

中国科学技术大学

University of Science and Technology of China

硕士学位论文



论文题目 基于多视图深度学习的非小细胞肺癌

组织亚型分类方法研究

作者姓名 高子玉

学科专业 生物医学工程

导师姓名 王明会 副教授

完成时间 二〇二四年五月

中国科学技术大学

硕士学位论文



基于多视图深度学习的非小细胞 肺癌组织亚型分类方法研究

作者姓名：高子玉

学科专业：生物医学工程

导师姓名：王明会 副教授

完成时间：二〇二四年五月二十五日

University of Science and Technology of China
A dissertation for master's degree



Study on Subtype Classification of Non-small Cell Lung Cancer Based on Multi- view Deep Learning

Author: Ziyu Gao

Speciality: Biomedical Engineering

Supervisors: Assoc Prof. Minghui Wang

Finished time: May 25th, 2024

摘 要

根据世界卫生组织的数据显示,肺癌是全球最主要的癌症死亡原因,对公众的生命健康构成严峻挑战。其中非小细胞肺癌占肺癌患者的 85%,且不同组织亚型患者的治疗方法和预后复发率存在较大差异。因此,从计算机断层扫描(CT)图像中准确和无创地检测组织学亚型对于为患者制定适当的治疗策略至关重要。当前研究表明,通过整合多个视图的数据能够实现对复杂疾病的全面理解和精确建模,从而有效提升模型的学习性能和泛化能力。然而,在实际应用场景中,多视图数据常常伴随着视图间差异与视图内干扰的问题,具体表现为视图间各向异性分辨率导致的分布差异以及各视图内在的背景噪声干扰,增加了对多视图表征进行有效学习的难度。此外,许多研究采用标签驱动的学习策略,导致亚型分类表现高度依赖大量肺癌患者的组织亚型标签作为监督信息。然而,作为监督信息的组织亚型标签匮乏,上述方法难以准确获取鲁棒的肿瘤特征,严重影响模型的性能。为此,本研究提出了一种基于多视图深度学习的非小细胞肺癌亚型分类方法,通过寻找各视图间的共性和特性,有效挖掘鲁棒的多视图表征,进而显著增强了模型在各类数据集上的分类性能。具体研究内容如下:

(1) 为了解决视图间差异与视图内干扰问题,提出了一种新的多视图解耦方法 MVD,该方法致力于用分而治之的策略寻找视图之间的共性和特性。具体来说,MVD 采用一种基于注意力的解耦机制,同时将所有视图投射到不同的视图不变和特定于视图的子空间中,从而为每个视图生成公共和特异表示。此外,设计了一个跨视图转换损失,以成功地缓解视图不变子空间中的视图间差异。同时,引入了跨亚型鉴别损失,以确保视图特定子空间中的特异表征仅捕获与亚型无关的信息,从而通过对抗学习策略有效地消除视图内干扰。广泛的实验证实,MVD 有效地解决了视图间差异和视图内干扰的挑战,在公共和内部非小细胞肺癌数据集上的性能始终显著优于其它的方法。

(2) 为了引入额外的监督信息,提出了一种基于跨模态对比学习的多视图解耦学习方法 CMCL-MVD,通过结合影像报告,有效提升了对医学图像中亚型相关语义信息的理解能力。具体来说,CMCL-MVD 首先运用混合注意力机制构建的解耦模块来实现对各视图潜在表征的有效分解。针对多切片场景,设计了一种基于多实例池化的多切片融合模块,能够动态优化各切片的权重分配以生成更具判别力的多视图表征。此外,通过引入语义对齐损失,确保图像与文本特征之间的语义一致性,从而提升对图像中亚型相关语义信息理解的准确性。大量的实验证据表明,在综合数据集上,CMCL-MVD 相较于当前同类技术展现出了显著的性能提升优势。

关键词：非小细胞肺癌；组织亚型分类；CT 影像；多视图学习；解耦

ABSTRACT

According to data from the World Health Organization, lung cancer is the leading cause of cancer death worldwide, posing a significant challenge to public health. Non-small cell lung cancer (NSCLC) accounts for over 85% of lung cancer cases, and there are substantial differences in treatment strategies and prognostic recurrence rates among NSCLC patients with different histological subtypes. Therefore, accurate and non-invasive classification of histological subtypes from computed tomography (CT) images is pivotal for devising appropriate treatment strategies for patients. Existing research suggests that integrating data from multiple views can achieve a comprehensive understanding and precise modeling of complex data, thus effectively improving the learning performance and generalization ability of models. However, in practical applications, multi-view data often come with challenges such as inter-view discrepancy and intra-view interference. Specifically, these challenges manifest as distribution differences caused by the anisotropic resolution between views, as well as background noise interference inherent within each view, significantly complicating the effective learning of multi-view representations. Additionally, many studies adopt a label-driven learning strategy, resulting in a high dependence on a large number of lung cancer patient tissue subtype labels as supervised information for subtype classification. However, due to the scarcity of tissue subtype labels as supervised information, the above-mentioned methods struggle to accurately capture robust tumor features, significantly impacting the performance of the models. To this end, this paper proposes multi-view deep learning-based methods for histologic subtype classification of NSCLC. By seeking commonality and peculiarity among different views, these methods can mine robust multi-view representations, result in significantly enhancing the classification performance on various datasets. The specific research content is outlined as follows:

(1) To tackle the issues of inter-view discrepancy and intra-view interference, this paper presents a novel multi-view decoupling (MVD) method dedicated to seeking commonality and peculiarity across views with a divide and conquer strategy. Specifically, MVD employs an attention-based decoupling mechanism that simultaneously projects all views into distinct view-invariant and view-specific subspaces, therefore generating both common and peculiar representation for each view. Moreover, a cross-view transformation loss is designed to successfully mitigate inter-

view discrepancy in view-invariant subspace. Meanwhile, a cross-subtype discrimination loss is introduced to ensure that peculiar representations in view-specific subspaces exclusively capture subtype-irrelevant information, thereby effectively eliminating intra-view interference through adversarial learning. Extensive experiments confirm that MVD effectively addresses the challenges of inter-view discrepancy and intra-view interference, consistently outperforming state-of-the-art approaches by a significant margin on public and in-house NSCLC datasets respectively.

(2) To introduce additional supervised information, this paper proposes a multi-view decoupling method based on cross-modal contrastive learning (CMCL-MVD), which effectively enhances the understanding of subtype-related semantic information in medical images by integrating radiology reports. Specifically, CMCL-MVD first utilizes a decoupling module constructed by a hybrid attention mechanism to effectively decouple the latent representations of each view. For multi-slice scenarios, a multi-slice fusion module based on multi-instance pooling is designed to dynamically optimize the weight distribution of each slice, which empowers a robust multi-view representation with strong discrimination power. Additionally, introducing a semantic alignment loss to guarantee the semantic consistency between image and text features, thereby improving the accuracy of understanding subtype-related semantic information in images. Extensive experimental evidence demonstrates that CMCL-MVD exhibits significant performance improvement advantages compared to the current state-of-the-art techniques in the same category on comprehensive datasets.

Key Words: Non-small cell lung cancer; Histologic subtype classification; CT images; Multi-view learning; Decoupling

目 录

第 1 章 绪论.....	1
1.1 研究背景及意义.....	1
1.2 研究现状及进展.....	3
1.2.1 基于影像组学的肺癌亚型分类研究.....	3
1.2.2 基于深度学习的肺癌亚型分类研究.....	6
1.3 研究内容与组织安排.....	10
第 2 章 研究基础	13
2.1 非小细胞肺癌 CT 影像数据	13
2.1.1 CT 影像数据集介绍.....	13
2.1.2 CT 影像的预处理.....	15
2.2 多视图学习.....	16
2.2.1 基本概念.....	16
2.2.2 多视图深度学习.....	18
2.2.3 基于多视图深度学习的解耦方法.....	20
2.3 本章总结.....	22
第 3 章 基于多视图解耦学习的肺癌亚型分类方法	23
3.1 MVD 分类方法	23
3.1.1 基于注意力的多视图解耦网络.....	23
3.1.2 跨视图转换损失.....	26
3.1.3 跨亚型鉴别损失.....	28
3.1.4 亚型分类预测.....	29
3.2 模型评价与训练.....	30
3.2.1 性能测试方法.....	30
3.2.2 实验环境及超参数设置.....	31
3.2.3 模型评价指标.....	32
3.3 性能评估与比较.....	32
3.3.1 解耦模块的有效性分析.....	32
3.3.2 损失函数的有效性分析.....	33
3.3.3 预测分数的比较分析.....	35
3.3.4 多视图与单视图的性能比较.....	36
3.3.5 与现有方法的性能比较.....	37
3.3.6 模型复杂度比较.....	40
3.3.7 可视化分析.....	41
3.4 本章总结.....	42

第 4 章 基于跨模态对比学习的多视图肺癌亚型分类方法	43
4.1 跨模态对比学习及其预训练模型	43
4.1.1 跨模态对比学习简介	43
4.1.2 基于跨模态对比学习的预训练模型	45
4.2 CMCL-MVD 方法	46
4.2.1 图像和文本编码器	47
4.2.2 基于注意力的解耦模块	49
4.2.3 多切片融合模块	49
4.2.4 语义对齐损失	51
4.2.5 亚型分类预测	52
4.3 数据处理与模型训练	53
4.4 性能评估与比较	55
4.4.1 模块的消融实验	55
4.4.2 损失的有效性分析	56
4.4.3 预测分数的比较分析	57
4.4.4 学习率调节实验	58
4.4.5 与现有方法的性能比较	59
4.4.6 可视化分析	60
4.5 本章总结	61
第 5 章 总结与展望	63
5.1 研究总结	63
5.2 研究展望	64
参考文献	65

图索引

图 1.1 鳞、腺癌患者的 CT 图像示例图	3
图 1.2 Song 等人网络框架示意图 ^[24]	4
图 1.3 Lin 等人网络框架示意图 ^[16]	5
图 1.4 Xiao 等人网络框架示意图 ^[14]	6
图 1.5 Chen 等人网络框架示意图 ^[8]	8
图 1.6 El-Regaily 等人网络框架示意图 ^[44]	9
图 1.7 Zhai 等人网络框架示意图 ^[48]	10
图 2.1 NSCLC-Radiomics 数据集网页	14
图 2.2 数据筛选流程图	15
图 2.3 CT 三视图切片示意图	16
图 2.4 多视图数据和基于 CCA 的子空间学习示意图 ^[65]	18
图 2.5 两种多视图深度学习框架示意图 ^[58]	19
图 2.6 单视图单网络机制框架示意图 ^[66]	20
图 2.7 Yue 等人网络框架示意图 ^[73]	22
图 3.1 MVD 方法示意图	24
图 3.2 AD 模块示意图	25
图 3.3 五折交叉验证流程图	30
图 3.4 α, β 不同组合时的性能	34
图 3.5 MVD 在公共数据集上的收敛实验	35
图 3.6 亚型分类器和跨亚型鉴别器的预测得分	36
图 3.7 多个和单个视图的 MVD 性能	37

图 3.8	采用五折交叉验证的公共数据集上各方法的 ROC 曲线	39
图 3.9	采用独立测试的自建数据集上各方法的 ROC 曲线	40
图 3.10	公共数据集上各方法的 Grad-CAM 可视化	41
图 4.1	Yan 等人网络框架示意图 ^[94]	44
图 4.2	Zhang 等人网络框架示意图 ^[103]	45
图 4.3	CMCL-MVD 方法示意图	47
图 4.4	同一 CT 影像的多个切片图	53
图 4.5	影像与对应的影像报告示例图	54
图 4.6	α, β 不同组合时的性能	57
图 4.7	亚型分类器的预测得分	58
图 4.8	不同学习率下的 CMCL-MVD 性能	59
图 4.9	综合数据集上各方法的 ROC 曲线	60
图 4.10	不同方法学习的图像—文本表征的 t-SNE 可视化	61

表索引

表 3.1	亚型分类器的具体信息.....	29
表 3.2	MVD 的超参数设置	31
表 3.3	AD 模块不同组件的对比结果.....	33
表 3.4	不同损失的对比结果.....	33
表 3.5	公共数据集上各方法的对比结果.....	38
表 3.6	自建数据集上各方法的对比结果.....	39
表 3.7	MVD 与 3D 方法的模型复杂度比较	41
表 4.1	亚型分类器的具体信息.....	52
表 4.2	CMCL-MVD 的超参数设置.....	55
表 4.3	不同模块的对比结果.....	55
表 4.4	不同损失的对比结果.....	56
表 4.5	综合数据集上各方法的对比结果.....	59

缩略语列表

英文缩写	英文全称	中文全称
CT	Computed Tomography	计算机断层扫描
NSCLC	Non-small cell lung cancer	非小细胞肺癌
ADC	adenocarcinoma	腺癌
SCC	squamous carcinoma	鳞癌
SVM	Support Vector Machine	支持向量机
RF	Random Forest	随机森林
ROC	Receiver Operating Characteristic	受试者操作特征
AUC	Area Under Curve	受试者操作特征曲线下的面积
CNN	Convolutional Neural Networks	卷积神经网络
GPU	graphics processing unit	高性能图形处理单元
CCA	Canonical Correlation Analysis	典型相关分析
VAE	Variational auto-encoders	变分自编码器
Grad-CAM	Gradient Weighted Class Activation Mapping	梯度加权类激活映射
CLIP	Contrastive Language-Image Pre-training	对比语言-图像预训练
MIL	Multiple Instance Learning	多实例学习
t-SNE	t-Distributed Stochastic Neighbor Embedding	t 分布随机近邻嵌入

第1章 绪论

根据世界卫生组织的数据显示,肺癌是全球最主要的癌症死亡原因,对公众的生命健康构成严峻挑战,并产生了沉重的社会与经济压力^[1]。当前组织病理学手段,包括活检或术后切片分析,是鉴别非小细胞肺癌的主要诊断依据,但存在创伤性和成本高的问题。在常规临床诊断路径中,计算机断层扫描(Computed Tomography, CT)技术凭借其无创性和高检测敏感度的特性,广泛应用于肺癌的诊断过程,但在鳞癌和腺癌的区分上仍存在挑战^[2]。人工智能技术在非小细胞肺癌亚型分类中展现出巨大应用价值,通过自动提取肿瘤 CT 影像的关键特征,进而精确区分非小细胞肺癌的不同组织学亚型,对指导医生实施个体化治疗方案、提升患者生存率和预后质量具有极其重大的临床价值。本章节首先将系统阐述非小细胞肺癌亚型分类研究的背景意义及其现状进展,最后将简明扼要地介绍本研究的研究内容以及各章节的具体结构布局。

1.1 研究背景及意义

根据国际癌症研究机构(International Agency for Research on Cancer, IARC)于2020年发布的全球癌症负担数据显示,全球新发肺癌病例约为220万例,肺癌导致的死亡病例为180万例,占据癌症死亡总例数的18%,成为癌症死亡的最主要原因^[1]。此外,我国肺癌的严峻状况同样引起了深度关注。自1988年至今,我国男性和女性的肺癌发病率和死亡率持续升高,且均高于世界水平。目前整体5年生存率仅为10%~15%,疾病负担沉重^[3,4]。据统计,在我国男性群体中,肺癌发病率和死亡率均居于首位,而在女性群体中,肺癌发病率居第二,且女性不吸烟罹患肺癌的发病率呈上升趋势,这一点构成了新的公共卫生难题^[5]。根据肺癌细胞的临床表现和组织学特征,肺癌可分为小细胞肺癌和非小细胞肺癌,其中非小细胞肺癌约占所有诊断肺癌的85%^[6,7]。2021年9月,国家卫生健康委办公厅印发《肿瘤诊疗质量提升行动计划》提出提升肿瘤诊疗能力,优化肿瘤诊疗模式。作为当前肺癌精准医学的研究热点,全球医疗与科研界均正积极投入到非小细胞肺癌的精准诊断研究中,致力于开发更敏感、特异的筛查方法与诊疗手段,这对于改善肺癌患者预后和降低死亡率具有至关重要的意义。

根据世界卫生组织(World Health Organization, WHO)肺癌组织学分型标准,非小细胞肺癌主要分为两大组织学亚型:腺癌(ADC,约占50%)和鳞癌(SCC,约占40%),该两种亚型的患者在治疗方法和预后复发率上存在较大差异^[8,9]。

在精准医疗时代,有效识别患者肿瘤的组织学和遗传特征,是优化肿瘤诊疗模式,提高科学决策水平的必然途径。例如,在化疗中,腺癌患者对培美曲塞二钠类药物更为敏感,且多数腺癌患者具有晚期表皮生长因子受体(Epithelial Growth Factor Receptor, EGFR)和间变性淋巴瘤激酶(Anaplastic Lymphoma Kinase, ALK)基因突变,可采用厄洛替尼、吉非替尼和阿法替尼等进行靶向治疗^[10]。相比之下,鳞癌患者常表现有出血症状,因此不能使用抗血管生成药物,且鳞癌患者的基因突变较少,难以进行靶向治疗,通常选择铂双重化疗和免疫检查点抑制剂等治疗方法^[11]。综上所述,两种非小细胞肺癌亚型患者应该采用不同的治疗策略,以期实现更优的临床疗效。因此,准确区分腺癌和鳞癌是指导患者治疗的基础,其重要性不容小觑。

目前,确定非小细胞肺癌亚型的金标准主要依赖于病理学家对支气管肺活检或手术切除的肿瘤组织切片进行病理分析。然而,这种方法存在一些固有的局限性和挑战。首先,该侵入性诊断手段可能导致患者承受显著疼痛,尤其对于晚期、无法手术的患者群体,其适用性大打折扣^[12,13]。其次,由于需在术后制作并深入探究病理组织切片,涉及复杂的分子检测技术流程,导致整个诊断过程不仅耗资巨大且耗时较长,有可能延误患者接受治疗的最佳时机窗口^[14]。再者,鉴于肺癌所具有的空间异质性特点,单一病理切片仅能反映肿瘤微环境中的一部分信息,难以完整揭示肿瘤的整体生物学特性,从而降低了诊断准确性。最后,不同病理学家在解读病理结果时可能存在主观判断偏差,增加了误诊的风险^[15]。因此,目前急切寻求一种能在术前实施的无创、低成本、快速精确的非小细胞肺癌亚型分类策略,以弥补现有病理诊断方法中存在的不足,提高临床诊疗效能。

近年来,CT作为一种非侵入性的影像学技术,在非小细胞肺癌的临床实践中彰显出重大价值,尤其在患者的精准诊断、分型评估、生存预测以及治疗效果监测等方面,起到了至关重要的作用。研究发现,肺癌CT影像中蕴含丰富的肿瘤病理学信息,能够弥补传统病理检测手段的局限性,并展现出广阔的应用潜力。具体来说,通过肺部CT图像可以精确解析肿瘤的位置、尺寸、形态学特征、坏死、钙化等细节,而不同非小细胞肺癌亚型在CT影像上有其特有的表型表达^[16]。例如,鳞癌与不规则边缘毛刺和空洞形成具有较高相关性,而腺癌则以磨玻璃影和空气支气管征为主要的CT成像特点,两种亚型的CT图像如图1.1所示^[17]。尽管如此,由于部分表型特征在CT图像中可能因模糊不清或与周围组织结构重叠而难以分辨,加之病理病灶区域微小且易于与背景组织混淆,使得放射科医生仅依赖视觉判断对非小细胞肺癌亚型进行区分时面临极大挑战,高度依赖于他们的专业知识、经验和精细解读能力。

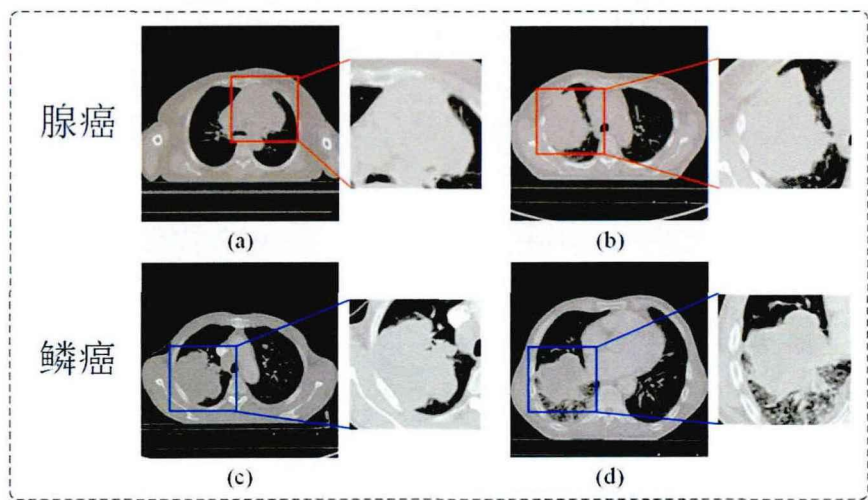


图 1.1 鳞、腺癌患者的 CT 图像示例图

在此背景下，人工智能技术展现出了巨大的应用价值，引领了非小细胞肺癌亚型分类迈入智能化的新纪元^[18-20]。目前，诸多科研团队已成功运用以深度学习为代表的新技术对非小细胞肺癌患者的 CT 影像进行深入研究和精准分析，显著提高了诊断效率和分类准确性，其结果不仅媲美，甚至超越了经验丰富的病理学家^[21]。这些人工智能系统在分析大量肺部 CT 图像时，能够捕捉到细微的特征和模式，从而更精准地判断肿瘤的类型和亚型。与传统的病理学手段相比，人工智能技术在处理影像数据时更为高效，且不受病理学家个体经验的影响，具备更广泛的适用性。人工智能技术的不断发展无疑为医学科研和临床实践带来了革命性的改变，有力地推动了肺癌精准医疗的发展进程，为全球肺癌防治工作注入强大动力，助力提升肺癌患者的生存率与生活质量，开启了智慧医疗的新篇章。

1.2 研究现状及进展

1.2.1 基于影像组学的肺癌亚型分类研究

近年来，基于 CT 影像的放射组学技术在无创性诊断领域已经引起了人们的关注^[22-24]。该技术核心在于运用定量解析手段，从医学图像中提取诸如像素强度、形态及纹理特征等富含诊断价值的高通量、多维度信息，实现对病变部位的无创、可重复性的精确表征。研究者通过对这些独特影像特征进行统计学习与分析，构建模型以诊断癌症的组织病理学信息，为精准医疗策略提供有力的决策支持工具。例如，Chen 等人^[23]从专业人员手动分割的 75 个结节的 CT 影像中提取了 750 个放射学特征，并运用非参数统计方法中的 Wilcoxon 秩和检验构建了一个由 4 个

具有显著区分效能的特征组成的标识符。基于此标识符，通过训练支持向量机（SVM）分类器，在良恶性分类任务上实现了 84% 的诊断准确率。然而，由于实验设计中未包含独立验证集以防止过拟合现象的发生，此结果的有效性和泛化性能尚存局限性。因此，Song 等人^[24]深刻认识到独立验证集在模型性能评估中的核心地位，并在其研究中特别强调了这一点。他们的研究方法包括从肿瘤感兴趣体积（Volume of Interest, VOI）中系统地提取 1409 个包含形状、强度和纹理等信息的特征，然后采用 L2 与 L1 范数进行特征选择，并探讨了所选特征在模型构建中的关键作用及重要性等级，流程如图 1.2 所示。为了对这些模型进行全面且客观的性能评估并考察其跨数据集的泛化能力，Song 等采用三个独立测试集，并计算各模型对应的平均接收者操作特性曲线下的面积（AUC）作为评价指标，以此确保了研究成果的稳定性和可靠性。

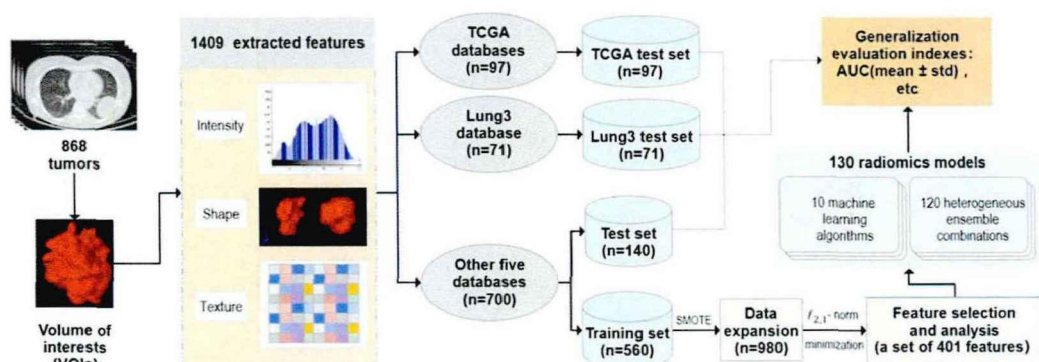
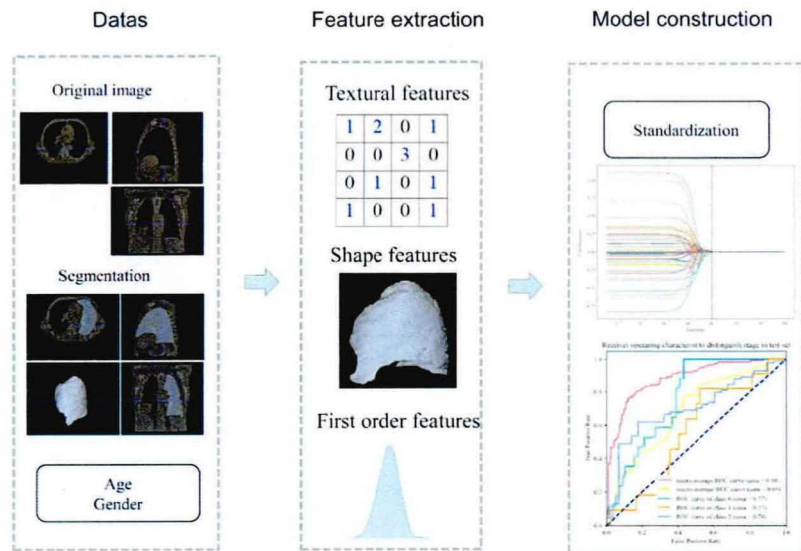


图 1.2 Song 等人网络框架示意图^[24]

研究表明，从影像检测图像中提炼出的放射组学特征，在非小细胞肺癌患者的组织分型诊断中具有显著效能，有望弥补传统分子生物学检测方法存在的局限性和不足。因此，科研人员不断探索并提取多元化的定量图像特征，在各类数据集上构建了众多精准且高效的放射组学分类模型，进一步推动了这一领域的快速发展。Khodabakhshi 等人^[25]从恶性病变的 CT 图像中提取了 1433 个放射组学特征，通过采用包装器算法以及多元自适应回归样条方法来识别有效特征。再基于所选特征，对 1000 个自举样本进行多变量多项逻辑回归分析，对非小细胞肺癌亚型进行分类。结果显示，纹理特征，特别是灰度大小区域矩阵特征（GLSZM），是非小细胞肺癌亚型的显著指标。

此外，Lin 等人^[16]也基于 CT 图像提取的放射学特征建立了非小细胞肺癌患者的组织分型和临床分期模型，流程如图 1.3 所示。该研究对多种分类算法在患者组织病理亚型划分及临床分期辨识上的效能进行了详尽比较。结果显示，相较于其他对比方法，随机森林（RF）模型在对非小细胞肺癌患者进行组织学分型和临床分期分类时表现出更为优越的性能。进一步探究发现，在随机森林模型所采

图 1.3 Lin 等人网络框架示意图^[16]

用的特征集里，一阶灰度统计特性以及纹理特征在提升模型预测精确性方面起到了关键作用，这些特征对于揭示潜在的影像模式并精准区分不同病程阶段具有显著价值。

上述成果显示，精确且高效地界定感兴趣体积是确保所得特征计算可靠性的核心环节。然而，对海量影像数据进行精准勾画是一项颇具挑战的任务。因此，D'Arnese 等人^[26]运用先进的自动多阶段图像分割技术，对肿瘤病灶区域进行精确量化分割，旨在减少人为因素引入的误差，提高数据预处理阶段的工作效率和准确性。其后，基于所得分割结果，挖掘有效的放射组学特征并将其输入到层级分类器中，以实现非小细胞肺癌不同组织学亚型的自动化鉴别。此方法在提升分析效能和诊断精确度方面展现出了显著优势，为临床病理分型提供了有力工具及理论支持。

然而，该领域仍面临一些亟待解决的关键问题和局限性。首要挑战在于缺乏标准化的方法来计算和定义作为模型输入的大规模放射组学特征，这直接影响了研究的稳健性和一致性^[24]。此外，在这些机器学习方法的实施过程中，仍高度依赖人工设计复杂的放射学特征，这一过程不仅耗时费力，而且可能导致特征提取和选择的不准确性^[27]。与此同时，由于当前样本容量有限以及普遍存在的单中心数据集偏倚，现有研究尚无法充分证实放射组学方法具备良好的泛化能力，即难以将训练出的模型有效推广至其他临床研究中心的应用场景。因此，亟需研发一种基于无创检测手段、能够智能提取深层组织学特征的非小细胞肺癌组织学分型模型，以有效突破现有瓶颈，并显著提升该模型在不同研究中心间的泛化预测效能及准确性。

1.2.2 基于深度学习的肺癌亚型分类研究

针对影像组学方法中依赖人工设计放射学特征的局限性问题,深度学习端到端的解决方案正逐步崭露头角,并展现出巨大潜力。深度神经网络具有强大的学习能力,能够自动生成高维非线性特征,从而极大地减轻了人力成本,并能深入挖掘传统方法难以捕捉的复杂内在关联^[28]。特别是在医学图像处理领域,卷积神经网络(Convolutional Neural Networks, CNN)作为一种广泛应用的深度学习技术,以其强大的自动化特征提取与精准分类能力,在肿瘤区域精确分割^[29,30]、生存率预测分析^[15]、基因表达推断^[31,32]等临床任务中均有出色表现,尤其在非小细胞肺癌亚型诊断方面成果斐然。

以 Han 等人^[33]的研究为例,他们对支持向量机、线性判别分析以及 VGG16 深度学习模型在肺癌亚型分类任务中的性能进行了对比实证。实验结果显示, VGG16 模型在区分腺癌和鳞癌这两种主要亚型的二元分类任务上,成绩远超同期的传统机器学习算法,有力彰显了深度学习技术在医学图像解析领域的颠覆性和优越性。除了 VGG16 这一基础模型之外,更多前沿且创新的研究正在不断涌现,持续推动着深度学习在医学影像分析领域的深化应用与发展。例如, Xiao 等人^[14]在基于 CT 影像轴向视图的前沿探索中,为非小细胞肺癌组织学亚型分类任务注入了新的动力。该团队创新性地提出了一种多特征多关注分类网络(MFMANet),其整体框架如图 1.4 所示。该网络核心包括两个关键模块:多尺度空间通道注意模块(MSAM)以及多特征融合全局局部关注模块(MFGLA)。实验表明,所提出的 MFMANet 网络架构在鳞腺癌诊断任务上展现出卓越的精准度和泛化能力。这些研究不仅各自展示了深度学习在非小细胞肺癌分类上的独特优势,为精准医疗决策提供了有力支持,同时也为医学影像分析领域开辟了多元化、深层次的研究视角。

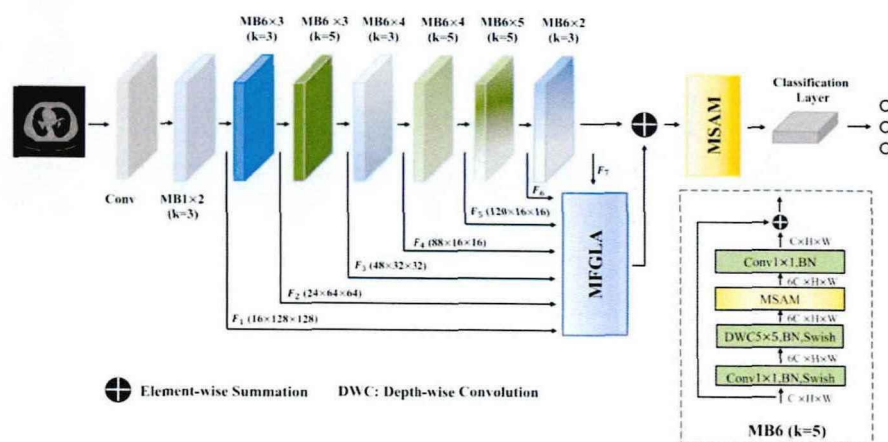


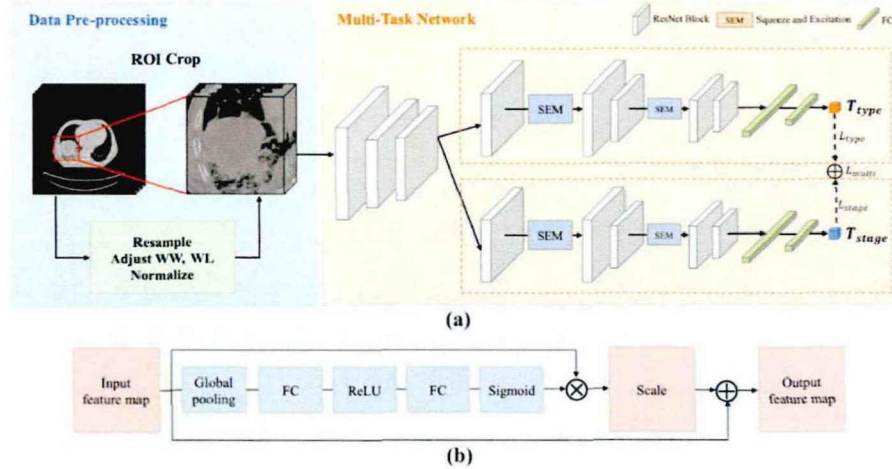
图 1.4 Xiao 等人网络框架示意图^[14]

值得注意的是, 尽管上述方法在非小细胞肺癌组织学亚型区分上取得了一定进展, 但其主要依赖于 CT 图像的轴向视图进行分析。此单一维度的图像虽然能够相对准确地描绘出肿瘤的形态、大小以及与周围组织的关系等基本信息, 但无法充分展示肿瘤组织的空间分布特性、生长模式及其三维结构等, 在全面展现肿瘤的空间结构特征与内在复杂性上仍存在局限性^[34]。例如, 肿瘤的浸润范围、边界模糊程度、内部密度不均匀性等立体特征, 在轴向视图中往往难以完整体现, 这一局限性无疑限制了对非小细胞肺癌病灶进行全面评估和精确分类的能力, 从而影响临床治疗策略的选择和预后评估的准确性。因此, 为了实现对非小细胞肺癌更深入细致的研究和个体化诊疗, 有必要结合先进的图像分析算法和技术手段, 充分挖掘并利用肿瘤的空间结构信息, 为临床实践提供更为全面且精准的决策支持依据。

在此背景下, 3D 深度神经网络显示出了良好的应用前景, 其能够立体解析 CT 影像数据, 对大量三维医学图像进行深度特征提取^[30,35-37]。这一技术有力地突破了传统轴向视图分析的局限性, 实现了对非小细胞肺癌组织空间分布、生长特性以及微观结构等多元复杂信息的全面捕捉与精准解读。例如, Guo 等人^[36]设计了一个基于胸部 CT 的 3D 深度学习模型, 通过保留和凸显肿瘤 CT 图像的空间特征, 实现了对数据内在三维结构信息的有效捕捉与高效利用, 从而精确识别并区分不同的肺癌组织学亚型。进一步, Yanagawa 等人^[37]构建了一个由七个卷积-池化层和两个最大池化层以及全连接层构成的 3D-CNN 模型, 并融入了批归一化、残差连接和全局平均池化技术。实验结果显示, 此 3D-CNN 模型能够显著增强医生对肺侵袭性腺癌的诊断准确性, 在保持诊断性能不变的前提下提升诊断精确度。

此外, 一些开创性的研究已采用更为精密的特征提取网络架构来应对医学影像分析中的复杂难题, 并为非小细胞肺癌的诊断与分类领域带来了新的突破。Chen 等人^[8]创新性地构建了一种基于 3D CT 图像的多任务学习框架, 其网络框架如图 1.5 所示。该模型整合了组织学亚型分类分支与肿瘤分期预测分支, 两者共享底层特征提取网络, 并进行联合训练。通过协同优化两个核心任务, 模型能够在无需依赖医生对肿瘤区域精准标注的前提下, 有效提升非小细胞肺癌组织学亚型分类的准确性。研究采用内部验证结合外部测试的严格评估策略, 以确保模型的稳定性和泛化能力得到充分验证。

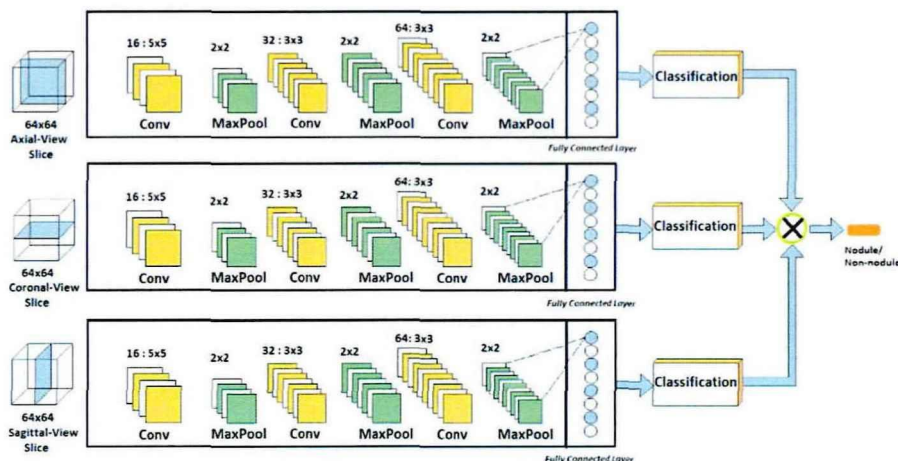
尽管 3D 深度学习网络能够有效地弥补单视图方法在表征立体信息方面的不足, 其不可避免地加剧了计算负担。在训练过程中, 此类网络对硬件资源的需求显著增加, 特别是对高性能图形处理单元 (GPU) 和大规模内存系统的占用量显著提升^[38,39]。这种额外的计算复杂度引入会导致网络陷入过拟合困境, 并限制其

图 1.5 Chen 等人网络框架示意图^[8]

分类性能的优化潜力，表现为次优的分类效果。对于这些问题，多视图学习策略被公认为一种颇具前景的解决方案。该方法通过整合多个视角的数据表征，在保持较低内存占用率的同时，有效缓解了过拟合现象，从而实现在有限资源条件下的高效且精确的分类任务目标，展现出卓越的模式识别能力。目前，多视图学习已广泛渗透并应用于众多科研领域中，特别是三维医学影像分析领域^[40-42]。该技术通过整合多元异构的数据视图，实现对复杂医学影像的深度挖掘和综合理解，从而在疾病诊断、治疗规划以及医学研究等方面发挥出显著优势，推动了三维医学影像分析领域的前沿发展与技术创新。

在一项针对卵巢癌残留疾病预测的前沿研究中，Gao 等人^[43]创新性地构建了一种多视图注意力学习模型，基于三维 CT 图像的轴位、冠状位和矢状位视图进行疾病检测。通过引入先进的注意力机制，该方法能够在各个视图层面选取更具诊断价值的切片，从而有效地提升 3D CT 图像的整体表征能力和疾病预测精确度。大量严谨的实验以及多元化的评估指标均有力验证了所提出方法在其领域的优越性能及潜在应用价值。此外，El-Regaily 等人^[44]同样采用了一种多视图卷积神经网络架构，提取轴向、冠状及矢状三种视角的图像信息以实现分类任务，如图 1.6 所示。研究人员利用 LIDC 数据集中包含的 650 例样本进行训练与测试，在算法性能和计算效率之间实现了良好的权衡，显著减少了计算时间。实验结果展示了所提出的基于二维多视图的卷积神经网络对孤立性结节、血管相关结节以及附着于肺壁的结节的强大检测能力，有效缓解了假阳性问题，展现出在肺部影像分析领域的广阔应用潜力。

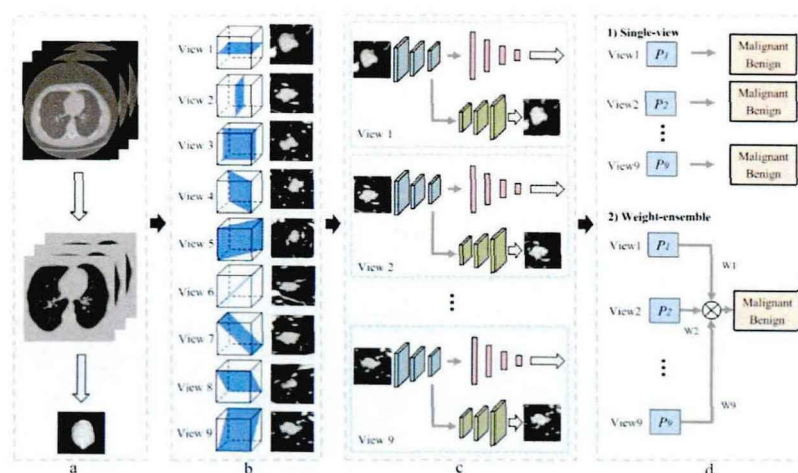
在基于 CT 影像的肺结节诊断研究领域中，除了常规地从三个基本解剖方位（冠状、矢状和轴向）提取图像特征外，部分研究进一步扩展了视图的获取范围，采用了九视图的分析策略^[45-47]。例如，Zhai 等人^[48]提出了一种基于多视图学习

图 1.6 El-Regaily 等人网络框架示意图^[44]

的二维多任务卷积神经网络（MT-CNN）框架来解决胸部 CT 扫描中良性与恶性肺结节识别过程中存在的高假阳性问题，整体网络框架如图 1.7 所示。MT-CNN 模型在数据集 LUNA-16 和 LIDC-IDRI 上进行了测试，并与当前最先进的模型进行了详尽对比，有力证明了所提方法的有效性和优越性。以上多视图学习方法通过整合多个视图的表示，不仅可以探索肿瘤区域丰富的空间属性，还可以显著降低过度拟合的风险，提高模型在测试阶段的泛化能力。综上所述，多视图学习策略能够有效提升肺部 CT 影像诊断系统的智能化水平和临床实用价值，为实现精准医疗和个性化诊疗方案的制定奠定了坚实的技术基础。

尽管以上基于肺部 CT 多视图的方法已取得了显著的研究成果，然而当前直接整合多个视图潜在特征的方法并非最优方案。究其原因，首先是简单融合忽视了各视图间内在的差异性。事实上，由于生成机制各异，不同视图的空间分辨率呈各向异性，导致各视图间的分布存在显著差异^[49]。该差异在有效整合与深度挖掘多个视图所蕴含信息时构成挑战，使得模型无法充分发掘和利用多源信息所带来的互补优势和增益效果，从而限制了模型性能和泛化能力的提升。其次，以往的研究大多致力于将来自各个视图的信息整合为统一的全局特征描述符，但这一策略无法完全避免视图内无关信息带来的干扰问题。具体而言，每个视图都可能存在一些与非小细胞肺癌亚型分类无关的背景或噪声信息，如背景中高密度的骨组织以及突出的血管结构，对学习有效的共同表征造成负面影响，阻碍了模型精准捕捉肿瘤的复杂特性，最终降低了模型对不同亚型的区分能力。

在此背景下，Gao 等人^[50]提出了一种基于多视图特征分解的深度学习框架。该方法利用多视图特征分解技术，对源自多视角的特征进行深度挖掘与处理，旨在滤除背景噪声对最终肿瘤识别特征的影响。尽管该研究在解决 CT 图像中肿瘤

图 1.7 Zhai 等人网络框架示意图^[48]

与背景混杂难题上取得了进展，但在实际应用中，其方法仍面临一些局限性。具体而言，尽管使用多视图策略有助于揭示肿瘤的空间特征，然而，各视图间的差异性并未得到充分考虑，这些差异在多视图特征分解过程中引入偏差，从而影响最终肿瘤特征提取的准确性。此外，视图间差异问题与视图内干扰问题存在着内在关联。视图间差异导致的特征分布不一致性，可能进一步加剧各视图内部信息的混淆，使得有效分离肿瘤信号与无关干扰变得更为困难，对后续的诊断与分析工作构成挑战。因此，亟待研发一种能够在抑制视图内干扰性信息的同时减轻视图间差异性的策略，以促使多视图模型充分发掘并精准利用多视图数据中的深层次肿瘤信息，提升对非小细胞肺癌组织学亚型分类的准确性和稳健性。

1.3 研究内容与组织安排

为了有效解决视图间差异和视图内干扰问题，本研究创新性地提出了一种基于多视图解耦学习的非小细胞肺癌组织学亚型分类方法 MVD (A novel multi-view decoupling method for histologic subtype classification of non-small cell lung cancer)，以实现对肿瘤轴位、冠状位及矢状位三个关键视图的有效处理。该方法采用基于注意力的解耦机制 AD (attention-based decoupling)，将所有视图投影到视图不变和视图特定的子空间中，并为每个视图生成公共和特异表征，从而以分而治之的策略解决上述两个问题，并习得鲁棒的多视图表征来实现精准的肺癌组织学亚型分类。此外，在视图不变子空间中设计了一个跨视图转换损失 CVT (cross-view transformation loss)，以有效地探索不同视图中分布一致性的公共表征。与此同时，在视图特定子空间中引入跨亚型鉴别损失 CSD (cross-subtype discrimination loss)，来促使特异表征捕捉与亚型分类无关的信息。通过在公共和

自建非小细胞肺癌数据集上实施的一系列实验,证实了 MVD 相较于现有先进方法的优越性能。在此基础上,本研究进一步提出了一种新颖的基于跨模态对比学习的多视图非小细胞肺癌组织学亚型分类方法 CMCL-MVD (Cross-Modal Contrastive Learning),对 CT 图像和诊断报告之间的复杂关系进行建模,从而获得更具表达力的特征表示。实验结果有力证明了深入挖掘各模态特征间的语义信息以及精确建模不同模态数据间的复杂关系,对进一步提升模型分类准确度的重要价值。本研究主要研究内容如下:

(1) 为解决多个视图间具有差异性和每个视图内存在干扰的两大难题,本研究提出了一种多视图解耦方法 MVD 对非小细胞肺癌组织学亚型进行准确分类。首先,该方法通过基于注意力的解耦模块,将不同的视图投影到一个视图不变子空间以及多个视图特定子空间中,从而有效地学习公共和特异表征。进一步,针对视图间的差异问题,MVD 设计了一个特制的跨视图转换损失函数,以确保公共表征在视图不变子空间内分布达到最大程度的一致性。此外,引入了基于对抗学习策略的跨亚型鉴别损失,以引导特异表征捕获亚型无关信息,从而有力地消除了视图内的干扰。通过在公共和自建数据集上进行的评估和验证,所提出的 MVD 方法在区分鳞癌和腺癌方面表现出了卓越的诊断准确度,在多个指标上均优于已建立的先进分类模型。

(2) 在上述研究的基础上,为了有效利用 CT 影像信息和影像报告文本信息,本研究提出了一种新颖的基于跨模态对比学习的多视图肺癌组织学亚型分类方法 CMCL-MVD,以实现对患者亚型的更精确预测。该方法沿用了 MVD 中的注意力解耦模块,随后引入基于多实例池化的多切片融合模块,融合出更具判别力的多视图特征。此外,设计跨模态语义对齐损失确保图像—文本间表征的语义一致性,进而有效增强与亚型分类相关的图像特征的表现力。最后将优化后的多视图特征输入到分类器中,以实现对患者组织学亚型的精确预测。实验结果表明,相较于现有方法,CMCL-MVD 在非小细胞肺癌组织学亚型分类任务上展现出优越的准确性。

本研究主要涵盖了五个研究章节:第一章阐述非小细胞肺癌亚型分类研究的理论背景及其实质性的临床意义,继而系统梳理了国内外在该领域的研究进展现状,并深度剖析现有研究成果的优势与不足之处。最后明确提出本研究的核心研究内容和各章节的具体组织结构。第二章专注与深度学习在亚型分类方法中的应用基础。首先,概述常用 CT 影像数据集的基本特性,随后阐明本研究所采用的 CT 图像数据采集策略以及预处理技术。在此基础上,进一步深入介绍多视图深度学习的关键概念和技术框架。第三章重点介绍名为 MVD 的基于多视图解耦学习的非小细胞肺癌亚型分类方法,并通过一系列严谨的实验设置及评价指标体系

对该方法进行性能评估验证。第四章细致阐述并实证分析基于跨模态对比学习的非小细胞肺癌组织学亚型分类新方法 CMCL-MVD。通过对现有先进方法的比较以及实验结果的有力展示，证明 CMCL-MVD 在提升分类准确性上的重要突破。第五章对全文的研究成果进行全面总结，并在此基础上前瞻性地展望未来可能的研究方向和潜在的工作领域，为非小细胞肺癌亚型分类研究的深化拓展提供新的思路和视角。

第2章 研究基础

本章节将系统阐述与本研究研究相关的核心基础与理论框架。首先介绍在基于 CT 影像的非小细胞肺癌亚型分类研究中广泛应用的癌症影像档案库 (the cancer imaging archive, TCIA) 及其所包含的关键数据集, 随后对 CT 图像的筛选流程及预处理步骤进行深度解读。进一步地, 聚焦多视图深度学习这一前沿概念及其在医学影像分析中的实际应用情境, 深入探讨其原理、架构以及在挖掘和整合多元异构医疗影像数据方面的独特优势。最后, 着力介绍基于多视图深度学习的解耦方法, 探究如何通过此类先进技术有效提取并解析来自不同视图的深层次特征信息, 从而实现了对非小细胞肺癌亚型更为精确的分类。

2.1 非小细胞肺癌 CT 影像数据

2.1.1 CT 影像数据集介绍

本研究采用了公共与自建的非小细胞肺癌 CT 影像数据集, 以验证所构建模型的有效性。首先介绍的是公开数据集所在的癌症影像档案库 TCIA^[51], 该数据库由美国国家癌症研究所 (NCI) 旗下的癌症成像计划资助, 并由弗雷德里克国家癌症研究实验室 (FNLCR) 管理。TCIA 作为一项开源、开放获取的科研资源, 旨在强化研究的可重复性及推动数据再利用, 以支持高级医学影像在癌症研究、开发和教育活动中的应用。TCIA 核心管理的数据涵盖了多种放射学图像资料, 包括计算机断层扫描、磁共振成像、正电子发射断层扫描等医学图像, 且均以通信标准 (digital imaging and communications in medicine, DICOM) 格式保存。除了原始影像资料, TCIA 还包括图像标注与其他注释信息, 例如感兴趣区域及关联统计参数、数字化病理切片及其对应的病理诊断结果、临床诊疗详情、基因组学数据, 以及经由专业医疗人员或定量分析算法衍生出的分析成果 (如肿瘤分割、放射组学特征、二次图像处理产品、放射科医师评估报告) 等多维度信息, 从而为开放科学及癌症研究提供全面的支持。目前, TCIA 通过严谨的数据管理体系, 将多个项目和试验的数据集整合, 构建起包含 3090 万幅放射学图像、覆盖约 37568 名受试者的大规模患者队列, 有力地激励并支撑了癌症领域的研究与发展。

针对非小细胞肺癌亚型分类研究, 本研究从 TCIA 数据库精选了两个关键的 CT 影像资源, 分别是 NSCLC-Radiomics^[52]和 NSCLC-Radiogenomics^[53,54]。NSCLC-Radiomics 数据集包含 422 名在荷兰 MAASTRO 诊所接受治疗的非小细

胞肺癌患者的治疗前 CT 图像。这些病例的肿瘤总体积由放射肿瘤专家进行了手动三维勾画，并附带了详尽的临床数据以及生存状况信息。此数据集中，CT 切片层厚统一为 3mm，单幅切片图像尺寸为 512×512 像素，像素大小为 0.977×0.977mm。对于 NSCLC-Radiogenomics 数据集，囊括了斯坦福大学医学院和帕洛阿尔托退伍军人事务医疗保健系统的 211 名早期非小细胞肺癌患者的影像学数据，其网页如图 2.1 所示。此外，还系统收集了一系列关键临床数据，诸如年龄、性别、体重、种族、吸烟状态、TNM 分期以及组织病理分级等信息。

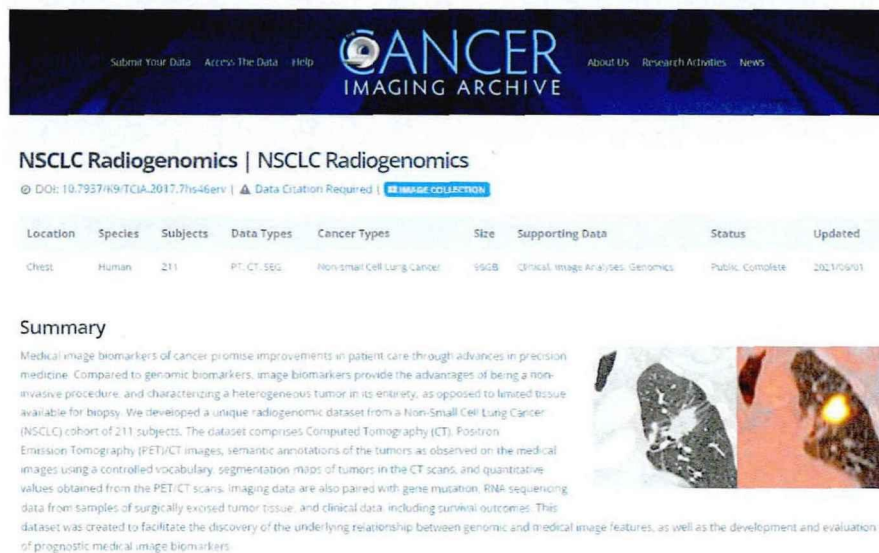


图 2.1 NSCLC-Radiomics 数据集网页

注：图片来源（<https://www.cancerimagingarchive.net/collection/nsclc-radiogenomics/>）

除了利用上述 TCIA 数据库中公开的数据集以外，本研究还从安徽医科大学第一附属医院影像科收集了 202 名肺癌病例的详细信息作为自建研究数据集。此集合包括通过西门子公司研发的 SOMATOM Definition AS 螺旋 CT 设备获取的肺部 CT 影像资料，以及相应的临床病理诊断报告等多维度医疗信息。对于每位肺癌患者的 CT 扫描，其层厚参数被精确设定在 3.0 至 5.0 毫米之间，确保了图像的质量和解析度。尤为重要的是，本数据集中所有非小细胞肺癌患者的 CT 影像数据，其肿瘤区域的标注工作均由安徽医科大学第一附属医院影像科两位资深医生依据严格的医学标准和专业判断共同完成，从而为后续的深度分析与研究提供了可靠且精准的靶向区域标注信息。

为了满足研究的严谨性要求，本研究对所获取的数据集进行了以下细致且专业的筛选流程^[16,24]，具体的数据筛选标准如图 2.2 所示。最终确定了包含完整组织病理学信息及精确肿瘤轮廓勾画的非小细胞肺癌患者的术前 CT 图像数据集。其中，来源于公共数据库的合格病例共 325 例涵盖了 146 例腺癌病例与 179 例鳞

癌病例。而自建数据库提供了 125 例详实的病例数据，其中包括 41 例腺癌患者和 84 例鳞癌患者信息，这些数据共同构成了本次研究的核心数据集。

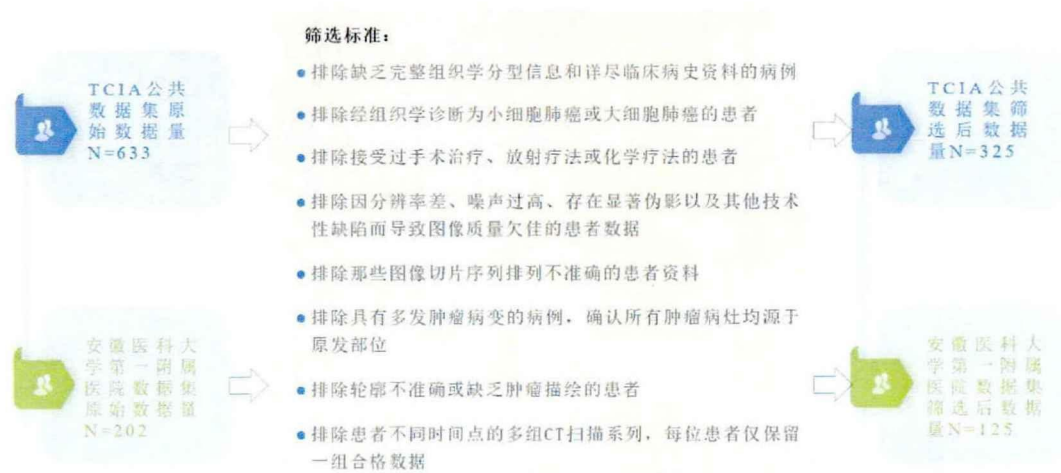


图 2.2 数据筛选流程图

2.1.2 CT 影像的预处理

鉴于多源 CT 影像数据的切片厚度与像素尺寸存在差异性，本研究采取了一系列标准化预处理步骤进行校正与优化。首先，运用三线性插值法对原始 CT 图像进行重采样处理^[47,55]，确保体素分辨率统一为 $1 \times 1 \times 1 \text{ mm}$ 。其次，为了实现对肿瘤三维空间特征的全方位获取与精准解析，本研究依据每个病例中肿瘤中心点的位置，在 CT 图像上选取了三个标准解剖平面，即轴向、冠状面和矢状面，并以肿瘤中心为参照点，裁剪出 128×128 像素大小的切片，具体展示如图 2.3 所示。同时，各个切片对应的标签信息均与相应的患者级别的病理标签保持高度一致。进一步地，考虑到 CT 成像原理，即不同器官和组织吸收 X 射线后的衰减系数各异，因此表现为不同的 CT 值（其单位为 Hounsfield Unit, HU）。密度较高的组织，由于 X 射线吸收强度大，对应的 CT 值呈现正值且在图像中表现为亮信号；反之，密度较低的组织，其 X 射线吸收较弱，HU 值则为负值，显示为暗信号，例如空气通常约为 -1000 HU，而骨骼通常大于 400 HU。因此，为了最大程度地减少 CT 影像中的噪声干扰并提高模型识别精度，参考相关研究实践^[32,56]，本研究将输入切片 CT 值的范围限定在 -1000 至 400 HU 之间，并进一步实施归一化处理，将像素强度映射到 $[0,1]$ 区间内，从而保证数据标准化的一致性和有效性，为后续深度学习模型训练提供高质量的数据支持。为了增强模型的泛化能力并有效防止过拟合问题^[57]，通过随机旋转和镜像翻转的方式对数据集进行了增强处理。

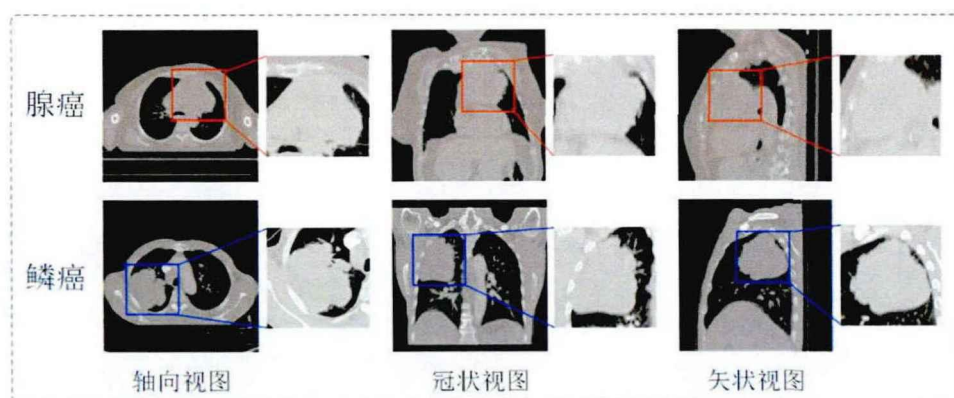


图 2.3 CT 三视图切片示意图

2.2 多视图学习

2.2.1 基本概念

在过去的数十年间，多视图数据在视频监控、社交网络以及医学影像检测等领域呈指数级增长态势，已然成为互联网核心数据类型之一^[58]。针对这一现象，多视图学习（Multi-view Learning, MVL）技术应运而生，其主要目标在于通过整合多种特征表达或异构数据源以揭示公共的特征空间或统一模式^[59]，从而提升分类、聚类等特定任务的表现性能。凭借从多个互补视角全面理解与建模复杂数据的能力，多视图学习展现出了广阔的应用潜力与研究价值。

在传统的机器学习算法领域，诸如支持向量机、判别分析、核方法及谱聚类等技术，倾向于将多源视图信息整合为单一视图以适应学习任务。然而，在训练样本数量有限的情况下，此类“视图拼接”策略容易导致过拟合现象，并且由于忽视了各视图间固有的统计特性差异，其处理方式缺乏物理意义。相较于上述单一视图学习范式，多视图学习作为一种新兴的学习模式，通过引入特定视图函数并协同优化所有相关功能，充分利用数据的多重互补视图信息，从而提升模型的学习性能和泛化能力。鉴于此，多视图学习因其独特的优越性与应用潜力，在机器学习和计算机视觉研究领域中已引起广泛关注，并取得了显著的理论进展与实践突破，进而推动了一系列极具前景的算法创新，如协同训练、多核学习以及子空间学习等^[60,61]。

协同训练（Co-training）^[59,62]是最早的多视角学习方案之一，其核心机制是在未标记数据集上运用两个迥异视图进行交替迭代训练，旨在最大程度地保证二者间的一致性。该方法主要基于分歧理论构建，设想每个数据可通过多种视角实

现分类, 并且不同视角所训练出的分类器之间能够形成互补, 从而有效提升分类精度。而多核学习 (Multiple Kernel Learning, MKL) [63] 最初为了解决如何在诸如线性核、多项式核以及高斯核等潜在核函数空间中提高搜索效率和泛化性能而发展起来, 如今已在处理多视图数据问题领域得到广泛应用。具体而言, 多核学习通常采用不同的预定义内核对各自的视图进行单独处理, 继而将这些内核以线性或非线性形式整合成一个统一的综合核函数。其中线性组合主要有两种基本策略: 直接求和核与加权求和核, 数学定义分别如公式 (2.1) 和公式 (2.2) 所示:

$$K(x_i, x_j) = \sum_{k=1}^M K_k(x_i, x_j) \quad (2.1)$$

$$K(x_i, x_j) = \sum_{k=1}^M d_k K_k(x_i, x_j) \quad (2.2)$$

由于基本核的线性组合有限, 因此可以通过其他非线性方式组合核来实现更丰富的表示。Varma 和 Babu [64] 尝试使用基核和其他正定核组合的乘积来进行多核学习, 如组合核的幂运算:

$$K(x_i, x_j) = \exp \left(- \sum_{k=1}^M d_k x_i^T A_k x_j \right) \quad (2.3)$$

或

$$K(x_i, x_j) = \left(d_0 + \sum_{k=1}^M d_k x_i^T A_k x_j \right)^n \quad (2.4)$$

此外, 基于子空间学习的理论框架认为, 各视图所包含样本的分布共同构成了一个多维样本空间, 其中每个样本可被视为该高维空间中的一个点, 且这些样本空间间存在着一个潜在的公共子空间 [65], 如图 2.4 所示。在这个公共子空间中可以更好地利用不同视图的数据, 为后续的关键任务, 诸如分类、聚类等提供强有力的计算基础。在众多子空间学习方法中, 典型相关分析 (Canonical Correlation Analysis, CCA) 占据重要地位。CCA 的核心理念在于通过优化算法最大化不同视图数据在映射至公共子空间后的相关系数, 旨在求解出将各个视图投射到公共空间的最优映射向量, 以此确保在公共子空间内, 各视图数据达到最大程度的相关性 [58]。CCA 的目标函数可以写作:

$$(w_1^*, w_2^*) = \arg \max_{w_1, w_2} \text{corr}(w_1^T X_1, w_2^T X_2) = \arg \max_{w_1, w_2} \frac{w_1^T C_{12} w_2}{\sqrt{(w_1^T C_{11} w_1)(w_2^T C_{22} w_2)}} \quad (2.5)$$

其中, C_{11} 和 C_{22} 分别为每个视图 X_1 和 X_2 的协方差, C_{12} 是 X_1 和 X_2 的交叉协方差。

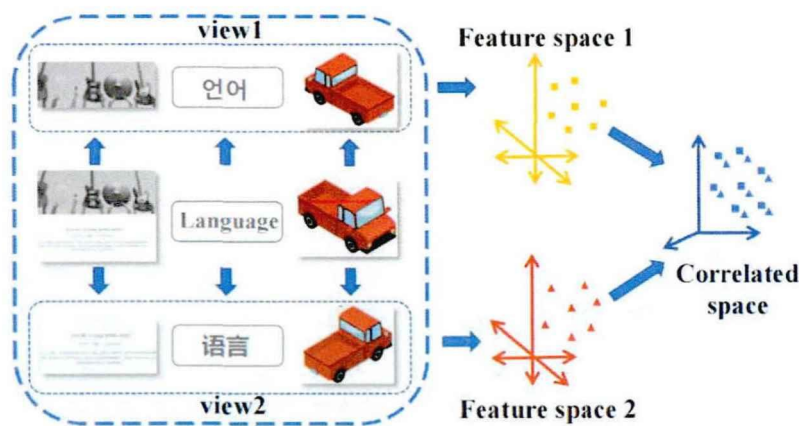


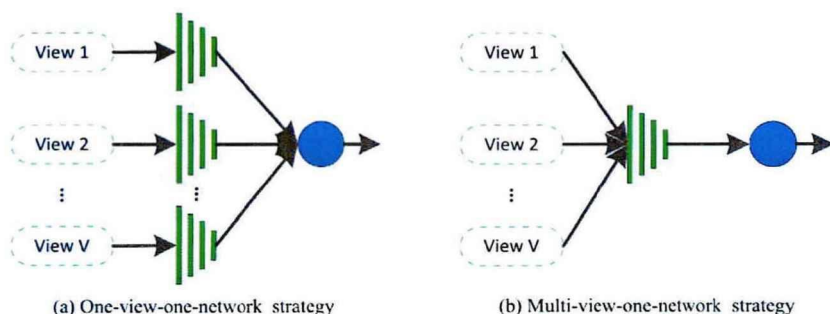
图 2.4 多视图数据和基于 CCA 的子空间学习示意图^[65]

综上所述，协同训练范式下的算法以独立并行方式处理各个视图，通过促使各模块在不同视图间达成一致性进行优化，是一种多视图后期融合机制。而多核学习算法则针对每个视图单独计算核函数，并将其有机整合至基于核的学习框架中，体现了多视图融合技术的特点，实时地将内核信息融入构建和优化过程。此外，基于子空间学习的方法假设多个输入视图来源于一个潜在共享的子空间，其核心目标是揭示并利用这一共享子空间，可视为多视图先验融合策略。尽管实现多视角融合以提升学习效能的策略存在显著差异，但其核心主要依赖于共识或互补原则来确保多视图学习的有效实施。总体来说，通过深入探究各视图间的一致性关系和互补性特征，多视图学习相较于单视图学习展现出了更高的性能、更广阔的前景以及更强的泛化性能优势。

2.2.2 多视图深度学习

近年来，深度学习技术因其强大的特征提取能力，在人脸识别、目标分类和目标检测等计算机视觉和人工智能领域实现了巨大成功。随着深度学习在众多应用场景中的广泛应用及显著成效，多视图深度学习策略亦逐步渗透并应用于医学诊断与分析领域，并已取得了一系列令人瞩目的科研突破。通过深度融合多视图学习算法和深度学习架构，可以有效地弥合不同视图间的异构性鸿沟，揭示视图间潜在的内在关系，从而获得更具判别力的共同表征。因此，在大规模图像分类任务上，构建一种能够挖掘多视图间内在关联的深度学习模型，无疑是一个极具研究价值和前瞻性的探索方向。

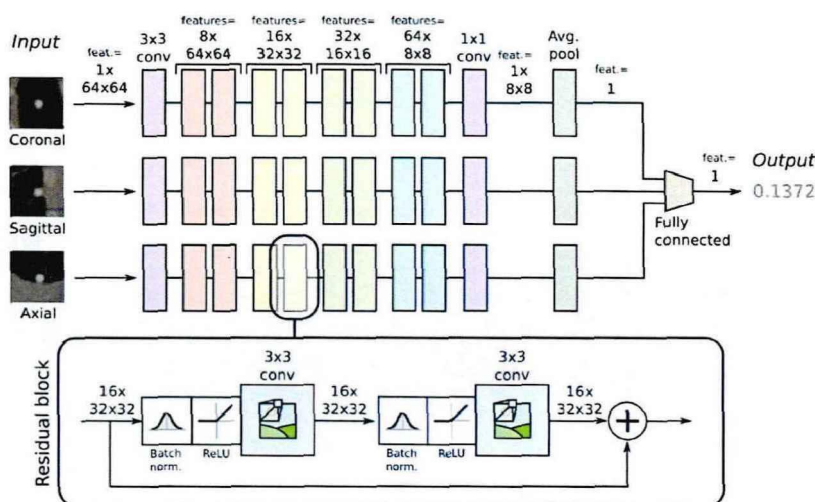
当前的多视图卷积神经网络架构主要分为两大类别：多视图单网络结构与单视图单网络结构^[58]，如图 2.5 所示。多视图一网机制将多视图数据馈送到同一网络中，以获得最终的表示。例如，Dou 等人^[56]研发了一种集成多层次上下文特征

图 2.5 两种多视图深度学习框架示意图^[58]

信息的三维卷积神经网络架构，该网络结构整合了三维卷积层、三维最大池化层以及全连接层等多组件，通过逐层递进的特征抽取过程，学习到能够精准描述目标结节深层特性的表征。该框架的各项性能在医学影像领域的 LUNA16 挑战赛上得到验证，结果显示其在抑制假阳性率方面取得了最优结果，证明了该方案的卓越性能和临床应用潜力，彰显了其在肺结节检测领域的先进性和实用性。

而基于单视图单网络机制的多视图网络框架，其核心思想是针对每个独立视图采用单独的 CNN 模型，以分别抽取各视图的独特表征，随后通过网络后续部分融合这些多元特征表示。Nibali 等人^[66]针对肺结节的恶性与良性鉴别问题，从冠状、矢状及轴向三个解剖平面提取特征，并采用三个独立的深度学习网络进行分析，如图 2.6 所示。研究人员以 ResNet 架构为基础，深入研究了课程学习、迁移学习和不同网络深度对分类准确性的影响。与 LIDC/IDRI 数据集上两个最先进的深度学习模型进行对比，该方法在相同的数据集和实验配置下，在灵敏度、特异性、精密度、受试者工作特征曲线下的面积（AUC）以及整体准确率等多项评价指标上均展现出最优性能。

由于数据样本能够通过多元视角进行多维度描述，各视图间自然蕴含着深层次的高级语义关联。当前的多视图学习研究大多立足于视图一致性原则，即假设多视图数据在本质上是严格对齐的。然而，在现实应用中，多视图数据集常常伴随着潜在噪声及信息分布的不均衡性，这会导致各个视图间产生信息偏差。因此，仅依赖简单的拼接或融合策略来直接学习跨视图共享特征的方法并不能确保稳定和理想的性能表现。同时，多个视图从本质上表示了同一数据集的不同和互补的信息，从而在多视图框架下涌现出的数据具备异构特征属性，这进一步凸显了各视图间的深层复杂交互机制。针对这一挑战，部分研究已提出了基于重建的融合机制。此类方法旨在从多个数据源提取公共表征，并进一步利用该公共表征去重建原始数据，从而在一定程度上缓解了视图间不一致性的问题。然而，这种重构策略迫使模型捕捉到数据的详尽细节信息，这意味着学习得到的表征可能会混杂噪声成分，进而影响模型的泛化能力。因此，设计一种创新的深度多视图学习

图 2.6 单视图单网络机制框架示意图^[66]

机制，高效发掘和提炼各视图间既具有一致性又富含信息的共享特征及特定视图独有的特性信息，将是未来研究的核心课题。

2.2.3 基于多视图深度学习的解耦方法

在人类认知过程中，个体通常运用先验知识对物体的各种属性（例如形状、尺寸、颜色等）进行解析和理解。然而，当前主流的深度神经网络高度依赖纹理和微观结构进行目标识别，而忽视了对全局形态信息的深邃挖掘与综合考量。此外，端到端深度学习模型容易在训练过程中捕获数据中蕴含的一些“无关”特征，例如环境光照变化、背景干扰等因素，这导致模型在面对分布外样本时泛化能力受限^[67]，无法有效揭示并模拟人类所具备的、能够跨越多种情境的类人泛化属性表示机制。为弥补这一理论与实践之间的鸿沟，研究者提出了一种至关重要的表征学习范式——解耦表征学习（Disentangled Representation Learning），并在学术界引起了广泛而深入的研究热潮^[68]。

解耦表征学习的核心目标在于开发一种先进的机器学习模型，能够挖掘并解析数据底层的各向异性的潜在变量，进而生成具有高度可解释性和结构性的表征，以实现复杂数据内在规律和驱动因素的有效揭示与解读。现有学术研究成果已证实^[68,69]，解耦表征学习在模拟人类认知模式和解析现实世界多元现象的本质特征方面具有显著潜力。其核心技术理念在于运用解耦机制，将复杂的语义信息拆解为一组相互独立且不重叠的因素子集，从而实现对现实世界观察数据的精细化理解和深层次解读。

鉴于解耦表征学习在挖掘具有可解释性、可控性及鲁棒性的高质量表征方面

展现出的卓越性能,该技术已广泛应用至计算机视觉,自然语言处理,推荐系统以及图神经网络等多个前沿领域,并显著提升了各类下游任务的表现力。在解耦表征学习这一研究领域中,最具有代表性的方法是基于生成模型的理论框架^[70]。其中,以 Kingma 等人^[71]提出的变分自编码器 (Variational auto-encoders, VAE) 为基础的生成模型方法在图像解耦的学习过程中扮演了至关重要的角色。该模型从极大似然的角度对真实数据进行表征建模,能够有效地对复杂图像数据进行分解与重构,其基础数学表达式如下:

$$\max_{\theta, \phi} \ln p_{\theta}(x) = \max_{\theta, \phi} \{KL(q_{\phi}(z|x) \| p_{\theta}(z|x)) + L(\theta, \phi; x, z)\} \quad (2.6)$$

式中,第一项为变分后验分布 $q_{\phi}(z|x)$ 与真实后验分布 $p_{\theta}(z|x)$ 间的 KL 散度 (Kullback-Leibler divergence)。由于此项非负,第二项 $L(\theta, \phi; x, z)$ 被称为真实数据的变分下界,因此 VAE 模型中新的优化目标函数为:

$$L(\theta, \phi; x, z) = E_{q_{\phi}(z|x)} \ln \frac{p_{\theta}(z|x)}{q_{\phi}(z|x)} = E_{q_{\phi}(z|x)} \ln p_{\theta}(z|x) - KL(q_{\phi}(z|x) \| p(z)) \quad (2.7)$$

式中,第一项 $\ln p_{\theta}(z|x)$ 称为数据的条件对数似然项,第二项 $KL(q_{\phi}(z|x) \| p(z))$ 常称为 KL 项,反映的是变分后验分布 $q_{\phi}(z|x)$ 与先验分布 $p(z)$ 间的相似性。在 VAE 模型中,由于人为选择的先验分布 $p(z)$ 通常满足独立特性,如高斯正态分布等,因此式(2.7)中的第二项 KL 项相当于对网络施加了一定程度的独立性约束,从而通过该优化函数训练出的模型具备了一定的解耦性能。

在实际应用过程中,尤其是医学研究领域,VAE 展现出广阔的应用潜力。例如,Cheng 等人^[72]提出了一个名为多模态解耦变分自编码器 (MMD-VAE) 的模型来执行神经胶质瘤分级的潜在特征学习任务。MMD-VAE 将不同模态下的潜在表征拆分为与胶质瘤分级相关的共享和独特信息,并力求最大化共享信息的一致性以及独特信息的差异性。此外,Yue 等人^[73]提出了一种基于纵向 CT 图像的多损失解耦表征学习方法来准确预测食管癌患者的完全病理缓解,该模型的核心架构由两个 VAE 构成,每个 VAE 内部均包含独立的编码器和解码器组件,如图 2.7 所示。此外,研究中还特别设计了交叉周期重建损失函数以及固有变分损失机制,旨在从纵向 CT 序列中高效提取并区分共享特征与互补特征。实验证明,该模型不仅在完全病理缓解预测任务上展现出卓越性能,而且在面对多中心、未标记数据集的预测分析时,亦能取得显著且稳健的效果。

通过理论分析与实践探索,解耦表征学习在三个核心维度上展现出显著的优势:首先,解耦表征具备对外部语义变化的高度稳健性,即表征的一个组成元素在面对语义层面的变化时能够保持内在一致性^[74];其次,所有解纠缠表征均能精确对应真实世界语义,并具备生成观察到的已知样本、尚未揭示的潜在样本以及

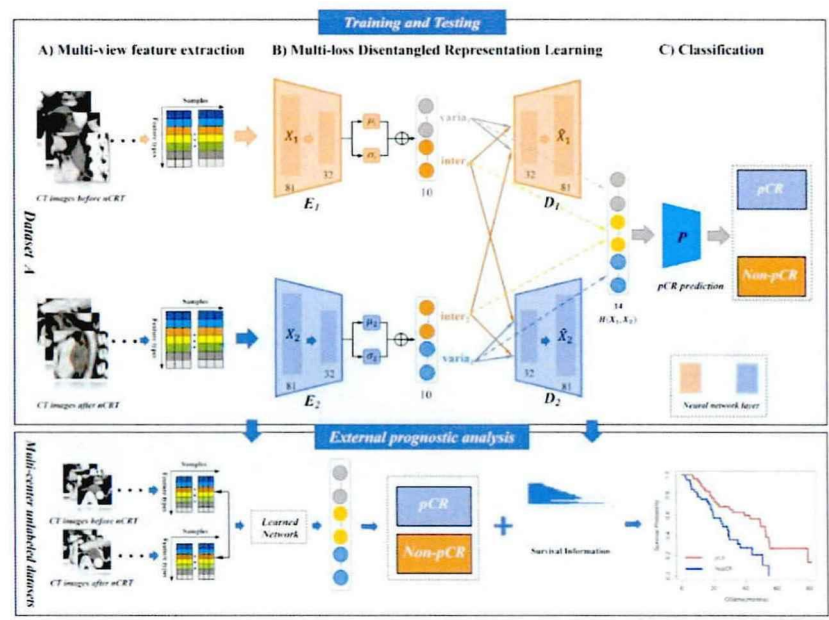


图 2.7 Yue 等人网络框架示意图^[73]

反事实假设样本的能力^[75]；再次，解耦表征学习能够从原始高度耦合的输入特征中提炼出针对特定任务至关重要的表征，该表征具有固有性和鲁棒性，能够有效避免捕获混淆或带有偏差的语义信息，从而实现更高的泛化性能及对未知数据的良好适应性^[76]；最后，解耦表征学习通过在潜在表征空间内操作和调整所获得的表征，实现了对生成过程的精细化调控，增强了模型预测结果的可控性和针对性，从而为临床决策提供更为精准的支持。

2.3 本章总结

本章节首先阐述了癌症图像资料库 TCIA 及其所涵盖的开源非小细胞肺癌 CT 影像数据集——NSCLC-Radiomics 与 NSCLC-Radiogenomics，以及安徽医科大学第一附属医院放射科采集的一系列 CT 影像数据。随后，制定了严谨统一的数据筛选准则以及 CT 影像预处理流程，以构建具有科研价值和实用性的高质量数据集。接下来，本章深入探讨了多视图学习的复杂范式，剖析了其在融合多元异构数据源以提升模型性能方面的核心机制和应用潜力。最后，进一步探究了解耦表征学习的基本原理和技术框架，并对其在医学影像领域中的实践应用进行了专业解读，展示了该技术如何有效地从高维、复杂的医学影像数据中提取出独立且具有生物学意义的底层特征，从而助力于非小细胞肺癌及其他疾病的精准诊断与研究。

第3章 基于多视图解耦学习的肺癌亚型分类方法

在非小细胞肺癌的个体化治疗过程中,从CT图像中精确且无创地辨别组织学亚型对于精准制定治疗方案至关重要。CT成像通常呈现多元化的断面视图,包括冠状面、矢状面、轴位面等,共同构成对肿瘤三维空间结构及病理特征的立体认知。然而,当前多视图深度学习模型在应用中仍面临视图间差异和视图内干扰的挑战,导致模型在处理多视图数据时,对关键病理特征的提取与识别能力减弱,进而制约了其泛化性能。为此,本研究创新性地提出了一种新颖的基于注意力机制解耦架构的MVD方法,运用分而治之的策略,深入挖掘并区分不同视图间的共性与特异性。通过引入基于注意力机制的解耦模块,模型能够敏锐捕捉并凸显关键病理特征,有效地增强公共表征,同时分离出各视图的特定特征,以消除潜在的噪声干扰。最后,通过对学习到的公共表征进行融合,生成鲁棒的多视图表征,用于后续的肺癌亚型分类决策。

3.1 MVD 分类方法

本研究提出了一种实现非小细胞肺癌组织学亚型的高精度分类的多视图解耦方法MVD,其整体架构如图3.1所示。该方法探讨了CT影像的三大核心视图维度:轴位、冠状位与矢状位,并通过寻找不同视图间的共性与特异性,有效应对视图间差异和视图内干扰问题。具体来说,首先运用三个编码器,针对各个视图提取潜在特征表示。随后,通过基于注意力的解耦机制将每个视图的表示投影到视图不变和视图特定子空间中。在视图不变子空间中,利用设计的跨视图转换损失,以全局优化的方式有效地减少不同视图之间的内在差异。与此同时,在视图特定子空间中,引入跨亚型鉴别损失来促使特异表征捕捉亚型无关信息。最后,将分布一致且鲁棒的多视图表征送入亚型分类模块中,以完成对非小细胞肺癌的精准诊断。下面将对MVD中的各个关键组件进行详尽解析与阐述。

3.1.1 基于注意力的多视图解耦网络

为有效应对视图间差异和视图内干扰的难题,MVD方法的核心策略在于运用多视图解耦机制,通过发掘不同视图CT图像间的共性和特异性,促使多视图模型提炼出具有高辨识力和鲁棒性的CT特征。具体来说,MVD作为一种集成学习框架,采用基于注意力的解耦机制,将各种视图投影至一个视图不变的空间

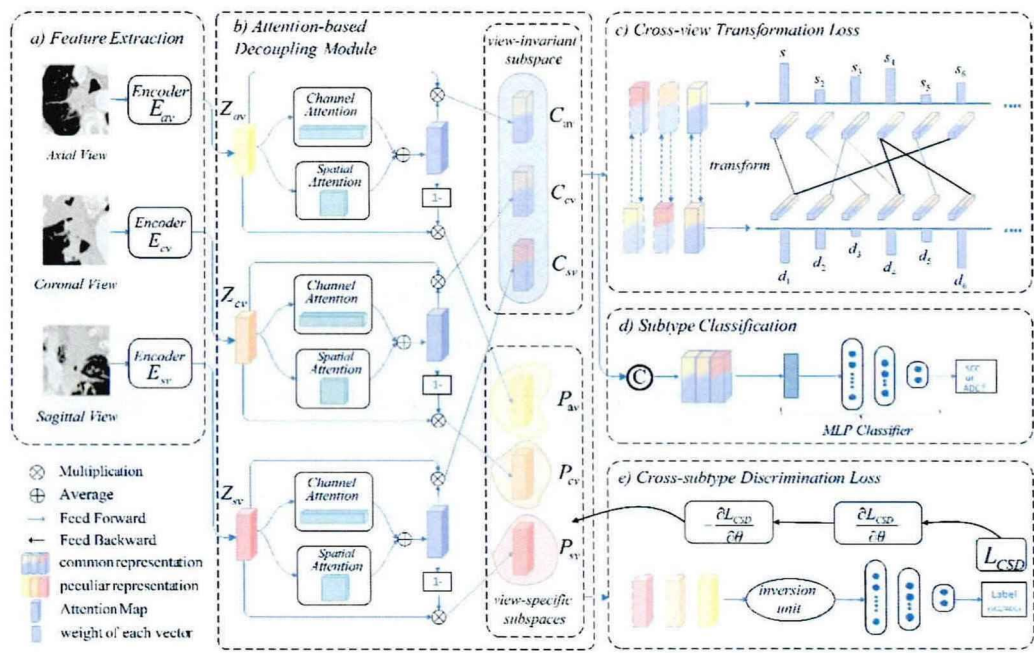


图 3.1 MVD 方法示意图

域，以缩小视图间差异，进而学习到一致性高的公共表征。同时，通过将各个视图映射到视图特定的子空间以提取其特异性的表达，从而独立地挖掘每个视图与肿瘤亚型无关的信息，并将其从共同表征的学习过程中分离出去，从而实现对视图内干扰的有效消除，提升模型预测的准确性与稳定性。

近年来，针对深度解耦结构的研究取得了一定进展^[72,77-79]。例如，Cheng 等人^[72]通过两个线性映射将不同视图的表示分解为共同表示和独特表示，但该方法存在以下局限性：1)仅在一维特征向量层面进行解耦操作，未能充分考虑肿瘤的空间结构信息；2)对于共享表示与独特表示的期望特性未给予明确定义；3)无约束的解耦过程不可避免地导致信息冗余。为有效解决上述缺陷，本研究设计了一种基于注意力的解耦（Attention-based Decoupling, AD）模块，将多个视图分别投射到视图不变和视图特定子空间中。具体实现步骤如下：对于三种 CT 图像的主要视图：轴向视图 X_{av} ，冠状视图 X_{cv} 和矢状视图 X_{sv} ，利用三个类似 ResNet 的编码器 E_{av} 、 E_{cv} 和 E_{sv} ，将每个视图转换为相应的表示 $Z_m \in \mathbb{R}^{C \times H \times W}$ ，即 $Z_m = E_m(X_m)$ ， $m \in \{av, cv, sv\}$ 。随后，引入混合空间一通道注意力模块，对与肺癌亚型高度相关的特征向量施加注意力，从而协同地构建出蕴含在视图不变子空间中的公共表征，这一流程如图 3.2 所示。

其中通道注意力模块通过平均池化和最大池化操作来压缩潜在表征 Z_m ，以凸显特征图的关键信息。随后，将聚合的特征 $Z_m^{cavg} \in \mathbb{R}^C$ 和 $Z_m^{cmax} \in \mathbb{R}^C$ 输入到多层感知机(MLP)以及 Sigmoid 函数，生成通道注意图 $A_m^C(Z_m)$ ，具体数学表达如下：

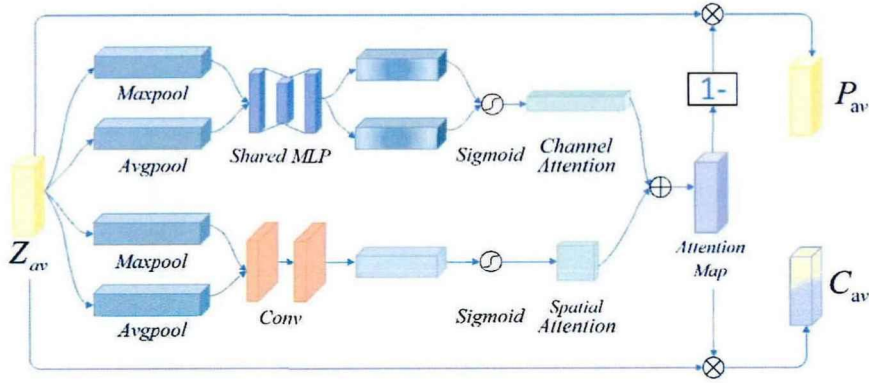


图 3.2 AD 模块示意图

$$A_m^C(Z_m) = \sigma \left(f(Z_m^{cavg}) + f(Z_m^{cmax}) \right) \quad (3.1)$$

式中, σ 为对注意力权值进行归一化的 sigmoid 激活函数, f 为多层感知机, Z_m^{cavg} 和 Z_m^{cmax} 分别为平均池化特征和最大池化特征。该通道注意模块着重强化了通道间的相互依赖关系, 从而提升了视图不变子空间中公共表征的表达效能。

此外, 空间注意模块运用最大池化和平均池化操作沿通道维度聚合特征, 进而获得空间特征 $Z_m^{smax} \in \mathbb{R}^{H \times W}$ 和 $Z_m^{savg} \in \mathbb{R}^{H \times W}$ 。接下来, 将这两个特征映射投影到联合空间中, 并经过一个卷积层及 Sigmoid 函数来习得空间注意力图 $A_m^S(Z_m)$:

$$A_m^S(Z_m) = \sigma \left(\text{Conv} \left(C(Z_m^{savg}, Z_m^{smax}) \right) \right) \quad (3.2)$$

式中, σ 为对注意力权值进行归一化的 sigmoid 激活函数, C 为融合特征的拼接操作, Conv 为卷积网络。此空间注意模块揭示了任意两个特征之间的空间关联性, 以突出肿瘤区域内的信息丰富部分, 进一步增强其特征的表达强度。最后, 混合空间—通道注意力图可以表示为:

$$A_m = (A_m^C(Z_m) + A_m^S(Z_m)) * 0.5, m \in \{av, cv, sv\} \quad (3.3)$$

其中, $A_m^C(Z_m)$ 和 $A_m^S(Z_m)$ 分别为通道注意特征图和空间注意力图。借助混合空间—通道注意力模块, MVD 方法能够充分考量不同通道间的相关性和各关键区域的重要性, 从而有效提升各视图间的共性。

通过运用公式 (3.3), 进一步将多元视图映射至一个视图不变的潜在子空间中 $S_{vi} \in \mathbb{R}^{C \times H \times W}$, 其中各个视图的公共表征可被表述为:

$$C_m = Z_m \otimes A_m \quad (3.4)$$

其中, $\otimes$ 表示逐元素相乘, 且 $m \in \{av, cv, sv\}$ 。直观上看, 那些未得到显著突出的特征通道和空间区域蕴含着与亚型分类无关的固有信息, 而这些信息揭示了

CT 视图间的独特性。因此，本研究引入剩余注意力图，其数学定义如下：

$$A_m^R = I - A_m \quad (3.5)$$

其中， I 表示单位矩阵。同样地，借助公式 (3.5)，将不同视图分别投影到各自视图特定子空间 $S_{vs} \in \mathbb{R}^{C \times H \times W}$ ，每个视图的特异表征可以表示为：

$$P_m = Z_m \otimes A_m^R = Z_m \otimes (I - A_m) \quad (3.6)$$

混合空间—通道注意力机制作为一种简洁而高效的解耦策略，能够协同并自适应地调控各通道间相关性和关键区域的重要性权重。通过同时考虑视图不变子空间和视图特定子空间，AD 模块运用分而治之的策略，有效地应对视图间差异性和视图内干扰问题，从而实现更为精确的多视图 CT 图像分析及亚型预测任务。

3.1.2 跨视图转换损失

尽管解耦架构的实施有助于识别视图之间的公共和特异表示，但仍存在一个亟待解决的问题：不同视图的公共表示可能在视图不变子空间中仍不完全一致，进而影响分类效能。因此，为解决不同视图间的差异性问题，本研究将其重新构建为在视图不变子空间中视图到视图的转换问题，从而对解决上述问题提供独特视角。具体而言，以轴向视图与冠状视图为例，假定空间中存在多条变换路径，轴向视图可沿这些路径映射至冠状视图，实现从一个视图到另一个视图的几何转换与重构，并通过预设成本函数 c_i 衡量转换过程。本质上，降低转换成本有助于提升表征分布的一致性，从而有效减小视图间异质性。基于上述假设，本研究在视图不变子空间内引入了跨视图转换损失函数，在全局范围内寻找最优转换路径，从而最大程度地减小各视图之间的差异。

受最优传输理论^[80]的启发，首先将 AD 模块生成的公共表征 $C_m \in \mathbb{R}^{C \times H \times W}$ 拆分为一组局部向量集合。例如，分解 C_{av} 为 $\{u_1, u_2, \dots, u_{HW}\}$ ，同样地，分解 C_{cv} 为 $\{v_1, v_2, \dots, v_{HW}\}$ ，其中每个向量 u_i 和 v_j 代表对应集合中的一个元素。同时，每个向量关联一个关键参数—转换权重，例如， s_i 和 d_j ，用以反映特征 u_i 和 v_j 在视图间转换过程中的相对重要性。向量 u_i 和 v_j 之间的转换质量表示为 x_{ij} ，而单位质量转换的成本则定义为 c_{ij} 。基于此，转换问题的目标是在两视图间寻找具有最低总成本的路径 $\tilde{X} = \{\tilde{x}_{ij} | i = 1, \dots, HW, j = 1, \dots, HW\}$ ，该目标可形式化表述为：

$$\min_{x_{ij}} \sum_{i=1}^{HW} \sum_{j=1}^{HW} c_{ij} x_{ij}$$

$$\text{subject to } x_{ij} \geq 0, i = 1, \dots, HW, j = 1, \dots, HW$$

$$\sum_{j=1}^{HW} x_{ij} = s_i, i = 1, \dots, HW$$

$$\sum_{i=1}^{HW} x_{ij} = d_j, j = 1, \dots, HW \quad (3.7)$$

其中, s_i 和 d_j 是控制整个转换路径中各向量的权值。因此, 遵循公式(3.7)设计了跨视图转换损失, 来寻求并优化视图 C_{cv} 与视图 C_{av} 之间的最优转换路径 $\tilde{x}$, 从而最大程度地降低整体转换成本。所述单位转换成本是通过向量 u_i 和 v_j 之间的差异进行量化而确定的:

$$c_{ij} = 1 - \frac{u_i^T v_j}{\|u_i\| \|v_j\|} \quad (3.8)$$

更一致的向量将产生相对更低的转换成本。

除了转换成本 c_{ij} 之外, 分配至各个向量的权重也至关重要, 例如, s_i 和 d_j , 揭示了从 $\sum_{j=1}^{HW} x_{ij}$ 到 $\sum_{i=1}^{HW} x_{ij}$ 的整个转换路径。直观上, 权重较高的向量在转换过程中扮演更为重要的角色, 而权重极低的向量对总体变换损失的影响则趋于微弱。为此, 采用视图不变子空间内各向量与平均向量的点积作为相关性得分以计算权重值, 其数学定义如下:

$$s_i = \max\left(u_i^T \cdot \frac{\sum_{j=1}^{HW} v_j}{HW}, 0\right) \quad (3.9)$$

其中, u_i 和 v_j 分别代表两个公共表征的集合中的向量。通过应用 $\max(\cdot)$ 函数确保权重始终保持非负。最后, 为了确保在视图到视图的转换过程中双方具有均衡的总权重, 利用如下公式将所有权重进行归一化处理:

$$\hat{s}_i = s_i \frac{HW}{\sum_{j=1}^k s_j} \quad (3.10)$$

为便于理解, 这里提供一个简化的实例: 在确定全局最优转换路径后, 可以运用公式(3.7)来估算跨视图转换损失, 最终表达式为:

$$L_{CVT} = \sum_{i=1}^{HW} \sum_{j=1}^{HW} \tilde{x}_{ij} c_{ij} \quad (3.11)$$

其中, $\tilde{x}_{ij}$ 表示从向量 u_i 映射至向量 v_j 的最小成本路径上的转换质量, c_{ij} 代表单位质量转换的成本。

总体而言, 本研究从全局最佳视图到视图转换的新颖视角出发, 提出一种新策略来应对视图差异带来的挑战。利用跨视图转换损失衡量将一个视图有效且精准地转换至另一个视图所需成本, 随着该损失的减小, 各视图在视图不变子空间内的公共表征将呈现出更为一致的分布特性, 进而可有效弥合不同视图间的异质性差距。

3.1.3 跨亚型鉴别损失

尽管跨视图转换损失确保了公共特征在视图不变子空间中保持一致性分布,但它无法保证视图特定子空间中的特异表征在解耦过程中仅捕捉与亚型分类无关的信息,从而无法完全消除视图内干扰。实际上,在仅对公共表征进行约束时,存在有用信息泄漏到特异表征中的风险。这种情况下,特异表征不恰当地包含了与肺癌组织学亚型分类相关的信息,这对于有效抑制干扰及精准预测肺癌病理分类十分不利。为应对这一挑战,引入了跨亚型鉴别损失,通过对抗性训练过程,确保特异表征仅包含与组织学亚型无关的信息,从而最大程度地减少信息泄露的问题,进而促进视图内干扰的有效消除。

跨亚型鉴别损失的核心思想在于利用对抗学习策略协同训练特征提取器(即AD模块)和亚型鉴别器,使特异表征无法判断组织学亚型。具体来说,鉴别器利用泄漏至特异表征中的公共信息对非小细胞肺癌亚型进行分类,从而最大限度地减少亚型分类错误。同时,特征提取器则通过减少特异表征中公共亚型相关信息的存在来欺骗鉴别器,以使鉴别器产生的误差最大化。通过这种协同对抗训练过程,特异表征得以纯粹地捕获与亚型无关的信息,从而有效地消除了视图内的干扰。因此,优化目标可以形式化表述为最小—最大优化问题:

$$\min_{\theta_E} \left(\max_{\theta_D} (y \log \varphi(P_m) + (1-y) \log(1 - \varphi(P_m))) \right) \quad (3.12)$$

其中, θ_E 和 θ_D 分别表示特征提取器和鉴别器的参数, $\varphi(\cdot)$ 表示鉴别器, y 表示真实的肺癌亚型标签。

在本研究中,通过整合梯度反转层构建了一个反演单元,对从鉴别器到特征提取器的梯度进行了反转操作,以实现最小—最大优化目标。在前向传播阶段,该反演单元维持输入信号的完整性,并将其无修改地传递至鉴别器,从而促使鉴别器在梯度下降过程中达到最小分类误差。然而,在反向传播环节,反演单元接收到下一层传回的梯度并进行翻转(乘 $-\lambda$)后再向上层传输,从而实质上最大化鉴别器的分类误差,实现对鉴别器的混淆策略。将此正向和反向传播过程数学化表达为:

$$G(P_m) = P_m \quad (3.13)$$

$$\frac{dG}{dP_m} = -\lambda I \quad (3.14)$$

式中, $G(\cdot)$ 为反演单元, I 为单位矩阵,而 λ 为参数,本研究设其为1。基于上述公式,设计了一个多视图跨亚型鉴别损失函数:

$$L_{CSD} = \sum_{m=av,cv,sv} \left(-y \log \varphi(G(P_m)) - (1-y) \log (1 - \varphi(G(P_m))) \right) \quad (3.15)$$

利用上述跨亚型鉴别损失,可巧妙地通过最小—最大博弈策略训练AD模块与亚型鉴别器,促进了特异特征对亚型无关信息的捕获,从而有效缓解视图内部的干扰问题。此外,这种对抗性学习机制有力地抑制了由于信息泄漏而导致的公共表征的退化,进而为肺癌组织学分类任务提供了稳健且抗干扰的多视图表征。

3.1.4 亚型分类预测

通过设计包含上述跨视图转换损失和跨亚型鉴别损失的解耦结构,本研究有效地将每个视图 X_m 的潜在表示 Z_m 分解为公共表征 C_m 和特异表征 P_m 。随后,将这些提取出的公共表征聚合以实现非小细胞肺癌组织学亚型的精确预测,聚合表达式为:

$$C_{comb} = Cat(C_{av}, C_{cv}, C_{sv}) \quad (3.16)$$

其中, $Cat(\cdot)$ 表示聚合, C_{comb} 表示聚合后的多视图表征。此外,依据真实的亚型标签 y (鳞癌或腺癌分别设为0或1),使用交叉熵计算分类损失 L_{cla} :

$$L_{cla} = -y \log f(C_{comb}) - (1-y) \log (1 - f(C_{comb})) \quad (3.17)$$

其中, $f(\cdot)$ 为亚型分类器,该模块由多个ReLU激活函数的FC层组成,表3.1展示了具体信息。

值得注意的是,尽管公式(3.15)和公式(3.17)在表达式结构上具有相似性,但它们接受不同的输入并进行独立训练优化。因此,在本研究中,综合考虑跨视图转换损失 L_{CVT} 、跨亚型鉴别损失 L_{CSD} 以及分类损失 L_{cla} ,构建了最终的目标函数如下:

表 3.1 亚型分类器的具体信息

Layer	Details	Output size
GAP	AdaptiveAvgPool(1,1)	384
FC1	Linear(384,256)	256
	ReLU	
FC2	Linear(256,128)	128
	ReLU	
FC3	Linear(128,2)	2
	Softmax	

$$L_{all} = L_{cla} + \alpha L_{cvr} + \beta L_{csp} \tag{3.18}$$

其中， α 和 β 分别代表调整不同损失项之间相对权重的权衡因子，确保模型在多任务学习环境中能够协同优化各项性能指标。

3.2 模型评价与训练

3.2.1 性能测试方法

为了严谨而全面地评估所提出方法的稳健性和泛化能力，本研究在公共数据集上采取了五折交叉验证这一经实践检验的有效策略，如图 3.3 所示。具体实施步骤如下：首先，对全体患者样本进行随机抽样，确保将整个数据集划分为五个既互斥又均衡的子集，每一子集中各类标签的分布特性与原始数据集一致，以维持样本分布的真实性及类别平衡性。在接下来的五轮迭代过程中，每一轮将其中一个子集作为独立的测试集，其余四个子集则联合构建训练集。为进一步抑制模型在复杂训练过程中可能出现的过拟合现象，在每一轮训练子集中遵循文献^[81]中的方案，抽取 75% 的患者样本进行训练，其余 25% 的患者样本作为验证集，实时监测并调整模型的学习过程，实现对模型泛化性能的有效调控。通过对五轮交叉验证结果的汇总分析，可以计算出各项关键性能指标如准确率、灵敏度、F1 值等（见 3.2.3）的平均值和标准差，从而对模型的整体效能进行深入、全面且客观的量化评估与剖析。

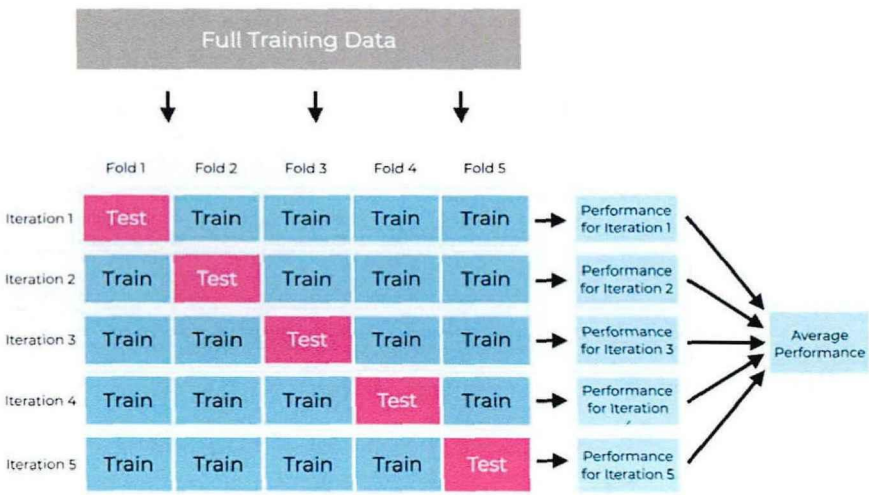


图 3.3 五折交叉验证流程图

注：图片来源（<https://www.sharpsightlabs.com/blog/cross-validation-explained/>）

此外，值得注意的是，除了在公共数据集上采用五折交叉验证法来严格评估模型在 MVD 任务上的表现，本研究还额外引入了一个自建数据集作为独立的外部测试集，从另一个角度验证模型在新环境下的稳定性和泛化能力，确保评估结果的有效性以及无偏性，进而为模型的实际应用提供更为坚实的实证支持。

3.2.2 实验环境及超参数设置

本研究所提出的神经网络架构是在配备 Intel Xeon CPU 4110 @ 2.10GHz 处理器和四块 Nvidia GeForce RTX 2080Ti GPU 的工作站平台上，运用深度学习框架 PyTorch 实现。本研究将初始学习率设定为 0.0001，每经过 20 个 epoch，学习率衰减 0.5 倍。实验过程中观察到该模型在不到 150 个 epoch 内即达到收敛状态，因此，将总训练 epoch 数定为 150，以确保模型训练充分而不过度。

在优化阶段，本研究采用了融合了动量算法和自适应学习率调整策略的 Adam 优化器，以期最大限度地降低 MVD 方法中的损失函数值。动量算法可视作为对传统梯度下降法的一种动态改进，通过引入惯性概念，在每一轮迭代过程中累积并整合历史梯度信息，并赋予一定的权重系数，从而实现更新路径的平滑化，进而有效提升收敛效率。另一方面，自适应学习率算法则针对每个参数独立设定学习率，根据梯度的历史变化自动调整参数的学习速率：对于具有较小梯度幅度的参数，降低其更新步长以防止过度调整；而对于梯度幅度较大的参数，则加快其学习速率以促进快速收敛。基于上述两种关键技术的有机结合，Adam 优化器能够针对神经网络中每一个独立参数动态计算出适应性的学习率，显著增强梯度下降过程的收敛速度与稳定性，进而高效优化模型参数配置，有力推动模型泛化性能的提升。本研究在表 3.2 详尽展示了 MVD 方法所采用的一系列超参数设定。

表 3.2 MVD 的超参数设置

Hyperparameter	Value
Learning rate	0.0001
Learning rate decay	0.5 / 20 epoch
Epoch	150
Batch size	20
Adam(ϵ)	10^{-8}
Adam(β_1)	0.9
Adam(β_2)	0.999

3.2.3 模型评价指标

针对所提出的 MVD 方法,本研究采用了一系列广泛使用的性能评估指标,包括准确率(Accuracy, Acc)、灵敏度(Sensitivity, Sen)、特异度(Specificity, Spe)以及 F1 值等^[82]。这些关键性能指标的计算遵循以下数学公式:

$$Acc = \frac{TP + TN}{TP + TN + FP + FN} \quad (3.19)$$

$$Sen = \frac{TP}{TP + FN} \quad (3.20)$$

$$Spe = \frac{TN}{TN + FP} \quad (3.21)$$

$$F1 = \frac{2TP}{2TP + FP + FN} \quad (3.22)$$

式中, TP 代表真正例数量,即被正确识别为正类别的样本个数; TN 则表示真负例数目,即成功鉴别为负类别的样本数。FP 和 FN 分别对应误报率或假阳性样本量,以及漏报率或假阴性样本量。为了更全面地评估模型性能,本研究引入了受试者操作特征(Receiver Operating Characteristic, ROC)曲线,并以其下方的面积(Area Under Curve, AUC)作为一项重要的附加性能度量指标。ROC 曲线下的 AUC 值能够直观、量化地反映模型在不同阈值设定下对正负样本分类的能力,从而提供了一个综合且具有普适性的评价标准。

3.3 性能评估与比较

3.3.1 解耦模块的有效性分析

为了深入探究 AD 模块中解耦机制及其各组成部分的有效性,本研究在公开的非小细胞肺癌数据集上实施了消融实验。本次实验对多种 MVD 方法的不同变体进行了评估:(i) Baseline: 未集成解耦机制的基线模型;(ii) +AD (CA) 变体: 在 Baseline 的基础架构上进行扩展,运用基于通道注意力(Channel Attention)机制的 AD 模块;(iii) +AD (SA) 变体: 同样在 Baseline 之上进行改良,运用基于空间注意力(Spatial Attention)机制的 AD 模块;(iv) +AD (HA) 变体: 此模型创新性地采用了混合注意力(Hybrid Attention)设计策略于 AD 模块中,实现了通道注意力与空间注意力机制的深度融合与协同作用。

表 3.3 AD 模块不同组件的对比结果

Methods	AUC	Acc	Sen	Spe	F1
Baseline	0.783±0.010	0.738±0.010	0.749±0.016	0.726±0.029	0.759±0.007
+AD (CA)	0.791±0.024	0.748±0.012	0.764±0.039	0.728±0.031	0.768±0.017
+AD (SA)	0.800±0.004	0.751±0.018	0.761±0.033	0.738±0.035	0.772±0.019
+AD (HA)	0.811±0.012	0.763±0.008	0.778±0.018	0.745±0.017	0.784±0.008

实验结果如表 3.3 所示，所有嵌入 AD 模块的变体模型相较于基础模型 Baseline 均获得了显著的性能提升，这一发现强有力地验证了在组织学亚型分类任务中，基于注意力机制的解耦策略所展现的核心作用。特别值得关注的是，融合了混合注意力机制的+AD(HA)变体相较于仅采用单一通道注意力或空间注意力的变体表现出更为卓越的性能。说明通过集成协同与自适应双重注意力机制，+AD(HA)模型不仅能够高效地汇聚并提炼出与亚型紧密相关的公共特征，而且能将亚型分类不相关的噪声信息精准地分离至各自独立的特异表征中。综上所述，上述实验结果充分证实了整合协同与自适应注意力机制对 AD 模块性能优化的重要推动作用，使其能够在复杂医学图像分析过程中采取一种分而治之的策略，成功应对因视图差异和内部干扰带来的挑战。

3.3.2 损失函数的有效性分析

为了深入探究并验证提出的跨视图转换损失与跨亚型鉴别损失在提升 MVD 分类性能上的有效性和优越性，本研究进行了单一损失和两种损失组合的消融实验。在这些实验中，“Baseline”表示仅利用分类损失的解耦模型；“+ L_{CVT} ”指在 Baseline 解耦模型上整合了跨视图转换损失；“+ L_{CSD} ”指具有额外的跨亚型鉴别损失的解耦模型。表 3.4 详尽展示了实验结果，有力证实了相较于简单的 Baseline 解耦模型，在引入跨视图转换损失 L_{CVT} 后，所有评估指标均有显著的性能提升。这表明了跨视图转换损失可以确保视图不变子空间中公共表征遵循一致分布，有

表 3.4 不同损失的对比结果

Methods	AUC	Acc	Sen	Spe	F1
Baseline	0.811±0.012	0.763±0.008	0.778±0.018	0.745±0.017	0.784±0.008
+ L_{CVT}	0.820±0.031	0.772±0.012	0.787±0.021	0.755±0.025	0.791±0.010
+ L_{CSD}	0.823±0.013	0.775±0.016	0.787±0.028	0.762±0.004	0.793±0.017
+ L_{CVT} + L_{CSD}	0.838±0.008	0.794±0.012	0.811±0.011	0.772±0.017	0.813±0.011

效地对抗了因视图间差异引发的负面影响，从而增强了模型对多视图数据内在联系的理解力和利用效率。同时，集成跨亚型鉴别损失 L_{CSD} 也带来了明显的性能增进效果。这主要得益于跨亚型鉴别损失能够有助于防止信息泄漏，并有效抑制无关干扰特征，使模型能够更精确地区分和捕获各类亚型的核心特征。最为突出的进步体现在将跨视图转换损失与跨亚型鉴别损失两者相结合的模型上，其在 AUC 和 Acc 上分别比 Baseline 提升了 2.7%和 3.1%。这一系列实证研究充分证明了提出的跨视图转换损失与跨亚型鉴别损失对于挖掘最优的多视图公共表征以及独特的特异表征具有重要作用，从而显著增强了 MVD 模型的判别准确度及泛化能力。

进一步，为了全面探究跨视图转换损失与跨亚型鉴别损失的权衡因子 $\{\alpha, \beta\}$ 对 MVD 性能的影响，本研究为两个变量 α 和 β 设置了相同的范围 $\{0.2, 0.4, \dots, 1.0\}$ ，在公共数据集上对不同 $\{\alpha, \beta\}$ 组合的 MVD 进行了评估实验。通过这一网格化的探索过程，可识别能够实现最佳性能的参数： $\alpha = 0.6$ 和 $\beta = 0.8$ 。在此最优参数配置下，MVD 方法展现出了卓越的性能表现：AUC 高达 0.838，准确率达到 0.794，而 F1 分数则达到了 0.813，有力证明了该模型在亚型分类领域的高效识别性能。此外，从图 3.4 所示的结果趋势图中可以观察到，无论变量 α 和 β 处于何种取值区间，MVD 模型始终保持了高度的鲁棒性，其 AUC 及准确率均稳定在一个较高水平。这种稳健的性能表现揭示了即使面临操作条件和环境变化，MVD 模型仍能保持良好的适应性和有效性。

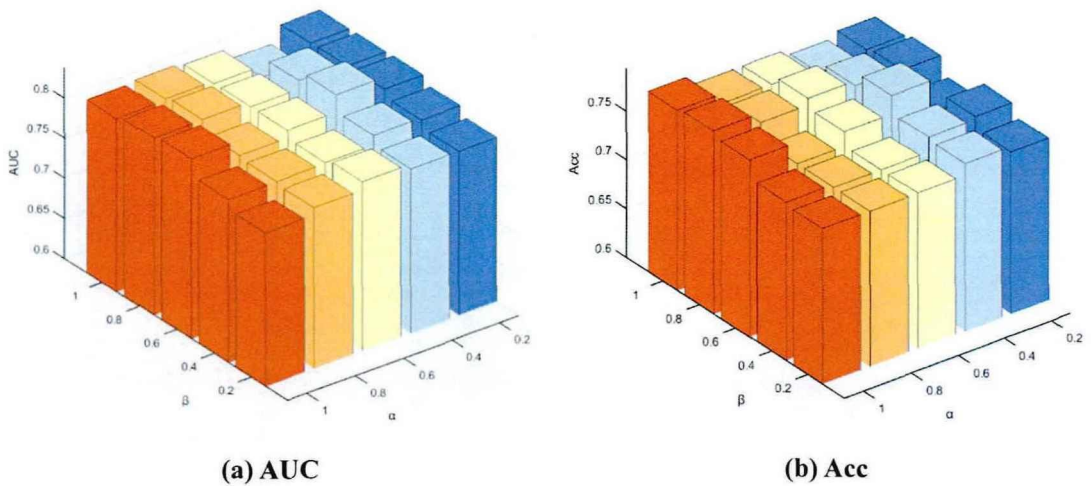


图 3.4 $\{\alpha, \beta\}$ 不同组合时的性能

另外，MVD 算法的有效性主要体现在对视图间差异和视图内干扰的有效处理上，但这两个关键挑战往往阻碍模型有效提取并整合各视图中的核心信息，从而影响模型的学习效能和泛化性能。针对这个问题，本研究进行了收敛性实验，并特别关注了损失函数在此过程中的内在作用机制。如图 3.5 所示，MVD 训练

进程中各个独立损失项及整体损失函数的收敛路径，生动地揭示了随着训练梯度的迭代，所有损失值均呈现出稳健递减的趋势。这不仅体现了模型学习能力的持续增强，更有力证明了设计的损失函数能够有效地指导模型逐步弥合视图间差异并消除视图内的冗余和噪声干扰，从而成功习得更具判别力和鲁棒性的多视图表征。此外，定量评估结果与损失收敛图相互支撑，共同突显出所提损失函数对于提升模型整体性能的关键驱动作用，展现了 MVD 在应对多视图学习环境中固有的复杂性和多样性难题方面卓越的适应性和解决能力。

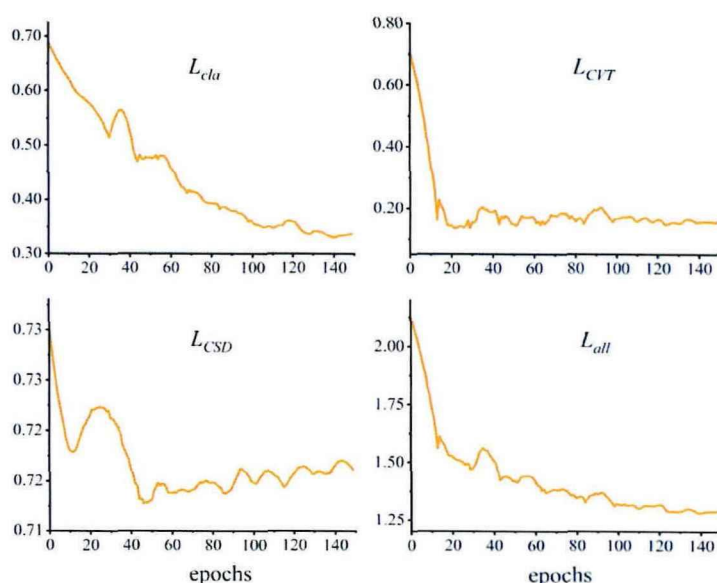


图 3.5 MVD 在公共数据集上的收敛实验

3.3.3 预测分数的比较分析

此外，本研究深度探究了 MVD 模型中亚型分类器与跨亚型鉴别器所输出的预测分数，揭示其内在的表征学习机制以及对非小细胞肺癌不同亚型有效区分的能力。如图 3.6 所示，亚型分类器生成的预测分数在鳞状细胞癌（SCC）和腺癌（ADC）两种亚型群体间呈现出显著差异性，统计学检验结果显示 p 值小于 0.001，这一显著性证据强烈表明，MVD 模型在视图不变子空间中学到的公共表征有效地嵌入了丰富的亚型相关信息，从而能够精确地区分非小细胞肺癌的主要亚型。然而，相比之下，源自跨亚型鉴别器的预测分数并未表现出在不同亚型之间的明显分布差异，对应的统计检验 p 值高达 0.890，这显示了输入鉴别器的特异表征成功地聚焦并捕获了那些对于亚型区分并不关键的视图特异性和潜在干扰成分。这一结果进一步验证了提出的解耦策略在挖掘与亚型分类紧密相关公共特征时的高效性和准确性，同时亦能有效地筛选并滤除影响或降低模型性能的干扰因素。

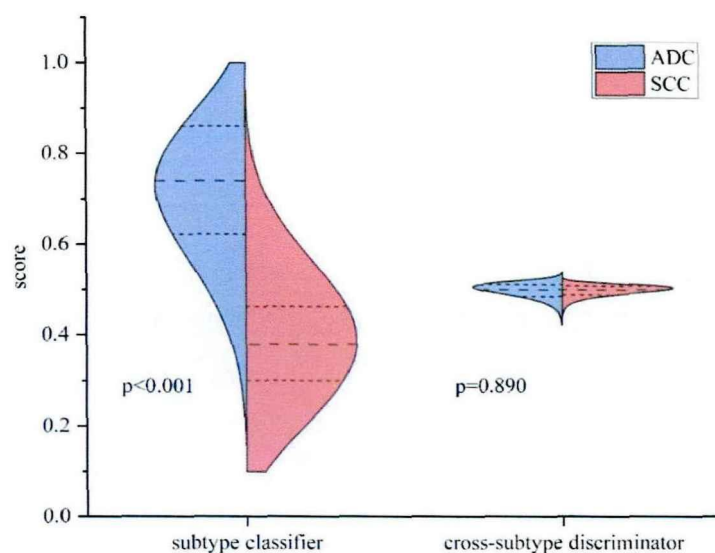


图 3.6 亚型分类器和跨亚型鉴别器的预测得分

3.3.4 多视图与单视图的性能比较

本研究针对多视图输入相较于单视图在亚型分类任务上的优越性进行了深入探究，并将该方法部署于四种不同的视图组合下，实验结果如图 3.7 所示。初步分析可得，由于各类视图间的成像质量差异普遍存在，这一因素直接影响了分类性能的输出结果。具体而言，轴向视图在各视图表现中脱颖而出，其准确率及诊断效能均显著优于冠状与矢状视图。此发现主要原因在于轴向扫描在大多数临床场景中分辨率更高，能够提供更为详尽的解剖结构细节，从而有力提升分类的精确度。进一步深入探究，将轴向、冠状以及矢状三种视图同时纳入模型输入时，MVD 模型实现了最优的分类性能指标。这一实证结果强有力地验证了该模型能够有效地发掘并融合来自 CT 数据不同视图图像中的互补性和多层次特征，进而推动整体诊断效能的提升。这也与当前非小细胞肺癌临床诊断实践中广泛遵循的标准操作规范相一致，在实际操作中，临床医师通常依据轴向视图获取主要诊断线索，同时利用冠状和矢状视图所提供的额外解剖视角和辅助信息，这些多元信息的整合对于确保诊断精度具有决定性作用。综上所述，通过对多视图输入与单视图输入进行对比分析，不仅验证了不同视图间存在差异，其中轴向视图在医学影像诊断中具有突出优势，同时也揭示了借助多视图数据可以有效提升深度学习模型的特征抽取能力和诊断精准性。

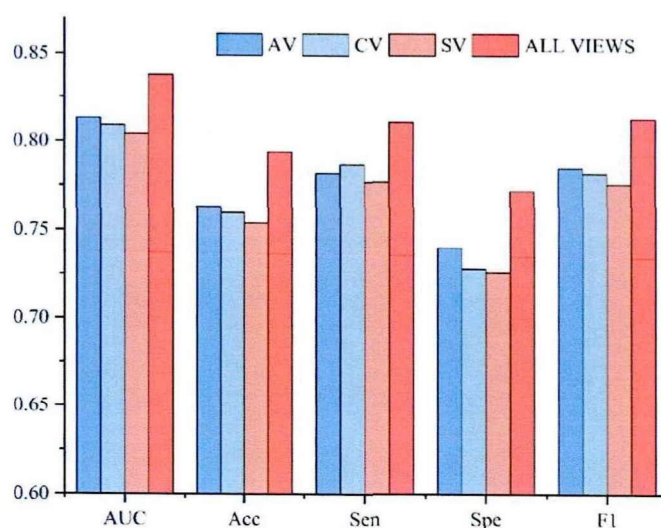


图 3.7 多个和单个视图的 MVD 性能

3.3.5 与现有方法的性能比较

为了深入探究 MVD 方法在非小细胞肺癌组织学亚型分类任务上的性能表现,本研究将其在公开数据集上与其他多种先进的分类方法进行了系统全面的对比与深度分析。在传统机器学习领域,选取了包含 RF^[83]和 SVM^[20]在内的两种经典算法,并采用 PyRadiomics 工具包提取的放射组学特征^[84]以实现肿瘤信息的量化和表征。同时,还选择一些近期深度学习技术进行对比,其中包括 Chaunzwa 等人^[18]以及 Marentakis 等人^[22]提出的单视图学习方法,还有 Yanagawa 等人^[37]与 Guo 等人^[36]提出的 3D 深度学习框架。此外,还引用了一系列基于多视图深度学习策略的研究成果,如 Su 等^[85]、Feng 等^[86]、Wu 等^[41]、Gao 等^[50]及 C. Li 等^[87]的工作。

表 3.5 详细列出了在公共数据集上的五折交叉验证结果。显然,深度学习方法在各指标上均优于传统方法,这突出了深度学习在提取和学习有效的肿瘤表征方面的有效性。Yanagawa 等人所提出的三维深度神经网络架构,在与 Chaunzwa 等人和 Marentakis 等人采用的单视图方法进行对比时,展现出了更为卓越的性能优势。然而,值得注意的是,包含 MVD 在内的多视图解决方案在整体效能上超越了单视图及 3D 模型方法,这一现象凸显了多视图学习在有效利用 CT 中全方位空间特征以及在缓解 3D 模型中潜在过拟合问题方面的优越性。此外,相较于其他多视图技术如 Su 等、Feng 等、Wu 等、Gao 等以及 C. Li 等的研究成果,MVD 实现了显著性能提升,其中准确率和特异性至少分别提高了 4.0% 和 3.9%,进一步证实了该技术在医学影像分析领域的创新性和高效性。

表 3.5 公共数据集上各方法的对比结果

Category	Methods	AUC	Acc	Sen	Spe	F1
Conventional	RF ^[83]	0.760±0.041	0.708±0.055	0.737±0.083	0.672±0.042	0.733±0.059
	SVM ^[20]	0.769±0.052	0.714±0.036	0.771±0.102	0.643±0.082	0.745±0.048
Deep Learning	Chaunzwa et al. ^[18]	0.780±0.021	0.735±0.031	0.744±0.021	0.724±0.053	0.757±0.025
	Marentakis et al. ^[22]	0.781±0.010	0.745±0.012	0.759±0.037	0.728±0.038	0.765±0.014
	Yanagawa et al. ^[37]	0.795±0.011	0.751±0.020	0.781±0.028	0.715±0.048	0.774±0.019
	Guo et al. ^[36]	0.779±0.010	0.729±0.016	0.751±0.050	0.702±0.061	0.751±0.020
	Su et al. ^[85]	0.792±0.017	0.748±0.029	0.771±0.037	0.719±0.034	0.771±0.027
Multi-view Deep Learning	Feng et al. ^[86]	0.783±0.017	0.751±0.018	0.771±0.022	0.726±0.031	0.773±0.017
	Wu et al. ^[41]	0.806±0.010	0.754±0.010	0.775±0.018	0.728±0.005	0.775±0.011
	Gao et al. ^[50]	0.809±0.009	0.751±0.012	0.765±0.013	0.733±0.023	0.772±0.011
	C. Li et al. ^[87]	0.784±0.012	0.754±0.014	0.776±0.036	0.726±0.051	0.751±0.038
	MVD (Ours)	0.838±0.008	0.794±0.012	0.811±0.011	0.772±0.017	0.813±0.011

除上述指标外，图 3.8 以 ROC 曲线的形式对比了各类方法的性能表现。值得强调的是，MVD 方法所呈现的 ROC 曲线显著地更趋于左上角，即理想的完美分辨点 (0,1)，这一现象验证了 MVD 在腺癌和鳞癌（分类任务中的卓越区分能力。具体来说，首先，该模型中基于注意力的解耦模块，通过分而治之的策略，有效地应对了存在于不同子空间中的视图间差异及视图内干扰问题，这种策略有助于模型从多元数据中提炼出更具判别力及泛化能力的特征。其次，MVD 的核心机制包含了两个精心设计的损失函数，二者协同作用，共同辅助模型学习到鲁棒且无干扰的公共表征。因此，结合 ROC 曲线以及上述解析，可以得出结论：MVD 方法凭借其模型设计和优化目标设定，在腺癌和鳞癌的精确区分任务上取得了显著优于现有其他方法的表现，彰显了其在医学影像分析领域的巨大潜力。

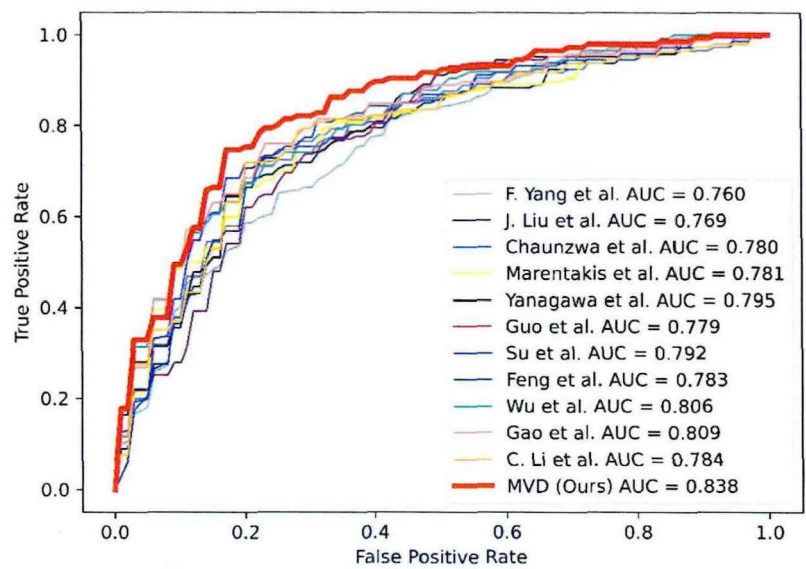


图 3.8 采用五折交叉验证的公共数据集上各方法的 ROC 曲线

为了确保模型评估的严谨性和无偏性，本研究不仅对公共数据集实施了严格的五折交叉验证策略，同时引入了一个独立且未参与训练环节的自建数据集作为外部测试集，从而进一步探究 MVD 方法在非小细胞肺癌组织学亚型分类任务中的泛化能力。表 3.6 详尽列出了这一系列对比实验的结果。尤其值得关注的是，MVD 方法在 AUC 指标上取得了 0.805 的优异表现，该指标作为衡量分类器整体

表 3.6 自建数据集上各方法的对比结果

Category	Methods	AUC	Acc	Sen	Spe	F1
Conventional	RF ^[83]	0.701	0.672	0.714	0.585	0.745
	SVM ^[20]	0.704	0.672	0.702	0.610	0.742
Deep Learning	Chaunzwa et al. ^[18]	0.723	0.688	0.714	0.634	0.755
	Marentakis et al. ^[22]	0.729	0.696	0.726	0.634	0.763
	Yanagawa et al. ^[37]	0.743	0.696	0.714	0.659	0.759
	Guo et al. ^[36]	0.717	0.680	0.714	0.610	0.750
	Su et al. ^[85]	0.740	0.704	0.726	0.659	0.767
Multi-view Deep Learning	Feng et al. ^[86]	0.746	0.712	0.738	0.659	0.775
	Wu et al. ^[41]	0.765	0.720	0.726	0.707	0.777
	Gao et al. ^[50]	0.764	0.720	0.738	0.683	0.780
	C. Li et al. ^[87]	0.753	0.712	0.726	0.683	0.772
	MVD (Ours)	0.805	0.760	0.774	0.732	0.812

性能的关键参数，相较于对照组方法显示出至少 4.0% 的显著提升；与此同时，在 F1 分数方面，MVD 实现了 0.812 的高分值，该分数整合了分类器的精度和召回率两个维度，相比同类竞争技术也展现了至少 3.5% 的优势增长。这些量化的优势凸显了 MVD 在识别和区分非小细胞肺癌不同组织学亚型方面的卓越性能。此外，图 3.9 所示的 ROC 曲线生动地描绘了 MVD 与其他方法在不同阈值设定下真假正例率之间的关系，其曲线位置明显优于其他方法，从而从可视化角度再次印证了 MVD 优异的预测性能。

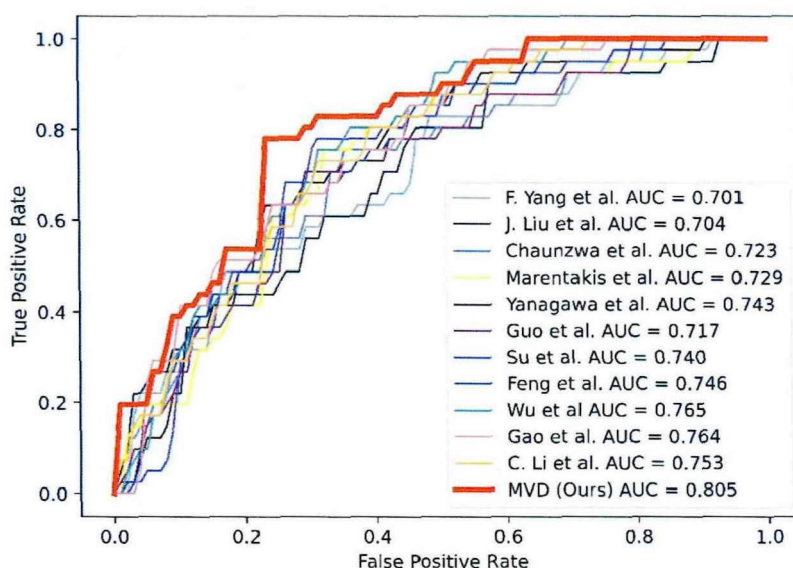


图 3.9 采用独立测试的自建数据集上各方法的 ROC 曲线

3.3.6 模型复杂度比较

为了深入评估 MVD 在模型复杂度方面的优越性，本研究将其与 3D 深度学习方法在模型参数量与计算复杂度层面进行了细致对比。如表 3.7 所示，MVD 模型的训练参数量为 2.187 百万 (million, M)，而 Yanagawa 等人及 Guo 等人的 3D 深度学习框架则分别高达 44.870M 和 41.537M，表明 MVD 的参数需求至少比这些 3D 模型精简了 19 倍之多。进一步，从计算复杂度的维度考量，采用每秒浮点运算 (floating-point operations per second, FLOPS) 作为度量标准，MVD 展现出显著低于 3D 方法的计算负担。这一显著差异主要归因于三维卷积神经网络因处理立体数据而拥有更多可训练权重，这不仅加剧了模型训练的难度，还显著提升了对 GPU 的需求。此外，大量可训练参数亦要求更为庞大的训练数据集以确保模型的有效收敛，进一步限制了 3D 模型在资源约束环境下的适用性和效率。因此，本研究通过采用二维切片来对三维病变结构进行建模，巧妙地平衡了模型

表 3.7 MVD 与 3D 方法的模型复杂度比较

Methods	Parameter	FLOPS
Yanagawa et al. [37]	44.870M	48.469G
Guo et al. [36]	41.537M	4.758G
MVD	2.187M	1.236G

复杂度与性能表现。这一策略不仅有效降低了内存占用，还在有限的硬件资源条件下实现了高效且精确的分类任务目标。

3.3.7 可视化分析

为了进一步探究 MVD 在识别和聚焦肿瘤关键信息区域的能力，本研究运用了先进的可视化技术——梯度加权类激活映射（Gradient Weighted Class Activation Mapping, Grad-CAM）^[88]进行验证。Grad-CAM 的核心机制在于其通过一种梯度驱动的方式量化并可视化神经网络模型中最后一层卷积层各特征图对于输出类别概率的影响程度。具体而言，该方法首先计算每一张特征图对应于特定类别输出的梯度权重，随后将这些权重与各自特征图进行点乘并求和，从而生成一张加权汇总图，并将其映射到原始输入图像中。其中，红色越深的区域表示其对模型最终预测结果的贡献度越高，相应地，这部分区域在进行特征解释时受

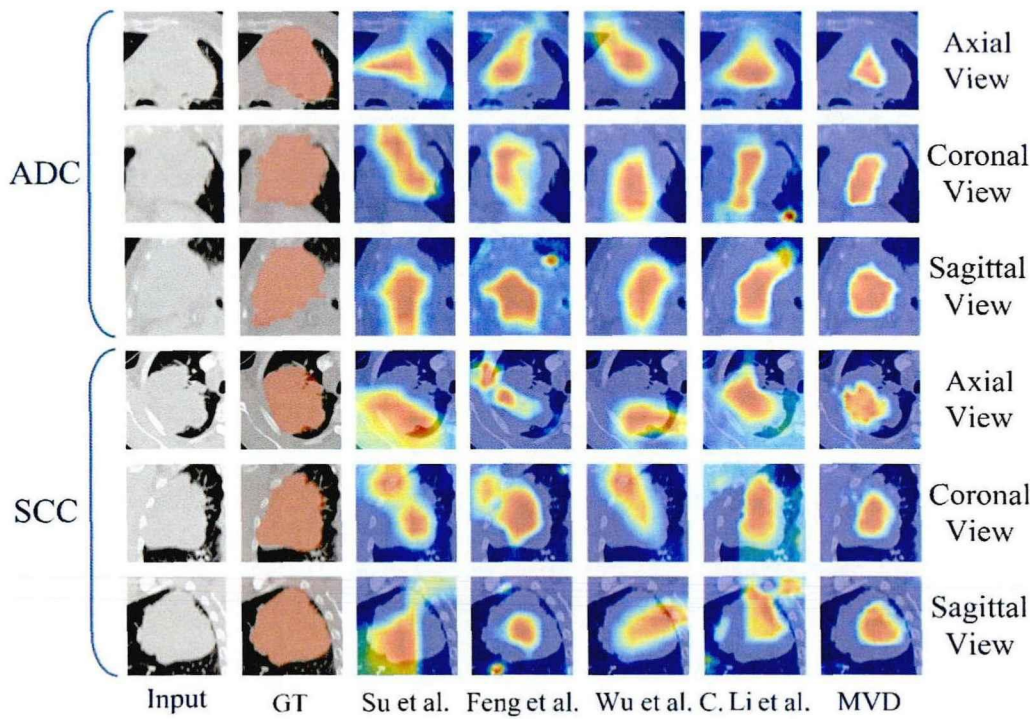


图 3.10 公共数据集上各方法的 Grad-CAM 可视化

到的关注度也更为显著。如图 3.10 所示,在采用 Wu 等人所构建的方法生成的激活热图中,可以注意到显著突出的区域不仅涵盖了肿瘤实体本身,还延伸至相邻的骨组织,这一现象揭示了该方法易收到背景因素的潜在干扰。此外, Feng 等人的研究方案在分析过程中额外注意到了血管,再次验证了其在排除无关背景噪声方面存在的局限性。与之相比, MVD 则展现出了显著的优势,尤其是在有效抑制视图内干扰方面表现卓越。通过 Grad-CAM 提供的可视化证据,可以明确观察到 MVD 能够在不同视图下的 CT 图像中精准定位并锁定肿瘤的核心区域,实现对肿瘤特征的精确提取和强化展示。这种多视图的直观证据强有力地证实了 MVD 在处理复杂的视图间差异及视图内干扰时,具备高度的选择性和辨别力,能够针对性地增强关键特征的表现强度,并同时有效地滤除冗余无意义的背景噪声,这对于提升诊断精度,应对三维医学影像分析中的挑战具有决定性的价值。

3.4 本章总结

本章节创新性地提出了一种名为 MVD 的多视图解耦算法,旨在有效地应对非小细胞肺癌多视图亚型分类中的两大难题:视图间的差异和视图内的干扰。MVD 核心在于其新颖的基于注意力机制的解耦模块,该模块能够将不同视图信息映射至一个视图不变的共享子空间以及一系列视图特定的子空间中,实现对潜在表征的解耦。这种解耦策略使得模型能够生成反映各视图共性的以及体现独特属性的特征。针对视图间的差异问题,本研究特别设计了跨视图转换损失,确保公共表征在视图不变子空间中的分布达到一致。同时,为了有效抑制视图内的干扰,引入基于对抗性学习策略的跨亚型鉴别损失,以引导特异表征捕捉与亚型无关的信息。因此, MVD 方法在提取多视图亚型分类任务所需表征方面表现出显著的优势。经过对公共数据集和自建数据集的严格实验验证和性能评估,所提出的 MVD 方法在区分肺腺癌(ADC)和肺鳞癌(SCC)这两种亚型时展现出了卓越的诊断精确度,并在多个评价指标上超越了一系列的同类分类模型。

第4章 基于跨模态对比学习的多视图肺癌亚型分类方法

尽管第三章中的MVD方法卓有成效,但要指出的是,其仅依赖于图像信息,而未能充分利用临床背景数据。事实上,在实际诊疗过程中,影像报告能够对影像学结果提供详尽解读,从而增强医生的诊断精确度,改善患者的治疗预后。与此同时,CLIP(Contrastive Language-Image Pre-training)大模型成功实现了视觉特征与人类语言描述之间的深度融合。然而,CLIP模型最初仅针对自然图像—文本对进行优化,所以在医学影像与医学知识特有属性方面存在局限性,导致其在医学场景应用中的性能不理想。基于此,本研究创新性地提出了一个基于跨模态对比学习的多视图方法CMCL-MVD,巧妙地引入影像报告中的视觉语义信息作为额外的监督,从而能够精细且深入的理解影像中的肿瘤信息,提高模型区分鳞腺癌的能力。相比于依赖单一图像模态的方法,通过运用跨模态对比学习范式,CMCL-MVD提升了图像与文本间的语义一致性,强化了跨模态深层语义关联,从而有力地促进了有效图像特征的提取过程。

4.1 跨模态对比学习及其预训练模型

4.1.1 跨模态对比学习简介

在医学图像处理领域,深度学习技术已逐步成为应对各类挑战的核心策略。然而,对于以学习为导向的人工智能来说,对医学图像的精准解析仍然是一项极具挑战性的任务,主要原因在于高质量、大规模且具有标注的医学图像数据集的稀缺性。此外,当前的深度学习框架无法模拟放射科医生经过长期教育与实践所形成的复杂推理能力^[89]。在实际临床教学情境中,资深医生通过深入解读影像报告中的病灶特征,并以此为指导线索帮助学习者建立疾病诊断逻辑,而非仅依赖学习者独立从原始影像资料中探寻诊断依据^[90]。这种专家引导的教学方式有助于学习者系统性地理解和掌握病灶表现与医学影像之间的内在联系,从而显著提升诊断的精确度和整体质量。因此,一个自然的解决方案是结合深度学习技术与专家知识引导机制,设计并开发能够模拟和借鉴资深医生诊断思维模式的新型智能医学影像分析系统。

现代医学实践在很大程度上倚重于多元信息和数据资源的深度融合,涵盖影像数据、结构化医疗记录、非结构化文本数据以及特定情境下的音频或观测数据。这些报告往往包含了由专业医师基于影像观察所得出的复杂诊断逻辑和深层次

病理理解，是疾病诊断和病情评估过程中不可或缺的重要组成部分。一项针对放射科医师的研究显示，绝大多数（87%）受访者认为临床信息对影像解释具有重大影响^[91]。然而，目前的深度学习模型主要聚焦于如何从像素级数据中提取潜在的结构和形态特征，而对影像报告中所蕴含的生物特征及临床语义信息的整合与利用尚显不足。这种现状限制了深度学习模型在全面理解并模拟人类医生诊疗思维过程的能力，进而可能影响到其在复杂疾病识别、病情进展预测以及个体化治疗方案推荐等方面的准确性和有效性。为了解决这一问题，亟需发展新型深度学习框架，将医学影像数据与相应的影像报告进行深度融合与联合建模，以期实现从跨模态数据中协同抽取关键信息，并进一步提升智能医疗系统的综合判断与决策能力^[92]。

近年来，整合文本和影像数据，并利用影像报告作为监督信息的深度学习模型，已逐渐成为研究热点^[89,90,93]。这种方法可以突破单一图像模型的局限性，并为影像分析模型赋予更为详尽的解释性支撑。目前，从报告中抽取监督信息以训练医学视觉表征模型的主流策略之一，是通过人为设计的规则从影像报告中提取疾病标签，随后将这些标签用于模型训练。例如，Yan 等人^[94]从与病变影像相对应的影像报告中抽取关联的语义标签，再基于所挖掘出的图像和文本双重标签资源，提出了一种以多标签卷积神经网络为基础的病灶标注网络（Lesanet），从而通过标签间存在的层次关系和互斥关系来优化标签预测准确性，其网络架构如图 4.1 所示。

然而，该方法仍面临一系列技术瓶颈。首要挑战源自高质量标注数据的稀缺及其品质不一问题，这对深度学习模型训练效率和性能优化构成了显著制约。其次，深度理解医学报告亦是关键挑战之一。由于影像报告蕴含高度专业性、灵活

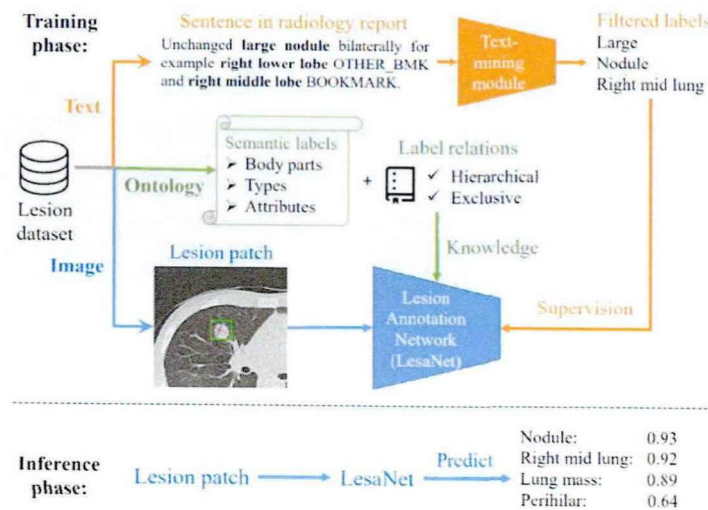


图 4.1 Yan 等人网络框架示意图^[94]

多变的句法结构以及大量使用的医学术语,因此制定标签规则时需要投入大量的人力与专业知识,从而导致其文本数据的有效利用率偏低^[95]。再者,提取的标签通常含有噪声,导致训练出的模型并非最优^[96]。此外,仅关注标签提取过程,不能充分利用报告中描述,限制了模型的全面理解与表现能力。相对普通自然语言处理技术,医学报告具有高度复杂性和专业特性,这极大地提高了对模型所需理解力和推理能力的要求。尤其是在跨机构、跨地域情境下,医生个体间书写风格和表达方式的异质性进一步提升了构建联合嵌入模型的复杂程度^[97,98],严重限制了模型在不同医疗学科领域及医疗机构间的有效泛化性能与普通适性。

4.1.2 基于跨模态对比学习的预训练模型

借鉴于从自然图像中进行无监督对比学习的前沿技术^[99-101],现有研究^[95,97,98,102]尝试将这一思路应用于医学影像领域,以有效利用未标注图像数据资源,并通过迁移 ImageNet 预训练模型权重来优化性能。然而,相较于识别自然图像物体所需的粗略视觉特征表示,医学影像理解往往要求更为精细且独特的视觉特征表达。同时,已有研究发现^[102],在针对医学影像编码器初始化的过程中,采用 ImageNet 预训练权重相较于随机初始化并未带来显著优势。

因此,当前研究的焦点转向了利用配对图像-文本预训练方法来增强对医学图像的视觉理解,通过跨模态对比学习策略优化潜在空间中视觉表征与文本表征的对应关系。这种方法无须依赖额外人工标注,能够深入挖掘并学习到蕴含在影像报告中的医学视觉特征,并将这些学到的特征有效迁移至下游任务中。例如,Zhang 等人^[103]创新性地提出了一种名为 ConVIRT 的预训练策略,其整体架构如图 4.2 所示。该方法摒弃了依赖报告生成分类标签的传统做法,转而利用医学影像与影像报告之间的自然配对关系来学习视觉表征。具体来说,利用影像报告所蕴含的语义信息作为弱监督信号来预训练图像模型,并进一步在有标记的下游任务上进行微调,从而实现了模型性能的提升或所需标注样本量的减少。ConVIRT 巧妙整合了大规模文本数据驱动学习和无监督技术,可以放大真实图像-文本配

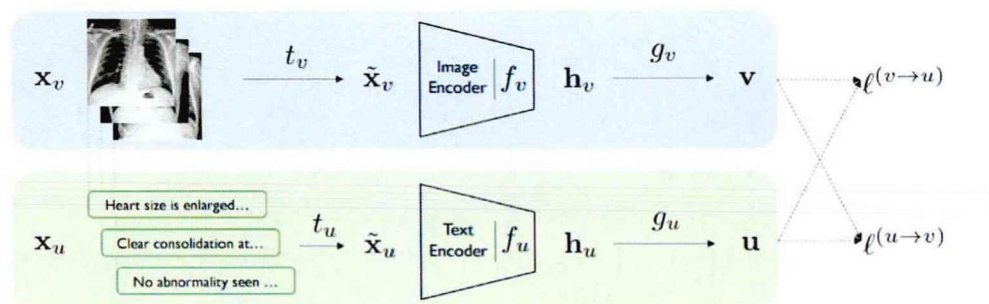


图 4.2 Zhang 等人网络框架示意图^[103]

对与随机图像对之间的一致性,从而极大地提升了医学影像领域内视觉表征学习的有效性和鲁棒性。实验结果显示,在多种情境下,ConVIRT 生成的图像表征相较于基准方法均表现出了显著优越性。

进一步,在 ConVIRT 研究的启发下,OpenAI 团队的 Radford 等人^[104]进一步探究了大规模自然语言监督训练在图像分类领域的应用,并构建出一种名为 CLIP (Contrastive Language-Image Pre-training) 的大模型。该模型基于跨模态对比学习框架,其核心思想是通过对比语言-图像,从丰富的自然语言标签中抽取有效知识。研究团队首先利用互联网上广泛可获取的大量(图像,文本)对数据资源,构建了一个包含四亿对配对数据的庞大预训练集。然后,在初始化阶段,CLIP 摒弃了基于 ImageNet 权重初始化图像编码器以及采用预训练权重初始化文本编码器的传统做法,选择从头开始同时训练两种编码器。此外,模型在融合不同模态表征时,摒弃了非线性投影方式,仅采用线性投影将每个编码器的输出映射至跨模态嵌入空间。同时,鉴于 CLIP 预训练数据集中许多(图像,文本)对仅包含单个句子,移除了 ConVIRT 中负责从文本中均匀抽样生成单个句子的文本转换函数 t_u 。图像变换函数 t_v 也作简化处理,仅保留了从调整尺寸后的图像中随机裁剪这一数据增强方法。最后,为了更精确地调控 softmax 操作中的对数范围参数 τ ,研究团队创新性地将其直接融入训练过程中进行优化,从而避免了 τ 作为超参数可能带来的潜在问题。

此后,一系列学术论文^[105-107]将跨模态对比学习应用于医学影像领域。这些研究大多借鉴并拓展了 CLIP 模型,通过最大化配对文本与图像样本之间的相似度,并同时最小化非配对元素间的相似性,来有效挖掘和构建跨模态的匹配关系。在此基础上,Yan 等人^[108]选取难度更高的负样本,从而优化类内差异特征的学习。Huang 等人^[109]以及 Seibold 等人^[110]则聚焦于对齐视觉区域与疾病标签来学习多粒度特征表示。此外,Eslami 等人^[111]根据医疗数据对其进行微调,使 CLIP 更好地适应医疗领域的下游任务。

总的来说,尽管上述方法在操作执行上存在差异,但其核心理念均在于挖掘医学图像与相应影像报告间的内在一致性。每种方法都在医学影像领域提供了独特的价值贡献,生动展现了 CLIP 预训练模型的卓越适应性和广阔潜力。

4.2 CMCL-MVD 方法

在现有工作的基础上,本研究提出了一种基于跨模态对比学习的多视图解耦学习方法 CMCL-MVD,来实现非小细胞肺癌组织学亚型分类任务,其整体架构如图 4.3 所示。该方法引入 CT 影像和对应的影像报告,深入探究了两者之间的

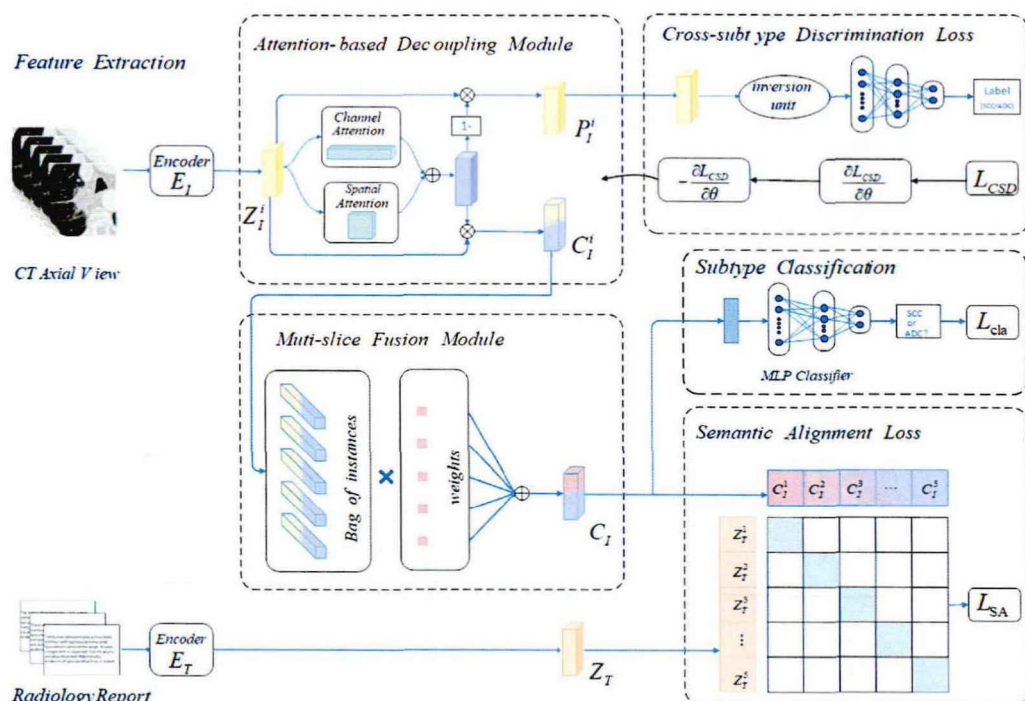


图 4.3 CMCL-MVD 方法示意图

跨模态信息交互，从而提升了对不同肺癌亚型的图像更细致的理解能力。具体来说，CMCL-MVD 采用了双编码器结构，首先通过图像编码器和文本编码器将 CT 影像及对应的影像报告映射至一个跨模态共享嵌入空间中，从而提取出蕴含丰富语义信息的跨模态潜在特征表示。其次，借助第三章中的基于注意力机制的解耦策略与跨亚型鉴别损失，将影像的潜在表征分解为公共表征和特异表征。随后，运用精心设计的多切片融合模块有效地整合了来自不同切片的多层次信息，并通过语义对齐损失增强图像和文本之间的语义对应关系。最后，将新的 CT 图像输入到已训练完成的分类器中，实现对非小细胞肺癌各亚型的精确识别与区分。下面将对 CMCL-MVD 中的核心组件及其功能进行深入阐述。

4.2.1 图像和文本编码器

本研究采用 Vision Transformer (ViT) 作为图像编码器^[112]。首先，将原始图像 $x \in \mathbb{R}^{C \times H \times W}$ 重塑为一个二维切片序列 $x_p \in \mathbb{R}^{N \times (P^2 \cdot C)}$ ，其中 (H, W) 为原始图像的分辨率， C 为通道数， (P, P) 为每个切片的分辨率， $N = HW/P^2$ 为生成的总切片数，即 Transformer 模型的有效输入序列长度。随后，运用可训练的线性映射将每个切片映射为一维向量，其数学表达式如下：

$$z_0 = [x_{class}; x_p^1 E; x_p^2 E; \dots; x_p^N E] + E_{pos} \quad (4.1)$$

式中, $E \in \mathbb{R}^{(P^2 \cdot C) \times D}$, $E_{pos} \in \mathbb{R}^{(N+1) \times D}$, $z_0^0 = x_{class}$ 是在切片序列前添加的一个可学习的嵌入。此外, 将标准的一维位置特征添加到切片特征中以保留位置信息, 并将整合后的特征向量序列作为编码器的输入。

编码器由多头自注意力(multiheaded selfattention, MSA)和 MLP 块的交替堆叠而成, 其数学表述为:

$$z'_l = \text{MSA}(\text{LN}(z_{l-1})) + z_{l-1} \quad (4.2)$$

$$z_l = \text{MLP}(\text{LN}(z'_l)) + z'_l \quad (4.3)$$

式中, LN 为层标准化 (LayerNorm), 且 $l = 1 \cdots L$ 。LN 层之后应用残差连接^[113]。MLP 块包含两个具有 GELU 激活函数的非线性层。最终, Transformer 编码器输出处 z_L^0 的状态 y 作为图像的高层次表征, 其公式表示为:

$$y = \text{LN}(z_L^0) \quad (4.4)$$

本研究采用一种改进的 Transformer^[114]作为文本编码器, 该架构具有 6300 万个参数, 12 层结构, 并生成 512 维的隐层向量, 配备 8 个注意头。此文本编码器对具有 49,152 词汇大小的文本的小写字节对编码 (byte pair encoding, BPE) 表示进行操作^[115]。为了优化计算效率, 最大序列长度限制为 76。处理过程中, 文本序列以特殊符号[SOS]和[EOS]进行封装, 在[EOS]位置处的向量作为文本的特征表示, 并经过层归一化处理后再通过线性投影映射至多模态嵌入空间。此外, 在文本编码器中融入了隐式自注意力机制, 以便于保留预训练语言模型初始化的优势或添加语言建模作为辅助任务的能力。

值得注意的是, 当应用于医学影像这样的专业领域时, CLIP 预训练模型展现出了显著的泛化性能。尽管其最初是在互联网图像及其相关文本描述数据集上进行训练, 但已有研究表明^[116], CLIP 已成功实现了对医学影像的识别和分类。因此, 为了充分挖掘预训练 CLIP 模型在临床环境下的潜在应用价值, 本研究在训练阶段选择保持 CLIP 预训练阶段所获得的图像和文本编码器参数的冻结状态, 以期实现对预训练知识的有效迁移与利用。具体来说, 图像特征可利用图像编码器提取:

$$Z_I^i = E_I(x_i^i) \quad (4.5)$$

而文本特征则通过文本编码器提取:

$$Z_T = E_T(x_T) \quad (4.6)$$

两式中, x_i^i 和 x_T 分别代表待处理的第 i 张医学影像切片和影像报告, E_I 和 E_T 则表示从 CLIP 预训练阶段继承且保持不变的图像或文本编码器。

4.2.2 基于注意力的解耦模块

尽管 CLIP 的图像-文本联合编码器表现出强大的功能，但其在区分医学影像中细微差异时仍存在局限性。实际上，在医学影像中，关键的病理特征往往局限于图像的小范围区域，而大部分图像区域可能为不含有效信息的背景，这对生成高质量的图像表征构成潜在干扰。当前研究工作^[110,117,118]普遍采用针对目标医学影像域的微调策略来优化预训练的 CLIP 模型，使模型能够有效适应并理解医学影像领域的特定背景特征，从而获得更高质量的鲁棒的图像表征。因此，基于第三章的研究成果，本研究引入了一种基于混合注意力的解耦机制，其数学表达式如下所示：

$$A_I = (A_I^C(Z_I^i) + A_I^S(Z_I^i)) * 0.5 \quad (4.7)$$

式中， $A_I^C(Z_I^i)$ 和 $A_I^S(Z_I^i)$ 分别为通道和空间注意力特征图。则解耦过程的表达式写作：

$$Z_I^i = \underbrace{Z_I^i \otimes A_I}_{C_I^i} + \underbrace{Z_I^i \otimes (1 - A_I)}_{P_I^i} \quad (4.8)$$

式中， $\otimes$ 为逐元素相乘， C_I^i 和 P_I^i 分别为第 i 张切片的公共表征和特异表征。此解耦策略能够更加精细化的处理医学图像，通过强化模型对局部关键区域的注意力分配，以提取出更具针对性的图像表示，从而更好地捕获 CT 影像中亚型分类任务所必需的视觉特征信息。

此外，引入与第三章一致的跨亚型鉴别损失，通过对抗性训练过程，确保特异表征仅包含与组织学亚型无关的信息，从而促进图像背景干扰的有效消除。跨亚型鉴别损失函数表达式如下：

$$L_{CSD} = -y \log \varphi(G(P_I^i)) - (1 - y) \log(1 - \varphi(G(P_I^i))) \quad (4.9)$$

式中， $G(\cdot)$ 为反演单元。利用跨亚型鉴别损失，巧妙地通过最小-最大博弈策略训练注意力解耦模块与亚型鉴别器，促进了特异特征对亚型无关信息的捕获，从而有效缓解背景干扰问题。

4.2.3 多切片融合模块

本研究中，每个病患 CT 影像的全部轴向切片被逐张输入模型进行分析，并据此形成最终诊断。然而，对每张切片进行精确标注是一项艰巨且耗时的任务。事实上，患者级别的标签并不适用于每一张切片，例如，肺部 CT 仅部分切片存在肿瘤信息，而非所有切片均能看到病变。由于数据标注的不完整性或低质量性

问题日益凸显,多实例学习(Multiple Instance Learning, MIL)逐渐成为研究焦点并在医学影像分析领域展现出了显著的应用潜力。本研究采用 MIL 策略,将每位患者的全部轴向切片视为一个“包”结构内的多个独立实例进行后续融合。其核心机制在于挖掘并识别对预测结果具有决定性影响的关键实例(Key instance),通过赋予这些关键实例更高的权重,以确保它们在融合过程中占据主导地位,从而提升整合后包级特征的表达力和诊断准确性。已有研究表明^[119],精准探测并有效利用关键示例对推进临床实践中的疾病诊断具有极高的价值和意义。

目前, MIL 方法主要分为两大类^[120]:一类是实例级方法,采用实例级分类器对每个实例生成分数,然后通过 MIL 池化操作整合各个实例得分以获得整体预测分数。然而,由于单个实例标签未知,可能导致实例级分类器训练不足,从而在最终预测中引入额外误差或不确定性;但该方法的可解释性强,能够提供有助于定位关键实例的信息。另一类是嵌入级方法,首先将实例映射至低维嵌入空间,通过 MIL 池化得到与包内实例数量无关的包级表示,再由包级分类器进一步处理以得出最终评分。嵌入级方法确保了包级别的联合表示,避免给包级分类器带来额外偏差,并在分类任务上通常能达到更高的精度,然而其可解释性相对较低。因此,综合以上两类方法,考虑引入一个合适的 MIL 池化函数,将可解释性纳入 MIL 嵌入级方法。

目前,神经网络中的 MIL 池化可以划分为不可训练池化和可训练池两类^[121]。不可训练池如最大池化、均值池化以及加和池化等,实施简单,但因不可训练限制了其适用性。相比之下,可训练池化如门控注意力池化、注意力池化、自适应池化等复杂度更高,具备更强的灵活性、适应性及可解释性,能在特定任务和数据集上获得更优结果。因此,本研究引入了一种基于完全可训练 MIL 池化的多切片融合(Multi-slice fusion, MSF)模块,从而更好地适应各个实例并进行有效融合。具体来说,采用基于注意机制的实例加权平均,其中权重由神经网络动态确定,并确保所有权重之和为1,以保持与“包”大小的一致性。设 $\{C_1^I, C_2^I, \dots, C_K^I\}$ 是一组包含K个切片公共表征的集合,则提出的 MIL 池化融合的表达形式如下:

$$C_I = \sum_{k=1}^K a_k C_I^k \quad (4.10)$$

其中

$$a_k = \frac{\exp\{w^T \tanh(Vh_k^T)\}}{\sum_{j=1}^K \exp\{w^T \tanh(Vh_j^T)\}} \quad (4.11)$$

式中,参数 $w \in \mathbb{R}^{L \times 1}$, $V \in \mathbb{R}^{L \times M}$ 。此外,利用双曲正切函数 $\tanh(\cdot)$ 逐元素非线性操作来包括负值和正值,确保梯度的正确传播,从而发现实例之间的相似性或差异性。。然而,值得注意的是,对于 $x \in [-1, 1]$ 范围内的输入, $\tanh(x)$ 表现出近似

线性的特性,这对其捕捉实例复杂关系方面产生局限性。因此,进一步引入门控机制^[122],并与 $\tanh(\cdot)$ 非线性单元结合,形成以下公式:

$$a_k = \frac{\exp\{w^T(\tanh(Vh_k^T) \odot \text{sigm}(Uh_k^T))\}}{\sum_{j=1}^K \exp\{w^T(\tanh(Vh_j^T) \odot \text{sigm}(Uh_j^T))\}} \quad (4.12)$$

式中,参数 $U \in \mathbb{R}^{L \times M}$, $\odot$ 表示为逐元素乘法, $\text{sigm}(\cdot)$ 代表 sigmoid 型非线性函数。门控机制通过引入一种可训练的非线性函数,能够减轻 $\tanh(\cdot)$ 在近似线性区域的限制。

从本质上讲,基于多实例学习池化的多切片融合模块能够对包内各个实例赋予不同的权重,从而融合生成富含信息的包级表示。换句话说,该方法能够自适应地发现并有效整合关键实例,为包级别的分类任务提供具有判别力的特征表达。

4.2.4 语义对齐损失

尽管图像编码器和文本编码器能够将图像和文本映射到跨模态共享嵌入空间,但未能充分建模并捕获两模态间的内在关系,导致无法将影像报告中的先验知识有效融入深度学习过程。为此,本研究提出了一种跨模态语义对齐损失函数,通过强化 CT 影像与其对应影像报告之间的语义一致性,并有效地抑制与其他患者无关报告之间的混淆,从而引导模型学习和理解更具判别力的视觉特征。语义对齐损失的核心思想是在训练阶段,在表征空间中拉近匹配的图像—文本配对(正样本,如图 4.3 中绿色方块显示)的距离,同时推远非匹配的图像—文本配对(负样本)的距离,实现正样本间的高相似性和负样本间的低相似性。

具体来说,对于批次大小为 N 的图像—文本对集合 $\{(x_I^j, x_T^j)\}$, x_I^j 和 x_T^j 表示为第 j 对图像和文本,可以构建出 N^2 个潜在的图像—文本对,其中包括 N 个正样本匹配对以及 (N^2-N) 个负样本不匹配对。令 C_I^j 和 Z_T^j 分别代表第 j 对的图像和文本特征向量,采用余弦相似度 $\text{sim}(\cdot)$ 作为衡量不同元素之间语义相似性的量化指标,其数学表达式如下:

$$\text{sim}(C_I, Z_T) = \frac{C_I \cdot Z_T^T}{\|C_I\| \cdot \|Z_T\|} \quad (4.13)$$

鉴于医学图像和对应的影像报告在病变语义层面应具有高度的一致性,因此力求最大化同一患者的影像与文本特征间的语义相似度,同时最小化不同患者的特征间的语义相似度。因此,最终形式化的语义对齐损失函数为:

$$L_{sa} = -\log \frac{\exp(\text{sim}(C_I, Z_T)/\tau)}{\sum_{k=1}^{2N} \mathbf{1} \exp(\text{sim}(C_I, Z_T)/\tau)} \quad (4.14)$$

式中, $\tau \in \mathbb{R}^+$ 表示温度参数, 用于调整相似度分布的平滑度。通过最小化上述语义对齐损失, 可确保图像—文本正样本在共享潜在空间中得到紧密对齐, 突显关键图像区域与文本描述之间的相互作用, 有利于深入挖掘疾病特征, 从而显著提升模型性能。

4.2.5 亚型分类预测

通过实施基于医学图像—文本对的跨模态语义对齐策略, 能够学习到富含语义信息且高质量的图像表征, 从而实现对非小细胞肺癌组织学亚型的精确预测。本研究提出的模型能够模拟放射科医师对医学影像资料进行深度解析的过程, 从而识别并整合影像报告中的关键描述性特征, 形成诊断结论。基于本文 3.1 节, 依据真实的亚型标签 y (鳞癌或腺癌分别设为 0 或 1), 使用交叉熵计算分类损失 L_{cla} :

$$L_{cla} = -y \log f(C_I) - (1 - y) \log(1 - f(C_I)) \quad (4.15)$$

其中, $f(\cdot)$ 为亚型分类器, 该模块由多个 ReLU 激活函数的 FC 层组成, 表 4.1 展示了具体信息。

在本研究中, 综合考虑跨亚型鉴别损失 L_{CSD} 、语义对齐损失 L_{sa} 以及分类损失 L_{cla} , 构建了最终的目标函数如下:

$$L_{all} = L_{cla} + \alpha L_{CSD} + \beta L_{sa} \quad (4.16)$$

其中, α 和 β 分别代表调整不同损失项之间相对权重的权衡因子, 确保模型在多任务学习环境中能够协同优化各项性能指标。

表 4.1 亚型分类器的具体信息

Layer	Details	Output size
GAP	AdaptiveAvgPool(1,1)	512
FC1	Linear(512,256)	256
	ReLU	
FC2	Linear(256,128)	128
	ReLU	
FC3	Linear(128,2)	2
	Softmax	

4.3 数据处理与模型训练

鉴于本章节的研究内容涉及到影像报告的使用,因此仅选取了包含详尽报告信息的 NSCLC-Radiogenomics^[53,54]数据库中的 117 例非小细胞肺癌患者数据。同时,本研究还从安徽省胸科医院影像科额外采集了 343 例患者的影像资料与配套的影像报告。所纳入的影像数据均来源于使用西门子公司生产的 SOMATOM Definition AS 螺旋 CT 设备进行的肺部 CT 扫描,其层厚设定为 2 毫米,空间分辨率为 1 毫米。遵循与第三章严格一致的病例入选标准与筛选流程,最终确定采用 232 例符合要求的非小细胞肺癌病例的数据作为有效样本。综上所述,本研究整合了公共数据集及新自建数据集,构建了一个共计包含 349 例符合研究条件的非小细胞肺癌患者的综合数据集,其中涵盖了 189 例腺癌病例以及 160 例鳞癌数据,作为本研究的核心实验数据资源。

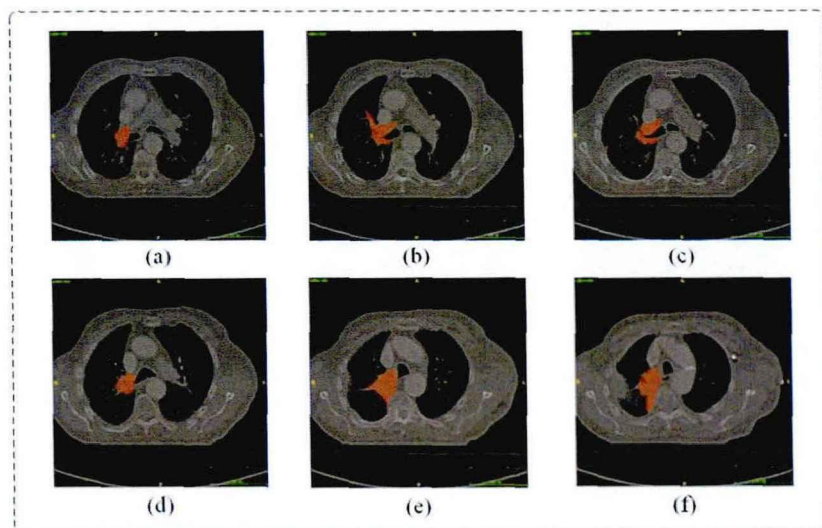


图 4.4 同一 CT 影像的多个切片图

此外,相较于第三章所采用的数据预处理策略,本研究在方法上有所优化。主要因为 CT 成像是沿身体顺序一层一层进行扫描,因此一个肿瘤会在多个连续的扫描切片中出现,并因其形态特征及切片厚度差异而在图像上产生细微形状变化。如图 4.4 所示,由于肿瘤轮廓不规则,在连续切片上其外观表现出动态变化。实际上,仅依赖单一 CT 切片,许多非肿瘤结构可能因形态相似而被误判为肿瘤组织。即使是经验丰富的放射科医师,也会系统地分析一系列轴向视图切片,以捕捉其间存在的特征差异,从而排除非肿瘤候选目标。因此,本研究借鉴临床实践,针对一组 CT 轴向切片进行探索并提取特征表达,使分类网络能够模拟放射科医生的诊断过程。通过整合多个连续切片信息,不仅能够获取器官在三维空间上的深度特征,还能充分考虑每个切片内像素间的空间相关性。基于这一策略,

本研究摒弃了为匹配肿瘤尺寸而对原始图像进行裁剪的做法，转而采取将完整肺部图像优化调整至 384×384 大小，以保留全肺影像的全局上下文信息。此方法避免了人工勾画肿瘤区域的需求，使得模型训练更为高效且贴近临床实际。

本研究构建的图像—文本对是由综合数据集上的 CT 数据及相应的影像报告构成。如图 4.5 所示，影像报告详尽记录了由专业放射科医师基于影像中观察到的多种医学征象，诸如 **pleural retraction**（胸膜牵拉）、**ground-glass opacity**（磨玻璃影）和 **cavitation**（空泡）等，为模型识别并解释图像中的相关特点提供了宝贵的标注信息。每份报告的篇幅介于 30 至 50 个单词之间，通常由多个句子组成，每个句子集中描述一个或多个解剖病理方面的特征，并与成对影像中的特定区域高度关联。在预处理阶段，本研究剔除了少于 3 个单词的句子，从而保证内容的有效性^[123]。

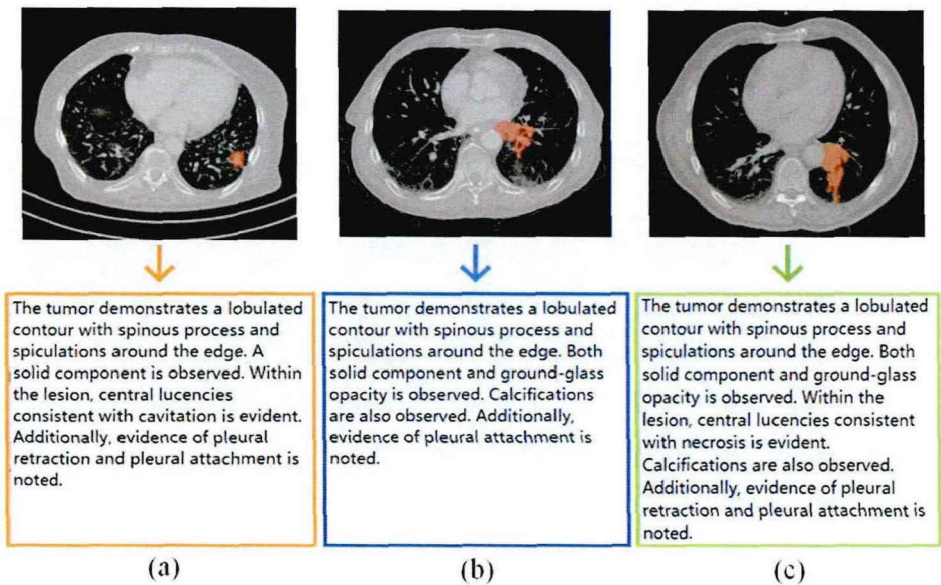


图 4.5 影像与对应的影像报告示例图

为了对所提方法的稳健性和泛化性能进行全面的评估，本研究采用了严格的五折交叉验证策略。这一策略旨在最大程度减少过拟合风险，保证模型在不同子集上的训练及测试结果具备高度稳定性，从而反映出方法在未知数据上的预测效果，体现其泛化能力。实验环境的搭建与配置严格遵循了第三章所述的标准和规范，确保了实验条件的一致性。同时，在模型性能的评价指标选取上，同样沿用了第三章中选定的评价指标，包括 AUC、准确率以及 F1 分数等，以全方位衡量和展现所提方法的实际效能。在模型参数设定方面，本研究部分延续了第三章中的超参数配置，并针对 CMCL-MVD 方法的独特特性与应用场景对部分超参数进行了针对性的调整与优化，这些差异化的参数设定在表 4.2 中进行了详细的列举。

表 4.2 CMCL-MVD 的超参数设置

Hyperparameter	Value
Learning rate	0.0001
Learning rate decay	0.5 / 20 epoch
Epoch	100
Batch size	8
Temperature parameter τ	0.07

4.4 性能评估与比较

4.4.1 模块的消融实验

为了深入探究 AD 模块和 MSF 模块在 CMCL-MVD 模型中的有效性, 本研究开展了一组消融实验。本次实验评估了几种不同的 CMCL-MVD 模型变体: (i) **Baseline**: 未集成解耦和融合机制的基线模型; (ii) **+AD** 变体: 在 **Baseline** 基础架构上进行扩展, 引入基于混合注意力机制的 AD 模块以实现特征解耦; (iii) **+MSF** 变体: 同样在 **Baseline** 结构上进行改进, 采用基于多实例池化的 MSF 模块以增强特征融合; (iv) **+AD+MSF** 变体: 此模型在 **Baseline** 基础上联合应用 AD 模块和 MSF 模块。

实验结果显示于表 4.3 中, 嵌入 AD 模块的模型变体相较于原始 **Baseline** 模型取得了显著的性能提升, 这一结论有力证明了在组织学亚型分类任务中, 基于注意力机制的解耦策略能够有效聚焦并强化对于识别肺癌亚型至关重要的公共特征表达。此外, 当在 **Baseline** 模型上添加 MSF 模块后, 模型对非小细胞肺癌亚型的区分准确率有所提高, 揭示了基于多实例池化的融合策略相比固定的特征均值操作能更有效地整合多切片特征信息。这是因为, 多实例池化具备可训练和

表 4.3 不同模块的对比结果

Methods	AUC	Acc	Sen	Spe	F1
Baseline	0.806±0.032	0.788±0.017	0.794±0.070	0.784±0.048	0.773±0.028
+AD	0.816±0.058	0.793±0.036	0.812±0.034	0.777±0.039	0.783±0.035
+MSF	0.814±0.057	0.796±0.037	0.819±0.036	0.777±0.039	0.787±0.037
+AD+MIL	0.825±0.036	0.803±0.031	0.813±0.034	0.795±0.031	0.790±0.032

自适应特性,能够通过对各实例(切片)权重的学习,动态构建出更具辨别力和综合性强的整体特征表达,从而获取更优的结果。尤为重要的是,同时整合混合注意力机制和多切片融合模块的变体模型相比于仅采用单一 AD 模块或 MSF 模块的变体展示出更为优越的性能表现。这一现象表明,通过协同优化训练两个模块,可以敏锐捕获并有效处理不同切片间的异质性和关联性,进而高效凝聚并提炼出与亚型高度相关的公共特征,从而提升模型在亚型分类任务上的精确度和泛化性能。

4.4.2 损失的有效性分析

为了深入探究并验证提出的跨亚型鉴别损失与语义对齐损失在提升 CMCL-MVD 分类性能上的有效性和优越性,本研究进行了单一损失和两种损失组合的消融实验。在实验设置中,“Baseline”表示仅利用基础分类损失的解耦模型;

“ $+L_{CSD}$ ”指在 Baseline 基础上引入额外的跨亚型鉴别损失;“ $+L_{SA}$ ”指在 Baseline 解耦架构中融合了语义对齐损失。表 4.4 详细呈现了实验结果,有力证明相较于朴素的 Baseline 解耦模型,加入跨亚型鉴别损失 L_{CSD} 能显著提升模型性能,主要原因在于该损失函数能够有效地抑制干扰特征,从而促使模型更加精准地识别并提炼各类亚型的核心特征。与此同时,当引入语义对齐损失 L_{SA} 后,所有评估指标均呈现出显著的性能跃升,这一现象揭示了语义对齐损失在深化模型对图像与文本间内在语义联系的认知和建模能力方面的重要价值。具体表现为模型不仅能于视觉空间内精确定位目标实体,还能在跨模态领域实现深层次的语义互动,进而使得图像特征与文本描述间的对应关系更为精确、紧密。实验数据显示,采用跨亚型鉴别损失与语义对齐损失相结合的双损失优化策略,在各项评估指标上均明显优于单一损失应用的情况,这充分验证了这两种损失函数在挖掘最优多视图公共表征层面所展现的有效互补性和协同效应。

进一步,为了全面探究跨亚型鉴别损失与语义对齐损失的权衡因子 $\{\alpha, \beta\}$ 对 CMCL-MVD 性能的影响,本研究为两个变量 α 和 β 在预设的相同参数范围

表 4.4 不同损失的对比结果

Methods	AUC	Acc	Sen	Spe	F1
Baseline	0.825±0.036	0.803±0.031	0.813±0.034	0.795±0.031	0.790±0.032
$+L_{CSD}$	0.833±0.027	0.811±0.014	0.812±0.013	0.810±0.025	0.798±0.012
$+L_{SA}$	0.837±0.033	0.806±0.015	0.818±0.013	0.795±0.035	0.794±0.011
$+L_{CSD} + L_{SA}$	0.847±0.007	0.817±0.011	0.825±0.015	0.811±0.011	0.805±0.012

{0.2, 0.4, ..., 1.0}内进行了详尽的实验评估。从图 4.6 所示的性能趋势分析图中，可以观察到随着两个权衡因子的连续调整，尽管模型性能呈现一定的波动性，但整体上展现出强大的稳健性和卓越的泛化能力。尤其是在找到最优权衡点为 $\alpha = 0.8$ 和 $\beta = 0.8$ 时，CMCL-MVD 模型在 AUC、准确率和灵敏度等核心评价指标上均达到了峰值，这有力地验证了该模型在处理复杂亚型分类任务时所具备的高度精确性和高效性。此外，值得注意的是，无论权衡因子处于何种取值区间，CMCL-MVD 模型始终保持其性能优势，能在各种参数配置下实现有效的工作性能。这一结果揭示出，即使面临操作条件及环境变化带来的挑战，CMCL-MVD 模型仍能展现极高的适应性和可靠性，从而确保其在不同场景下的稳定表现和应用价值。

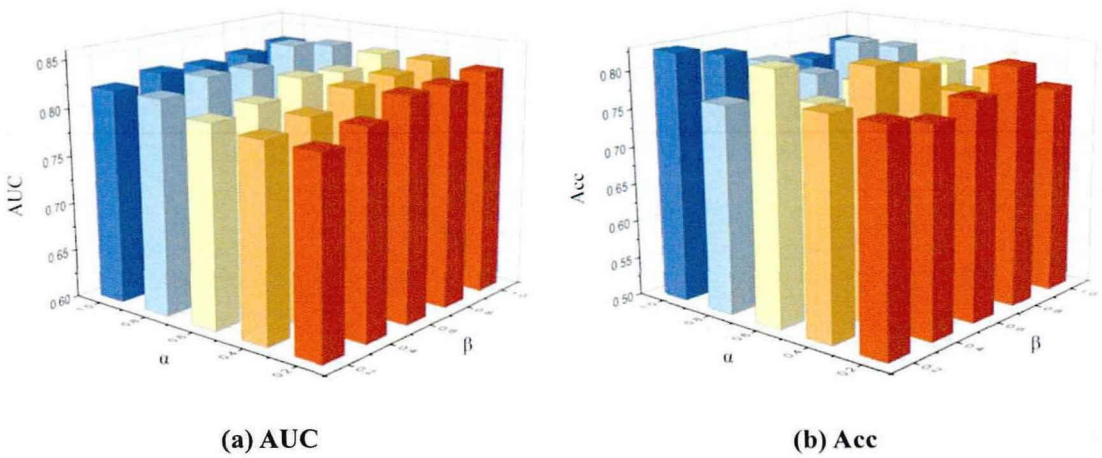


图 4.6 $\{\alpha, \beta\}$ 不同组合时的性能

4.4.3 预测分数的比较分析

本研究对 CMCL-MVD 模型所采用的亚型分类器产生的预测分数进行了深度分析，进一步验证该模型对于非小细胞肺癌不同亚型之间实现有效区分的能力。如图 4.7 所示，亚型分类器生成的预测分数在鳞癌（SCC）和腺癌（ADC）亚型群体间的分布呈现出显著差异性，这一结论得到了统计学检验结果的支持，其 p 值小于 0.001，远低于常规显著性水平阈值 0.05，充分证明了观察到的差异并非偶然。深入剖析数据分布特性，可以观察到腺癌样本的预测分数中位数在 0.75 左右，说明大多数腺癌样本的得分集中在一个较高的水平上。相反，鳞癌样本的预测分数中位数较低，且四分位间距较窄，箱体长度短，这直观地反映出鳞癌样本内在变异性相对较小，预测分数分布紧凑且集中，进一步证明模型在鳞癌样本上的分类效能更加卓越，能够稳健而精确地刻画其内在肿瘤表征。以上证据证实了 CMCL-MVD 模型的独特优越性，即其成功学习到了富含生物学语义信息的公共表征，能从复杂的肿瘤影像数据中提取出关键而独特的差异化特征，精准地捕

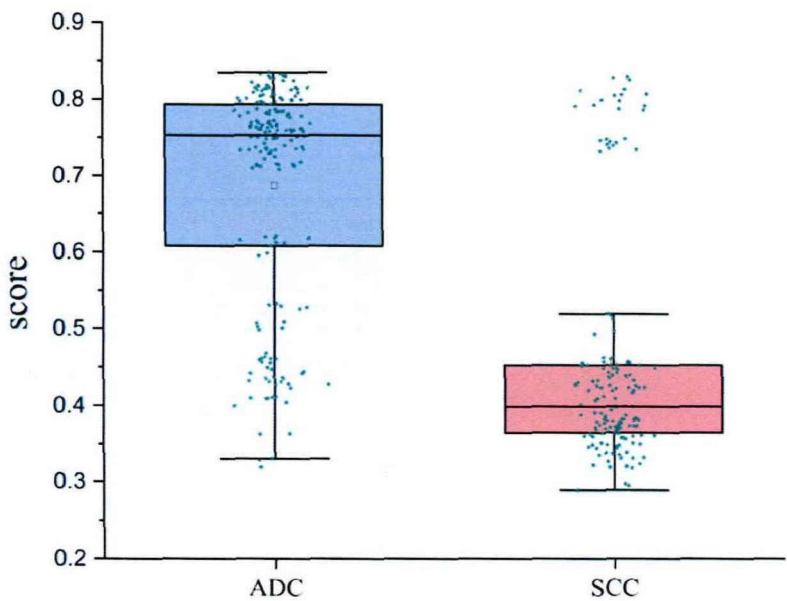


图 4.7 亚型分类器的预测得分

捉到鳞癌与腺癌这两种非小细胞肺癌亚型间微妙却又至关重要的病理异质性。综上所述，CMCL-MVD 模型凭借其先进的深度学习性能，有效地实现了对肿瘤组织图像的多层次深度解析，提取出了能够鉴别和识别非小细胞肺癌不同亚型的多视图表征。

4.4.4 学习率调节实验

在对学习率对 CMCL-MVD 分类性能的深度探究中，本研究精心设计并执行了一系列调节实验。具体而言，本研究选取了五个具有代表性的初始学习率值进行系统性研究，即 $1e-5$ 、 $5e-5$ 、 $1e-4$ 、 $2e-4$ 以及 $3e-4$ 。图 4.8 直观地展示了在不同学习率下训练 CMCL-MVD 时对应的 AUC、Acc 性能曲线，清晰地揭示出一种趋势规律：随着学习率的逐步增大，模型的性能表现经历了先增后减的过程，在学习率设定为 $1e-4$ 时达到了最优峰值。这一观察结果强有力地验证了在运用 CMCL-MVD 进行非小细胞肺癌亚型分类任务训练的过程中，学习率的选择展现出显著的敏感性和优化阈值特性。学习率过小可能导致模型在训练过程中容易陷入局部最优解，而难以跳出局部最优解继续寻找全局最优解，使得模型性能无法达到最优水平。而学习率过大也可能导致模型参数的更新步长过大，可能导致模型在参数空间中来回震荡，无法稳定地收敛到最优解。因此过低或过高的学习率配置均可能对模型的分类效能产生不利影响。基于上述严谨的实验数据分析和最优结果推断，本研究决定将初始学习率设定为 $1e-4$ ，在 CMCL-MVD 训练初期以此实现模型分类性能的最大化优化。

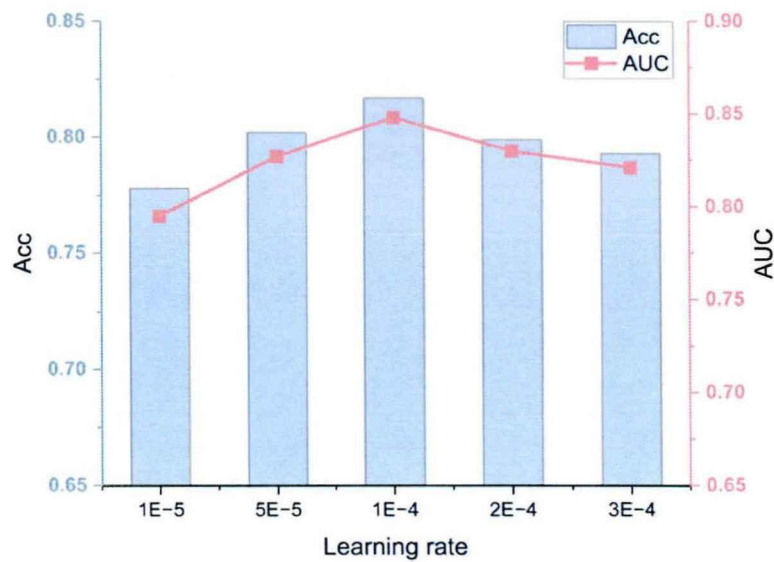


图 4.8 不同学习率下的 CMCL-MVD 性能

4. 4. 5 与现有方法的性能比较

为了深入探究 CMCL-MVD 方法在非小细胞肺癌组织学亚型分类任务上的性能表现,本研究将其在综合数据集上与多种先进的多视图分类方法进行了细致的对比与分析,其中包括 Su 等^[85]、Feng 等^[86]、Wu 等^[41]、Gao 等^[50]以及 C. Li 等^[87]的工作,并同步对比了第三章中的 MVD 模型。表 4.5 展示了上述方法在综合数据集上实施五折交叉验证的实验结果。通过对比可以明显看出,相较于 Su 等、Feng 等、Wu 等、Gao 等以及 C. Li 等的研究方案, MVD 和 CMCL-MVD 在准确率上至少提高了 2.3%和 6%,有力地证明了解耦技术在医学影像分析领域的优越

表 4.5 综合数据集上各方法的对比结果

Methods	AUC	Acc	Sen	Spe	F1
Su et al. ^[85]	0.793±0.024	0.754±0.033	0.756±0.087	0.752±0.031	0.738±0.011
Feng et al. ^[86]	0.784±0.023	0.745±0.023	0.769±0.025	0.725±0.052	0.735±0.020
Wu et al. ^[41]	0.808±0.018	0.757±0.034	0.775±0.031	0.742±0.045	0.746±0.020
Gao et al. ^[50]	0.812±0.004	0.757±0.016	0.769±0.025	0.747±0.027	0.743±0.016
C. Li et al. ^[87]	0.793±0.016	0.745±0.040	0.763±0.032	0.730±0.046	0.734±0.029
MVD (Ours)	0.836±0.012	0.780±0.019	0.794±0.025	0.768±0.030	0.767±0.019
CMCL-MVD (Ours)	0.847±0.007	0.817±0.011	0.825±0.015	0.811±0.011	0.805±0.012

性和有效性。进一步探究发现,CMCL-MVD 在 MVD 架构基础上融合文本信息,并实现图像特征与文本信息的深层次语义对齐后,各个评价指标均有明显的优化。这意味着,通过引入影像报告所蕴含的丰富文本语义知识,模型能够有效地指导和强化对复杂医学影像特征的理解与学习,从而显著增强了模型在复杂医学影像识别中的鲁棒性。

除上述指标外,图 4.9 以 ROC 曲线的形式对比了各类方法的整体表现,进一步揭示了各模型在不同决策阈值下的性能差异。从图中清晰可见,CMCL-MVD 的 ROC 曲线相较于其他对比方法更接近理想点(左上角),直观地体现了 CMCL-MVD 在亚型分类任务上的卓越优势,其 AUC 值也明显优于其他方法。综上所述,通过将预训练跨模态模型与 MVD 结构相融合,并巧妙利用影像报告内的文本信息对图像特征进行深度语义对齐,CMCL-MVD 模型成功在核心评价指标上取得进步,证实了集成文本信息以及解耦技术在医学影像分析领域的巨大潜力和实用价值。

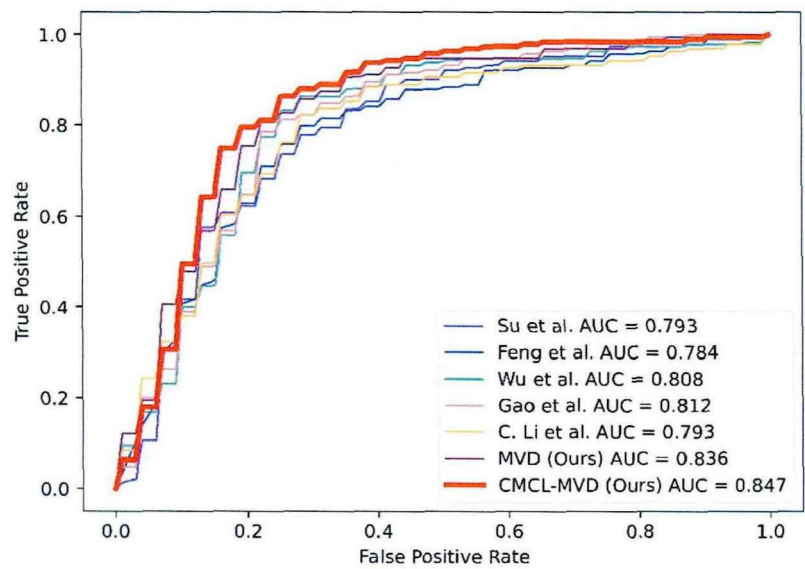


图 4.9 综合数据集上各方法的 ROC 曲线

4.4.6 可视化分析

为了直观地展示 CMCL-MVD 架构从医学影像数据中抽取富含语义信息特征的能力,本研究采用 t 分布随机近邻嵌入 (t-Distributed Stochastic Neighbor Embedding, t-SNE) 技术^[124]对高维特征空间中的特征进行降维处理,并将其可视化在二维图谱上。图像特征(绿色圆点)与相应文本特征(粉色圆点)的邻近程度越高,表明所学习到的多视图影像特征含有的语义信息越多。如图 4.10 所

示,观察到在图(a)中,影像特征与文本特征在二维投射上有部分紧凑的聚类现象,这验证了 CLIP 预训练阶段的图像和文本编码器已具备一定的语义关联性,从而突显了 CLIP 模型的有效性。而在图(b)中,经过 CMCL-MVD 微调后的特征映射,医学影像特征与其相应的文本特征整体上呈现出更高的紧凑性和一致性,实现了更为精确的跨模态对齐。这一现象有力证明了,通过运用跨模态对比学习策略,CMCL-MVD 在提取具有丰富语义内涵特征方面的优越性能,同时也为 CMCL-MVD 在亚型分类任务中取得较高准确率提供了合理的解释。

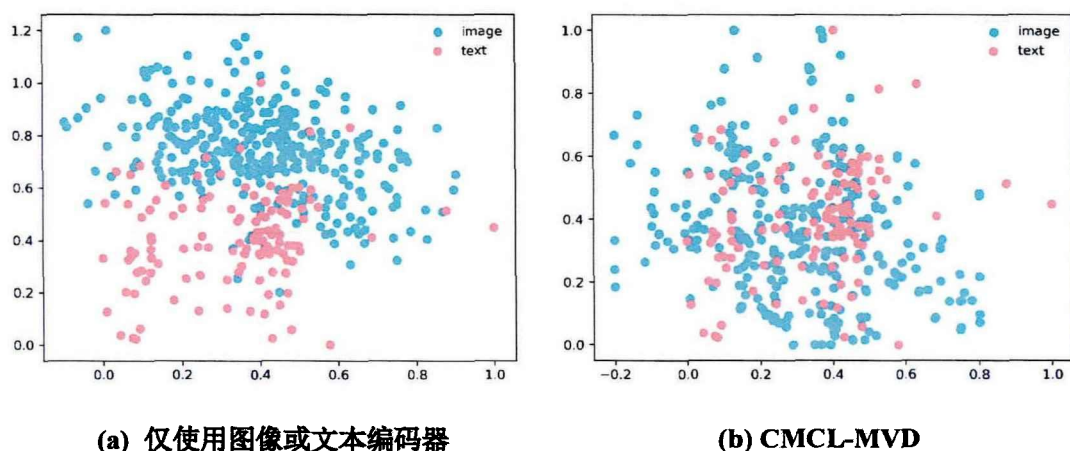


图 4.10 不同方法学习的图像—文本表征的 t-SNE 可视化

4.5 本章总结

本章节创新性地提出了一种名为 CMCL-MVD 的多视图解耦方法,巧妙地整合了医学影像数据和影像报告,以实现肺癌 CT 影像更为深入的理解。具体来说,CMCL-MVD 首先基于混合注意力机制的解耦模块来实现对各视图潜在表征的有效分解,使得模型能够同时提炼出反映多个视图的公共表征以及体现独特属性的特异特征。针对所研究的多切片场景,本研究设计了一种基于多实例池化的多切片融合策略,此策略能够动态优化并调整各个切片的权重分配,从而生成具备高度辨别能力的综合多视图特征,进一步提升了分类任务的准确性。此外,引入语义对齐损失函数,以确保图像特征域与文本特征域之间的语义一致性,从而引导和强化了图像亚型特征的学习过程。大量的实验证据表明,在涵盖广泛病例的综合性数据集上,CMCL-MVD 相较于当前最先进的同类技术展现出性能提升优势,彰显其技术先进性和应用潜力。

第5章 总结与展望

5.1 研究总结

肺癌是全球癌症死亡的首要原因，其5年生存率不足15%。根据临床表现和组织学特征，肺癌主要分为小细胞肺癌和非小细胞肺癌，其中非小细胞肺癌占比高达85%。在精准医疗领域内，精确鉴别非小细胞肺癌的不同组织学亚型是构建患者个体化治疗策略的基础，并有望大幅提升患者的预后效果及整体生存率，从而具有显著的科学价值和迫切的社会需求。当前，基于CT影像分析的无创性诊断技术已成为获取肿瘤组织病理学信息的重要手段，它有效弥补了传统分子生物学方法存在的侵入性高、耗时长、风险高等固有缺陷。其中，深度学习技术因具备强大的自动化特征提取与分类能力，在非小细胞肺癌亚型分类领域得到了广泛应用。然而，大部分现有研究仅利用CT轴向视图进行模型训练，未能充分利用肿瘤的三维结构信息，导致难以提取形态各异的肿瘤的鉴别信息。因此，本研究整合来自不同视图的信息，利用多视图深度学习方法来辅助医生快速准确识别非小细胞肺癌亚型。本文主要包括以下内容：

(1) 首先阐述了对非小细胞肺癌亚型进行无创分类的重要意义，系统梳理了国际和国内相关领域的研究成果，并分析了当前研究存在的局限性。随后，简述了本研究所使用的公共数据集（源自NSCLC-Radiomics、NSCLC-Radiogenomics数据库）及自建数据集（源于安徽医科大学第一附属医院影像科）的基本情况，并介绍了数据筛选和预处理工作的具体细节，以为后续实验提供可靠的数据支持。

(2) 为了有效应对多视图数据集中普遍存在的视图间差异和视图内干扰问题，本研究创新设计了一种基于多视图解耦学习的非小细胞肺癌亚型分类方法MVD。具体来说，MVD采用基于注意力机制的解耦模块，将不同视图投影到一个视图不变子空间及多个视图特定子空间中，从而生成公共表征和特异表征。针对视图间差异问题，本研究提出了跨视图转换损失函数，确保公共表征在视图不变子空间中的分布一致。同时，通过引入对抗性的跨亚型鉴别损失，保证特异表征仅捕获与亚型无关的信息，从而有效地消除了视图内的干扰问题。在公共数据集和自建数据集上，MVD方法经过严格评估验证。实验结果表明，MVD在区分非小细胞肺癌亚型方面具有卓越的诊断准确性，并在各种性能指标上始终优于其他先进模型。

(3) 在此基础上，本研究进一步提出了基于跨模态对比学习的多视图肺癌

亚型分类方法 CMCL-MVD。该方法巧妙地融合了医学图像影像报告中的语义知识,引导多视图解耦框架对影像内容进行更精细、深入的理解。该模型首先利用基于混合注意力机制的解耦模块,对各视图潜在表征进行有效分解和独立提取。同时,本研究还引入了语义对齐损失,确保从 CT 图像中提取的肿瘤信息与文本特征域中的描述之间达到高度的语义一致性,从而使得模型能够更好的提取富含语义的视觉特征。在综合数据集上,CMCL-MVD 相较于当前最先进的同类技术展现出了显著的性能提升优势。

5.2 研究展望

尽管本研究提出的多视图深度学习方法在非小细胞肺癌亚型分类任务中取得了当前最优性能,推动了医学影像深度学习模型的构建与优化,但该工作尚存在几方面局限性。

首先,尽管当前模型在非小细胞肺癌亚型分类领域取得了进展,然而其对高质量标注数据具有依赖性,限制了模型对大量未标记数据资源的有效利用。此类高质量标注数据获取过程往往伴随着高昂的人力投入、时间消耗及经济成本,从而制约了模型的扩展与广泛应用。为了缓解当前模型对高质量标注数据的依赖,未来研究将着力探究无监督学习与弱监督学习这两种策略。通过运用这些前沿技术手段,有望使模型在应对实际医疗场景中疾病表型的复杂多变性时,展现出更强的自适应能力和泛化性能,从而提升其在非小细胞肺癌亚型分类任务上的精确度与效率。

其次,第四章中提出的跨模态对比学习方法主要依托基于自然图像—文本数据集预训练的 CLIP 模型,这可能对其在医学临床实践中的表现构成一定限制。已有研究表明,针对特定目标领域进行预训练的 CLIP 模型,能够有效地适应各类医学影像场景,凸显了特定领域预训练的重要性。尤其在疾病之间需进行细粒度区分的情况下,专门预训练的 CLIP 模型展现出优于通用模型的性能。因此,期望未来的研究能够积累更多医学影像和报告的成对数据,研发更为复杂的 CLIP 式预训练方法,以解决医学领域泛化性不够的问题。

最后,本研究仅运用了影像报告作为先验知识,而未充分利用患者的基本信息。然而,患者的基本信息通常包含了丰富且与视觉形态学密切相关的信息,如年龄等属性。已有研究表明,这些常见属性对于脑组织分割等任务具有重要意义。因此,未来的研究可以探索将患者的基本信息整合到模型中的方法,将患者个体特征纳入模型建模框架。通过这种方式,能够增强模型对数据的理解能力,从而实现更准确、可靠和可解释的医学图像分析结果,推动该领域的深度发展。

参 考 文 献

- [1] Sung H, Ferlay J, Siegel R L, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries[J]. CA: A Cancer Journal for Clinicians, 2021, 71(3): 209-249.
- [2] Liu H, Jiao Z, Han W, et al. Identifying the histologic subtypes of non-small cell lung cancer with computed tomography imaging: A comparative study of capsule net, convolutional neural network, and radiomics[J]. Quantitative Imaging in Medicine and Surgery, 2021, 11(6): 2756-2765.
- [3] She J, Yang P, Hong Q, et al. Lung cancer in China: Challenges and interventions[J]. CHEST, 2013, 143(4): 1117-1126.
- [4] Nie Y, Cai Z, Zhao S. Early lung cancer baseline screening: Preliminary study with low-dose spiral CT[J]. Chinese Journal of Radiology, 2002, 36(3):230-234.
- [5] Chen P, Liu Y, Wen Y, et al. Non-small cell lung cancer in China[J]. Cancer Communications, 2022, 42(10): 937-970.
- [6] Moitra D, Kr. Mandal R. Classification of non-small cell lung cancer using one-dimensional convolutional neural network[J]. Expert Systems with Applications, 2020, 159: 113564.
- [7] Dwivedi K, Rajpal A, Rajpal S, et al. An explainable AI-driven biomarker discovery framework for Non-Small Cell Lung Cancer classification[J]. Computers in Biology and Medicine, 2023, 153(20):106544.
- [8] Chen K, Wang M, Song Z. Multi-task learning-based histologic subtype classification of non-small cell lung cancer[J]. La radiologia medica, 2023, 128(5): 537-543.
- [9] Zhu X, Dong D, Chen Z, et al. Radiomic signature as a diagnostic factor for histologic subtype classification of non-small cell lung cancer[J]. European Radiology, 2018, 28(7): 2772-2778.
- [10] Howlader N, Forjaz G, Mooradian M J, et al. The effect of advances in lung-cancer treatment on population mortality[J]. New England Journal of Medicine, 2020, 383(7): 640-649.
- [11] Duma N, Santana-Davila R, Molina J R. Non-small cell lung cancer: Epidemiology, screening, diagnosis, and treatment[J]. Mayo Clinic Proceedings, 2019, 94(8): 1623-1640.
- [12] Yang H, Chen L, Cheng Z, et al. Deep learning-based six-type classifier for lung cancer and mimics from histopathological whole slide images: A retrospective study[J]. BMC Medicine, 2021, 19(1): 80.
- [13] Kanavati F, Toyokawa G, Momosaki S, et al. A deep learning model for the classification of

- indeterminate lung carcinoma in biopsy whole slide images[J]. *Scientific Reports*, 2021, 11(1): 8110.
- [14] Xiao H, Liu Q, Li L. MFMANet: Multi-feature multi-attention network for efficient subtype classification on non-small cell lung cancer CT images[J]. *Biomedical Signal Processing and Control*, 2023, 84(3): 104768.
- [15] Wu Y, Ma J, Huang X, et al. DeepMMSA: A novel multimodal deep learning method for non-small cell lung cancer survival analysis[C]. *IEEE International Conference on Systems, Man, and Cybernetics*. 2021: 1468-1472.
- [16] Lin J, Yu Y, Zhang X, et al. Classification of histological types and stages in non-small cell lung cancer using radiomic features based on CT images[J]. *Journal of Digital Imaging*, 2023, 36(3): 1029-1037.
- [17] Bashir U, Kawa B, Siddique M, et al. Non-invasive classification of non-small cell lung cancer: A comparison between random forest models utilising radiomic and semantic features[J]. *The British Journal of Radiology*, 2019, 92(1099): 20190159.
- [18] Chaunzwa T L, Hosny A, Xu Y, et al. Deep learning classification of lung cancer histology using CT images[J]. *Scientific Reports*, 2021, 11(1): 5471.
- [19] Xu Z, Ren H, Zhou W, et al. ISANET: Non-small cell lung cancer classification and detection based on CNN and attention mechanism[J]. *Biomedical Signal Processing and Control*, 2022, 77: 103773.
- [20] Liu J, Cui J, Liu F, et al. Multi-subtype classification model for non-small cell lung cancer based on radiomics: SLS model[J]. *Medical Physics*, 2019, 46(7): 3091-3100.
- [21] Pang S, Meng F, Wang X, et al. VGG16-T: A novel deep convolutional neural network with boosting to identify pathological type of lung cancer in early stage by CT images[J]. *International Journal of Computational Intelligence Systems*, 2020, 13(1): 771-780.
- [22] Marentakis P, Karaikos P, Kouloulis V, et al. Lung cancer histology classification from CT images based on radiomics and deep learning models[J]. *Medical & Biological Engineering & Computing*, 2021, 59(1): 215-226.
- [23] Chen C H, Chang C K, Tu C Y, et al. Radiomic features analysis in computed tomography images of lung nodule classification[J]. *PLOS ONE*, 2018, 13(2): e0192002.
- [24] Song F, Song X, Feng Y, et al. Radiomics feature analysis and model research for predicting histopathological subtypes of non-small cell lung cancer on CT images: A multi-dataset study[J]. *Medical Physics*, 2023, 50(7): 4351-4365.
- [25] Khodabakhshi Z, Mostafaei S, Arabi H, et al. Non-small cell lung carcinoma histopathological subtype phenotyping using high-dimensional multinomial multiclass CT

- radiomics signature[J]. Computers in Biology and Medicine, 2021, 136: 104752.
- [26] D'Arnese E, Donato G W D, Sozzo E D, et al. On the automation of radiomics-based identification and characterization of NSCLC[J]. IEEE Journal of Biomedical and Health Informatics, 2022, 26(6): 2670-2679.
- [27] Li H, Song Q, Gui D, et al. Reconstruction-assisted feature encoding network for histologic subtype classification of non-small cell lung cancer[J]. IEEE Journal of Biomedical and Health Informatics, 2022, 26(9): 4563-4574.
- [28] Domingues I, Pereira G, Martins P, et al. Using deep learning techniques in medical imaging: A systematic review of applications on CT and PET[J]. Artificial Intelligence Review, 2020, 53(6): 4093-4160.
- [29] Liu Z, Wei J, Li R, et al. Learning multi-modal brain tumor segmentation from privileged semi-paired MRI images with curriculum disentanglement learning[J]. Computers in Biology and Medicine, 2023, 159(6): 106927.
- [30] Yu X, Jin F, Luo H, et al. Gross tumor volume segmentation for stage III NSCLC radiotherapy using 3D ResSE-Unet[J]. Technology in Cancer Research & Treatment, 2022, 21(5):1-12.
- [31] Wang C, Xu X, Shao J, et al. Deep learning to predict EGFR mutation and PD-L1 expression status in non-small-cell lung cancer on computed tomography images[J]. Journal of Oncology, 2021, 2021(14): 1-11.
- [32] Gui D, Song Q, Song B, et al. AIR-Net: A novel multi-task learning method with auxiliary image reconstruction for predicting EGFR mutation status on CT images of NSCLC patients[J]. Computers in Biology and Medicine, 2022, 141: 105157.
- [33] Han Y, Ma Y, Wu Z, et al. Histologic subtype classification of non-small cell lung cancer using PET/CT images[J]. European Journal of Nuclear Medicine and Molecular Imaging, 2021, 48(2): 350-360.
- [34] Wu R, Huang H. Multi-scale multi-view model based on ensemble attention for benign-malignant lung nodule classification on chest CT[C]. International Congress on Image and Signal Processing, BioMedical Engineering and Informatics. 2022: 1-6.
- [35] Tomassini S, Falcionelli N, Sernani P, et al. Cloud-YLung for non-small cell lung cancer histology classification from 3D computed tomography whole-lung scans[C]. Annual International Conference of the IEEE Engineering in Medicine & Biology Society. 2022: 1556-1560.
- [36] Guo Y, Song Q, Jiang M, et al. Histological subtypes classification of lung cancers on CT images using 3D deep learning and radiomics[J]. Academic Radiology, 2021, 28(9): e258-

- e266.
- [37] Yanagawa M, Niioka H, Kusumoto M, et al. Diagnostic performance for pulmonary adenocarcinoma on CT: Comparison of radiologists with and without three-dimensional convolutional neural network[J]. *European Radiology*, 2021, 31(4): 1978-1986.
 - [38] Dong X, Xu S, Liu Y, et al. Multi-view secondary input collaborative deep learning for lung nodule 3D segmentation[J]. *Cancer Imaging*, 2020, 20(1): 1-13.
 - [39] Zhai P, Cong H, Zhu E, et al. MVCNet: Multiview contrastive network for unsupervised representation learning for 3-D CT lesions[J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2022, 99: 1-15.
 - [40] Li Z, Zhao S, Chen Y, et al. A deep-learning-based framework for severity assessment of COVID-19 with CT images[J]. *Expert Systems with Applications*, 2021, 185(8): 115616.
 - [41] Wu X, Hui H, Niu M, et al. Deep learning-based multi-view fusion model for screening 2019 novel coronavirus pneumonia: A multicentre study[J]. *European Journal of Radiology*, 2020, 128: 109041.
 - [42] Li R, Xiao C, Huang Y, et al. Deep learning applications in computed tomography images for pulmonary nodule detection and diagnosis: A review[J]. *Diagnostics*, 2022, 12(2): 298.
 - [43] Gao X, Ruan S, Shi J, et al. Multi-view attention learning for residual disease prediction of ovarian cancer[C]. *IEEE International Conference on Systems, Man, and Cybernetics*. 2023: 653-658.
 - [44] El-Regaily S A, Salem M A M, Abdel Aziz M H, et al. Multi-view convolutional neural network for lung nodule false positive reduction[J]. *Expert Systems with Applications*, 2020, 162: 113017.
 - [45] Yang Y, Li X, Fu J, et al. 3D multi-view squeeze-and-excitation convolutional neural network for lung nodule classification[J]. *Medical Physics*, 2023, 50(3): 1905-1916.
 - [46] Setio A A A, Ciompi F, Litjens G, et al. Pulmonary nodule detection in CT Images: False positive reduction using multi-view convolutional networks[J]. *IEEE Transactions on Medical Imaging*, 2016, 35(5): 1160-1169.
 - [47] Xie Y, Xia Y, Zhang J, et al. Knowledge-based collaborative deep learning for benign-malignant lung nodule classification on chest CT[J]. *IEEE Transactions on Medical Imaging*, 2019, 38(4): 991-1004.
 - [48] Zhai P, Tao Y, Chen H, et al. Multi-task learning for lung nodule classification on chest CT[J]. *IEEE Access*, 2020, 8(2): 180317-180327.
 - [49] Kang H, Xia L, Yan F, et al. Diagnosis of coronavirus disease 2019 (COVID-19) with structured latent multi-view representation learning[J]. *IEEE Transactions on Medical*

- Imaging, 2020, 39(8): 2606-2614.
- [50] Gao H, Wang M, Li H, et al. A multi-view feature decomposition deep learning method for lung cancer histology classification[C]. International Conference on Graphics and Image Processing. SPIE, 2023: 369-377.
 - [51] Clark K, Vendt B, Smith K, et al. The Cancer Imaging Archive (TCIA): Maintaining and operating a public information repository[J]. Journal of Digital Imaging, 2013, 26(6): 1045-1057.
 - [52] Aerts H J W L, Velazquez E R, Leijenaar R T H, et al. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach[J]. Nature Communications, 2014, 5(1): 4006.
 - [53] Bakr S, Gevaert O, Echegaray S, et al. A radiogenomic dataset of non-small cell lung cancer[J]. Scientific Data, 2018, 5(1): 1-9.
 - [54] Gevaert O, Xu J, Hoang C D, et al. Non-small cell lung cancer: Identifying prognostic imaging biomarkers by leveraging public gene expression microarray data—methods and preliminary results[J]. Radiology, 2012, 264(2): 387-396.
 - [55] Hosny A, Parmar C, Coroller T P, et al. Deep learning for lung cancer prognostication: A retrospective multi-cohort radiomics study[J]. PLOS Medicine, 2018, 15(11): e1002711.
 - [56] Dou Q, Chen H, Yu L, et al. Multilevel contextual 3-D CNNs for false positive reduction in pulmonary nodule detection[J]. IEEE Transactions on Biomedical Engineering, 2017, 64(7): 1558-1567.
 - [57] Wang D, Li M, Ben-Shlomo N, et al. Mixed-supervised dual-network for medical image segmentation[C]. International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, 2019: 192-200.
 - [58] Yan X, Hu S, Mao Y, et al. Deep multi-view learning methods: A review[J]. Neurocomputing, 2021, 448(1): 106-129.
 - [59] Sun S. A survey of multi-view machine learning[J]. Neural Computing and Applications, 2013, 23(7): 2031-2038.
 - [60] Li Y, Yang M, Zhang Z. A survey of multi-view representation learning[J]. IEEE Transactions on Knowledge and Data Engineering, 2019, 31(10): 1863-1883.
 - [61] Xu C, Tao D, Xu C. A survey on multi-view learning[J]. arXiv preprint arXiv:1304.5634, 2013.
 - [62] Zhao J, Xie X, Xu X, et al. Multi-view learning overview: Recent progress and new challenges[J]. Information Fusion, 2017, 38: 43-54.
 - [63] Bennett K P, Momma M, Embrechts M J. MARK: A boosting algorithm for heterogeneous

- kernel models[C]. International Conference on Knowledge Discovery and Data Mining. ACM, 2002: 24-31.
- [64] Varma M, Babu B R. More generality in efficient multiple kernel learning[C]. International Conference on Machine Learning. ACM, 2009: 1065-1072.
- [65] Guo C, Wu D. Canonical correlation analysis (CCA) based multi-view learning: An overview[J]. arXiv preprint arXiv:1907.01693, 2019.
- [66] Nibali A, He Z, Wollersheim D. Pulmonary nodule classification with deep residual networks[J]. International Journal of Computer Assisted Radiology and Surgery, 2017, 12(10): 1799-1808.
- [67] Geirhos R, Jacobsen J H, Michaelis C, et al. Shortcut learning in deep neural networks[J]. Nature Machine Intelligence, 2020, 2(11): 665-673.
- [68] Bengio Y, Courville A, Vincent P. Representation learning: A review and new perspectives[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2013, 35(8): 1798-1828.
- [69] Lake B M, Ullman T D, Tenenbaum J B, et al. Building machines that learn and think like people[J]. Behavioral and Brain Sciences, 2017, 40: e253.
- [70] Wang X, Chen H, Tang S, et al. Disentangled representation learning[J]. arXiv preprint arXiv:2211.11695, 2022.
- [71] Wen Z, Wang J, Rui, et al. A review of disentangled representation learning[J]. Acta Automatica Sinica, 2022, 48(2): 351-374.
- [72] Cheng J, Gao M, Liu J, et al. Multimodal disentangled variational autoencoder with game theoretic interpretability for glioma grading[J]. IEEE Journal of Biomedical and Health Informatics, 2022, 26(2): 673-684.
- [73] Yue H, Liu J, Li J, et al. MLDRL: Multi-loss disentangled representation learning for predicting esophageal cancer response to neoadjuvant chemoradiotherapy using longitudinal CT images[J]. Medical Image Analysis, 2022, 79(6):102423.
- [74] Chen R T Q, Li X, Grosse R B, et al. Isolating sources of disentanglement in variational autoencoders[J]. Advances in Neural Information Processing Systems, 2018, 31.
- [75] Yang M, Liu F, Chen Z, et al. CausalVAE: Disentangled representation learning via neural structural causal models[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition. IEEE, 2021: 9593-9602.
- [76] Lee J, Kim E, Lee J, et al. Learning debiased representation via disentangled feature augmentation[J]. Advances in Neural Information Processing Systems, 2021, 34: 25123-25133.
- [77] Hu D, Zhang H, Wu Z, et al. Disentangled-multimodal adversarial autoencoder: Application

- to infant age prediction with incomplete multimodal neuroimages[J]. *IEEE Transactions on Medical Imaging*, 2020, 39(12): 4137-4149.
- [78] Hazarika D, Zimmermann R, Poria S. MISA: Modality-invariant and -specific representations for multimodal sentiment analysis[C]. *International Conference on Multimedia*. ACM, 2020: 1122-1131.
- [79] Jin C, Yu H, Ke J, et al. Predicting treatment response from longitudinal images using multi-task deep learning[J]. *Nature Communications*, 2021, 12(1): 1851.
- [80] Cao K, Gong Q, Hong Y, et al. A unified computational framework for single-cell data integration with optimal transport[J]. *Nature Communications*, 2022, 13(1): 7419.
- [81] Kasinathan G, Jayakumar S, Gandomi A H, et al. Automated 3-D lung tumor detection and classification by an active contour model and CNN classifier[J]. *Expert Systems with Applications*, 2019, 134(5): 112-119.
- [82] Li X, Guo R, Lu J, et al. Causality-driven graph neural network for early diagnosis of pancreatic cancer in non-contrast computerized tomography[J]. *IEEE Transactions on Medical Imaging*, 2023, 42(6): 1656-1667.
- [83] Yang F, Chen W, Wei H, et al. Machine learning for histologic subtype classification of non-small cell lung cancer: A retrospective multicenter radiomics study[J]. *Frontiers in Oncology*, 2021, 10: 608598.
- [84] Van Griethuysen J J M, Fedorov A, Parmar C, et al. Computational radiomics system to decode the radiographic phenotype[J]. *Cancer Research*, 2017, 77(21): e104-e107.
- [85] Su H, Maji S, Kalogerakis E, et al. Multi-view convolutional neural networks for 3D shape recognition[C]. *International Conference on Computer Vision*. IEEE, 2015: 945-953.
- [86] Feng Y, Zhang Z, Zhao X, et al. GVCNN: Group-view convolutional neural networks for 3D shape recognition[C]. *IEEE/CVF Conference on Computer Vision and Pattern Recognition*. IEEE, 2018: 264-272.
- [87] Li C, Xu J, Liu Q, et al. Multi-view mammographic density classification by dilated and attention-guided residual learning[J]. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 2021, 18(3): 1003-1013.
- [88] Selvaraju R R, Cogswell M, Das A, et al. Grad-CAM: Visual explanations from deep networks via gradient-based localization[J]. *International Journal of Computer Vision*, 2020, 128(2): 336-359.
- [89] Wang X, Peng Y, Lu L, et al. TieNet: Text-image embedding network for common thorax disease classification and reporting in chest X-Rays[C]. *IEEE/CVF Conference on Computer Vision and Pattern Recognition*. IEEE, 2018: 9049-9058.

- [90] Zhang Z, Chen P, Sapkota M, et al. TandemNet: Distilling knowledge from medical images using diagnostic reports as optional semantic references[C]. International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, 2017, 10435: 320-328.
- [91] Boonn W W, Langlotz C P. Radiologist use of and perceived need for patient data access[J]. Journal of Digital Imaging: The official journal of the Society for Computer Applications in Radiology, 2009, 22(4): 357-362.
- [92] Huang S C, Pareek A, Seyyedi S, et al. Fusion of medical imaging and electronic health records using deep learning: A systematic review and implementation guidelines[J]. npj Digital Medicine, 2020, 3(1): 136.
- [93] Chauhan G, Liao R, Wells W, et al. Joint modeling of chest radiographs and radiology reports for pulmonary edema assessment[C]. International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, 2020, 12262: 529-539.
- [94] Yan K, Peng Y, Sandfort V, et al. Holistic and comprehensive annotation of clinically significant findings on diverse CT images: Learning from radiology reports and label ontology[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition. IEEE, 2019: 8515-8524.
- [95] Wang X, Peng Y, Lu L, et al. ChestX-ray8: Hospital-scale chest X-Ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition. IEEE, 2017: 2097-2106.
- [96] Liang G, Greenwell C, Zhang Y, et al. Contrastive cross-modal pre-training: A general strategy for small sample medical imaging[J]. IEEE Journal of Biomedical and Health Informatics, 2022, 26(4): 1640-1649.
- [97] Irvin J, Rajpurkar P, Ko M, et al. CheXpert: A large chest radiograph dataset with uncertainty labels and expert comparison[C]. AAAI Conference on Artificial Intelligence. AAAI, 2019, 33(01): 590-597.
- [98] Wang X, Xu Z, Tam L, et al. Self-supervised image-text pre-training with mixed data in chest X-rays[J]. arXiv preprint arXiv:2103.16022, 2021.
- [99] Grill J B, Strub F, Altché F, et al. Bootstrap your own latent: A new approach to self-supervised learning[J]. Advances in Neural Information Processing Systems, 2020, 33: 21271-21284.
- [100] He K, Fan H, Wu Y, et al. Momentum contrast for unsupervised visual representation learning[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition. IEEE,

2020: 9726-9735.

- [101] Chen T, Kornblith S, Norouzi M, et al. A simple framework for contrastive learning of visual representations[C]. International Conference on Machine Learning, 2020: 1597-1607.
- [102] Raghu M, Zhang C, Kleinberg J, et al. Transfusion: Understanding transfer learning for medical imaging[J]. Advances in Neural Information Processing Systems, 2019, 32.
- [103] Zhang Y, Jiang H, Miura Y, et al. Contrastive learning of medical visual representations from paired images and text[C]. Machine Learning for Healthcare Conference. PMLR, 2022: 2-25.
- [104] Radford A, Kim J W, Hallacy C, et al. Learning transferable visual models from natural language supervision[C]. International conference on machine learning. PMLR, 2021: 8748-8763.
- [105] Rizvi S A, Tang R, Jiang X, et al. Local contrastive learning for medical image recognition[C]. American Medical Informatics Association, 2023: 1236-1245.
- [106] Wu C, Zhang X, Zhang Y, et al. MedKLIP: Medical knowledge enhanced language-image pre-training for X-ray diagnosis[C]. IEEE/CVF International Conference on Computer Vision. IEEE, 2023: 21315-21326.
- [107] Li Y, Liang F, Zhao L, et al. Supervision exists everywhere: A data efficient contrastive language-image pre-training paradigm[J]. arXiv preprint arXiv:2110.05208, 2021.
- [108] Yan A, He Z, Lu X, et al. Weakly supervised contrastive learning for chest X-Ray report generation[J]. arXiv preprint arXiv:2109.12242, 2021.
- [109] Huang S C, Shen L, Lungren M P, et al. GLoRIA: A multimodal global-local representation learning framework for label-efficient medical image recognition[C]. IEEE/CVF International Conference on Computer Vision. IEEE, 2021: 3942-3951.
- [110] Seibold C, Reiß S, Sarfraz M S, et al. Breaking with fixed set pathology recognition through report-guided contrastive training[C]. International Conference on Medical Image Computing and Computer-Assisted Intervention, 2022: 690-700.
- [111] Eslami S, Meinel C, de Melo G. PubMedCLIP: How much does CLIP benefit visual question answering in the medical domain?[C]. Annual Meeting of the Association for Computational Linguistics. ACL, 2023: 1181-1193.
- [112] Dosovitskiy A, Beyer L, Kolesnikov A, et al. An image is worth 16x16 words: Transformers for image recognition at scale[J]. arXiv preprint arXiv:2010.11929, 2020.
- [113] Wang Q, Li B, Xiao T, et al. Learning deep transformer models for machine translation[J]. arXiv preprint arXiv:1906.01787, 2019.
- [114] Radford A, Wu J, Child R, et al. Language models are unsupervised multitask learners[J].

- OpenAI blog, 2019, 1(8): 9
- [115] Sennrich R, Haddow B, Birch A. Neural machine translation of rare words with subword units[J]. arXiv preprint arXiv:1508.07909, 2015.
- [116] Zhao Z, Liu Y, Wu H, et al. CLIP in medical imaging: A comprehensive survey[J]. arXiv preprint arXiv:2312.07353, 2023.
- [117] Tiu E, Talus E, Patel P, et al. Expert-level detection of pathologies from unannotated chest X-ray images via self-supervised learning[J]. Nature Biomedical Engineering, 2022, 6(12): 1399-1406.
- [118] Mishra A, Mittal R, Jestin C, et al. Improving zero-shot detection of low prevalence chest pathologies using domain pre-trained language models[J]. arXiv preprint arXiv:2306.08000, 2023.
- [119] Liu G, Wu J, Zhou Z H. Key instance detection in multi-instance learning[C]. Asian Conference on Machine Learning. PMLR, 2012: 253-268.
- [120] Han Z, Wei B, Hong Y, et al. Accurate screening of COVID-19 using attention-based deep 3D multiple instance learning[J]. IEEE Transactions on Medical Imaging, 2020, 39(8): 2584-2594.
- [121] Bhattacharjee K, Pant M, Zhang Y D, et al. Multiple instance learning with genetic pooling for medical data analysis[J]. Pattern Recognition Letters, 2020, 133: 247-255.
- [122] Dauphin Y N, Fan A, Auli M, et al. Language modeling with gated convolutional networks[C]. International Conference on Machine Learning. PMLR, 2017: 933-941.
- [123] Wang Z, Wu Z, Agarwal D, et al. MedCLIP: Contrastive learning from unpaired medical images and text[J]. arXiv preprint arXiv:2210.10163, 2022.
- [124] Maaten L, Hinton G E. Visualizing data using t-SNE[J]. Journal of Machine Learning Research, 2008, 9(2605):2579-2605.