

## 引文格式:

谢卓恒, 伊鸣, 黄新瑞. 多实例学习在医学图像分析中的应用进展 [J]. 集成技术, 2025, 14(2): 24-32.

Xie ZH, Yi M, Huang XR. Application progress of multi-instance learning in medical image analysis [J]. Journal of Integration Technology, 2025, 14(2): 24-32.

## 多实例学习在医学图像分析中的应用进展

谢卓恒<sup>1</sup> 伊鸣<sup>2,3\*</sup> 黄新瑞<sup>1,3\*</sup>

<sup>1</sup>(北京大学基础医学院生物物理学系 北京 100191)

<sup>2</sup>(北京大学神经科学研究所 北京 100191)

<sup>3</sup>(神经科学教育部重点实验室/国家卫生健康委员会神经科学重点实验室 北京 100191)

**摘要** 多实例学习 (multiple-instance learning, MIL) 是一种弱监督学习方法, 近年来广泛应用于医学图像分析领域。本文综述了 MIL 在全切片图像中的应用进展, 详细分析了其在肿瘤检测、亚型分级和生存预测中的作用。MIL 在弱监督学习中具有独特优势, 可通过引入新机制进行优化和拓展, 以适应更多的应用场景。本文首先综述了部分应用广泛或独具优势的 MIL 模型, 并详细介绍了它们的技术特点和具体应用场景; 其次, 介绍了 MIL 在多模态医学图像分析中的应用进展和技术进步; 最后, 总结了 MIL 目前的研究进展, 并展望了其未来发展。

**关键词** 图像分析; 多实例学习; 医学图像; 机器学习; 深度学习

中图分类号 R319 文献标志码 A doi: 10.12146/j.issn.2095-3135.20241111001

CSTR: 32239.14.j.issn.2095-3135.20241111001

Application Progress of Multi-instance Learning in  
Medical Image Analysis

XIE Zhuoheng<sup>1</sup> YI Ming<sup>2,3\*</sup> HUANG Xinrui<sup>1,3\*</sup>

<sup>1</sup>(Department of Biophysics, School of Basic Medical Sciences, Peking University, Beijing 100191, China)

<sup>2</sup>(Neuroscience Research Institute, Peking University, Beijing 100191, China)

<sup>3</sup>(Key Lab for Neuroscience, the Ministry of Education of China; Key Lab for Neuroscience, the National Health and Family Planning Commission of China, Beijing 100191, China)

\*Corresponding Authors: mingyi@hsc.pku.edu.cn; huangxr@pku.edu.cn

**Abstract** Multiple-instance learning (MIL), as a weakly supervised learning method, has been widely applied in the field of medical image analysis in recent years. The paper reviews the progress of MIL applications in whole slide images, with a detailed analysis of its roles in tumor detection, subtype

收稿日期: 2024-11-11 修回日期: 2024-12-28

基金项目: 北京大学医学部教育教学研究课题项目 (2022YB17); 北京市自然科学基金面上项目 (4242004); 国家蛋白质科学研究 (北京) 设施北京大学分中心开放课题项目 (KF-202402); 国家自然科学基金项目 (32271053); 北京市自然科学基金-海淀原始创新联合基金项目 (L222016)

作者简介: 谢卓恒, 本科生, 研究方向为临床医学; 伊鸣 (通讯作者), 博士, 研究员, 研究方向为疼痛与认知相关的神经机制和行为学, E-mail: mingyi@hsc.pku.edu.cn; 黄新瑞 (通讯作者), 博士, 讲师, 研究方向为生物医学多模态成像及其数据图像处理, E-mail: huangxr@pku.edu.cn.

classification, and survival prediction. MIL holds unique advantages in weakly supervised learning, which can be optimized and extended through the introduction of new mechanisms to adapt to a broader range of application scenarios. The paper first reviews some widely used or uniquely advantageous MIL models, elaborating on their technical features and specific application contexts. Secondly, it introduces the application and technology advancements of MIL in multimodal medical image analysis. Finally, the current research progress of MIL is summarized, and its future development prospects are explored.

**Keywords** image analysis; multiple-instance learning; medical images; machine learning; deep learning

**Funding** This work is supported by Peking University Health Science Center Medical Education Research Funding Project (2022YB17), Beijing Natural Science Foundation (4242004), Open Research Fund of the National Center for Protein Sciences at Peking University in Beijing (KF-202402), National Natural Science Foundation of China (32271053), and Beijing Natural Science Foundation - Haidian Original Innovation Joint Fund (L222016)

## 1 引言

1997年, Dietterich等<sup>[1]</sup>提出了多实例学习(multi-instance learning, MIL)。其最早用于解决药物活性预测中的问题,核心思想如下:数据由“包(bag)”组成,每个包包含多个实例(instances),但只有包有标签,实例标签信息不明确。这种学习框架适用于无法明确标注单个实例的任务,因此非常适合处理复杂数据集。目前, MIL已经得到更广泛的应用,并在医学图像分析领域取得一定研究进展。本文旨在综述 MIL在医学图像分析领域,尤其是医学图像分类领域的应用进展,总结 MIL的具体应用领域,并提供应用参考,展望其未来研究方向。

在早期发展历程中, Dietterich等<sup>[1]</sup>提出的轴平行矩形方法是 MIL的首个应用,它能利用几何方法划分特征空间中的正负类实例。随后,基于支持向量机的 MI-SVM模型和基于集成学习的 boosting算法被引入 MIL中<sup>[2]</sup>。这些算法通过优化实例和包的分类边界提高 MIL在不同领域的应用效果。

随着深度学习的兴起, MIL逐渐与深度卷积神经网络结合,形成深度 MIL模型。该模型能通过深度卷积神经网络提取实例的高维特征,然后使用池化或注意力机制将这些特征进行聚合,生成包的特征表示。目前,深度学习与 MIL结合的应用已十分广泛。例如,深度 MIL(deep MIL)应用于全切片图像(whole slide images, WSI)不仅能分类图像,还能学习包内实例(如图像中的局部区域或像素)的特征<sup>[3]</sup>。

近年来,关键技术的突破促使 MIL技术在医学图像分类中显示出更强大的应用潜力,尤其是处理 WSI、弱监督学习和多模态数据融合等方面的应用,进展显著。

## 2 多实例学习在全切片图像分析领域的研究进展

MIL是一种弱监督学习方法,尤其适用于 WSI分析任务。WSI的分辨率极高,每幅图像由大量像素构成,因此无法精确标注每个像素或小块。而 MIL将这些高像素图像视为一个“包”,

对于其中包含的多个“实例”，模型仅须标记整个袋子，无须逐个标记每个实例。换句话说，只要袋子中至少包含一个阳性实例，那么整个袋子就会被标记为阳性。这使得 MIL 尤其适用于 WSI 中的肿瘤检测和分类等任务。

随着计算病理学的发展，MIL 在 WSI 分类中的应用被广泛关注和研究<sup>[4]</sup>。许多先进的 MIL 框架已经在 WSI 任务中展现出出色的性能。例如：基于注意力的 MIL (attention-based MIL, Attention-MIL) 通过引入注意力机制，自动聚焦图像中最具诊断价值的区域，如肿瘤核心或淋巴结转移等，以提高病变区域的识别精度；在乳腺癌或肺癌的 WSI 分析中，Attention-MIL 能精准识别微小病变，即使标注稀缺或图像复杂，也能有效聚焦关键区域；Attention-MIL 生成的热力图可以直观展示关注区域，帮助医生了解病情，以取得较好的诊断效果。

检索增强 MIL (retrieval-augmented MIL) 也在 WSI 领域表现出色。在医学图像分析中，尤其是癌症诊断，标注数据通常非常有限，检索增强 MIL 能从大规模图像库中检索与当前病例相似的病理切片图像，帮助检索增强 MIL 学习更多潜在的病变特征。检索增强 MIL 通过检索到的相似图像提供的额外信息，提高识别微小或早期病变的能力，可增强检索增强 MIL 的鲁棒性<sup>[5]</sup>。

MIL 主要用于肿瘤检测及分类、癌症亚型分级和生存预测。在肿瘤的检测和分类方面，MIL 在没有像素级标注的情况下，通过弱监督学习的方式，可以有效检测 WSI 中的癌变区域。关于癌症亚型分级，MIL 模型能结合图像特征和分子信息，帮助医生进行更细粒度的分类，进行癌症亚型和组织分级。就生存预测而言，MIL 主要通过结合多模态数据 (如影像、基因表达数据等)，达到预测患者的生存率和治疗效果的目的<sup>[6]</sup>。

MIL 在 WSI 中的应用不仅提升了肿瘤检测的精度，还拓展了其在癌症分类、亚型分级和生

存预测等方面的应用前景<sup>[6]</sup>。未来，随着对注意力机制、对比学习和多模态数据处理的进一步优化，MIL 在 WSI 肿瘤检测中的应用将更广泛和深入。

### 3 多实例学习在弱监督学习中的技术进步与拓展

MIL 在弱监督学习中的应用日益受到关注，尤其是在医学图像分析领域。弱监督学习通过使用不完全或不精确的标签，减少对大量精确标注数据的依赖，尤其适用于弱监督学习和其他复杂数据的分析。除了在弱监督学习中的独特优势外，MIL 发展出的诸多新方法也在不断寻求新的技术突破，并引入其他相关技术，达到拓宽其应用场景的目的。

#### 3.1 多实例学习在弱监督学习中的优势

弱监督学习利用有限或粗粒度标签 (如包级别标签) 进行学习，无须实例级别的精确标注。MIL 提供了一种自然的框架，以处理这一挑战，因为 MIL 的基本假设是通过包标签进行实例分类。每个包包含多个实例，包的标签仅提供包内某些实例的信息，而非具体的实例标签<sup>[7]</sup>。这极大地降低了标注成本，同时，利用 MIL 模型能自动定位和检测病灶。

#### 3.2 关键技术进展

##### 3.2.1 注意力机制

注意力机制是弱监督 MIL 模型中的一种关键技术，允许模型自动关注包内关键实例，并根据这些实例的权重进行分类，因此 MIL 在不依赖实例标签的情况下，自动定位对分类任务至关重要的区域或实例。注意力机制广泛应用于肿瘤检测任务中，通过对图像中重要区域的自动识别，提高模型的分类精度。研究表明，基于注意力的弱监督 MIL 在 WSI 分类中表现出色，尤其是在病灶的定位上<sup>[8]</sup>。

### 3.2.2 自监督学习与对比学习

近年来, 自监督学习和对比学习也逐渐被引入 MIL 的弱监督框架中。与传统的监督学习不同, 自监督学习无须人工标注数据, 而是通过设计一些简单的任务帮助模型学习数据的结构。例如, 模型可以通过学习如何预测图像中的遮挡部分或判断图像的旋转角度学习有分辨价值的特征。因此模型能通过上述任务自动提取数据中的关键信息, 进而提升对其他任务(如分类、检测等)的表现。

通过自监督学习, MIL 模型可以从未标注的数据中学习有用的特征表示, 对比学习则通过最大化正负实例的差异提升模型的分​​类性能。一些研究提出将对​​比学习与 MIL 结合, 通过结合自监督对比学习, MIL 模型能在学习后更鲁棒。对比学习尤其适用于肿瘤分类等任务, 可以在没有实例级标签的情况下提升分类效果<sup>[9]</sup>。此外, 半监督学习、部分或不完整标签学习等方法也逐渐被引入 MIL 的具体应用中, 这为 MIL 提供了新的处理方法和应用场景<sup>[10]</sup>。

### 3.2.3 少样本学习与迁移学习

在医学图像分类中, 弱监督学习还结合了迁移学习和少样本学习技术。通过将预训练的模型迁移到医学图像分类任务上, 可以在有限的标注数据下快速训练出高性能模型。例如, 利用预训练的 ResNet 模型进行乳腺癌 X 射线图像的检测和分级, 可减少对大规模标注数据的需求<sup>[11]</sup>。

## 3.3 模型优化与拓展

MIL 在弱监督学习中的应用主要集中于 WSI 分析。由于 WSI 的规模大且难以逐像素标注, 因此 MIL 在 WSI 分类任务中显示出独特优势。通过包级别的标签, MIL 模型能自动检测和分类切片中的关键区域, 尤其适用于肿瘤检测。

在弱监督学习的应用过程中, 科研工作者通过改进模型使 MIL 的应用范围更广。为适应

WSI 对医学图像分类的不同需求, MIL 在研究中逐渐发展, 形成了不同的技术改进形式和满足相应需求的适应性策略。目前已有一定数量的改进后的 MIL 方法, 其命名主要在 MIL 之前添加前缀, 如表 1 所示。这些方法的共同点是它们都致力于解决 WSI 分类中的高分辨率、数据异质性和弱标注等挑战性问题, 通过不同技术手段提高模型的准确性。

例如, 基于注意力的 MIL 通过弱监督的方式, 仅依赖图像级标签, 无须实例级标注。通过注意力机制, MIL 模型能自动识别最相关的实例(图像块), 并基于这些实例进行分类。基于注意力的 MIL 适用于没有充分标注的 WSI 分类任务<sup>[3, 12-13]</sup>。而双流 MIL 结合了自监督学习, 能在弱监督环境下有效训练模型。通过使用对比学习技术, 双流 MIL 能自动提取有用的特征, 增强实例级和包级分类能力, 在 WSI 癌症检测中表现突出<sup>[14]</sup>。基于对比学习框架的 MIL 允许模型在没有标注数据的情况下进行预训练, 提高了其在 WSI 中的分类性能。基于对比学习框架的 MIL 先对未标注的图像块进行特征提取, 后进行弱监督分类<sup>[15-16]</sup>。端到端 MIL 是典型的扩展实例, 通过引入解释性模块和不确定性估计, 提高 WSI 分类中模型的透明性和可靠性<sup>[17]</sup>。掩码难实例挖掘 MIL 通过掩蔽易分类实例, 专注于困难实例的学习, 增强了模型的准确性<sup>[18]</sup>。上述 MIL 模型通常利用准确率和受试者工作特征曲线下面积 (area under curve, AUC) 评估分类性能。此外, 更多优化拓展模型在科研人员的努力中不断问世, 拓展了 MIL 的应用场景。

综上所述, MIL 及其优化拓展模型为弱监督学习提供了强大工具, 特别是在处理 WSI 等高级数据时, 能通过包级别的标签实现精准分类。随着机制模型的引入, MIL 得到有效优化和拓展, 在弱监督学习中的应用前景将更广阔, 尤其是在医学图像分析和多模态数据处理等领域。

表 1 多实例学习的优化拓展模型举例

Table 1 Examples of optimization and expansion model for multi-instance learning

模型名称	基本原理	应用领域	模型评估
基于注意力的 MIL <sup>[5,12-13]</sup>	常见的 MIL 方法之一, 通过为每个实例分配注意力权重, 自动聚焦最有可能含有病灶的区域	WSI; 弱监督学习	使用肺癌 WSI 数据集评估, AUC 为 0.7143 <sup>[3]</sup>
检索增强 MIL <sup>[5]</sup>	通过引入最优传输距离解决跨域泛化问题, 提升模型在不同医院数据集上的表现	多模态数据融合; 弱监督学习	使用乳腺癌 (Camelyon16) WSI 数据集评估, AUC 为 0.945 1, 准确率为 0.920 0
双流 MIL <sup>[14]</sup>	结合自监督对比学习和实例级、包级分类器, 用于聚合不同尺度的实例特征, 提升分类精度	多模态数据融合; 弱监督学习	使用癌症基因组图谱 (the cancer genome atlas, TCGA) 中的肺癌 WSI 数据集和 Camelyon16 中的乳腺癌 WSI 数据集评估, 对应的 AUC 分别为 0.963 3 和 0.894 4, 对应的准确率分别为 0.919 0 和 0.868 2
基于对比学习框架的 MIL <sup>[15-16]</sup>	利用对比学习框架进行预训练, 以增强 MIL 模型的特征提取能力, 特别是用于无标签或弱标签数据集的任务	弱监督学习	在 5 种医学图像领域 (临床皮肤病学摄影、眼底成像、数字病理学、胸部 X 射线和乳腺 X 射线) 的未标注数据集、领域内数据集和领域外数据集构建, 与强监督基线相比, AUC 均提高 <sup>[15]</sup>
端到端 MIL <sup>[17]</sup>	主要由 3 个模块组成: 细节感知注意力蒸馏、结构感知注意力精炼和不确定性感知实例分类器	WSI; 弱监督学习	用两个肾癌 (TEAH & TCGA) 和肺癌 (TCGA) WSI 数据集评估, 对应的 AUC 分别为 0.952 0、0.988 0 和 0.929 0, 对应的准确率分别为 0.852 0、0.927 0 和 0.859 0
掩码难实例挖掘 MIL <sup>[18]</sup>	该方法通过屏蔽易分类的实例, 强制模型专注于难分类区域, 提升了模型的泛化能力	WSI; 弱监督学习	使用 TCGA 中的肺癌 WSI 数据集和 Camelyon16 中的乳腺癌 WSI 数据集评估, 对应的 AUC 分别为 0.955 3 和 0.964 9, 对应的准确率分别为 0.900 2 和 0.924 8
可训练原型增强 MIL <sup>[19-20]</sup>	通过结合原型学习, 将所有实例划分为原型, 提升分类器对实例的区分能力, 并增强模型的解释性	WSI; 弱监督学习	使用肺癌 (TCGA) 和脑癌 (TCGA) WSI 数据集评估, 对应的 AUC 分别为 0.979 9 和 0.941 7, 对应的准确率分别为 0.942 7 和 0.831 6 <sup>[20]</sup>
基于图神经网络的 MIL <sup>[21-22]</sup>	引入图神经网络, 将 WSI 中的实例构建为图节点, 增强了不同实例之间的关联表示能力	多模态数据融合	使用非小细胞癌 (TCGA) 和肾细胞癌 (TCGA) WSI 数据集评估, 对应的 AUC 分别为 0.952 0 和 0.992 0, 对应的准确率分别为 0.902 0 和 0.954 0 <sup>[22]</sup>
跨尺度蒸馏 MIL <sup>[23]</sup>	该方法结合了不同尺度的特征, 通过蒸馏学习技术提升分类精度	WSI; 弱监督学习	使用乳腺癌 (Camelyon16) WSI 数据集评估, AUC 为 0.973 0, 准确率为 0.945 0

## 4 多实例学习应用于多模态医学图像分析的研究进展

MIL 在多模态医学图像分析中得到广泛应用, 在融合不同成像模式以提高诊断精度和临床决策支持方面表现尤其出色。

### 4.1 多模态医学图像融合的优势

目前, MIL 已经在计算机断层扫描 (computed tomography, CT)、核磁共振成像 (magnetic resonance imaging, MRI) 和 B 型超声 (brightness-mode ultrasound) 等医学影像检查的图像分类上取得了一定进展, 主要通过其特征性的数据处理方式解决医学图像标注不完全的问题。例如, 在 CT 和 MRI 的医学图像分类领域, MIL 被广泛应用于肿瘤检测<sup>[24]</sup>, 而针对 MRI 的组

织高分辨率特性, MIL 在阿尔茨海默病的诊断上也取得了一定技术进展<sup>[25]</sup>; 在超声方面, MIL 已被用于先天缺陷诊断和病灶识别领域, 并展现出显著优势<sup>[26]</sup>。近年来, 科研人员已在这些传统模态中进行了大量 MIL 模型改进工作<sup>[27]</sup>, MIL 在这些领域的应用也逐渐落实到具体临床问题上。但 MIL 对医学图像的分析大多局限于单一模态, 这可能导致患者的医学影像信息缺乏全面分析, 继而影响模型诊断的准确性。

相比之下, 多模态医学图像结合了不同类型的医学影像数据, 为医生提供了更全面的患者医学影像信息。每一种医学影像都具有独特的特征和优势: 在骨骼和硬组织的显影上, CT 不可替代; MRI 能实现对软组织高分辨率的显影; 正电子发射断层扫描可以利用放射性示踪剂检测体内

的代谢活动。此外, 还有更多适用于专科的医学影像数据。而多模态 MIL 能将不同模态的数据作为实例进行处理, 可以有效提取每个模态的独特特征, 同时实现特征融合, 以提高诊断准确性。例如, 在自闭症预测任务中, 研究者利用多模态图生成图卷积网络, 通过不同模态的图融合(如脑功能连接图和表型数据)进行预测<sup>[28]</sup>。在癌症生存预测任务中, 研究者利用 MIL 结合病理图像和基因组学数据, 生成更全面的肿瘤特征<sup>[29]</sup>。

#### 4.2 多实例学习应用于多模态医学图像分析的技术支持

近年来, 深度学习模型在多模态医学图像分析中的应用越来越多。基于对抗生成网络的方法用于 MRI 和 CT 图像的相互转化<sup>[30]</sup>, 解决了多模态数据分布不匹配的问题<sup>[31]</sup>。通过这种技术, MIL 能在无配对数据的情况下学习不同模态图像之间的转换关系, 从而提高多模态图像的融合效果。

为保护患者隐私, 联邦学习与多模态医学图像分析结合也是研究热点。已有研究提出了跨模态垂直联邦学习的框架, 通过在不同机构之间共享学习模型而不是数据, 实现了 MRI 重建任务中的多模态特征提取和融合<sup>[32]</sup>。这种方法有效解决了跨数据源的域偏移问题, 同时确保了数据隐私。通过这项技术, MIL 在处理多模态数据时不但能有效提高模型精度和鲁棒性, 还能保护数据隐私, 并支持跨机构合作。

#### 4.3 多实例学习在多模态医学图像分析领域的具体应用

在多模态医学图像分析中, 使用 MIL 的方法能结合不同类型的医学数据(如病理图像、基因组数据等)进行更准确的疾病分类和生存预测。

经典的多模态 MIL 是一种结合多个数据模式的 MIL 框架, 通常用于整合不同的医学数据来源(如病理图像和基因数据)。这种方法可以同时分析病理图像中的组织形态和基因变异信息, 从

而提供更全面的疾病诊断。当其用于癌症生存预测和肿瘤分类时, 尤其是在结合组织学图像和分子数据(如基因表达、DNA 突变等)时, 能提高预测准确性<sup>[33]</sup>。最近, 多模型 MIL 在心电图的分类和心血管类疾病的诊断方面也有许多创新应用<sup>[34-35]</sup>; 而其与视觉转换器(vision Transformer, ViT)结合的基于视觉转换器的多模态 MIL(vision Transformer-based multimodal MIL)通过对病理图像块进行自注意力机制的处理, 能更好地理解 WSI 的局部与全局信息。在多模态框架中, ViT 可以结合图像和其他数据(如基因数据), 从多个角度进行数据融合。基于视觉转换器的多模态 MIL 主要应用于癌症亚型分类、肿瘤分级和基因变异预测等<sup>[36-37]</sup>。

图神经网络(graph neural network, GNN)与多模态 MIL 结合, 可用于构建和分析实例间的关系。GNN 是一类专门处理图结构数据的深度学习模型。在图结构中, 数据通过节点和边表示, 通过消息传递机制(message passing), 让每个节点的呈现不仅依赖于自身的特征, 还依赖于其邻居节点的信息。通过多次迭代的传播过程, 节点逐步汇聚来自邻居节点的信息, 最终达到反映节点间复杂关系的目的。GNN 能将不同模态的数据(如影像和基因信息)以图的形式表示, 并通过图的传递机制捕捉更丰富的特征。在生存预测和基因变异分析中, 这项技术能结合病理图像和基因组数据, 通过多模态数据图的表示, 实现更精准的分类<sup>[21-22]</sup>。

对比学习结合 MIL, 可以整合不同类型医学图像(如 CT、MRI 及病理图像), 已在医学图像分类中应用<sup>[38]</sup>。进行多模态数据处理时, 在大量标注数据的情况下, 通过对比学习的自监督学习预训练模型, 提取各模态间的相关特征, 然后整合不同类型的医学图像数据与其他模态(如基因数据、临床信息), 已应用于生存预测和癌症分类<sup>[39-40]</sup>。

这些多模态 MIL 方法通过结合不同类型的医学数据,进一步提升对复杂疾病的分类、分级和生存预测的准确性,拓展了医学图像分析的应用范围。

## 5 结论

首先,本文总结了 MIL 在医学图像分类中的应用进展,特别是在 WSI 中的应用;其次,本文对多个优化后的 MIL 模型进行了详细分析,探讨了它们的技术优势和具体应用场景,如 Attention-MIL、掩码难实例挖掘 MIL 和双流 MIL 等,这些模型利用注意力机制、对比学习和自监督学习技术,提高了分类的准确性和模型的可解释性<sup>[28]</sup>;最后,本文总结了 MIL 在多模态医学图像分析中的应用,展示了它在结合不同数据模态(如病理图像和基因组数据)时的巨大潜力。

未来, MIL 的发展方向主要集中在以下几个方面:首先,进一步优化注意力机制和自监督学习方法,以应对标注数据稀缺的挑战;其次,多模态数据融合技术需进一步完善,特别是通过结合多种医学图像和临床数据,为复杂疾病的诊断提供更全面的支持;最后,多模态数据融合技术还面临跨数据源整合等调整,体现在数据异质性、对齐问题、缺失数据、标注不一致和融合策略的选择上。不同模态数据的特性不同,导致它们的处理和融合变得更复杂。如何有效对齐不同数据源、处理缺失数据和噪声、统一不同的标注体系、选择合适的融合策略,以及保证模型的计算效率和准确性,是当前多模态数据融合面临的主要问题。这些问题需结合先进的算法和技术解决,尤其是在医学等复杂领域,如何高效且准确地整合不同类型的数据仍然是一个亟待攻克的难题。

MIL 需要在保护数据隐私的前提下,实现更精确的医学图像分类和个性化医疗决策<sup>[31-32]</sup>。这

对提升模型的自适应能力和跨模式的数据融合提出了新要求。为满足未来发展需求,科研人员已取得部分技术突破,群体学习 (swarm learning) 结合区块链的分布式安全性和边缘计算的低延迟特性,形成了一种无中心服务器的架构,适合多机构数据的安全协作<sup>[41]</sup>。此外,个性化学习方法(如 pFLFE)以特征增强为基础,能自适应地提升局部模型的准确性,尤其适用于异构医学数据<sup>[42]</sup>。MIL 能进一步完善医学图像分类领域的相关技术,提升多模态数据综合分析能力,并实现个性化的医学图像分类与医疗决策,为人工智能的临床应用提供更强的技术支持。

## 参考文献

- [1] Dietterich TG, Lathrop RH, Lozano-Pérez T. Solving the multiple instance problem with axis-parallel rectangles [J]. *Artificial Intelligence*, 1997, 89(1-2): 31-71.
- [2] Auer P, Ortner R. A boosting approach to multiple instance learning [J]. *Lecture Notes in Computer Science*, 2004, 3201: 63-74.
- [3] Yao JW, Zhu XL, Jonnagaddala J, et al. Whole slide images based cancer survival prediction using attention guided deep multiple instance learning networks [J]. *Medical Image Analysis*, 2020, 65: 101789.
- [4] Huang XR, Li ZT, Zhang MH, et al. Fusing hand-crafted and deep-learning features in a convolutional neural network model to identify prostate cancer in pathology images [J]. *Frontiers in Oncology*, 2022, 12: 994950.
- [5] Cui YF, Liu ZQ, Chen YX, et al. Retrieval-augmented multiple instance learning [C] // *Proceedings of the 37th Conference on Neural Information Processing Systems (NeurIPS 2023)*, *Advances in Neural Information Processing Systems*, 2023: 24859-24878.
- [6] Li XY, Yang B, Chen TD, et al. Multiple instance learning-based two-stage metric learning network for whole slide image classification [J]. *The Visual*

- Computer, 2024, 40(8): 5717-5732.
- [7] Park S, Kim H, Kim M, et al. Normality guided multiple instance learning for weakly supervised video anomaly detection [C] // Proceedings of the 2023 IEEE/CVF Winter Conference on Applications of Computer Vision, 2023: 2664-2673.
- [8] Ilse M, Tomczak JM, Welling M. Attention-based deep multiple instance learning [C] // Proceedings of the 35th International Conference on Machine Learning, 2018: 3376-3391.
- [9] Zhang SW, Ke W, Yang L. Discovery-and-selection: towards optimal multiple instance learning for weakly supervised object detection [Z/OL]. arXiv Preprint, arXiv: 2110.09060, 2021.
- [10] Nguyen T, Raich R. Incomplete label multiple instance multiple label learning [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022, 44(3): 1320-1337.
- [11] Zhang XQ, Liu C, Zhu HT, et al. A universal multiple instance learning framework for whole slide image analysis [J]. Computers in Biology and Medicine, 2024, 178(6): 108714.
- [12] Chen JH, Zeng HY, Zhang C, et al. Lung cancer diagnosis using deep attention-based multiple instance learning and radiomics [J]. Medical Physics, 2022, 49(5): 3134-3143.
- [13] Xiang HC, Shen JY, Yan QG, et al. Multi-scale representation attention based deep multiple instance learning for gigapixel whole slide image analysis [J]. Medical Image Analysis, 2023, 89(6): 102890.
- [14] Li B, Li Y, Eliceiri KW. Dual-stream multiple instance learning network for whole slide image classification with self-supervised contrastive learning [C] // Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021: 14318-14328.
- [15] Sharma S, Siddiqui MK, Akhtar MS, et al. Domain-aware self-supervised pre-training for label-efficient meme analysis [C] // Proceedings of the 2nd Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics and the 12th International Joint Conference on Natural Language Processing, 2022: 792-805.
- [16] Azizi S, Culp L, Freyberg J, et al. Robust and data-efficient generalization of self-supervised machine learning for diagnostic imaging [J]. Nature Biomedical Engineering, 2023, 7(6): 756-779.
- [17] Shi JB, Li C, Gong TL, et al. E<sup>2</sup>-MIL: an explainable and evidential multiple instance learning framework for whole slide image classification [J]. Medical Image Analysis, 2024, 97(4): 103294.
- [18] Tang WH, Huang S, Zhang XX, et al. Multiple instance learning framework with masked hard instance mining for whole slide image classification [C] // Proceedings of the 2023 IEEE/CVF International Conference on Computer Vision, 2023: 4055-4064.
- [19] Qu LH, Ma YF, Luo XY, et al. Rethinking multiple instance learning for whole slide image classification: a good instance classifier is all you need [J]. IEEE Transactions on Circuits and Systems for Video Technology, 2024, 34(10): 9732-9744.
- [20] Yang LT, Mehta D, Liu SD, et al. TPMIL: trainable prototype enhanced multiple instance learning for whole slide image classification [C] // Proceedings of Machine Learning Research, 2023: 1655-1665.
- [21] Wang J, Mao Y, Guan N, et al. Advances in multiple instance learning for whole slide image analysis: techniques, challenges, and future directions [Z/OL]. arXiv Preprint, arXiv: 2408.09476, 2024.
- [22] Guan YH, Zhang J, Tian K, et al. Node-aligned graph convolutional network for whole-slide image representation and classification [C] // Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022: 18813-18823.
- [23] Bontempo G, Porrello A, Bolelli F, et al. DAS-MIL: distilling across scales for MIL classification of histological WSIs [C] // Proceedings of the 26th International Conference on Medical Image Computing and Computer-Assisted Intervention, 2023: 248-258.



- [24] Qi SL, Xu CW, Li C, et al. DR-MIL: deep represented multiple instance learning distinguishes COVID-19 from community-acquired pneumonia in CT images [J]. *Computer Methods and Programs in Biomedicine*, 2021, 211(2): 106406.
- [25] Zhu WY, Sun L, Huang JS, et al. Dual attention multi-instance deep learning for Alzheimer's disease diagnosis with structural MRI [J]. *IEEE Transactions on Medical Imaging*, 2021, 40(9): 2354-2366.
- [26] Yin S, Peng QM, Li HM, et al. Multi-instance deep learning of ultrasound imaging data for pattern classification of congenital abnormalities of the kidney and urinary tract in children [J]. *Urology*, 2024, 142: 183-189.
- [27] Struski Ł, Janusz S, Tabor J, et al. Multiple instance learning for medical image classification based on instance importance [J]. *Biomedical Signal Processing and Control*, 2023, 91(1-2): 105874.
- [28] Mao JB, Liu J, Lin HH, et al. Multi-modal multi-kernel graph learning for autism prediction and biomarker discovery [Z/OL]. *arXiv Preprint, arXiv: 2303.03388*, 2023.
- [29] Peng SJ, Cai YH, Yao ZJ, et al. Weakly-supervised video anomaly detection via temporal resolution feature learning [J]. *Applied Intelligence*, 2023, 53(22): 1-19.
- [30] Li ZT, Huang XR, Zhang ZR, et al. Synthesis of magnetic resonance images from computed tomography data using convolutional neural network with contextual loss function [J]. *Quantitative Imaging in Medicine and Surgery*, 2022, 12(6): 3151-3169.
- [31] Xu Y. Deep learning in multimodal medical image analysis [C] // *Proceedings of the Health Information Science*, 2019: 193-200.
- [32] Yan YL, Wang H, Huang YW, et al. Cross-modal vertical federated learning for MRI reconstruction [J]. *IEEE Journal of Biomedical and Health Informatics*, 2024, 28(11): 6383-6394.
- [33] Cai W, Gao HY, Chu CC, et al. Engineering phototheranostic nanoscale metal-organic frameworks for multimodal imaging-guided cancer therapy [J]. *ACS Applied Materials and Interfaces*, 2016, 9(3): 2040-2051.
- [34] Han HZ, Lian C, Zeng ZG, et al. Multimodal multi-instance learning for long-term ECG classification [J]. *Knowledge-Based Systems*, 2023, 270(3): 110555.
- [35] Huang Z, Yu XW, Wessler BS, et al. Semi-supervised multimodal multi-instance learning for aortic stenosis diagnosis [Z/OL]. *arXiv Preprint, arXiv: 2403.06024*, 2024.
- [36] Bi Q, Sun X, Yu S, et al. MIL-ViT: a multiple instance vision Transformer for fundus image classification [J]. *Journal of Visual Communication and Image Representation*, 2023, 97(12): 103956.
- [37] Han X, Wang YT, Feng JL, et al. A survey of Transformer-based multimodal pre-trained models [J]. *Neurocomputing*, 2022, 515(2): 89-106.
- [38] Zhu ZH, Yu LQ, Wu W, et al. MuRCL: multi-instance reinforcement contrastive learning for whole slide image classification [J]. *IEEE Transactions on Medical Imaging*, 2022, 42(5): 1337-1348.
- [39] Jiang CY, Xu HY, Dong MF, et al. Hallucination augmented contrastive learning for multimodal large language model [C] // *Proceedings of the 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024: 27036-27046.
- [40] Wang YC, Luo Y, Li B, et al. Multi-modality fusion based lung cancer survival analysis with self-supervised whole slide image representation learning [C] // *Proceedings of the 2023 Pattern Recognition and Computer Vision: 6th Chinese Conference*, 2023: 333-345.
- [41] Koutsoubis N, Waqas A, Yilmaz Y, et al. Future-proofing medical imaging with privacy-preserving federated learning and uncertainty quantification: a review [Z/OL]. *arXiv Preprint, arXiv: 2409.16340*, 2024.
- [42] Xie LY, Lin MQ, Liu SY, et al. pFLFE: cross-silo personalized federated learning via feature enhancement on medical image segmentation [C] // *Proceedings of the Medical Image Computing and Computer Assisted Intervention*, 2024: 559-610.