

DOI:10.3969/j.issn.1005-202X.2025.02.006

医学影像物理

# ABMIL-BiGRU: 基于双向门控循环注意力多示例学习的乳腺癌淋巴结转移可解释性预测

李波<sup>1</sup>, 杨艳斌<sup>2</sup>, 李帅<sup>3</sup>, 梁美彦<sup>3</sup>

1. 山西省荣军医院物理治疗科, 山西 太原 030031; 2. 山西省民政康复医院康复治疗科, 山西 太原 030032; 3. 山西大学物理电子工程学院, 山西 太原 030006

**【摘要】**针对千兆像素乳腺癌病理全切片图像分类和病灶定位的问题,提出基于双向门控循环注意力多示例学习(ABMIL-BiGRU)模型对H&E染色的乳腺癌淋巴结转移图像进行可解释性预测。该方法通过两个正交的双向门控循环单元分别建立图像块行方向和列方向特征间的长短距离依赖关系,从而实现图像块空间位置和上下文信息的嵌入,再通过注意力多示例池化来量化每个特征表示的注意力分数,从而实现全切片图像级特征聚合并生成可解释性热图。研究结果表明,ABMIL-BiGRU模型在乳腺癌转移数据集上的平均准确率和AUC值分别为0.9186和0.9467,不仅实现全切片图像的高精度预测和感兴趣区域定位,而且还提供图像块级别的人类可解释的特征,该模型在一定程度上解决了“准确性-可解释性权衡”问题,其优越性能够为计算机辅助诊断和智能系统的临床应用提供新的范式。

**【关键词】**乳腺癌;淋巴结转移;精准诊断;双向门控循环单元;上下文信息;可解释性;全切片图像

**【中图分类号】**R318;TP183

**【文献标志码】**A

**【文章编号】**1005-202X(2025)02-0175-09

## ABMIL-BiGRU: bidirectional gated recurrent unit attention based multi-instance learning for interpretable prediction of sentinel lymph node metastasis in breast cancer

LI Bo<sup>1</sup>, YANG Yanbin<sup>2</sup>, LI Shuai<sup>3</sup>, LIANG Meiyan<sup>3</sup>

1. Department of Physical Therapy, Shanxi Rongjun Hospital, Taiyuan 030031, China; 2. Department of Rehabilitation, Shanxi Civil Administration Rehabilitation Hospital, Taiyuan 030032, China; 3. College of Physics and Electronic Engineering, Shanxi University, Taiyuan 030006, China

**Abstract:** Aiming at the classification and lesion localization of giga-pixel pathology whole slide images of breast cancer, a bidirectional gated recurrent unit attention based multi-instance learning (ABMIL-BiGRU) model is proposed for interpretable prediction of H&E stained breast cancer lymph node metastasis images. The method uses two orthogonal bidirectional gated recurrent units to establish the long-short distance dependencies between the features in the row and column directions of the image block, thereby embedding the spatial position and context information of the image block, and then quantifies the attention score of each feature representation through attention multi-instance pooling, thereby achieving whole slide image-level feature aggregation and generating interpretable heat maps. The results show that ABMIL-BiGRU model has an average accuracy of 0.918 6 and an AUC of 0.946 7 on the breast cancer metastasis dataset, realizing high-precision prediction of whole slide images and localization of regions of interest, and also providing human-interpretable features at the image block level. The proposed model solves the "accuracy-interpretability trade-off" problem to a certain extent, and its superior performance provides a new paradigm for the clinical application of computer-aided diagnosis and intelligent systems.

**Keywords:** breast cancer; lymph node metastasis; precise diagnosis; bidirectional gated recurrent unit; contextual information; interpretability; whole slide image

**【收稿日期】**2024-10-11

**【基金项目】**国家自然科学基金(11804209);山西省自然科学基金(202103021223411)

**【作者简介】**李波,硕士,主管治疗师,研究方向:神经康复及肿瘤康复, E-mail: mvplibo@163.com

**【通信作者】**梁美彦,博士,硕士生导师,研究方向:医学图像处理, E-mail: meiyianliang@sxu.edu.cn

## 前言

世界卫生组织国际癌症研究机构(IARC)发布数据显示,2020年全球女性乳腺癌新增病例数达到226万人,已经超越肺癌成为全球第一大癌症。其中,乳腺癌导致的死亡人数为68.5万人,占女性每年癌症

死亡人数的15.5%<sup>[1]</sup>。精准诊断有助于在癌症早期阶段发现病变,及时制定个性化治疗及康复方案,提高治疗成功率和患者生存率。目前,组织病理学检查依然是乳腺癌及淋巴结转移的“金标准”。然而,单张病理全切片图像存在尺寸较大、不能压缩的特点,因此完全依靠医生肉眼阅片的方式,不仅检测周期长而且严重依赖医生的主观判断,导致检测结果存在不同程度的误差,这种方法对重症病人来说容易错过最佳治疗时机<sup>[2]</sup>。因此,迫切需要研发计算机辅助智能诊断系统,以实现具有千兆像素病理学全切片图像的自动分析和病灶定位。

现有对病理全切片图像进行自动分析的方法包括监督学习和弱监督学习。监督学习算法使用像素级别的标签来构建深度学习模型,其在大量的计算机视觉任务包括图像分割和病灶边界定位中已取得最先进的性能<sup>[3-6]</sup>。然而其需要通过手工标注不仅费时费力,而且所需要的像素级标记很难在临床中应用。基于弱监督学习的算法,仅需要全切片图像级的粗粒度标注,就可以通过学习算法实现标签预测和定位,并通过热图的方式为识别结果给出解释性,因此基于弱监督学习的框架已经成为计算病理学领域的研究热点<sup>[7-9]</sup>。在全切片图像的预测和分类任务中,常用的弱监督学习方法是基于多示例学习(Multiple Instance Learning, MIL)的病理全切片预测,然而大多数基于MIL方法都基于示例独立同分布假设,未能利用示例间的空间关系形成有效的上下文感知信息,导致模型只能定位部分高分示例,这将导致模型的可解释性较差,与病理医生的检查结果不一致。因此,如何有效地挖掘并利用图像块之间隐含的关联性建立上下文信息,对构建更加精确的病理全切片推理预测模型具有重要的意义和临床实用价值。

## 1 国内外研究现状及分析

为了有效地建立并利用示例间的依赖关系,研究人员尝试在MIL模型中分别引入示例级约束、多尺度信息,以模拟全切片图像中示例之间的局部和全局相关性,获取图像块隐含的上下文信息从而改进模型的整体性能。(1)研究者试图在MIL模型中引入示例级约束来提升模型的识别效果<sup>[10-16]</sup>。该方法是在模型迭代过程中,选择一定数量的特征图像块并加入约束条件来辅助模型分类。然而,利用该方法建模时,实际参与训练的图像块只占病理全切片的一小部分,因此需要大量的全切片图像才有可能得到一个性能较好的模型。2019年,Campanella等<sup>[17]</sup>提出基于CNN-RNN结构的MIL模型,该方法挑

选出一定数量的高分图像块送入RNN网络聚合图像块特征并对全切片图像标签做出预测。2020年,Zhao等<sup>[18]</sup>提出将VAE-GAN作为图像块特征提取器,通过选择具有区分性的图像块特征实现结肠病理学图像淋巴结转移预测。2021年,Lu等<sup>[19]</sup>提出基于聚类约束的数据高效性注意力MIL框架,利用注意力机制来学习全切片图像中具有高诊断价值的关键示例,采用示例级约束为网络提供一定的上下文信息。(2)受病理专家阅片流程的启发,通过多个尺度的特征作为辅助信息来提高MIL框架的预测性能。2020年,Hashimoto等<sup>[20]</sup>提出一种结合多尺度、领域对抗性的淋巴瘤亚型分类的MIL方法,在196例临床病例组成的数据集上的平均准确率和召回率分别为0.871和0.813。2021年,Li等<sup>[21-22]</sup>提出一种双流自监督对比学习的MIL网络,根据金字塔融合机制将来自不同尺度的高级特征进行特征映射和融合,进行全切片图像级特征聚合和分类,分类准确率和AUC值分别为0.8992和0.9165。然而,基于多尺度信息辅助的MIL方法,其性能在很大程度上取决于多尺度信息的表示方法和融合规则。也有相关研究表明,当多尺度信息不能充分利用时,模型性能反而会变差,因此不同尺度特征的有效结合也是一个重要的问题<sup>[23-25]</sup>。

综上所述,在全切片图像标签预测和感兴趣区域定位任务中,研究的目标都是如何有效地利用每个图像块的上下文关系来驱动模型获得更高的预测精度和更准确的定位精度。为此,有些学者提出基于图卷积网络的MIL方法对数字病理学全切片图像识别<sup>[26]</sup>,该方法可以有效利用全切片图像中每个图像块的上下文关系,随着图卷积层的加深动态调整每个示例与一阶邻居的连接强度关系,因而得到较高的识别性能。然而,这种方法在网络持续加深的时候,存在节点间的标签过度平滑问题,不能很好地建立图像块之间的长期依赖关系,这大大限制网络的示例级推理能力。因此,现有的图网络的MIL框架一般采用2~3层的结构。然而,这种结构又不能很好地捕获全切片图像的语义特征。为了更充分地建立图像块的上下文关系,进一步提高数字病理学全切片图像的识别与分类效果,本文提出基于双向门控循环注意力多示例学习(Bidirectional Gated Recurrent Unit Attention-Based Multiple Instance Learning, ABMIL-BiGRU)的病理全切片图像标签预测与病灶定位方法。该方法将图像块看作水平和竖直方向的图像块序列,利用双向门控循环单元(GRU)分别建立水平和竖直方向相邻图像块特征之间的长短距离依赖关系,对两个方向图像块提取的

特征进行拼接后,再利用基于门控注意力机制的多示例池化网络量化每个图像块拼接特征的注意力分数并进行全切片图像级特征融合,最终实现乳腺病理学全切片图像的高精度识别与分类。

本文方法的创新点如下:(1)提出基于ABMIL-BiGRU的乳腺癌淋巴结转移预测方法,来模拟整个切片图像中病理补丁的局部和全局拓扑结构,它仅通过两个互相正交的双向GRU就可以充分建立病理全切片中每个图像块与其一阶以及高阶邻居的空间关系,使得模型可以有效利用上下文信息,来对全切片图像进行建模,预测性能也得到较大的提升。(2)在MIL中,示例独立同分布假设前提下,单个补丁的错误分类可能会改变切片标签的预测,从而导致大量的假阴性和假阳性。本文采用ABMIL-BiGRU框架为每个补丁嵌入相邻示例序列的空间位置,可以通过示例序列在两个方向的长短距离依赖机制降低实例的假阳性率和假阴性率。(3)ABMIL-BiGRU模型将全切片图像的上下文位置嵌入和基于注意力机制的池化相结合,通过梯度后向传播算法,得到可解释性热图,因此该模型可以有效地解决“精度-可解释性”的权衡问题。结果表明,H&E染色病理全切片中的感兴趣区域与像素级注释具有良好的一致性。

## 2 方法

### 2.1 预处理

给定一个全切片图像训练集  $D = \{X_i, Y_i\}_{i=1}^N$ , 其中  $X_i$  和  $Y_i$  分别是第  $i$  张病理全切片图像和对应的类标签,对于二分类问题有  $Y_i \in \{0, 1\}$ ,  $N$  是训练集全切片图像的数目。为了构造全切片图像  $X_i$  对应的多示例学习包,首先对H&E染色的全切片图像进行下采样并转换到HSV颜色空间,然后在饱和通道上使用大津阈值法对全切片图像进行自动前景分割,将H&E染色组织从背景中分离出来并通过倍率转换和图像配准得到前景,随后将全切片图像的前景分割为大小为  $256 \times 256$  的互不重叠的图像块,同时该过程还保留图像块在原全切片图像上的坐标。图1a为输入全切片图像的整个预处理过程。通过数据预处理可以得到第  $i$  个全切片图像的特征表示  $X_i = \{x_{i,j}\}_{j=1}^P$ , 其中  $x_{i,j}$  表示全切片图像  $X_i$  的第  $j$  个示例(即图像块),每个全切片图像可以被切分为  $P$  个图像块( $j=1, \dots, P$ )。全切片图像中所有图像块经过一个预训练的截断ResNet50进行特征提取后在潜在空间中的特征表示为  $H_i = \{h_{i,j}\}_{j=1}^P$ , 其中  $H_i \in R^{P \times 1024}$ ,  $h_{i,j} \in R^{1 \times 1024}$ 。如图1b所示,逐行、逐列扫描全切片图

像并对相应的图像块特征  $H_i$  按扫描顺序重排序得到行特征  $H_i^l = \{h_{i,j}^l\}_{j=1}^P$  和列特征  $H_i^c = \{h_{i,j}^c\}_{j=1}^P$ , 其中  $H_i^l \in R^{P \times 1024}$ ,  $h_{i,j}^l \in R^{1 \times 1024}$ ,  $H_i^c \in R^{P \times 1024}$ ,  $h_{i,j}^c \in R^{1 \times 1024}$ 。

### 2.2 ABMIL-BiGRU 模型

ABMIL-BiGRU模型的结构如图1c所示,该网络首先采用两个并行的GRU模块分别沿行方向(GRU 1)和列方向(GRU 2)进行图像块的单个方向上下文特征的提取,然后将这两个方向得到的特征进行融合,形成全切片图像在二维空间上的上下文特征,再将该特征输入基于注意力机制的多示例池化模块,即得到病理全切片的预测。在ABMIL-BiGRU模型中,行特征  $H_i^l$  和列特征  $H_i^c$  分别经过两个平行的双向GRU模块后得到  $H_i^{ol}$  和  $H_i^{oc}$ 。图2中以行方向特征为例阐述该方向的特征提取过程,列方向与该过程类似。如图2a所示,双向GRU模块由前向层和反向层组成。前向层和反向层由相同数量的门控循环单元组成,图中的GRU表示为  $A^1$ (或者  $A^2$ , 其中  $A^1$  和  $A^2$  结构相同)。GRU由重置门和更新门组成,直观上认为重置门决定如何将新的输入信息与前面的记忆相结合,更新门定义前面的记忆在当前时间的遗忘程度。

如图2b所示,给定输入  $H_i^l = \{h_{i,j}^l\}_{j=1}^P$ , 正向输入时记为  $H_i^{l \rightarrow} = \{h_{i,j}^{l \rightarrow}\}_{j=1}^P$ , 更新门  $z_{i,j}$  和重置门  $r_{i,j}$  的输出分别为:

$$z_{i,j} = \sigma(h_{i,j}^{l \rightarrow} W^{(z)} + h_{i,j-1}^{ol} U^{(z)}) \quad (1)$$

$$r_{i,j} = \sigma(h_{i,j}^{l \rightarrow} W^{(r)} + h_{i,j-1}^{ol} U^{(r)}) \quad (2)$$

其中,  $z_{i,j} \in R^{1 \times 256}$ ,  $r_{i,j} \in R^{1 \times 256}$ 。其中,  $W^{(z)} \in R^{1024 \times 256}$ ,  $U^{(z)} \in R^{256 \times 256}$  和  $W^{(r)} \in R^{1024 \times 256}$ ,  $U^{(r)} \in R^{256 \times 256}$  分别为GRU单元中更新门和重置门的权重参数,  $h_{i,j-1}^{ol} \in R^{1 \times 256}$  为GRU前一个输入的内存信息。GRU利用更新门和重置门存储并过滤图像块特征信息,更新门决定了将多少当前和上一个图像块特征信息继续传递,重置门决定了当前和上一个图像块特征信息的遗忘程度。GRU使用重置门计算属于当前输入的新记忆内容,新记忆内容的表达式为:

$$\tilde{h}_{i,j}^{l \rightarrow} = \tanh(h_{i,j}^{l \rightarrow} W^{(m)} + (r_{i,j} \odot h_{i,j-1}^{ol}) U^{(m)}) \quad (3)$$

其中,  $\tilde{h}_{i,j}^{l \rightarrow} \in R^{1 \times 256}$ ,  $W^{(m)} \in R^{1024 \times 256}$  和  $U^{(m)} \in R^{256 \times 256}$  为GRU计算新记忆内容时的权重参数。GRU最终的输出可记为:  $H_i^{ol \rightarrow} = \{h_{i,j}^{ol \rightarrow}\}_{j=1}^P$ 。其中:

$$h_{i,j}^{ol \rightarrow} = (1 - z_{i,j}) \odot h_{i,j-1}^{ol} + z_{i,j} \odot \tilde{h}_{i,j}^{l \rightarrow} \quad (4)$$

其中,  $H_i^{ol \rightarrow} \in R^{P \times 256}$ ,  $h_{i,j}^{ol \rightarrow} \in R^{1 \times 256}$ 。

综上,  $A^1$  和  $A^2$  分别接收行方向(或列方向)正向输入和反向输入的图像块特征,输出  $H_i^{ol \rightarrow}$  和  $H_i^{ol \leftarrow}$  并将两者拼接得到行方向(或列方向)双向GRU的



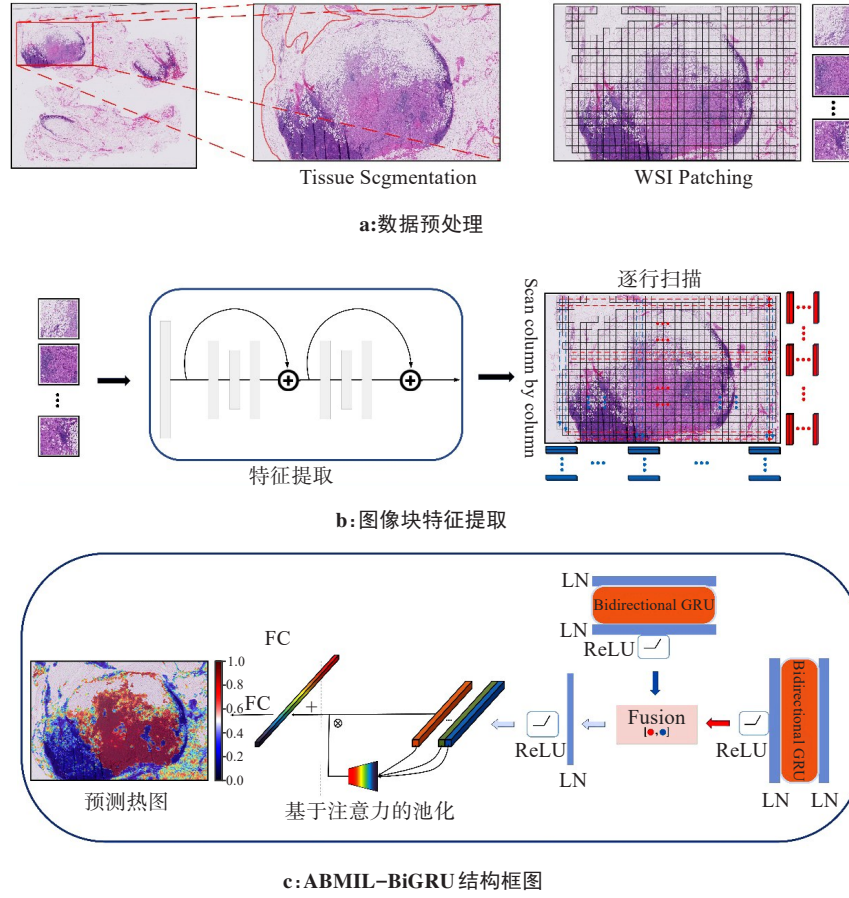


图1 数据预处理、图像块特征提取和ABMIL-BiGRU结构框图

Figure 1 Data preprocessing, image block feature extraction and ABMIL-BiGRU structure diagram

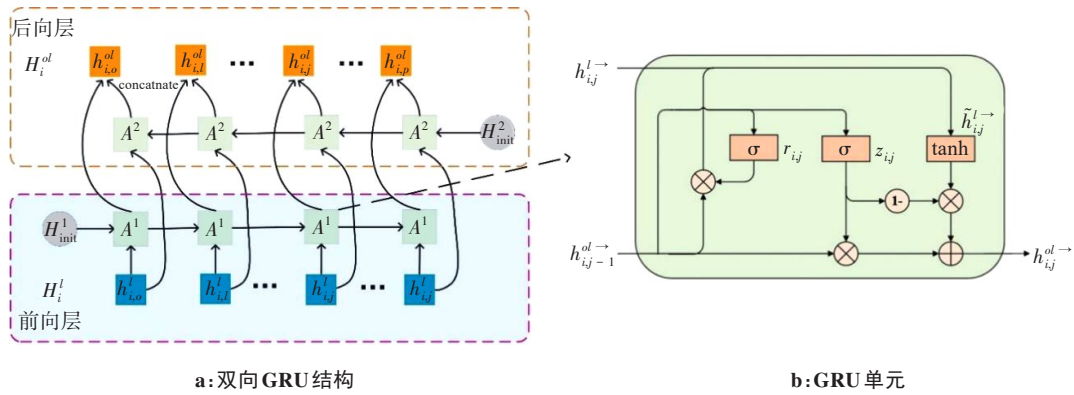


图2 双向GRU原理图

Figure 2 Schematic diagram of bidirectional GRU

输出:

$$[H_i^{ol \rightarrow}; H_i^{ol \leftarrow}] = \{h_{i,j}^{ol}\}_{j=1}^P \quad (5)$$

其中,  $[H_i^{ol \rightarrow}; H_i^{ol \leftarrow}] \in R^{P \times 512}$ , 其经过 LN 层和 ReLU 激活函数层之后便得到:

$$H_i^{ol} = \text{ReLU}(\text{LN}([H_i^{ol \rightarrow}; H_i^{ol \leftarrow}])) \quad (6)$$

同理, 按照上述计算方式可得到图像块特征按列扫描输入时的输出:

$$H_i^{oc} = \text{ReLU}(\text{LN}([H_i^{oc \rightarrow}; H_i^{oc \leftarrow}])) \quad (7)$$

其中,  $H_i^{ol}, H_i^{oc} \in R^{P \times 512}$ ,  $[\cdot; \cdot]$  代表特征拼接操作,  $H_i^{ol \rightarrow}, H_i^{ol \leftarrow}, H_i^{oc \rightarrow}, H_i^{oc \leftarrow} \in R^{P \times 256}$ 。  $H_i^{ol \rightarrow}$  和  $H_i^{ol \leftarrow}$  分别为图 1c 中双向 GRU 1 正反向两个方向的输出特征,  $H_i^{oc \rightarrow}$  和  $H_i^{oc \leftarrow}$  分别为图 1c 中双向 GRU 2 正反向两个方向的输出特征,  $H_i^{oc \rightarrow}, H_i^{oc \leftarrow}, H_i^{ol \rightarrow}$  和  $H_i^{ol \leftarrow}$  的计算方式相同。

将  $H_i^{ol}$  和  $H_i^{oc}$  两部分信息拼接后得到:

$$H_i^o = [H_i^{ol}; H_i^{oc}] \quad (8)$$

其中,  $H_i^o \in R^{P \times 1024}$ , 如图 1c 中融合操作所示。本文提出的 ABMIL-BiGRU 模型中双向 GRU 1 和 GRU 2 分别处理逐行和逐列扫描得到的图像块特征, 在较长距离上保留图像块的原始空间位置信息, 有效地建立图像块之间的依赖关系, 实现图像块空间位置信息嵌入。  $H_i^o$  通过一个全连接层和 ReLU 激活函数层后得到:

$$H_i^{out'} = \{h_{i,j}^{out'}\}_{j=1}^P = \text{ReLU}(\text{FC}(H_i^o)) \quad (9)$$

其中,  $H_i^{out'} \in R^{P \times 512}$ ,  $h_{i,j}^{out'} \in R^{1 \times 512}$ 。ABMIL-BiGRU 模型生成的属于  $X_i$  的全切片图像级特征可表示为:

$$C_i = \sum_{j=1}^P a_{i,j} h_{i,j}^{out'} \quad (10)$$

$$a_{i,j} = \frac{\exp\{w^T (\tanh(W h_{i,j}^{out'} T) \odot \text{sigm}(U h_{i,j}^{out'} T))\}}{\sum_{j=1}^P \exp\{w^T (\tanh(W h_{i,j}^{out'} T) \odot \text{sigm}(U h_{i,j}^{out'} T))\}} \quad (11)$$

其中,  $C_i \in R^{1 \times 512}$ ,  $a_{i,j} \in R$  表示图像块特征  $h_{i,j}^{out'}$  的注意力分数, 该注意力分数为可视化全切片图像解释性热图提供依据,  $w \in R^{256 \times 1}$ ,  $U \in R^{256 \times 512}$  和  $W \in R^{256 \times 512}$  为参数矩阵,  $\odot$  和  $\text{sigm}(\cdot)$  分别代表元素级乘法和 Sigmoid 函数, 上标 T 表示矩阵转置操作。全切片图像级特征表示  $C_i$  通过一个全连接层之后得到全切片图像  $X_i$  的预测结果  $\hat{Y}_i$ , 二元交叉熵损失函数被用来最小化预测标签  $\{\hat{Y}_i\}_{i=1}^N$  与真实标签  $\{Y_i\}_{i=1}^N$  之间的误差。

### 2.3 数据集与实现细节

本文实验用到的数据集是 Camelyon16, 共包括由内梅亨大学医学中心和乌得勒支大学医学中心收集到的 399 张乳腺病理学全切片图像, 包括苏木素和伊红染色全切片图像中淋巴结转移的像素级注释。训练用了全切片图像为 270 张, 其中正常组织切片 159 张, 淋巴结转移切片 111 张, 阳性和阴性切片按照 8.5:1.5 的比例分为训练集和验证集。然后用 129 张全切片图像对模型性能进行测试, 包括 80 张正常组织切片和 49 张淋巴结转移切片。在迭代过程中, 使用初始学习率为  $1 \times 10^{-4}$ 、权重衰减率为  $1 \times 10^{-6}$  的 Adam 优化器对模型进行优化。为了缓解 ABMIL-BiGRU 模型的过拟合, 采用早停策略。本文的模型是在 INVIDA RTX 3090 上训练得到的, 训练次数为 100, 批量大小为 1。

### 2.4 评价指标

本文用准确率 (Accuracy)、精确度 (Precision)、敏感度 (Sensitivity)、特异度 (Specificity) 和 F1 分值来对模型进行评价, 这些评价指标的公式如下:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{TN} + \text{FN}} \quad (12)$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (13)$$

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (14)$$

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}} \quad (15)$$

$$\text{F1} = \frac{2\text{Precision} \cdot \text{Sensitivity}}{\text{Precision} + \text{Sensitivity}} \quad (16)$$

其中, TP (True Positive) 表示正确分类的正样本; FP (False Positive) 表示错误分类的正样本; TN (True Negatives) 表示正确分类的负样本; FN (False Negatives) 表示错误分类的负样本。

此外, 采用评价指标 ROC 能够更直观地解释分类器的性能, 它是根据一系列不同的分界值或决定阈, 以真阳性率 (Sensitivity) 为纵坐标, 假阳性率 (1-Specificity) 为横坐标绘制的曲线。AUC 是 ROC 曲线下面积。

## 3 结果与分析

表 1 给出 ABMIL-BiGRU 模型在 Camelyon16 数据集上的识别与分类效果, 表中所有模型均重复实验 5 次。与其他模型相比, ABMIL-BiGRU 模型在表 1 给出的 6 个评价指标中有 5 个表现较好, 其平均准确率、AUC 值和 F1 分数值较优化后的 ABMIL-GCN 模型分别提高 0.97%、3.18% 和 1.08%。此外, ABMIL-BiGRU 模型的平均精确率和特异度均达到 0.98 以上, 较优化后的 ABMIL-GCN 模型分别提高 4.18% 和 2.19%, 其中特异度的平均值可达到 0.993 8。上述指标对比均表明本文提出的 ABMIL-BiGRU 模型在识别乳腺病理学全切片图像中的优越性。

表 1 还比较了 ABMIL-BiGRU 模型与现有的弱监督 MIL 方法的性能。例如基于 MIL 的 Mean-pooling 和 Max-pooling、ABMIL、CLAM 和 ABMIL-GCN 在乳腺淋巴结转移数据集上的性能。结果表明, 使用 Mean-pooling 和 Max-pooling 等传统池化算子进行 MIL 模型不能达到很好的性能, 这是因为传统方法在对全切片图像中的图像块进行处理时, 示例的独立同分布假设未能建立示例的上下文关系。而当采用基于注意力的 MIL 池化 (ABMIL) 时, 它可以通过实例间的注意力机制来提高模型的性能。对于 CLAM 模型, 采用实例级聚类约束在一定程度上建立示例的上下文关系, 与基于注意力机制的池化操作相结合, 可以提高对全切片图像的预测性能。然而, 在实际中每个全切片图像的图像块在空间上都有固定的排列, 这意味着相邻的图像块在空间上是相关的。基于示例独立同分布假设的 MIL 框架并

表 1 Camelyon16数据集不同方法结果对比  
Table 1 Results comparison of different methods on Camelyon16 dataset

方法	准确率	AUC	精确度	敏感度	特异度	F1 分数
Mean-pooling	0.598 4±0.031 2	0.538 7±0.003 2	0.499 0±0.110 2	0.261 2±0.111 3	0.805 0±0.116 9	0.314 9±0.084 6
Max-pooling	0.641 9±0.029 2	0.701 1±0.013 1	0.764 2±0.167 0	0.197 3±0.171 0	0.937 5±0.056 8	0.264 0±0.196 6
ABMIL	0.823 3±0.011 4	0.858 1±0.013 4	0.940 0±0.011 9	0.571 4±0.034 2	0.977 5±0.005 0	0.710 1±0.025 5
CLAM <sup>[19]</sup>	0.829 5±0.008 5	0.872 9±0.012 4	0.919 9±0.054 0	0.608 2±0.027 1	0.965 0±0.027 8	0.730 4±0.007 6
ABMIL-GCN(没有洪水水位) <sup>[26]</sup>	0.882 2±0.009 0	0.893 1±0.009 9	0.905 0±0.023 8	0.771 4±0.015 3	0.950 0±0.013 7	0.832 6±0.012 3
ABMIL-GCN(有洪水水位) <sup>[26]</sup>	0.908 9±0.003 3	0.914 9±0.004 1	0.946 3±0.009 3	0.806 1±0.010 2	0.971 9±0.005 4	0.870 5±0.005 1
ABMIL-BiGRU	0.918 6±0.006 7	0.946 7±0.005 9	0.988 1±0.020 6	0.795 9±0.020 4	0.993 8±0.010 8	0.881 3±0.010 4

不完全适用于全切片图像的预测。而 ABMIL-GCN 虽然可以通过图结构建立全切片图像中图像块的上下文关系,但是图神经网络是通过邻接矩阵来表示全切片图像中每个图像块与周围一阶邻居图像块的连接关系,对于图像块二阶邻居以及高阶邻居的关系,随着图神经网络深度的增加,节点之间的标签会表现出平滑的趋势。而全切片图像行方向和列方向的像素都在 10 K 以上,因此图神经网络不能充分利用图像块的上下文特征来进行标签预测和感兴趣区域定位,因而也不完全适用于全切片图像的预测。而 ABMIL-BiGRU 只通过两个相互正交的双向 GRU 分别建立图像块行方向和列方向特征间的长距离依赖关系,从而实现图像块上下文信息的嵌入,这种方法只采用两层网络就可以实现较大的感受野,因此,ABMIL-BiGRU 的性能明显优于 CLAM 和 ABMIL-GCN。为了进一步验证 ABMIL-BiGRU 模型的优越

性,下面将主要从混淆矩阵图、TSNE 特征可视化结果、ROC 曲线图和模型可解释性 4 个方面依次展开对比并讨论。

图 3 为 ABMIL-BiGRU 模型、ABMIL-GCN(有洪水水位)模型和 CLAM 模型的混淆矩阵图。从图 3 可知,CLAM 的效果相对较差,其原因是 CLAM 中的是在示例独立同分布假设的基础上加入示例级约束,而这种约束不是所有节点都有效的。对于 ABMIL-BiGRU 模型和 ABMIL-GCN 模型,二者的假阴性率相同,而这两个模型的假阳性率分别为 0 和 2.5%,即 ABMIL-BiGRU 模型的假阳性率较 ABMIL-GCN 模型降低 2.5%。比较重要的是,ABMIL-BiGRU 模型对测试集中所有阴性全切片图像的预测结果均正确。在临床应用中,ABMIL-BiGRU 模型能极大地降低误诊率,这对乳腺病理学全切片图像的高精度分类具有重要意义。

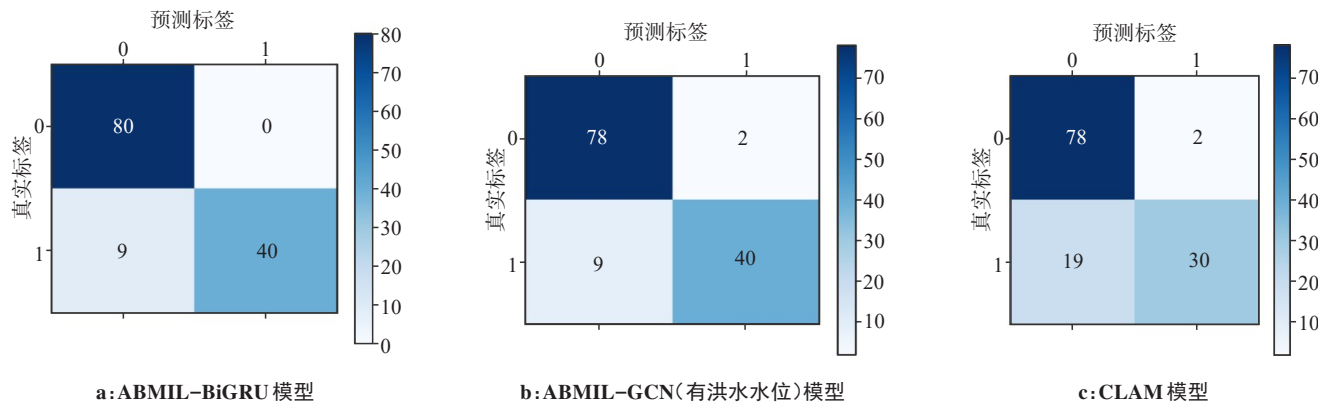


图 3 ABMIL-BiGRU 模型、ABMIL-GCN(有洪水水位)模型和 CLAM 模型的混淆矩阵图  
Figure 3 Confusion matrixes of ABMIL-BiGRU, ABMIL-GCN (w/flooding) and CLAM models

测试全切片图像在潜在空间中的分布是衡量模型分类性能的重要指标,它可以通过降维的方法可视化模型的分类性能。图 4 为 ABMIL-BiGRU 模型、

ABMIL-GCN(有洪水水位)模型以及 CLAM 模型的 TSNE 结果图。由图可知,CLAM 模型的预测精度虽然也可以达到 82.95%,然而在潜在空间中的同一类



并没有聚合在一起,而是距离较远的两个分簇,因而而这种方法的预测性能和感兴趣区域定位的性能有限。而 ABMIL-BiGRU 模型的 TSNE 结果图中两类别特征点之间的距离更加紧凑,且克服 ABMIL-GCN 模型中正常乳腺病理学全切片图像间距离较远的缺点,从而印证表 1 中 ABMIL-BiGRU 模型的平均特异度和召回率高的结果。从图 4 可以看出,ABMIL-BiGRU 模型得到的结果图类内距离更小,类间距离更大,这

是因为 ABMIL-BiGRU 模型在行、列两个方向使用双向 GRU 网络建立图像块特征间的联系,该结构较 GCN 网络能获得更大的感受野,更丰富上下文信息。双向 GRU 网络将输入的图像块特征看作一个序列,能够有效地建立图像块特征之间的长短距离依赖关系、降低分析图像块特征时的假阳性率和假阴性率,并提升癌症区域面积较小的阳性全切片图像的识别与分类效果。

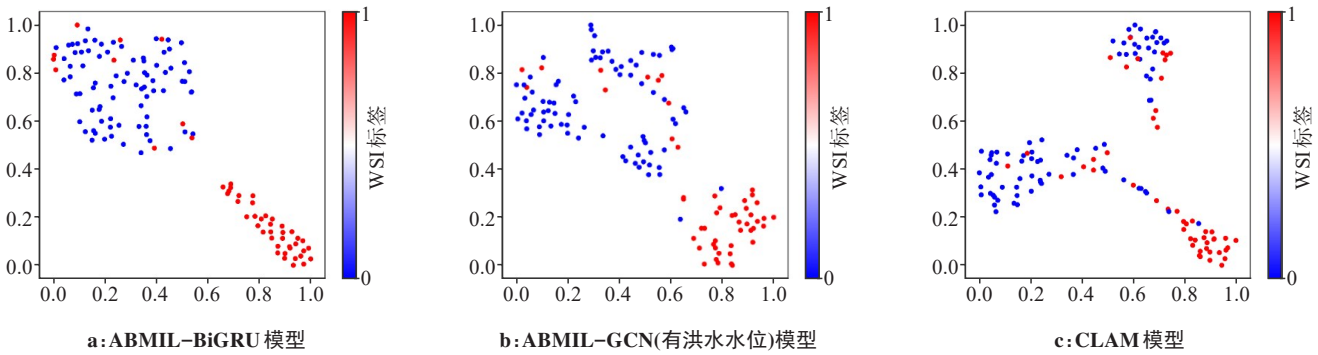


图 4 ABMIL-BiGRU 模型、ABMIL-GCN(有洪水水位)模型和 CLAM 模型的 TSNE 结果图  
Figure 4 TSNE results of ABMIL-BiGRU, ABMIL-GCN (w/flooding) and CLAM models

图 5 所示为表 1 中所有模型的 ROC 曲线,通过计算 AUC 可知,ABMIL-BiGRU 模型的 AUC 值最高可以达到 0.956 6,较优化后 ABMIL-GCN 模型提高 0.034 7,表明 ABMIL-BiGRU 模型较 ABMIL-GCN 模型能获得更多的上下文感知信息,进一步体现 ABMIL-BiGRU 模型的优越性。从图 5 可知,空间位置信息在乳腺病理学全切片图像识别与分类任务中的重要性,例如加入示例级约束的 CLAM 和 ABMIL 模型的 AUC 值均在 0.85 以上,而基于示例独立同分布的 Mean-pooling 和 Max-pooling 方法模型的 AUC 值均在 0.75 以下。

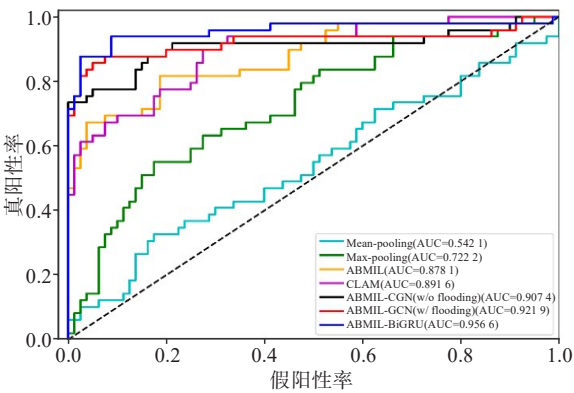


图 5 所有模型的 ROC 曲线图  
Figure 5 ROC curves of these models

图 6 为 ABMIL-BiGRU、ABMIL-GCN 和 CLAM 模型生成的可解释性热图以及对应全切片图像的真实标签图。图 6 中从上至下依次为 ABMIL-BiGRU 模型、ABMIL-GCN 模型、CLAM 模型生成的可解释性热图以及全切片图像的真实标签图,图 6 中第 2 列和第 4 列分别为第 1 列和第 3 列中红色框选中区域的放大图。上述 3 种模型对第 1 列中全切片图像的预测结果均为阳性,预测结果均正确,但通过对第 1 列和第 2 列图像观察可知,ABMIL-BiGRU 模型生成的热图与真实标签最吻合,ABMIL-GCN 模型次之。ABMIL-BiGRU 模型生成热图的淋巴结转移区域与正常组织区域之间的边界更清晰,淋巴结转移区域内图像块都能被准确定位。这表明 ABMIL-BiGRU 模型中两个正交方向的双向 GRU 网络结构能够在最大程度上保留图像块在原始全切片图像中的空间位置信息,并实现图像块特征的长短距离依赖关系和空间位置信息嵌入。虽然 ABMIL-GCN 模型也利用 GCN 建立图像块之间的联系并得到一定的上下文感知信息,但是这种联系是短距离的,得到的空间位置信息也是有限的,导致模型不足以实现更为精确的预测结果。此外,CLAM 模型对第 3 列中阳性全切片图像的预测结果为阴性,ABMIL-BiGRU 模型和 ABMIL-GCN 模型对其预测结果为阳性。同时,

ABMIL-BiGRU 模型不仅能够有效地预测淋巴结转移区域较大的全切片图像,而且对于淋巴结转移区域较小的全切片图像,其依然能够区分淋巴结转移

区域与正常组织区域间的微小差异并给出正确的预测结果,且比 ABMIL-GCN 模型的预测和定位结果更加精确。

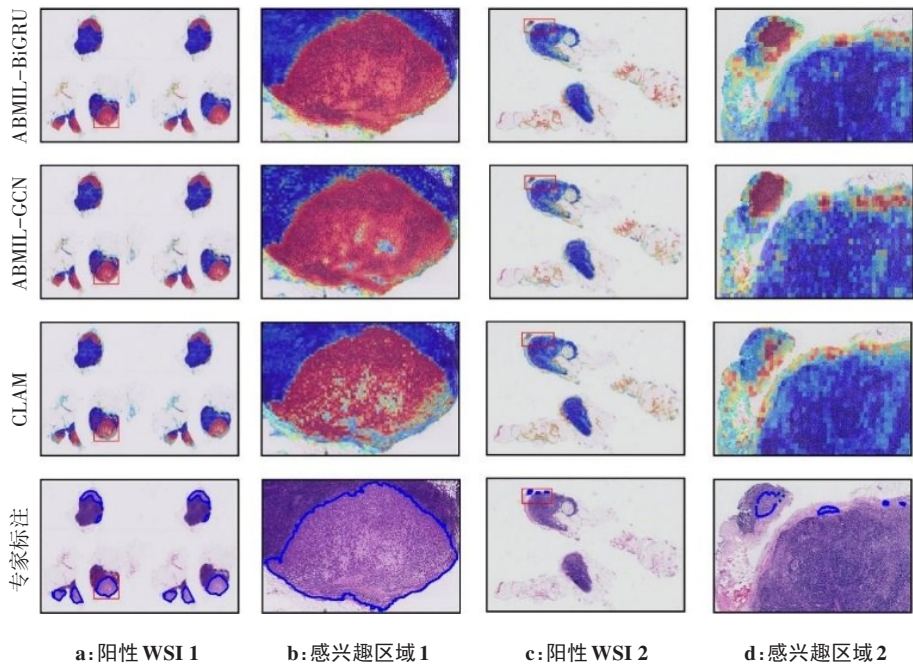


图6 全切片图像解释性热图  
Figure 6 Interpretative heat maps of whole slide images

4 结 论

本文将病理全切片中的图像块分别看成行方向和列方向的两个序列,提出基于并行双向GRU的特征提取方法,分别建立图像块行方向和列方向上每个示例的上下文依赖关系,再将这两个并行双向GRU得到的特征,通过融合机制来得到每个图像块在行方向和列方向的上下文关系,从而实现图像块的特征嵌入。最后再结合基于门控注意力机制的池化网络量化每个图像块拼接特征的注意力分数,实现全切片图像级特征融合,在得到全切片图像淋巴结转移识别与分类结果的同时生成可解释性热图。通过与其他弱监督MIL方法的对比,ABMIL-BiGRU模型的5次实验平均准确率、AUC值、精确度、特异度和F1分数值分别达到0.9186、0.9467、0.9881、0.9938和0.8813。ABMIL-BiGRU模型生成的可解释性热图与病理学专家诊断以及真实标签更加吻合,进一步证明ABMIL-BiGRU模型对乳腺病理学全切片图像识别与分类的有效性。

【参考文献】

[1] Sung H, Ferlay J, Siegel RL, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries[J]. CA Cancer J Clin, 2021, 71(3): 209-249.

[2] Marini N, Marchesin S, Otálora S, et al. Unleashing the potential of digital pathology data by training computer-aided diagnosis models without human annotations[J]. NPJ Digit Med, 2022, 5(1): 102.

[3] van Rijthoven M, Balkenhol M, Siliņa K, et al. HookNet: multi-resolution convolutional neural networks for semantic segmentation in histopathology whole-slide images[J]. Med Image Anal, 2021, 68: 101890.

[4] Pedersen A, Smistad E, Rise TV, et al. Hybrid guiding: a multi-resolution refinement approach for semantic segmentation of gigapixel histopathological images[EB/OL]. (2021-12-07). <https://arxiv.org/abs/2112.03455>.

[5] Feng YB, Hafiane A, Laurent H. A deep learning based multiscale approach to segment cancer area in liver whole slide image[EB/OL]. (2020-07-25). <https://arxiv.org/abs/2007.12935>.

[6] Feng RW, Liu XC, Chen JT, et al. A deep learning approach for colonoscopy pathology WSI analysis: accurate segmentation and classification[J]. IEEE J Biomed Health Inform, 2021, 25(10): 3700-3708.

[7] Lu MY, Chen TY, Williamson DF, et al. AI-based pathology predicts origins for cancers of unknown primary[J]. Nature, 2021, 594(7861): 106-110.

[8] Yao JW, Zhu XL, Jonnagaddala J, et al. Whole slide images based cancer survival prediction using attention guided deep multiple instance learning networks[J]. Med Image Anal, 2020, 65: 101789.



- [9] Tu M, Huang J, He XD, et al. Multiple instance learning with graph neural networks [EB/OL]. (2019-06-12). <https://arxiv.org/abs/1906.04881>.
- [10] Wang SJ, Zhu YX, Yu LQ, et al. RMDL: recalibrated multi-instance deep learning for whole slide gastric image classification[J]. Med Image Anal, 2019, 58: 101549.
- [11] Xu G, Song ZG, Sun Z, et al. CAMEL: a weakly supervised learning framework for histopathology image segmentation[C]//2019 IEEE/CVF International Conference on Computer Vision (ICCV). Piscataway, NJ, USA: IEEE, 2019: 10681-10690.
- [12] Lerousseau M, Vakalopoulou M, Classe M, et al. Weakly supervised multiple instance learning histopathological tumor segmentation[C]//Medical Image Computing and Computer Assisted Intervention - MICCAI 2020. Cham: Springer International Publishing, 2020: 470-479.
- [13] Chikontwe P, Kim M, Nam SJ, et al. Multiple instance learning with center embeddings for histopathology classification [C]//Medical Image Computing and Computer Assisted Intervention - MICCAI 2020. Cham: Springer International Publishing, 2020: 519-528.
- [14] Cheng SH, Liu SB, Yu JY, et al. Robust whole slide image analysis for cervical cancer screening using deep learning[J]. Nat Commun, 2021, 12(1): 5639.
- [15] Kanavati F, Toyokawa G, Momosaki S, et al. Weakly-supervised learning for lung carcinoma classification using deep learning[J]. Sci Rep, 2020, 10(1): 9297.
- [16] Ianni JD, Soans RE, Sankarapandian S, et al. Tailored for real-world: a whole slide image classification system validated on uncurated multi-site data emulating the prospective pathology workload[J]. Sci Rep, 2020, 10(1): 3217.
- [17] Campanella G, Hanna MG, Geneslaw L, et al. Clinical-grade computational pathology using weakly supervised deep learning on whole slide images[J]. Nat Med, 2019, 25(8): 1301-1309.
- [18] Zhao Y, Yang F, Fang YQ, et al. Predicting lymph node metastasis using histopathological images based on multiple instance learning with deep graph convolution[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ, USA: IEEE, 2020: 4836-4845.
- [19] Lu MY, Williamson DF, Chen TY, et al. Data-efficient and weakly supervised computational pathology on whole-slide images[J]. Nat Biomed Eng, 2021, 5(6): 555-570.
- [20] Hashimoto N, Fukushima D, Koga R, et al. Multi-scale domain-adversarial multiple-instance CNN for cancer subtype classification with unannotated histopathological images [C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ, USA: IEEE, 2020: 3851-3860.
- [21] Li B, Li Y, Eliceiri KW. Dual-stream multiple instance learning network for whole slide image classification with self-supervised contrastive learning[C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ, USA: IEEE, 2021: 14313-14323.
- [22] Xiang TG, Song Y, Zhang CY, et al. DSNet: a dual-stream framework for weakly-supervised gigapixel pathology image analysis[J]. IEEE Trans Med Imaging, 2022, 41(8): 2180-2190.
- [23] Marini N, Otálora S, Ciompi F, et al. Multi-scale task multiple instance learning for the classification of digital pathology images with global annotations [C]//Proceedings of the MICCAI Workshop on Computational Pathology. Chia Laguna Resort, Sardinia, Italy: PMLR, 2021: 170-181.
- [24] Thandiackal K, Chen BQ, Pati P, et al. Differentiable zooming for multiple instance learning on whole-slide images[EB/OL]. (2022-07-26). <https://arxiv.org/abs/2204.12454>.
- [25] Hou WT, Yu LQ, Lin CX, et al. H<sup>2</sup>-MIL: exploring hierarchical representation with heterogeneous multiple instance learning for whole slide image analysis[C]//Thirty-Sixth AAAI Conference on Artificial Intelligence Thirty-Fourth Conference on Innovative Applications of Artificial Intelligence the twelfth Symposium on Educational Advances in Artificial Intelligence. Palo Alto, CA, USA: AAAI Press, 2022: 933-941.
- [26] Liang MY, Chen QH, Li B, et al. Interpretable classification of pathology whole-slide images using attention based context-aware graph convolutional neural network[J]. Comput Methods Programs Biomed, 2023, 229: 107268.

(编辑:陈丽霞)