

文章编号: 1006-3080(2025)04-0505-09

DOI: 10.14135/j.cnki.1006-3080.20241104002

## 融合病理图像和基因组学多模态的癌症生存预测

张雪芹<sup>1</sup>, 李悦欣<sup>1</sup>, 刘畅<sup>1</sup>, 周韵斓<sup>2</sup>

(1. 华东理工大学信息科学与工程学院, 上海 200237; 2. 上海交通大学医学院附属新华医院  
临床检验实验室, 上海 200092)

**摘要:**融合病理图像和基因组学多模态数据进行生存预测有助于提高癌症患者生存预测的准确性, 从而为个性化医疗和精准治疗提供更加可靠的依据。为了提高生存预测的准确性, 针对病理图像和基因组学两个模态数据, 提出了一种基于中期特征融合的生存预测方法, 从“全局-局部-全局”3个层面来挖掘多模态数据间的潜在关系。该方法采用多示例学习, 基于ResNet50网络提取全尺寸病理图像示例级特征, 采用自归一化网络提取基因组学特征; 使用相似性度量方法学习模态间的全局相似语义信息, 利用双向交叉注意力模块挖掘模态间的密集局部联系, 通过最优传输方法捕获模态间的全局结构一致性, 同时使用基于Transformer编码器和门控注意力池化层构建的聚合器聚合形成包级特征, 最后采用估计危险函数预测得到癌症患者生存风险。在膀胱尿道上皮癌(BLCA)、肺腺癌(LUAD)和子宫体内膜癌(UCEC)这3个公共全尺寸病理图像数据集上的实验结果表明, 本文所提方法优于其他对比方法, 能够有效融合病理图像和基因组学数据, 显著提高生存预测的准确性。

**关键词:**病理图像; 基因组学; 多模态数据融合; 深度学习; 生存预测

**中图分类号:**TP391

**文献标志码:**A

目前, 癌症在全球各个国家都是导致死亡的主要病因。预计至2040年, 全球新增癌症病例数将达2840万人, 2050年时可能上升至3500万人<sup>[1]</sup>。生存预测主要用于估计癌症患者后期死亡的相对风险, 是临床预后研究中的重要内容。提高癌症的诊断精度, 准确进行生存预测, 能够帮助医生更好地评估疾病进展, 及时调整并优化治疗策略<sup>[2]</sup>。

病理图像评估在临床诊断中一直是癌症预后的金标准<sup>[3]</sup>, 通过对病理图像中细胞的形状、大小、纹理和像素强度等病理特征进行分析, 可以进行生存预测, 如NSCLC<sup>[4]</sup>、WSISA<sup>[5]</sup>等。同时, 癌症进展和相关的生物过程与基因表达和突变密切相关, 基因组数据近年来也被开始应用于生存预测, 如DeepSurv<sup>[6]</sup>、GCGCN<sup>[7]</sup>等。然而, 上述研究仅关注了

单一模态的医学数据, 忽略了整合来自不同医学模态的有用信息, 存在预测精度不够高的问题。

目前, 一些研究者开始结合组织学的形态信息和基因组的分子信息进行生存预测。有效的多模态融合方式使模型充分利用各个模态自身信息以及模态之间提供的互补信息, 从而提高模型性能<sup>[8-10]</sup>。根据不同的信息融合阶段和融合对象, 可以将多模态数据融合划分为早期数据融合、后期决策融合以及中期特征融合3种方式<sup>[11]</sup>。早期数据融合将多模态数据在输入模型前进行融合, 后期决策融合则是在每个模态数据独立通过模型处理后进行融合, 但这两种方法都存在无法充分捕捉模态间的复杂关系, 容易导致信息冗余或丢失的问题。

中期特征融合是在模型的中间层次对多模态数

收稿日期: 2024-11-04

基金项目: 国家社会科学基金重大项目(23&ZD142)

作者简介: 张雪芹(1972—), 女, 浙江台州人, 工学博士, 教授, 主要研究方向为计算机视觉、人工智能、信息安全。E-mail: zxq@ecust.edu.cn

通信联系人: 刘畅, E-mail: changliu20241218@163.com

引用本文: 张雪芹, 李悦欣, 刘畅, 等. 融合病理图像和基因组学多模态的癌症生存预测[J]. 华东理工大学学报(自然科学版), 2025, 51(4): 505-513.

Citation: ZHANG Xueqin, LI Yuexin, LIU Chang, et al. Cancer Survival Prediction by Integrating Multimodal Pathological Images and Genomics[J]. Journal of East China University of Science and Technology, 2025, 51(4): 505-513.

据进行融合,能够分析控制模态间的交互信息,是一种更优的融合策略。Xie 等<sup>[12]</sup>使用预训练的 ResNet50 提取全尺寸病理图像(Whole Slide Image, WSI)<sup>[13]</sup>的切块特征,通过计算病理图像和基因表达数据两种模态之间的共注意力分数,构建基于 Transformer 和注意力池化层的多模态特征融合网络,得到包括病理和遗传信息的特征向量。Li 等<sup>[14]</sup>提出的 HFBSurv 方法采用基于注意力的因子化双线性模型,对病理图像、拷贝数变异和基因表达数据进行逐层融合,发掘模态间的关系信息。Lyu 等<sup>[15]</sup>提出的 TransSurv 方法,使用 Transformer 从不同尺度提取病理图像特征和基因特征,利用交叉注意力机制,进行跨类融合和跨特征融合的两级融合,从而提供更佳和更全面的融合特征表示。Chen 等<sup>[16]</sup>提出的 MCAT 方法利用 ResNet 的多示例学习提取 WSI 示例特征,并通过自归一化网络提取 6 种基因组特征。通过共注意力模块,建立基因与病理图像特征的联系,增强基因表达与病理表型的关联。Jaume 等<sup>[17]</sup>通过转录组学中的生物通路特征和共注意力机制,探索病理图像与生物通路的多模态相互作用,进一步增加了模态互补信息。然而目前中期特征融合策略通常仅限于对单一层次、单一维度的特征提取和融合。

因此,为了更好地进行多模态数据融合,提高对

患者生存预测的准确性,本文提出了一种基于多层次中期特征融合的多模态生存预测框架:首先,通过 ResNet50 和自归一化网络(Self-Normalizing Networks, SNN)<sup>[18]</sup>分别提取模态自身特征,然后基于相似度衡量的方法对模态间的相似语义信息进行分析,采用双向交叉注意力模块挖掘模态间的局部关系,并通过最优传输学习模态间存在的潜在全局结构一致性;然后,通过“全局-局部-全局”的层级分析,可以充分学习模态间的相互作用和关联,从而得到丰富的多模态融合特征;最后,通过采用基于多层感知机的预测层实现了精度较高的生存预测结果。

## 1 多模态生存预测框架

### 1.1 整体框架设计

为了结合组织病理图像和基因组学数据 2 种数据模态特征,使患者生存状态的预测更加准确,本文提出了多模态生存预测总体框架,如图 1 所示。

多模态生存预测框架主要分为 WSI 示例特征和基因组特征提取、多模态信息融合,以及特征聚合和生存预测 3 个部分:

(1) WSI 示例特征和基因组特征提取。基于多示例学习方法构建 WSI 包和基因包。对 WSI 数据,

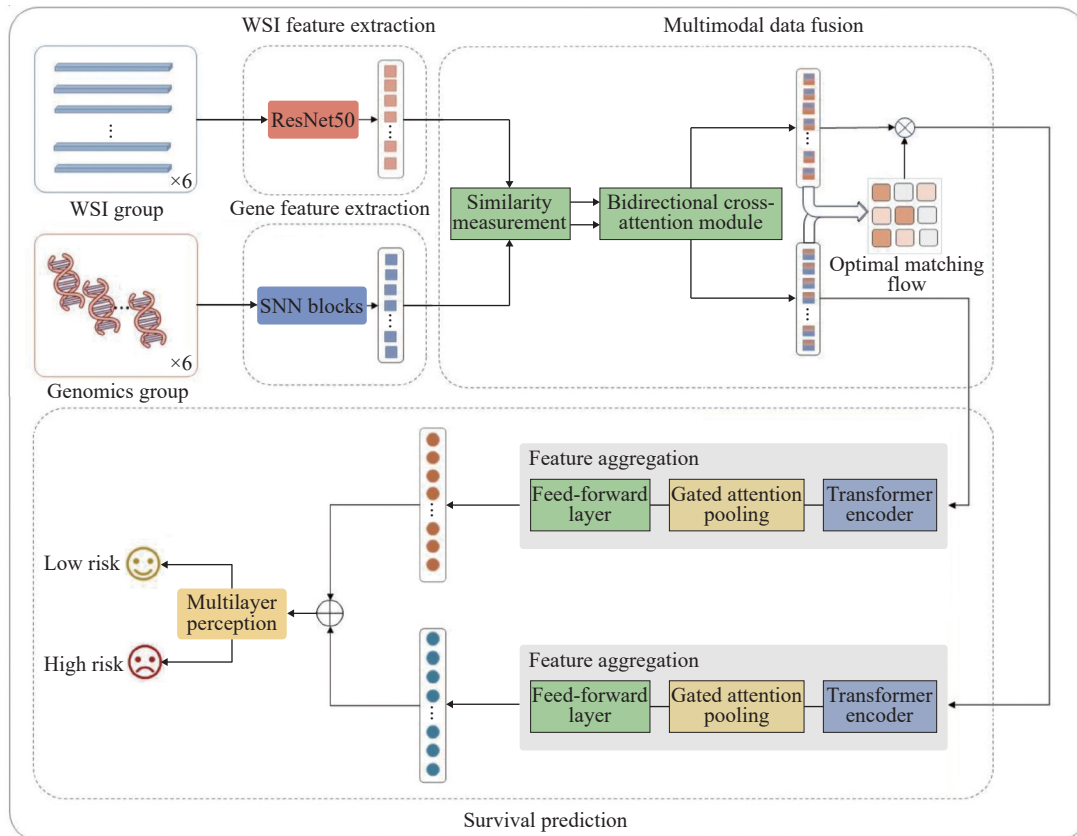


图 1 多模态生存预测模型框架

Fig. 1 Multimodal survival predication model framework

采用 ImageNet 上预训练的 ResNet50 提取 WSI 示例特征;对基因功能组数据,采用 SNN 网络提取基因组特征。

(2)多模态信息挖掘。为了使模型学习既能学习模态内的独有特征,同时获得模态间的相互关系,采用余弦相似度对两种模态数据进行相似性度量,在初步阶段进行模态全局对齐;采用双向交叉注意力模块进行局部关联性学习;采用最优传输方法,学习模态间的全局结构一致性。

(3)特征聚合和生存预测。将最优传输模块输出的两部分信息分别送入包级特征聚合器,其输出相连后经过多层感知机得到生存风险分数。

1.1.1 多示例学习 多示例学习 (Multiple Instance Learning, MIL)<sup>[19]</sup> 是一种不精确的弱监督学习方法,其训练数据由包(Bag)和示例(Instance)构成。在模型训练时所需的数据标签仅为包级标签,因此在分析粗粒度标签对象上, MIL 具有独特的优势。

给定一个包  $X = \{x_1, x_2, \dots, x_K\}$ , 其中  $K$  为包中示例的数量。对于每个示例  $x_k$  ( $k \in 1, 2, \dots, K$ ), 均具有潜在的标签  $y_k$ ,  $y_k = 1$  时代表示例为正,  $y_k = 0$  时则表明示例为负。在 MIL 的标准假设中,包中示例的标签未知,在训练时已知的信息仅有包级标签,若包中至少含有一个正示例,则此包的标签为正;当包中全部是负示例时,包的标签才为负<sup>[20]</sup>。MIL 的目标就是要检测出包中是否存在正示例。通常, MIL 利用适当的转换函数  $f$  和排列不变的聚合函数  $g$  来预测包标签  $Y_{\text{pre}}$ :

$$Y_{\text{pre}} = g(f(x_0), f(x_1), \dots, f(x_K)) \quad (1)$$

在基于嵌入空间的 MIL 方法中,  $f$  是一种示例级的特征提取器,将每个示例映射为低维特征嵌入,在低维空间进行训练和推理;  $g$  为聚合函数,将提取的示例特征整合得到包嵌入表示,并通过包分类器输出包的概率分数。

1.1.2 提取 WSI 示例特征 根据 WSI 中是否含有病变区域,可以很容易获得 WSI 级的标签。但是,由于 WSI 的尺寸过大,通常难以获取切片片级的精确标注。为了解决这个问题,采用多示例学习方法来提取 WSI 示例级特征,即将每张全尺寸病理图像看作一个包,从病理组织中获取的固定尺寸切片看作示例,提取 WSI 示例特征的过程可分为组织区域提取、示例提取和 WSI 示例特征提取。

#### (1)组织区域提取

WSI 中包含较多空白的背景区域,在提取示例前需要将其滤除,由于组织区域的颜色饱和度与背景相差较大,因此,首先选择将 WSI 转换至 HSV (Hue Saturation Value)颜色空间;在 HSV 的饱和度通道应用中值滤波,突出组织区域;然后采用二值化分割生成组织区域掩膜,根据掩膜来提取前景轮廓线,并过滤组织区域内的空洞;最后仅保留 WSI 的组织区域。此过程如图 2 所示。

#### (2)示例提取

在 20×放大倍数下,将每张 WSI 的组织区域划分为一系列不重叠的切片,切片大小为 256×256。由于部分切片所包含的组织区域较为稀疏,即病理信息表达不充分。因此,设置切片像素值总和的阈值,将低于此阈值的切片进行滤除。

#### (3)WSI 示例特征提取

经过组织区域提取、示例提取后,对于给定的 WSI 包可表示:

$$X^p = \{x_i^p\}_{i=1}^{M_p} \quad (2)$$

其中,  $p$  代表组织病理模态,  $M_p$  为示例数量,  $1 \leq i \leq M_p$ 。

为了提取示例特征,参考文献 [16] 采用在 ImageNet 上预训练的 Resnet50 对单个示例进行特征提取。ResNet50 通过引入残差连接,能够在深度网络中避免梯度消失问题,其强大的特征提取能力使

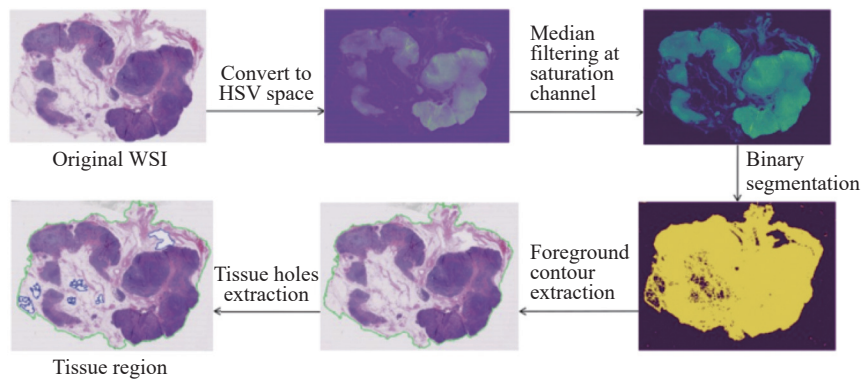


图 2 组织区域提取过程

Fig. 2 Tissue region extraction process

得模型能够很好地理解和表示图像信息,在图像分类、目标检测等诸多计算机视觉任务中表现出色。

得到 WSI 包的  $d$  维示例级特征  $\mathbf{B}^p$  为:

$$\mathbf{B}^p = \{b_i^p\}_{i=1}^{M_p} \quad (3)$$

其中,  $\mathbf{B}^p \in \mathbb{R}^{M_p \times d}$ ,  $b_i^p = f_p(x_i^p)$  为提取的单个示例特征提取函数。

为了匹配基因的功能组数量,对 WSI 包中示例特征随机分组,并经过线性层映射,最终 WSI 包的示例级特征可表示为:

$$\mathbf{B}^p = \{c_i^p\}_{i=1}^{C_p} \quad (4)$$

其中,  $c_i^p$  表示将示例特征  $b_i^p$  划分后的特征组,  $C_p$  表示组数,此时  $\mathbf{B}^p \in \mathbb{R}^{C_p \times d}$ 。

**1.1.3 提取基因组特征** 基因组图谱是个体最敏感和最具有识别性的信息,具有如突变状态、转录丰度、拷贝数变异等分子特征。根据不同分子特性对生物功能的影响,可将基因组学数据划分为 6 大功能类别,分别与肿瘤抑制、肿瘤发生、蛋白激酶、细胞分化、转录,以及细胞因子和生长相关<sup>[21]</sup>。

由于基因数据样本数量较少,但提取的特征通常有数百至数千维,因此传统前馈神经网络容易出现过拟合的问题。SNN 是 Günter Klambauer 等<sup>[18]</sup>提出的一种前馈神经网络,在高维小样本数据的学习上具有良好的效果。它引入了一种叫做缩放指数线性单元(Scaled Exponential Linear Units, SeLU)的激活函数来缓解梯度消失和爆炸问题,同时采用了 Alpha Dropout 来防止过拟合。基于上述特性,本文采用 SNN 对 6 个功能类别的基因组学数据进行特征提取。

设基因包可表示为:

$$\mathbf{X}^g = \{x_j^g\}_{j=1}^{M_g} \quad (5)$$

其中,  $g$  表示基因组学模态,  $M_g$  为包中的基因功能组数,  $1 \leq j \leq M_g$ ,  $\mathbf{X}^g \in \mathbb{R}^{M_g \times d_j}$ 。

采用 SNN 网络来提取每个基因组的  $d_j$  维特征嵌入,对于第  $j$  个基因组,设  $b_j^g = f_g^j(x_j^g)$ ,  $f_g^j(\cdot)$  为特征提取函数。则整个基因组特征可表示为:

$$\mathbf{B}_n^g = \{b_j^g\}_{j=1}^{M_g} \quad (6)$$

其中,  $\mathbf{B}_n^g \in \mathbb{R}^{M_g \times d}$ 。

## 1.2 基于相似性度量的多模态全局对齐

相似性度量可以使模型学习到更好表示模态间相似语义信息的共享特征空间,这样既保留每个模态的独特之处,也可以捕获模态间的语义相关性<sup>[22-23]</sup>。余弦相似度能够衡量高维空间中特征向量的方向性,而不受特征向量的幅值大小影响,可使两类模态

的特征分布不仅在数值尺度上有所差异,同时在语义上的相似性也能通过向量方向得到精确捕捉。因此,在得到了 WSI 示例特征和基因组特征后,为了促使两类特征能够进行初步的全局匹配,采用余弦相似度来捕捉模态间的全局语义一致性,如式 (7) 所示,并计算余弦相似度损失  $L_{\text{sim}}$ 。

$$S_{\cos}(\mathbf{B}^p, \mathbf{B}^g) = \frac{\mathbf{B}^p \cdot \mathbf{B}^g}{\|\mathbf{B}^p\| \cdot \|\mathbf{B}^g\|} \quad (7)$$

$$L_{\text{sim}} = 1 - S_{\cos}(\mathbf{B}_n^p, \mathbf{B}_n^g) \quad (8)$$

其中,  $S_{\cos}(\cdot)$  表示余弦相似距离,  $\|\cdot\|$  表示向量的模。

## 1.3 基于双向交叉注意力的局部关联学习

在有癌症病变的状态下,病理图像中对应组织区域的形态会发生显著变化,与之相关的基因也会发生异常表达。然而,在判定为恶性的全尺寸病理图像中,通常也只有小部分区域与病变直接相关。因此,为了聚焦病变区域挖掘基因数据与病理图像数据之间的局部关联,本文引入双向交叉注意力模块进行局部关联学习。双向交叉注意力模块的结构如图 3 所示。双向交叉注意力模块以两种模态信息互为引导,在两者之间进行信息的交叉和传递。它通过在模态间建立双向的注意力连接,使得模型能够将一个模态中发现的重要特征与另一个模态中的相关特征进行关联,并通过加权融合进一步强化该相关特征的表示,从而更加准确地表达两种模态之间的局部关联信息。

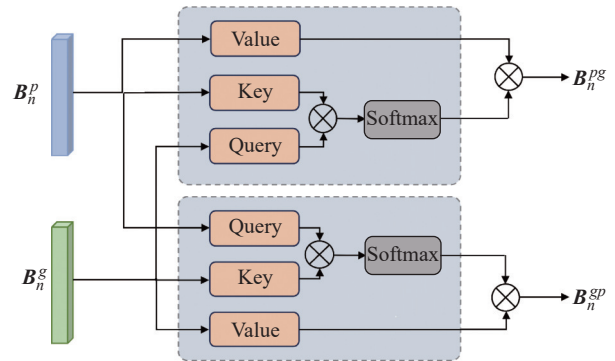


图 3 双向交叉注意力模块

Fig. 3 Bidirectional cross-attention module

WSI 示例特征  $\mathbf{B}^p$  和基因组学特征  $\mathbf{B}^g$  经过线性变换生成各自的查询,键和值,分别表示为  $\mathbf{Q}_p$ ,  $\mathbf{Q}_g$ ,  $\mathbf{K}_p$ ,  $\mathbf{K}_g$ ,  $\mathbf{V}_p$ ,  $\mathbf{V}_g$ 。二者的交叉注意力分数  $A^{pg}$  和  $A^{gp}$  分别通过由式 (9) 和 (10) 计算得到。

$$A^{pg} = \frac{\mathbf{Q}_p \mathbf{K}_g^T}{\sqrt{d_k}} \quad (9)$$

$$A^{gp} = \frac{\mathbf{Q}_g \mathbf{K}_p^T}{\sqrt{d_k}} \quad (10)$$

其中,  $d_k$  为注意力头的维度。交叉注意力分数对值进行加权求和, 得到的输出如式 (11) 和式 (12) 所示。

$$\mathbf{B}^{pg} = \text{Softmax}(\mathbf{A}^{pg})\mathbf{V}_g \quad (11)$$

$$\mathbf{B}^{gp} = \text{Softmax}(\mathbf{A}^{gp})\mathbf{V}_p \quad (12)$$

其中,  $\mathbf{B}^{pg}$  和  $\mathbf{B}^{gp}$  分别表示从病理图像到基因组学以及从基因组学到病理图像的特征关联注意力输出。

在学习过程中采用双向交叉注意力模块, 可以使模型重点学习病理图像特定区域, 并在基因组学和病理图像两种模态数据之间互相传递互补信息, 有效建立局部联系。

#### 1.4 基于最优传输匹配的全局一致性学习

肿瘤微环境内部的相互作用是影响生存结果的重要指标。肿瘤发生时, 通常伴随着基因的共表达, 基因与基因之间存在着生物通路结构。不同类型的组织细胞之间也会互相作用, 存在共发生现象<sup>[24]</sup>。这种共表达和共发生在一定程度上表明了基因和组织细胞具有全局关联性。因此, 在基于多模态数据进行生存预测时, 除了需要重点关注局部病变区域外, 还要关注不同模态数据间的全局结构一致性, 才能更加全面地融合病理图像与基因组学之间的特征。

最优传输是一种结构匹配方法, 可以提供具有全局感知的实例匹配过程。该方法既可以保留多模态数据间的潜在全局结构关联, 同时又能减轻模态间的异质性<sup>[25]</sup>。因此, 本文提出基于组织学和基因组学特征之间的成对代价, 采用最优传输法寻求总体上的最小匹配代价, 生成最优匹配解。

设经过局部密集相似性学习的 WSI 示例特征为  $\mathbf{B}^{pg} = [b_1^{pg}, b_2^{pg}, \dots, b_{M_p}^{pg}]$ , 基因组学特征为  $\mathbf{B}^{gp} = [b_1^{gp}, b_2^{gp}, \dots, b_{M_g}^{gp}]$ , 采用离散 Kantorovich 公式, 来寻找  $\mathbf{B}^{pg}$  和  $\mathbf{B}^{gp}$  之间的整体最佳匹配流  $\mathbf{P}$ , 如式 (13) 所示:

$$W(\mathbf{B}^{pg}, \mathbf{B}^{gp}) = \min_{\mathbf{P} \in \Pi(\mu_{pg}, \mu_{gp})} \langle \mathbf{P}, \mathbf{C} \rangle_F \quad (13)$$

其中,  $\mathbf{C} \geq 0 \in \mathbb{R}^{M_p \times M_g}$  是代价矩阵, 由  $\mathbf{C}^{u,v} = c(b_u^{pg}, b_v^{gp})$  计算得到,  $c(\cdot)$  为距离度量矩阵,  $\langle \cdot \rangle_F$  是指 Frobenius 点积。  $\Pi(\mu_{pg}, \mu_{gp}) = \{\mathbf{P} \in \mathbf{R}_+^{M_p \times M_g} | \mathbf{P}\mathbf{1}_{M_g} = \mu_{pg}, \mathbf{P}^T\mathbf{1}_{M_p} = \mu_{gp}\}$  为病理特征和基因组特征两个边际经验分布之间的联合概率耦合集, 其中,  $\mathbf{1}_k$  为  $k$  维全 1 向量, 它描述了当 WSI 示例特征  $\mu_{pg}$  和基因组学特征  $\mu_{gp}$  边际分布之间的总质量相等时, 达到的边际约束条件。获得最佳匹配流  $\mathbf{P}$  后, WSI 示例特征表示为  $\hat{\mathbf{B}}^{pg} = \mathbf{P}^T \mathbf{B}^{pg}$ , 使组织病理与基因特征的一致性结构分布得到对齐。

#### 1.5 生存预测

1.5.1 包级特征聚合 经过前面的中期融合学习, 得

到了丰富的示例级多模态特征表示。为获取最终包级的特征表示, 进一步构建两个聚合器  $T_p(\cdot)$  和  $T_g(\cdot)$ , 其结构相同, 均由 Transformer 编码器、门控注意力池化层 (Gated Attention Pooling, GAP) 和前馈神经网络层 (Feed-Forward Layer, FFL) 构成。将 WSI 示例特征  $\hat{\mathbf{B}}^{pg}$  和基因组学特征  $\mathbf{B}^{gp}$  分别送入  $T_p(\cdot)$  和  $T_g(\cdot)$ , 得到第  $n$  个病人的最终包级特征  $\mathbf{H}_n^p$  和  $\mathbf{H}_n^g$ :

$$\mathbf{H}_n^p = T_p(\hat{\mathbf{B}}_n^{pg}) \quad (14)$$

$$\mathbf{H}_n^g = T_g(\mathbf{B}_n^{gp}) \quad (15)$$

在完成包级特征的聚合之后, 得到了具有高度信息浓缩的多模态特征表示, 这为下一步的生存风险评估奠定了基础。

1.5.2 生存预测 生存预测的目的是预测一个结果事件在一个特定时间之前发生的风险概率。给定一组具有  $N$  对的“病理-基因”训练集  $\mathbf{D} = \{(\mathbf{X}_n^p, \mathbf{X}_n^g), c_n, t_n\}_{n=1}^N$ , 其中  $c$  表示观察状态,  $c=1$  表示发生了右截尾事件,  $c=0$  表示没有发生。  $t \in \mathbb{R}^+$  指以月为单位的总生存时间。将  $\mathbf{H}_n^p$  和  $\mathbf{H}_n^g$  直接相连, 获得最终用于生存风险预测的多模态融合特征  $\mathbf{H}_n$ :

$$\mathbf{H}_n = \mathbf{H}_n^p \oplus \mathbf{H}_n^g \quad (16)$$

危险函数用于描述在时间节点  $t$  附近发生死亡事件的概率, 则本文采用危险函数  $h_n(t|\mathbf{H}_n)$  来预测生存风险:

$$h_n(t|\mathbf{H}_n) = h_n(T = t | T \geq t, \mathbf{H}_n) \in [0, 1] \quad (17)$$

考虑到生存预测任务是通过累积生存函数来估计顺序死亡风险, 因此, 最终的生存概率:

$$S_n(t|\mathbf{H}_n) = \prod_{z=1}^t (1 - h_n(z|\mathbf{H}_n)) \quad (18)$$

#### 1.6 损失函数

参考文献 [16], 负对数似然函数 (Negative Log-Likelihood, NLL) 可以有效处理生存数据中的截尾现象, 同时最大化模型预测的生存概率, 因此生存预测模型训练过程中的损失采用 NLL 构建, 如式 (19) 所示:

$$L_{\text{surv}} = -c_n \cdot \lg(S_n(t_n|\mathbf{H}_n)) - (1 - c_n) \cdot \lg(S_n(t_n - 1|\mathbf{H}_n)) - (1 - c_n) \cdot \lg(h_n(t_n|\mathbf{H}_n)) \quad (19)$$

此外, 为了更好地捕捉不同模态数据间的潜在关联, 引入组织病理和基因组学特征的初步相似性度量损失  $L_{\text{sim}}$ , 因此训练过程中的总体损失如式 (20) 所示:

$$L_{\text{total}} = L_{\text{surv}} + L_{\text{sim}} \quad (20)$$

## 2 实验结果与分析

### 2.1 数据集描述

实验采用来自 TCGA 平台的 3 个大型癌症数据集:膀胱尿道上皮癌(Bladder Urothelial Carcinoma, BLCA)<sup>[26]</sup>,肺腺癌(Lung Adenocarcinoma, LUAD)<sup>[27]</sup>和子宫体内膜癌(Uterine Corpus Endometrial Carcinoma, UCEC)<sup>[28]</sup>,病例数分别为 371、450 例和 475 例。按照 256×256 进行切块, BLCA 的切块数为 29844424 个, LUAD 为 26563550 个, UCEC 为 29217248 个。

### 2.2 评价标准

#### (1) *c*-index

在生存预测任务中,主要采用一致性指数(*c*-index)来评估生存分析模型的性能表现,即衡量一个模型预测的生存时间顺序是否与实际观察到的顺序一致。一致性指数可采用下式计算:

$$c\text{-index} = \frac{1}{n(n-1)} \sum_{i=1}^n \sum_{j=1}^n I(T_i < T_j)(1 - c_j) \quad (21)$$

其中,  $n$  表示病例数量,  $T_i$  和  $T_j$  分别代表第  $i$  位和第  $j$  位病人的生存时间。  $I(\cdot)$  是指示函数,在条件为真时取 1,反之取 0。  $c_j$  代表观察状态。  $c$ -index 的值越高,即表示预测结果越准确。

#### (2) Kaplan-Meier 生存曲线

Kaplan-Meier 生存函数  $S(t)$  表示在时间  $t$  之前个体存活的概率,其计算公式为:

$$S(t) = \prod \left( \frac{n_j - d_j}{n_j} \right) \quad (22)$$

其中,  $d_j$  表示在时间间隔  $j$  内的死亡人数,  $n_j$  表示实际有风险的个体数量。通常情况下,生存函数  $S(t)$  随着时间  $t$  的增加而递减,  $S(t) = 1$  表示在开始时刻所有个体都存活,最终  $S(t)$  会趋向于 0。

### 2.3 实验环境和参数设置

实验采用的 CPU 为 Intel(R) Xeon(R) Platinum 8352V CPU @ 2.10GHZ,内存为 90 GB;系统为 ubuntu 18.04, GPU 为 NVIDIA GeForce RTX 4090,需至少 3 TB 大小的硬盘。

在生存预测模型的训练中,使用 Adam 优化器,初始学习率设为  $2 \times 10^{-4}$ ,权重衰减为  $1 \times 10^{-5}$ 。由于包的大小不同,将 batch size 设为 1,并带有 32 步梯度累积。

所有实验采用 5 折交叉验证的方式,有效地保证每个数据样本都能参与训练和测试,提高了模型的泛化能力。结果采用 *c*-index 作为评估指标,并计算其平均值和标准差,以充分反映模型在不同数据

集上的表现和稳健性。

### 2.4 实验结果及分析

2.4.1 对比实验 该实验在上述 3 个数据集上对比不同生存预测方法的准确性。基线方法包括单模态和多模态方法。其中,在单模态方法中, SNN<sup>[18]</sup>和 SNNTrans<sup>[18]</sup>仅采用基因组学数据, AttnMIL<sup>[29]</sup>、 CLAM<sup>[30]</sup>和 TransMIL<sup>[31]</sup>仅采用病理图像数据。多模态方法包括 MCAT<sup>[16]</sup>、 Porpoise<sup>[32]</sup>、 HFBSurv<sup>[18-19]</sup>以及 DualTrans<sup>[24]</sup>,同时使用病理图像数据和基因组学数据进行生存分析。实验中,所有方法采用相同的特征包构建方法和训练参数。对比实验结果如表 1 所示。表中,“P”表示组织病理模态,“G”表示基因组学模态,最佳的结果使用粗体表示,次优结果使用下划线表示。

表 1 3 个数据集上不同生存预测方法的对比结果

Table 1 Comparison results of different survival prediction methods on three datasets

Method	Model		<i>c</i> -index		
	P	G	BLCA	LUAD	UCEC
SNN		√	0.606±0.041	0.645±0.038	0.671±0.037
SNNTrans		√	0.653±0.020	<u>0.682±0.040</u>	0.661±0.047
AttnMIL	√		0.545±0.041	0.578±0.057	0.618±0.076
CLAM	√		0.560±0.023	0.582±0.055	0.605±0.078
TransMIL	√		0.594±0.042	0.585±0.035	0.672±0.041
Porpoise	√	√	0.636±0.024	0.647±0.031	<u>0.695±0.032</u>
HFBSurv	√	√	0.639±0.027	0.650±0.049	0.672±0.019
MCAT	√	√	0.660±0.020	0.667±0.029	0.664±0.047
DualTrans	√	√	<u>0.683±0.019</u>	0.665±0.037	0.675±0.038
Ours	√	√	<b>0.731±0.034</b>	<b>0.706±0.020</b>	<b>0.712±0.046</b>

从表 1 可以看到,本文所提方法的性能在 3 个数据集上均是最佳。总体而言,基于多模态数据进行生存预测的准确度要优于单模态方法,证明了融合多种模态数据在生存预测任务上具有一定优势。与单模态方法进行比较,本文所提方法与单模态方法中整体表现最好的 SNNTrans 方法对比, BLCA、 LUAD、 UCEC 的 *c*-index 分别提高了 0.078、 0.024 和 0.051。同时,可以看到,在单模态方法中,基于基因组学预测方法的准确度总体要高于基于病理图像的方法,在 BLCA、 LUAD、 UCEC 3 个数据集上,平均分别高出 0.063、 0.082 和 0.034,说明基因组学在生存分析中具有重要作用。

与多模态方法进行比较,在 BLCA 数据集上,本文所提方法与次优方法 DualTrans 相比, *c*-index 提高了 0.048;在 LUAD 数据集上较次优方法 SNNTrans

提升了 0.024; 在 UCEC 数据集上, 较次佳方法 Porpoise 提高了 0.017。分析原因, DualTrans 虽然也利用了 Transformer 层分别整合两种模态的内部信息并将其相连, 但在模型学习中没有挖掘模态间的关联特性。MCAT 利用基因作为指导, 关联考虑了与之相关度高的病理区域信息, 但忽略了全局一致性信息。Porpoise 忽略了模态间的关系学习, 丢失了重要的生存预测信息; HFBsurv 采用了基于因子化双线性模型的层级特征融合方式, 但相较于基于深度学习技术的特征提取方法, 对模态特征学习还不充分。本文所提方法对基因和病理图像两种模态数据进行了全局对齐, 深入挖掘了模态间的局部和全局关系, 获得了最佳性能。

**2.4.2 消融实验** 本实验用于验证本文所提方法中, 相似性度量、双向交叉注意力和最优传输 3 个主要组成模块的有效性, 实验结果如表 2 所示, 其中, w/o SimMe 为在本文方法的基础上, 移除相似性度量模块; w/o BDCA 为在本文方法的基础上, 移除双向交叉注意力模块; w/o OT 为在本文方法的基础上, 移除最优传输模块。

从表 2 可以看到, 当移除 SimMe 模块时, 变体模型在 BLCA、LUAD 和 UCEC 这 3 个数据集的 *c*-index 与原模型相比分别下降了 0.040、0.013 和 0.032。当缺少 BDCA 模块时, 分别降低了 0.066、0.037 和 0.043。当没有 OT 模块时, 则分别降低了 0.046、0.022 和 0.032。可见, SimMe 模块将 2 种模态数据进行初步全局对齐, 可以使模型更好地学习模态间相似语义信息, 因而提高生存预测的准确率。而通过添加 BDCA 模块, 能够有效挖掘基因组学与病理图像中特定区域间潜在的局部关联关系, 实验数据表明, 该模块对 *c*-index 的提升作用最为明显。OT 模块用于学习两种模态间潜在的全局结构一致性, 进一步提高了预测准确率。总体而言, 本文设计的 3 种模态间关系分析和挖掘模块, 对提升生存预测准确率

表 2 3 个数据集上的方法模块消融实验结果

Table 2 Ablation study results of method modules on three datasets

Model	<i>c</i> -index		
	BLCA	LUAD	UCEC
w/o SimMe	0.691±0.029	0.693±0.016	0.680±0.043
w/o BDCA	0.665±0.011	0.669±0.019	0.669±0.037
w/o OT	0.685±0.036	0.684±0.017	0.680±0.046
Ours (All components)	<b>0.731±0.034</b>	<b>0.706±0.020</b>	<b>0.712±0.046</b>

均有效。

**2.4.3 Kaplan-Meier 生存曲线分析** Kaplan-Meier 算法是一种常用的用于估计患者生存率的方法<sup>[33]</sup>, 根据该方法的估计结果可以绘制出生存曲线。为了进一步验证本文所提生存预测模型的有效性, 采用患者的生存时长、存活状态以及风险分数进行 Kaplan-Meier 生存曲线分析, 可视化不同风险下的患者生存事件。使用 R 语言的生存分析包 maxstat, 将所有病人分为低风险组和高风险组。在 BLCA、LUAD 和 UCEC 3 个数据集上获得的低风险组分别为 257、198 人和 345 人, 高风险组分别为 114、252 人和 130 人。

图 4 给出了基于 Kaplan-Meier 在 3 个数据集上绘制的生存曲线, 图中横坐标代表生存时间或随访时间, 纵坐标代表患者总体生存率。生存曲线一般是向水平延伸, 当某个时间节点出现病例死亡时, 曲线便会垂直下降形成阶梯状。曲线上的竖线代表删失病例, 即失去随访的病例。由图 4 可以看到, 在 BLCA 和 UCEC 数据集上, 高低风险组的 Kaplan-Meier 曲线之间有明显间隔, 在任一短期观察时间内, 低风险组患者的生存率均明显高于高风险组患者。在 LUAD 数据集上, 生存曲线虽有部分重叠, 但整体上看, 高低风险组的生存曲线也存在较为明显的差异。

此外, 实验还采用了 Logrank 检验, 以 *P*-value 来衡量低风险组和高风险组之间是否具有统计学的显

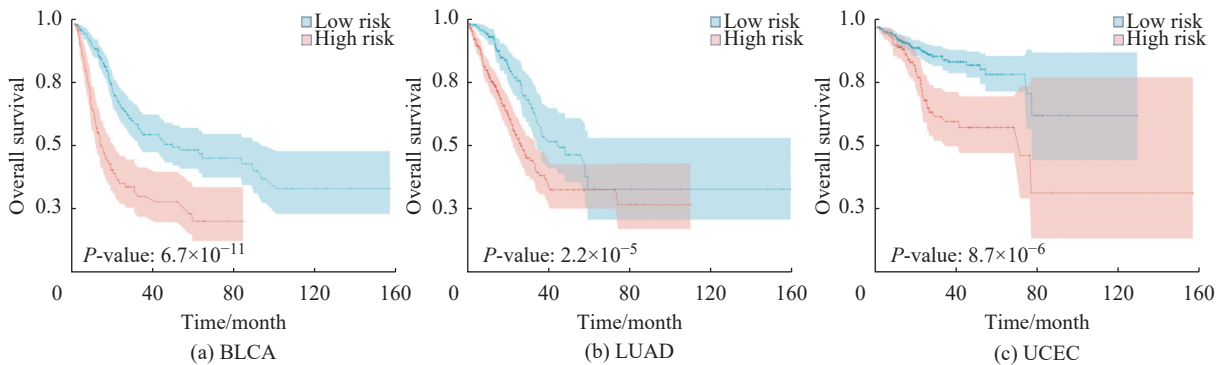


图 4 Kaplan-Meier 生存分析曲线和 Logrank 检验

Fig. 4 Kaplan-Meier survival analysis curves and Logrank test

著差异性。当  $P\text{-value} \leq 0.05$  时,通常认为是具有统计学上的显著差异。由图 4 可知,在 BLCA、LUAD 和 UCEC 数据集上的  $P\text{-value}$  分别为  $6.7 \times 10^{-11}$ 、 $2.2 \times 10^{-5}$  和  $8.7 \times 10^{-6}$ ,均明显低于 0.05,表明高低风险组患者的生存曲线之间具有统计学显著差异。因此,本文所提出的生存分析模型能够较为准确地预测不同风险下的患者生存率。

### 3 结 论

为提高癌症患者生存预测的准确性,本文提出一种结合全尺寸病理图像和基因组学多模态数据中期融合的生存预测方法。基于多示例学习方法,采用 ResNet50 提取病理图图像特征,采用 SNN 网络提取基因组学特征。为了充分挖掘两种模态数据之间的相关关系,基于相似度衡量的方法对不同模态间的相似语义信息进行分析;采用双向交叉注意力模块挖掘不同模态间的局部密集关联;通过最优传输模块学习模态间存在的潜在全局结构一致性。通过从“全局-局部-全局”3 个层级挖掘,充分学习了两种模态间的相互关系。最后通过分别聚合 WSI 示例特征和基因组特征并相连,得到最终包级特征表示,估计危险函数来预测癌症患者生存风险。实验证明,本文所提方法能够提高生存预测的准确性。未来还可以将更多的临床诊断数据,如放射学影像、临床诊断报告等数据整合到生存预测研究中。同时,可以探索在兼顾效率的情况下,进一步融合早期、中期和晚期的特征来提高生存预测准确度。

#### 参考文献:

- [ 1 ] SIEGEL R L, MILLER K D, WAGLE N S, *et al.* Cancer statistics, 2023[J]. *CA: A Cancer Journal for Clinicians*, 2023, 73(1): 17-48.
- [ 2 ] GOSPODAROWICZ M, O'SULLIVAN B. Prognostic factors in cancer[C]//*Seminars in Surgical Oncology*. Hoboken: Wiley Subscription Services, Inc. A Wiley Company, 2003, 21(1): 13-18.
- [ 3 ] DE HAAN K, ZHANG Y, ZUCKERMAN J E, *et al.* Deep learning-based transformation of H&E stained tissues into special stains[J]. *Nature Communications*, 2021, 12(1): 1-13.
- [ 4 ] WANG H, XING F, SU H, *et al.* Novel image markers for non-small cell lung cancer classification and survival prediction[J]. *BMC Bioinformatics*, 2014, 15: 1-12.
- [ 5 ] ZHU X, YAO J, ZHU F, *et al.* Wsisa: Making survival prediction from whole slide histopathological images[C]//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. New Jersey, USA: IEEE, 2017: 7234-7242.
- [ 6 ] KATZMAN J L, SHAHAM U, CLONINGER A, *et al.* DeepSurv: Personalized treatment recommender system using a Cox proportional hazards deep neural network[J]. *BMC Medical Research Methodology*, 2018, 18: 1-12.
- [ 7 ] WANG C, GUO J, ZHAO N, *et al.* A cancer survival prediction method based on graph convolutional network[C]//*IEEE Transactions on Nanobioscience*. New Jersey, USA: IEEE, 2019, 19(1): 117-126.
- [ 8 ] WAQAS A, TRIPATHI A, RAMACHANDRAN R P, *et al.* Multimodal data integration for oncology in the era of deep neural networks: A review[J]. *Frontiers in Artificial Intelligence*, 2024, 7: 1408843.
- [ 9 ] CUI C, YANG H, WANG Y, *et al.* Deep multimodal fusion of image and non-image data in disease diagnosis and prognosis: A review[J]. *Progress in Biomedical Engineering*, 2023, 5(2): 022001.
- [ 10 ] CUI C, LIU H, LIU Q, *et al.* Survival prediction of brain cancer with incomplete radiology, pathology, genomic, and demographic data[C]//*International Conference on Medical Image Computing and Computer-Assisted Intervention*. Cham: Springer Nature Switzerland, 2022: 626-635.
- [ 11 ] LIPKOVA J, CHEN R J, CHEN B, *et al.* Artificial intelligence for multimodal data integration in oncology[J]. *Cancer Cell*, 2022, 40(10): 1095-1110.
- [ 12 ] XIE Y, NIU G, DA Q, *et al.* Survival prediction for gastric cancer via multimodal learning of whole slide images and gene expression[C]//*2022 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*. Las Vegas, NV, USA: IEEE, 2022: 1311-1316.
- [ 13 ] HANNA M G, PARWANI A, SIRINTRAPUN S J. Whole slide imaging: Technology and applications[J]. *Advances in Anatomic Pathology*, 2020, 27(4): 251-259.
- [ 14 ] LI R, WU X, LI A, *et al.* HFBSurv: Hierarchical multimodal fusion with factorized bilinear models for cancer survival prediction[J]. *Bioinformatics*, 2022, 38(9): 2587-2594.
- [ 15 ] LYU Z, LIN Y, YAN R, *et al.* TransSurv: Transformer-based survival analysis model integrating histopathological images and genomic data for colorectal cancer[J]. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 2022, 20(6): 3411-3420.
- [ 16 ] CHEN R J, LU M Y, WENG W H, *et al.* Multimodal co-attention transformer for survival prediction in gigapixel whole slide images[C]//*Proceedings of the IEEE/CVF International Conference on Computer Vision*. Piscataway, NJ: IEEE, 2021: 4015-4025.
- [ 17 ] JAUME G, VAIDYA A, CHEN R J, *et al.* Modeling dense multimodal interactions between biological pathways and histology for survival prediction[C]//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Seattle, USA: The MIT Press. 2024: 11579-11590.
- [ 18 ] KLAMBAUER G, UNTERTHINER T, MAYR A, *et al.* Self-Normalizing Neural Networks[M]. Massachusetts, USA: The MIT Press, 2017.

- [19] ILSE M, TOMCZAK J, WELLING M. Attention-based deep multiple instance learning[C]//International Conference on Machine Learning. New York: ACM, 2018: 2127-2136.
- [20] QU L, LUO X, LIU S, *et al.* Dgmil: Distribution guided multiple instance learning for whole slide image classification[C]//International Conference on Medical Image Computing and Computer-Assisted Intervention. Cham: Springer Nature Switzerland, 2022: 24-34.
- [21] MYERS C L, BARRETT D R, HIBBS M A, *et al.* Finding function: Evaluation methods for functional genomic data[J]. *BMC Genomics*, 2006, 7: 1-15.
- [22] LI J, SELARAJU R, GOTMARE A, *et al.* Align before fuse: Vision and language representation learning with momentum distillation[J]. *Advances in Neural Information Processing Systems*, 2021, 34: 9694-9705.
- [23] WANG F, ZHOU Y, WANG S, *et al.* Multi-granularity cross-modal alignment for generalized medical visual representation learning[J]. *Advances in Neural Information Processing Systems*, 2022, 35: 33536-33549.
- [24] ZHOU F, CHEN H. Cross-modal translation and alignment for survival analysis[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. Piscataway, NJ: IEEE, 2023: 21485-21494.
- [25] JIN L, LANG D, LEI N. An optimal transport view of class-imbalanced visual recognition[J]. *International Journal of Computer Vision*, 2023, 131(11): 2845-2863.
- [26] 鲍振宇. 膀胱癌转录组数据中链非编码 RNA 的挖掘和预后分析模型的构建 [D]. 上海: 华东师范大学, 2017.
- [27] 熊雷, 易茜, 许明芳, 等. MRPL12 在肺腺癌中的表达和预后分析 [J]. *上海交通大学学报 (医学版)*, 2021, 41(8): 1033-1040.
- [28] LIU J, GENG R, NI S, *et al.* Pyroptosis-related lncRNAs are potential biomarkers for predicting prognoses and immune responses in patients with UCEC[J]. *Molecular Therapy-Nucleic Acids*, 2022, 27: 1036-1055.
- [29] ILSE M, TOMCZAK J, WELLING M. Attention-based deep multiple instance learning[C]//International Conference on Machine Learning. New York, NY: ACM, 2018: 2127-2136.
- [30] LU M Y, WILLIAMSON D F K, CHEN T Y, *et al.* Data-efficient and weakly supervised computational pathology on whole-slide images[J]. *Nature Biomedical Engineering*, 2021, 5(6): 555-570.
- [31] SHAO Z, BIAN H, CHEN Y, *et al.* Transmil: Transformer based correlated multiple instance learning for whole slide image classification[J]. *Advances in Neural Information Processing Systems*, 2021, 34: 2136-2147.
- [32] CHEN R J, LU M Y, WILLIAMSON D F K, *et al.* Pan-cancer integrative histology-genomic analysis via multimodal deep learning[J]. *Cancer Cell*, 2022, 40(8): 865-878.
- [33] GOEL M K, KHANNA P, KISHORE J. Understanding survival analysis: Kaplan-Meier estimate[J]. *International Journal of Ayurveda Research*, 2010, 1(4): 274.

## Cancer Survival Prediction by Integrating Multimodal Pathological Images and Genomics

ZHANG Xueqin<sup>1</sup>, LI Yuexin<sup>1</sup>, LIU Chang<sup>1</sup>, ZHOU Yunlan<sup>2</sup>

(1. School of Information Science and Engineering, East China University of Science and Technology, Shanghai 200237, China; 2. Department of Clinical Laboratory, Xinhua Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai 200092, China)

**Abstract:** Integrating data from pathology images and genomics enhances cancer patient survival prediction accuracy, offering solid foundation for personalized medicine and precision treatment. To enhance predictive accuracy, we propose an intermediate fusion-based method to explore the latent relationships between histopathological images and genomic data across global and local levels. Whole Slide Image (WSI) features are extracted using ResNet50 and genomic features are extracted by Self-Normalizing Networks (SNN). Similarity measures are used to learn enable global semantic similarities across modalities. A bidirectional cross-attention module identifies dense local connections. Optimal transport methods capture global structural consistency between modalities. These features are aggregated through a Transformer encoder and Gated Attention Pooling (GAP) to form bag-level representations. The model estimates the hazard function to predict cancer survival risk. Experimental results on the BLCA, LUAD, and UCEC WSI datasets demonstrate that the proposed method surpasses other comparative methods, effectively integrating pathology images and genomic data to significantly improve survival prediction accuracy.

**Key words:** pathology images; genomics; multimodal data fusion; deep neural network; survival prediction

(责任编辑: 王晓丽)