像中自动提取高层语义特征提供了强大工具⁷。循环神经网络（RNNs）及

其变体（如 LSTMs, GRUs）则有效处理了临床记录中的时序数据和医学文本等序列信息⁷。这些基础的深度学习模型为各个单模态的特征表示学习奠定了坚实基础，是后续进行有效多模态融合的前提。

- **注意力机制与 Transformer 模型的出现:** 注意力机制，特别是自注意力机制及其核心应用 Transformer 架构，彻底改变了序列数据处理和特征融合的方式。它们能够动态地权衡不同信息来源或特征内部不同部分的重要性，有效捕捉长距离依赖关系，从而实现更智能、更高效的跨模态信息整合与交互¹⁵。¹⁷ 将 Transformer 描述为“特别具有革命性”的技术，并提及其已从自然语言处理扩展到多模态任务。¹⁵ 则强调自注意力机制是整合异构数据源的“强大解决方案”。¹⁸ 和¹⁹ 更是详细阐述了 Transformer 模型（如 Swin Transformer）如何通过自注意力机制对多模态特征（如 CT 影像和临床文本）进行融合，以捕捉全局相关性。这一创新使得模型能够学习“关注什么”以及“如何融合”，是多模态融合领域的一大飞跃。
- **图神经网络（GNNs）的应用:** GNNs 为建模医学实体（如基因、蛋白质、药物、疾病、患者）之间的复杂关系网络提供了全新的范式。它们能够直接在图结构数据上进行学习，捕捉拓扑结构信息和节点属性，从而在药物发现、疾病关联预测、医学知识图谱推理等任务中展现出独特优势⁹。GNNs 使得将结构化的关系知识整合到学习系统中成为可能，弥补了传统模型在这方面的不足。
- **自监督学习（Self-supervised Learning）与预训练模型（Pre-trained Models）的兴起:** 鉴于医学领域高质量标注数据往往稀缺且昂贵，自监督学习通过设计巧妙的代理任务（pretext tasks）从大量无标注数据中学习有用的特征表示，极大地缓解了对标注数据的依赖。在此基础上发展起来的预训练模型（如在大型通用数据集 ImageNet 上预训练的 CNN 模型，或在大量文本语料上预训练的 BERT 等语言模型），可以通过迁移学习的方式快速适应特定的医学任务，显著提升了模型在小样本医学数据集上的性能和泛化能力³。例如，¹⁹ 提及的 CheXzero 就是一个在未标记 X 射线影像上进行自监督学习的模型。这些技术使得在数据有限的医学场景下训练强大的深度学习模型成为可能。

4.3 标志性研究成果与趋势

基于上述理论与模型的创新，多模态医学 AI 研究取得了一系列标志性的成果，并呈现出一些明显的发展趋势：

- **多模态模型性能优势的广泛证实:** 大量的比较研究表明，在各种不同的医学诊断、预后预测、治疗反应评估等任务中，精心设计的多模态模型通常能够取得优于仅使用单一模态数据的模型的性能^{8, 10}。系统综述发现，在其分析的 97 项研究中，高达 91% (88 项) 的研究显示多模态方法优于单模态算法。⁸ 中 MMHA 算法和¹⁵ 中 MFCN 模型也都报告了其多模态设计超越了单模态基线。这些经验证据为多模态方法的整体有效性提供了有力支持。
- **特定模态组合效用的深入探索:** 研究者们不再满足于简单地堆砌所有可用的数据模态，而是开始系统地探索不同模态组合的互补性和冗余性，以期找到针对特定临床问题的最优或最具成本效益的模态子集³。例如，³ 的分析 (图 2B) 显示，在已发表的多模态研究中，“放射影像与文本”是最常见的组合，其次是“组学数据与病理数据”。这表明研究正在识别具有高协同价值的模态配对。
- **融合机制向更复杂、更智能的方向演进:** 数据融合技术已从早期的简单特征拼接或加权平均，发展到基于深度学习的端到端联合学习，特别是广泛采用基于注意力机制的动态加权融合、基于图结构的交互建模以及基于 Transformer 的深层跨模态交互等更为复杂和智能的融合网络架构⁸。这反映了该领域在模型设计复杂度和精细度上的不断提升。
- **对临床转化、可解释性和可信赖性的日益关注:** 随着模型性能的不断提高，研究的焦点也逐渐从单纯追求更高的预测准确率，转向如何使这些先进的 AI 模型能够真正应用于临床实践。这意味着需要更加关注模型的鲁棒性、泛化能力、可解释性（即模型为何做出特定决策），以及模型的公平性、安全性和伦理合规性，以建立临床医生和患者对 AI 系统的信任，并确保其能够无缝集成到复杂的医疗工作流程中⁴。

这些里程碑式的进展共同描绘了多模态医学 AI 领域一幅充满活力的发展图景。早期在基础多模态整合（例如简单融合 CNN 特征）上取得的初步成功，起到了概念验证的作用，

吸引了更多的数据资源和计算能力投入。这反过来又促进了更复杂模型（如 Transformer 和 GNN）的开发和应用，这些模型带来了更显著的性能提升，从而进一步加速了研究的步伐、数据收集的努力（例如 Tempus 建立的大型数据库²¹）以及对临床转化的兴趣。这种成功孕育更多资源和创新，进而带来更大成功的正反馈循环，构成了该领域研究的“飞轮效应”，驱动了³所观察到的“多模态 AI 研究的激增”。

五、产业化案例

随着多模态医学 AI 技术的不断成熟，其商业化应用和产业化进程也初具规模。一批创新企业和科技巨头纷纷布局，致力于将多模态 AI 的潜力转化为实际的医疗产品和服务，覆盖了从辅助诊断、治疗决策到新药研发的多个环节。

5.1 领先企业及其多模态 AI 平台与解决方案

- **SOPHiA GENETICS:** 该公司专注于通过其 SOPHiA DDM™ (Data-Driven Medicine) 平台实现多模态数据的整合与分析，尤其在肿瘤学领域¹²。其平台的核心技术在于对基因组学、影像学（包括放射影像和数字病理）、临床记录和生物学数据进行统一的协调、标准化处理，并利用 AI 算法进行深度挖掘，以生成新的生物学洞见、支持队列分析和开发预测模型¹²。SOPHiA DDM™ for Multimodal 旨在通过促进多中心研究合作，加速数据驱动的精准医疗落地。其应用案例包括参与 DEEP-Lung-IV 国际观察性研究（针对 IV 期非小细胞肺癌的多模态数据分析）和 UroPredict 项目（与法国肾癌研究网络合作，利用真实世界数据预测肾癌复发风险）¹²。这表明该公司正通过其商业平台，将多模态数据整合能力应用于肿瘤研究和临床预测。
- **IQVIA:** 作为一家全球领先的生命科学领域服务提供商，IQVIA 将人类数据科学（human data science）与其在临床开发方面的专长相结合，利用多模态数据和 AI 技术推动精准健康和药物研发²²。虽然部分关于其具体多模态 AI 策略的直接信息受限（如²⁶无法访问），但其提供的产品和服务，如医院信息系统、移动与家庭健康解决方案、以及其在肿瘤临床试验领域的合作（例如与 Sarah Cannon 研究所的合作¹³），都暗示了其对多模态数据的整合与应用能力。²⁵ 也指出多模态 AI 正在从诊断到

治疗等多个方面变革医疗保健行业，这与 IQVIA 的业务范围高度吻合。IQVIA 的广泛业务组合表明，多模态 AI 可能被整合到其医疗运营、临床研究和药物安全监测等多个环节。

- **Tempus AI:** Tempus AI 以其构建的庞大、高质量的多模态数据库为核心竞争力，致力于推动 AI 在精准医疗中的应用¹⁴。其数据库据称包含超过 4000 万条研究记录，整合了临床数据、基因组数据（包括全外显子组和转录组测序）、影像数据等²¹。基于此数据库，Tempus 推出了“Fuses”项目，旨在利用其基础模型（foundation model）开发 AI 驱动的诊断、预后和预测模型，以实现高度个性化的患者护理²¹。此外，其“Loop”平台则专注于通过整合真实世界数据（RWD）、人类生物模型和 CRISPR 筛选技术，加速肿瘤治疗新靶点的发现和验证¹⁴。Tempus 的实践清晰地展示了一家公司如何围绕其核心的多模态数据资产，构建从基础研究到临床应用的 AI 赋能体系。
- **PathAI:** PathAI 专注于将 AI 技术应用于病理学领域，其核心平台 AISight®是一个云原生的数字病理图像管理和智能工作流程解决方案²³。PathAI 不仅开发用于肿瘤检测、组织学量化（如其定量多模态各向异性成像 QMAI 技术²³）等任务的 AI 算法，还积极与 Precision for Medicine 等机构合作，将 AI 能力扩展到生物标志物发现、空间生物学分析和临床试验服务中，通过整合数字病理图像与其他临床或分子数据，增强对复杂组织生物学的理解²⁴。这体现了以关键影像模态（病理）为切入点，逐步向多模态整合分析扩展的商业策略。
- **Google (Alphabet):** 作为科技巨头，Google 也通过其健康部门（Google Health）和云计算平台（Google Cloud）积极布局多模态医疗 AI。其推出的 MedGemma 是一个面向医疗保健领域的开放 AI 模型，而 Gemma 3n 则是针对移动端优化的轻量级多模态模型，强调隐私和实时性²⁵。其研发的 AMIE（Advanced Medical Intelligence Agent）项目旨在通过理解视觉医学信息来改进诊断对话²⁵。Google Cloud 也提供符合医疗数据安全与隐私保护要求的存储和分析平台，为多模态数据的管理和 AI 模型开发提供基础设施支持⁴。这代表了大型科技公司凭借其在 AI 基础研究和云计算方面的优势，向医疗这一专业领域渗透的趋势。

- **Inka Health (Onco-Innovations Limited):** Inka Health 发起了 PROmAI (Predicting Oncology Outcomes using Multimodal AI) 联盟, 旨在联合制药公司、学术机构和数据科学专家, 共同推动 AI 在肿瘤预测研究中的应用¹³。该联盟的核心目标是开发和验证能够整合真实世界数据和临床试验数据 (涵盖分子、影像、临床等多种模态) 的下一代 AI 方法, 以期提高肿瘤学预测的准确性、模型的可解释性、可移植性以及监管要求的相关性¹³。其自有的 SynoGraph 平台据称利用 AI 驱动的因果推断技术来分析多模态数据。PROmAI 联盟的模式体现了通过行业协作共同应对多模态 AI 在复杂疾病领域 (如肿瘤) 所面临挑战的一种努力。

5.2 在肿瘤学、放射学、药物研发等领域的具体应用

上述企业的实践以及更广泛的研究进展, 已在多个医学细分领域催生了具体的应用场景:

- **肿瘤学 (Oncology):** 这是多模态 AI 应用最为活跃和深入的领域之一。应用范围涵盖了从癌症的早期筛查与诊断、精确的肿瘤分型与分期、治疗反应 (如对免疫治疗、化疗的反应) 的预测、复发风险评估, 到最终的个性化治疗方案推荐等全周期管理⁴。例如, SOPHiA GENETICS 的 DEEP-Lung-IV 研究和 UroPredict 计算器¹², Tempus 的 Fuses 项目²¹, 以及 Inka Health 的 PROmAI 联盟¹³, 都将肿瘤学作为其核心应用场景。癌症的高度异质性和复杂性使其成为多模态数据分析的理想试验田。
- **放射学 (Radiology):** 多模态 AI 正在革新医学影像的解读与分析。通过融合 CT、MRI、X 射线、超声以及数字病理切片等影像数据与临床笔记、实验室结果、基因组信息等非影像数据, AI 能够实现更精准的病灶自动检测与分割、良恶性鉴别、疾病分级, 以及更全面的影像报告生成³。¹⁷ 特别指出, 在放射学领域, 结合影像和临床元数据的多模态模型正变得日益流行。PathAI 在数字病理影像 AI 方面的创新²³ 也是一个典型代表。
- **药物研发 (Drug Development):** 多模态数据分析为加速药物发现、优化临床试验设计和实现药物重定位提供了强大工具⁹。通过整合药物的分子结构、作用机制、在细胞和动物模型中的实验数据、以及在人体临床试验中的表现 (包括疗效和安全性数据), AI 可以更有效地识别潜在的药物靶点, 预测候选药物的成药性, 筛选对特定

患者亚群可能有效的药物，甚至发现已有药物的新用途。

- **临床决策支持与患者监测 (Clinical Decision Support and Patient Monitoring):** AI 系统可以通过整合来自可穿戴设备、床旁监护仪、实验室信息系统 (LIS) 和电子健康记录 (EHR) 的实时多模态数据 (如生命体征、检验结果、活动数据、睡眠模式、甚至语音和文本输入的情绪信息)，为临床医生提供实时的决策支持和早期预警¹。这在重症监护 (ICU)、慢性病 (如糖尿病、高血压、精神健康障碍) 管理以及远程医疗场景中尤为重要，有助于实现更主动、更连续的患者照护。

5.3 商业模式与市场前景

多模态医疗 AI 领域的商业模式呈现多样化趋势，主要包括：

- **平台即服务 (PaaS):** 企业提供一个集成了数据存储、管理、多模态数据整合、AI 模型开发与验证工具的云平台，供医院、研究机构或制药公司使用，如 SOPHiA DDM™、Tempus 的平台产品、PathAI 的 AISight®。
- **软件即服务 (SaaS):** 企业开发针对特定临床问题或工作流程的 AI 应用程序或算法模块，以订阅或许可的方式提供服务，例如用于特定癌症亚型诊断的算法、优化临床试验招募的工具等。
- **数据驱动的洞察服务:** 拥有大规模、高质量多模态数据库的企业 (如 Tempus、IQVIA)，可以基于其数据资产和分析能力，为制药公司提供药物研发相关的洞察、为医疗机构提供临床决策支持相关的分析报告或咨询服务。
- **合作研发与授权:** AI 公司与制药企业、医疗器械公司或大型医疗集团合作，共同开发新型的诊断工具、治疗方案或数字疗法，并通过技术授权、里程碑付款或销售分成等方式实现商业回报。

市场前景方面，全球医疗 AI 市场，特别是涉及复杂数据分析的多模态 AI 细分市场，预计将保持高速增长态势。驱动因素包括精准医疗的持续需求、AI 技术的不断进步、医疗数据的日益丰富以及各国政府对医疗创新和数字化的政策支持。肿瘤学、神经系统疾病 (如阿尔茨海默病)、心血管疾病等复杂疾病领域，因其对多维度信息分析的迫切需求，

将成为多模态 AI 应用的主要增长点和投资热点。

然而，该领域的商业化也面临挑战，包括漫长且严格的医疗器械或软件的监管审批流程、大规模临床验证的成本和复杂性、医疗数据的标准化和互操作性问题、以及如何将 AI 解决方案有效集成到现有复杂的医疗 IT 系统和临床工作流程中，并确保其得到临床医生的广泛接受和正确使用。

在这些产业化案例中，一个值得关注的现象是“数据网络效应”的显现，这可能成为该领域企业构建核心竞争壁垒的关键。像 Tempus 这样的公司¹⁴，通过长期投入构建了包含数千万研究记录的专有大规模多模态数据库。数据的广度、深度和多样性直接决定了其训练出的 AI 模型的性能和泛化能力。更优越的模型能够吸引更多的用户（如医院、医生）和合作伙伴（如制药公司），而这些用户和合作伙伴的使用过程又可能（直接或间接地）贡献更多的数据，从而进一步优化模型，形成一个“数据越多 -> 模型越好 -> 用户越多 -> 数据越多”的正向循环。这种数据驱动的飞轮一旦转动起来，将为先发者构建起强大的竞争壁垒，因为后来者难以在短时间内复制如此庞大且经过精心整理和标注的多模态数据资产，尤其是在开发需要海量数据训练的医学基础模型方面。Tempus 在其愿景中也提到，“通过提供随着公司收集更多数据而学习的工具，让每位患者都能从前辈的治疗中受益”²¹，这清晰地描述了这种由数据驱动的学习和改进循环。

下表概览了部分在医疗 AI 领域积极部署多模态技术的企业及其相关信息：

Table 3: 医疗 AI 领域多模态技术产业化案例概览 (Overview of Industrialization Cases for Multimodal Technology in Medical AI)

企业/平台 (Company/Platform)	主要多模态技术/ 侧重点 (Key Multimodal Technologies/Focus)	主要医学应用领域 (Primary Application Areas in Medicine)	商业模式特点 (Business Model Aspects)	文献参考 (Key Snippet References)



SOPHiA GENETICS (SOPHiA DDM™ for Multimodal)	整合基因组、影像、临床、生物学数据；AI 驱动的数据协调、标准化、预测模型。	肿瘤学研究（如 NSCLC、肾癌）、精准医疗。	PaaS, 合作研究。	12
IQVIA	整合人类数据、数据科学进行临床开发；云解决方案；精准健康。	临床试验优化、医疗信息系统、药物安全、真实世界证据研究。	PaaS, SaaS, 数据洞察服务, 合作研发。	13
Tempus AI (Fuses, Loop)	大规模多模态数据库（临床、基因组、影像）；AI 基础模型；靶点发现。	肿瘤学个性化护理、药物研发、伴随诊断。	PaaS, 数据洞察服务, 合作研发。	14
PathAI (AISight®)	AI 驱动的数字病理解决方案；定量多模态各向异性成像 (QMAI)；生物标志物发现。	数字病理、肿瘤检测、纤维化研究、临床试验生物样本分析。	PaaS, SaaS, 合作研发。	23
Google (MedGemma, AMIE)	医疗 AI 开放模型；端侧多模态模型；诊断对话智能体；安全云平台。	医疗影像分析、诊断支持、慢性病预测、药物研发。	PaaS (Google Cloud), 模型/技术输出。	4

Inka Health (PROmAI Consortium, SynoGraph)	整合真实世界与 临床试验多模态 数据（分子、影 像、临床）；AI 因果推断。	肿瘤学预测研 究、个性化癌症 护理、临床药物 开发。	联盟合作模式, PaaS (SynoGraph)。	13
---	--	-------------------------------------	---------------------------------	----

六、总结与展望

6.1 对当前研究和应用的总结

多模态医学数据融合与理解无疑已成为当前医学人工智能领域最具活力和潜力的前沿方向之一。通过整合来自影像、基因组学、临床记录、病理报告、生理信号等多种来源的异构信息，多模态 AI 系统旨在构建对患者健康状况和疾病过程更为全面、精准和动态的认知。现有研究已充分证明，相较于传统的单模态分析方法，多模态策略在提升疾病诊断的准确性（尤其是在复杂和早期病例中）、实现更精细的患者分层与预后预测、指导个性化治疗方案的选择以及加速创新药物的发现与开发等方面，均展现出显著的优势和巨大的应用价值。

在技术层面，以深度学习为核心的建模方法占据主导地位。特别是图神经网络（GNNs）凭借其对实体间复杂关系的强大建模能力，以及 Transformer 模型依靠其自注意力机制在捕捉长距离依赖和上下文信息方面的卓越表现，已成为处理和融合多模态医学数据的两大新兴利器。这些先进模型与各种融合框架（早期、中期、晚期及混合融合）相结合，在多个医学应用场景中取得了令人鼓舞的成果。

然而，尽管进展显著，多模态医学 AI 的发展仍面临诸多严峻挑战。数据层面的异构性、模态缺失、小样本问题、维度不平衡以及数据标注和质量控制等问题，持续制约着模型的性能和泛化能力。模型层面的可解释性不足、鲁棒性有待提升以及最优融合策略的选择困难，也限制了其在临床实践中的可信度和应用推广。此外，系统与伦理层面的数据隐私安全、算法偏见、责任归属、计算资源需求以及与现有医疗体系的集成等问题，更是关系到该技术能否健康、可持续发展的关键。

在产业化方面，全球范围内已涌现出一批专注于多模态医疗 AI 解决方案的创新企业和科技巨头。它们通过构建大规模多模态数据库、开发专用 AI 平台和算法，积极推动技术在肿瘤学、放射学、病理学、药物研发等领域的实际应用和商业转化，市场前景广阔，但也面临着监管、验证和临床接受度等方面的考验。

6.2 未来发展方向、潜在突破点及建议

展望未来，多模态医学数据融合与理解领域将在克服现有挑战的基础上，朝着更智能、更可靠、更易于临床应用的方向持续演进。

模型层面的未来发展将更加注重：

- **可解释 AI (Explainable AI, XAI):** 开发具有更高透明度和可解释性的多模态模型，使其决策过程能够被临床医生理解、验证和信任，是提升临床接受度的关键。这可能涉及设计内生可解释的模型架构，或结合有效的后验解释技术。
- **因果推断 (Causal Inference):** 超越当前模型主要学习的相关性，探索和引入因果推断能力，以更深入地理解疾病的发生发展机制、不同因素对结局的真实影响以及治疗干预的实际效果。¹³ 中提及 Inka Health 的 SynoGraph 平台已在尝试利用 AI 驱动的因果推断，这代表了一个重要的方向。
- **自监督、弱监督与无监督学习:** 进一步研究和应用这些能够有效利用大量未标注或弱标注数据的学习范式，以缓解对昂贵且耗时的人工标注的依赖，特别是在医学影像和文本等领域。
- **联邦学习与隐私增强计算:** 在严格保护患者数据隐私和符合法规要求的前提下，实现跨机构、跨地域的多中心数据协作和模型训练，从而利用更广泛的数据资源提升模型性能和泛化能力。
- **医学基础模型 (Foundation Models for Medicine):** 借鉴自然语言处理领域大型预训练模型的成功经验，构建能够适应多种下游医学任务的大规模、多模态医学基础模型。这类模型通过在海量、多样化的医学数据上进行预训练，学习通用的医学知识和特征表示，然后通过少量任务相关数据进行微调，快速应用于特定的临床问题。

Tempus 公司在其“Fuses”项目中致力于构建大型诊断模型²¹，正是这一趋势的体现。

数据层面的努力将集中于：

- **高质量多模态数据集的构建与共享:** 推动建立标准化的、大规模的、具有丰富标注和临床关联信息的开放或可控共享的多模态医学数据库，为算法研发和验证提供坚实基础。
- **模态缺失与数据不平衡的处理技术:** 开发更先进、更鲁棒的算法来有效处理真实世界医学数据中普遍存在的模态缺失、数据噪声和类别不平衡等问题。

应用层面的重点在于：

- **临床工作流程的深度集成:** 设计和开发能够无缝、高效地嵌入现有临床工作流程的 AI 工具，使其真正成为临床医生的得力助手，而不是额外的负担。这需要充分考虑用户体验和人机交互。
- **多学科协作 (Multidisciplinary Collaboration):** 加强临床医生、AI 研究人员、数据科学家、生物信息学家、伦理学家、社会学家和政策制定者之间的紧密合作与对话，共同推动技术的健康发展和负责任应用。
- **监管科学与标准化:** 建立和完善针对多模态医疗 AI 产品（特别是作为医疗器械的软件 SaMD）的评估、验证、审批和监管框架与标准，确保其安全性、有效性和临床价值。

潜在的突破点可能出现在：

- 在特定复杂疾病（如阿尔茨海默病、帕金森病、自身免疫性疾病、精神分裂症等）的多模态生物标志物的发现与验证上取得重大进展，从而实现更早期的诊断和更精准的干预。
- 真正实现基于连续多模态数据监测的动态、实时个性化治疗调整，形成“智能医疗闭环”。
- AI 驱动的、从靶点发现到临床前候选药物筛选的全自动化或高度自动化药物发现流

程。

为了促进该领域的健康发展，兹建议：

- **鼓励跨学科研究和复合型人才培养:** 支持医学、计算机科学、统计学、生物信息学等多学科交叉的研究项目，培养兼具多领域知识的专业人才。
- **加大对基础研究和开放数据集建设的投入:** 政府、科研机构和企业应共同努力，增加对多模态 AI 基础理论、核心算法以及高质量、标准化医学数据集建设的长期投入。
- **建立健全的伦理规范和监管政策:** 在鼓励技术创新的同时，必须同步建立和完善相关的伦理指导原则、数据治理框架和动态调整的监管政策，确保技术发展始终以患者福祉为中心，并防范潜在风险。

一个贯穿未来发展的重要主题是实现真正的“**转化型多模态 AI**” (Translational Multimodal AI) —— 即不仅要持续开发更先进、性能更优的算法（这是当前研究的主要焦点），更要系统性地解决阻碍这些强大模型从实验室走向临床、从代码惠及患者的诸多实际障碍。这些障碍包括模型的可解释性不足⁷、与现有临床系统的集成困难⁴、潜在的伦理风险和数据偏见¹⁷、以及漫长而复杂的监管审批流程等。正如 Inka Health 的 Paul Arora 所言，“下一代医学 AI 的定义将不仅仅取决于算法本身，更取决于我们如何将这些算法嵌入到药物开发和患者护理的真实世界复杂性中”¹³。PROmAI 联盟将“可解释性、可移植性和监管相关性”¹³ 作为其关注重点，也印证了这一转化需求。因此，未来的突破将不仅仅是算法的突破，更需要一个涵盖临床验证、伦理监督、工作流程优化、标准制定和政策支持的整体解决方案。只有这样，多模态医学 AI 的巨大潜力才能真正转化为改善全球人类健康的实际成果。

七、参考文献

1. Comprehensive Review of Multimodal Medical Data Analysis: Open Issues and Future Research Directions - Acta Informatica Pragensia, 6. 11, 2025, <https://aip.vse.cz/pdfs/aip/2022/03/09.pdf>
2. (PDF) Comprehensive Review of Multimodal Medical Data Analysis ..., 6. 11, 2025, https://www.researchgate.net/publication/366601357_Comprehensive_Review_of_Multimodal_Medical_Data_Analysis_Open_Issues_and_Future_Research_Directi

[ons](#)

3. Advancements in Medical Radiology Through Multimodal Machine Learning: A Comprehensive Overview - PMC, 6. 11, 2025, <https://pmc.ncbi.nlm.nih.gov/articles/PMC12108733/>
4. Multimodal AI: Revolutionising the Healthcare Industry, 6. 11, 2025, <https://www.octopusintelligence.com/what-are-the-most-promising-applications-of-multimodal-ai-in-healthcare/>
5. A systematic review of challenges and proposed solutions in modeling multimodal data, 6. 11, 2025, <https://arxiv.org/html/2505.06945v1>
6. A systematic review of challenges and proposed solutions in modeling multimodal data - arXiv, 6. 11, 2025, <https://arxiv.org/pdf/2505.06945>
7. The Role of AI in Hospitals and Clinics: Transforming Healthcare in the 21st Century - PMC, 6. 11, 2025, <https://pmc.ncbi.nlm.nih.gov/articles/PMC11047988/>
8. 基于层次注意力的多模态数据融合算法在疾病诊断与癌症亚型预测中 ..., 6. 11, 2025, <https://www.ebiotrade.com/newsf/2025-5/20250522115925959.htm>
9. AMFGNN: an adaptive multi-view fusion graph neural network ..., 6. 11, 2025, <https://pmc.ncbi.nlm.nih.gov/articles/PMC12066569/>
10. Is Multimodal Better? A Systematic Review of Multimodal versus ..., 6. 11, 2025, <https://www.medrxiv.org/content/10.1101/2025.03.12.25322656v1.full>
11. 多模态大数据智能应用引领精准医疗时代到来, 神州医疗南方医院项目二期启动, 6. 11, 2025, <http://www.dhctech.com/gongsixinwen/397.html>
12. SOPHiA DDM™ for Multimodal - SOPHiA GENETICS, 6. 11, 2025, <https://www.sophiagenetics.com/sophia-ddm-for-multimodal/>
13. Inka Health Launches Global AI-Oncology Consortium to Advance ..., 6. 11, 2025, <https://trial.medpath.com/news/655d43ab7d111ffe/inka-health-launches-global-ai-oncology-consortium-to-advance-predictive-cancer-research>
14. Tempus Introduces Loop, an AI-Powered Target Discovery and Validation Platform, 6. 11, 2025, <https://www.tempus.com/news/tempus-introduces-loop-an-ai-powered-target-discovery-and-validation-platform/>
15. MedMimic: A Physician-Inspired Multimodal Fusion Framework for Early Diagnosing Fever of Unknown Origin - arXiv, 6. 11, 2025, <https://arxiv.org/html/2502.04794v1>
16. Medical Heterogeneous Graph Transformer for Disease Diagnosis - Engineering Letters, 6. 11, 2025, https://www.engineeringletters.com/issues_v32/issue_12/EL_32_12_11.pdf
17. The future of multimodal artificial intelligence models for integrating imaging and clinical metadata - Diagnostic and Interventional Radiology, 6. 11, 2025, <https://dirjournal.org/pdf/beb8919b-f013-4ea1-b1c8-40332e840fe1/articles/dir.2024.242631/DIR-2024-242631-AYDA.pdf>
18. A multi-modal deep learning solution for precise pneumonia diagnosis: the PneumoFusion-Net model - PMC, 6. 11, 2025,



- <https://pmc.ncbi.nlm.nih.gov/articles/PMC11937601/>
19. MULTIMODAL LARGE LANGUAGE MODELS FOR AUTOMATED DIAGNOSIS AND CLINICAL DECISION SUPPORT, 6. 11, 2025, <https://aircconline.com/csit/papers/vol15/csit150515.pdf>
 20. Multimodal Data Fusion for Bioinformatics Artificial Intelligence 9781394269938, 6. 11, 2025, <https://dokumen.pub/multimodal-data-fusion-for-bioinformatics-artificial-intelligence-9781394269938.html>
 21. Tempus Introduces Fuses, A Program Designed to Transform ..., 6. 11, 2025, <https://investors.tempus.com/news-releases/news-release-details/tempus-introduces-fuses-program-designed-transform-therapeutic>
 22. AI Driven Healthcare: Leveraging Multimodal Data for Precision Health - IQVIA, 6. 11, 2025, <https://www.iqvia.com/locations/middle-east-and-africa/blogs/2025/04/ai-driven-healthcare-leveraging-multimodal-data-for-precision-health>
 23. PathAI Showcases AI-Powered Pathology Innovations at the United States, 6. 11, 2025, <https://www.pathai.com/resources/pathai-showcases-ai-powered-pathology-innovations-at-the-united-states-and-canadian-academy-of-pathology-uscap-114th-annual-meeting/>
 24. Precision for Medicine and PathAI Announce Strategic Collaboration ..., 6. 11, 2025, <https://www.prnewswire.com/news-releases/precision-for-medicine-and-pathai-announce-strategic-collaboration-to-advance-ai-powered-clinical-trial-services-and-biospecimen-products-302438551.html>
 25. Multimodal Intelligence Surges: From AI Assistants to Precision Medicine and Global Logistics, 6. 11, 2025, <https://www.geekynews.org/?newsletter=multimodal>
 26. 1. 1, 1970, <https://www.iqvia.com/locations/middle-east-and-africa/blogs/2025/04/ai-driven-healthcare-leveraging-multimodal-data-for-precision-health>