

基于分级注意力的多示例口腔癌病理分类

江慧敏, 方立铭, 陶龙

皖南医学院医学影像学院, 安徽 芜湖 241000

【摘要】针对病理数据超大尺寸、标注成本高昂等问题带来的病理分类准确率较低的问题,基于多示例网络,引入分级注意力模块,设计一种兼顾示例级和图像块级损失的病理分类算法。回顾性收集皖南医学院第一附属医院口腔颌面外科186例口腔癌(126例鳞癌、60例腺癌),其数字病理切片划分为验证集、测试集及训练集。首先对病理图像进行前后背景分割,去掉背景中的噪声部分,然后采用ResNet50对分割后的病理图像提取特征,并将特征输入第一级注意力网络,得到基于图像块的注意力得分和损失,再根据注意力得分对图像块进行排序重置标签输入第二级注意力网络,得到基于示例级别的损失,最后将两级注意力的损失作为模型的总损失,通过训练最终网络,得到口腔癌分类结果。实验结果表明,使用两级注意力的多示例网络准确率为78.95%,AUC为0.8430,相较于基线模型均有更高表现。

【关键词】口腔癌;多示例学习;两级注意力;病理图像

【中图分类号】R780.2;R318

【文献标志码】A

【文章编号】1005-202X(2024)08-0946-07

Pathological classification of oral cancer based on multi-instance network and two-level attention

JIANG Huimin, FANG Liming, TAO Long

Department of Medical Imaging, Wannan Medical College, Wuhu 241000, China

Abstract: To address the problems of the low accuracy of pathological classification caused by the large size of pathological data and the high cost of labeling, a pathological classification algorithm for oral cancer is designed based on multi-instance network and two-level attention module, which takes losses at the instance level and image block level into account. A retrospective analysis is conducted on 186 cases of oral cancer (126 cases of squamous cell carcinoma and 60 cases of adenocarcinoma) in the Department of Oral and Maxillofacial Surgery, the First Affiliated Hospital of Wannan Medical College, and the digital pathological sections are divided into training set, verification set and test set. The foreground and background segmentations are performed on the pathological image, and the noise is removed from the background. ResNet50 is used to extract features from the segmented pathological images, and the features are input into the first-level attention network to obtain the attention score and loss based on image block. Then, the image blocks are sorted according to the attention score, and the reset labels are input into the second-level attention network to obtain the loss based on the instance level. The loss of the two-level attention is taken as the total loss of the model, and the prediction result is obtained by training the final network. The experimental results show that the multi-instance network using two-level attention achieves an accuracy of 78.95% and AUC of 0.843 0, demonstrating superior performance than the baseline models.

Keywords: oral cancer; multi-instance learning; two-level attention; pathological image

前言

口腔癌约占全身肿瘤的5%,并且约有10%的患者在初诊时已存在远处转移现象,病死率高^[1-2],主要

原因在于口腔癌早期缺乏明显特征,导致部分患者确诊时已处于进展期或晚期,使得治疗方式受限,预后不良,另一方面由于鳞癌和腺癌不同的形成原因和作用机理,针对两种不同癌变的治疗方式也有很大区别^[3-4],尽早确认癌症的种类对于治疗有重大意义。活检是临床上针对鳞癌和腺癌筛查诊断最常用的分析方法,病理图像分析在现代医学特别是癌症治疗方面至关重要,一般将病理结果作为临床诊断的金标准^[5-6]。组织病理切片分析是一项耗时耗力的工作,一方面对于组织病理分析者有一定的要求,另

【收稿日期】2024-02-15

【基金项目】国家自然科学基金(11801199);安徽省自然科学基金(1908085QA30);皖南医学院中青年重点科研项目(WK2023ZZD04)

【作者简介】江慧敏,讲师,研究方向:医学图像,E-mail: 814891714@qq.com

一方面低分化患者从病理图像上较难分辨。

全景数字化病理切片(Whole Slide Image, WSI)的普及为计算机辅助诊疗开拓新局面^[7-8]。近年来,深度学习技术已成为分析医学图像的首选方法,包括病理图像分类^[9]、病理图像分割^[10]等。组织病理学数字化使使用深度学习技术对日常工作流程进行自动化分析成为可能,如癌症区域检测^[11]、细胞形态分类^[12]和患者生存预测^[13]等。数字化病理超高的分辨率和大量伪影的存在为计算机辅助病理治疗带来巨大的挑战^[14]。在实际的医疗环境中,临床数据集中的WSI标签存在于患者级别,注释很少甚至没有,但是大量的数据是深度学习的基础^[15]。大部分病理图像分析采用多实例方法^[16-18],在多实例学习中,每一个WSI是一个包括了来自同一个切片的数千张没有标签的实例的包,在这个包中若有实例被诊断为阳性则该包被看作阳性,否则看作阴性。多实例学习作为一种弱监督方法,在临床中取得较好的效果,但其在实例级别准确率较低,同时该方法无法捕获较长距离的实例依赖关系,这使得将模型真正应用于临床中缺乏一定的可解释性^[19]。针对该问题,近年来有研究者通过将池化层替换为线性注意力池化层^[20]或者聚类注意力池化层^[21],通过注意力模块将传统模型中不可以训练的全局池化改变成一个权重可以训练的可加权平均的聚合层^[22],使得网络可以更多关注全局信息。一般的注意力机制主要是衡量一个小块和整个小块序列嵌入之间的重要性程度,而病理分析往往需要建立基于示例和图像块之间的关联,因此传统的注意力机制不能满足需求,需通过引入分级注意力模块,以第一级注意力得分作为指导,第二级注意力为连接,兼顾病理切分后基于示例和图像块的损失,有效提高病理分类的准确性。

近年来组织病理图像的深度学习模型日益丰富,但是大多数研究主要针对乳腺癌^[23]、前列腺癌^[24]、胃癌^[25]、肺癌^[26]等,关于口腔的组织病理图像研究较少。目前已有部分研究实现人工智能对口腔组织病理图像诊疗辅助的工作^[27],如从原位口腔组织病理图像中自动识别角化和角蛋白珍珠区域^[28],使用线性层神经网络分类器进行区域分类^[29]。Gupta等^[30]使用迁移学习概念对4种类型的口腔鳞状细胞癌(Oral Squamous Cell Carcinoma, OSCC)进行分类,并使用4种预训练模型的深度卷积神经网络(Deep Convolutional Neural Network, DCNN)模型,在该模型中替换了全连接层并进行随机权重初始化,以便从口腔组织病理学图像中重新学习。上述研究主要是针对囊肿等口腔病变,目前关于口腔癌的智能病理诊断仍有待探索。本研究回顾性收集

2017~2023年皖南医学院第一附属医院口腔颌面外科186例患者病理切片,构建基于分级注意力机制的口腔癌病理分类网络有效捕获图像块与全切片中关于口腔癌的相关特征,有效提高口腔癌患者诊断准确率。

1 资料与方法

1.1 临床资料

本研究使用的数据来自皖南医学院第一附属医院,回顾性收集2017~2023年在口腔颌面外科接受口腔癌手术的186例患者病理切片602张,研究通过皖南医学院伦理委员会批准[(2022)伦审研第98号]。以患者病理报告中的检查结果作为鳞癌、腺癌判断的金标准,其中60例腺癌(病理切片210张)、126例鳞癌(病理切片392例)。患者染色样本均为苏木精伊红(H&E)染色,根据10年以上临床经验病理医生依据组织的情况将患者分成鳞癌或腺癌。样品通过滨松数字扫描仪扫描成WSI,保存为.ndpi格式。

1.2 方法

1.2.1 前背景分割与预处理 对病理图像进行前后背景分割,图1显示了WSI预处理流程。首先,将.ndpi格式文件输入网络,将原有色彩空间转化变成HSV色彩空间,通过色彩空间转变使得病灶区域更为突出。然后,使用openslide中的通道阈值和中值滤波方法计算出二值掩模图,光滑组织边缘,最终得到病理图像前景和背景的分割。其中,通道阈值分别为孔洞过滤阈值(默认值为10)过滤前景区域天然孔洞和细分阈值(默认值为8)调整前景与背景检测比例,值越高背景检测越多。由于病理图像尺寸巨大,即使前后背景分割后也无法满足网络输入需求。针对该问题,本研究首先将初始的WSI分割成大小为256×256的图像块;然后根据四角法对图像块进行筛选,即将图像块4个顶角与前一步骤勾画的病灶区域对比,如有任一角落在轮廓内则保留图像块,否则丢弃;最终,将挑选后的图像块和其对应的坐标保存在.h5文件中。将处理好的数据采用随机抽样的方式选择80%作为训练集,10%用作验证集,10%用作测试集,并采用5折交叉验证进行训练。

1.2.2 基于ResNet的特征图像提取 随着网络深度的不断增加,模型拟合数据的效果也随之提高;在传统的深度学习中,随着网络深度的逐渐加深,网络存在的问题主要是梯度消失和梯度爆炸。ResNet的提出主要是在现有网络深度的基础上,通过构建残差连接有效解决梯度问题,使网络可以随着深度的增加效果也有所提升。

采用ResNet^[31]提取经预处理后的病理图像的特

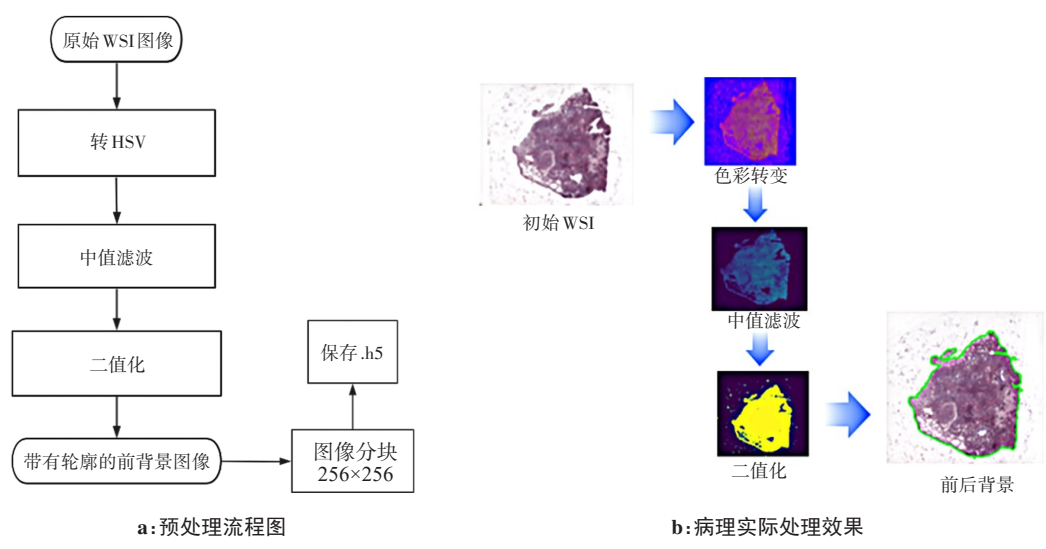


图 1 预处理流程图病理实际处理效果

Figure 1 Pretreatment flowchart and the actual pathological treatment outcome

征。考虑到样本数据较小,过深的网络层数不仅会使计算复杂度明显提升,还会由于样本规模问题导致模型拟合效果不够好,因此在特征提取模块中,本研究尝试的深度分别为 34、50、101,其模型架构如表 1 所示。这 3 种模型的网络架构类似但是略有不同,在 ResNet34 模型中,残差块没有上下采样的过程,采用的是 Basicblock,而 ResNet50 和 ResNet101 采用的则是 Bottleneck,见图 2。本模型使用的是 ImageNet 的预训练权重,不加微调,直接对病理图像提取特征,并将提取到的特征保存成 .pt 格式。

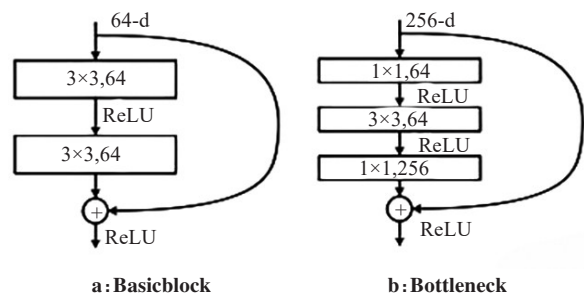


图 2 Basicblock 及 Bottleneck

Figure 2 Basicblock and Bottleneck

表 1 ResNet 不同模型架构图		
Table 1 Architecture diagram of different models of ResNet		
ResNet34	ResNet50	ResNet101
7×7, 64, stride 2		
3×3, max pool, stride 2		
$\begin{bmatrix} 3 \times 3 & 64 \\ 3 \times 3 & 64 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1 & 64 \\ 3 \times 3 & 64 \\ 1 \times 1 & 256 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1 & 64 \\ 3 \times 3 & 64 \\ 1 \times 1 & 256 \end{bmatrix} \times 3$
$\begin{bmatrix} 3 \times 3 & 128 \\ 3 \times 3 & 128 \end{bmatrix} \times 4$	$\begin{bmatrix} 1 \times 1 & 128 \\ 3 \times 3 & 128 \\ 1 \times 1 & 512 \end{bmatrix} \times 4$	$\begin{bmatrix} 1 \times 1 & 128 \\ 3 \times 3 & 128 \\ 1 \times 1 & 512 \end{bmatrix} \times 4$
$\begin{bmatrix} 3 \times 3 & 256 \\ 3 \times 3 & 256 \end{bmatrix} \times 6$	$\begin{bmatrix} 1 \times 1 & 256 \\ 3 \times 3 & 256 \\ 1 \times 1 & 1024 \end{bmatrix} \times 6$	$\begin{bmatrix} 1 \times 1 & 256 \\ 3 \times 3 & 256 \\ 1 \times 1 & 1024 \end{bmatrix} \times 23$
$\begin{bmatrix} 3 \times 3 & 512 \\ 3 \times 3 & 512 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1 & 512 \\ 3 \times 3 & 512 \\ 1 \times 1 & 2048 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1 & 512 \\ 3 \times 3 & 512 \\ 1 \times 1 & 2048 \end{bmatrix} \times 3$
平均池化,全连接,softmax		

1.2.3 基于分级注意力的聚类分类方法 在传统的神经网络中,每个神经元的输出只依赖于前一个神经

元的输出,导致远距离的神经元被忽略,使得网络的学习只集中于相近的几层中。提出注意力机制后,每个神经元的输出不仅取决于前一层的所有神经元的输出,还可以根据权重对之前学习到的内容进行叠加,从而增加知识的利用。对于病理图像而言,注意力机制的引入尤为为必要。病理图像由于其自身图像尺寸的问题,采用分割的方式进行预处理,这种分割的方式使得原本各个小块之间的信息连接断裂。具体来说,肿瘤由于切分被分散于多张图像块中,而最终给出病理分类分期的结果需要综合肿瘤的总体信息,需在病理图像中引入注意力机制来弥补各个图像块被分割后导致的信息损失。传统的多示例的分类方法主要是将切块后的图像给定统一的标签,当该集合中有一个被判定为阳性则标注结果为阳性,否则为阴性,但是上述的处理过程会受到噪声干扰,主要原因在于即使在癌症的病理图像中,癌变的组织区域只占区域的一部分,大部分的组织为正常组织,但是在标记时采用了阳性标签,为分类带来困难。

本研究采用分级注意力机制。第一级是基于图像块级别的注意力机制,旨在学习图像块级别的损失,在这个阶段对于不同块给出注意力得分,通过注意力方式给出最终分类标签并得到基于图像块级别的损失。第二级采用基于示例级别的注意力机制,

根据第一级中得到的注意力分数进行分类,分数高的给标签阳性反之给阴性,将其输入二分类网络中给出标签和基于示例级别的损失,最终依据第一级注意力得分融合两级损失最终给出整个病理图像的标签,具体操作见图3。

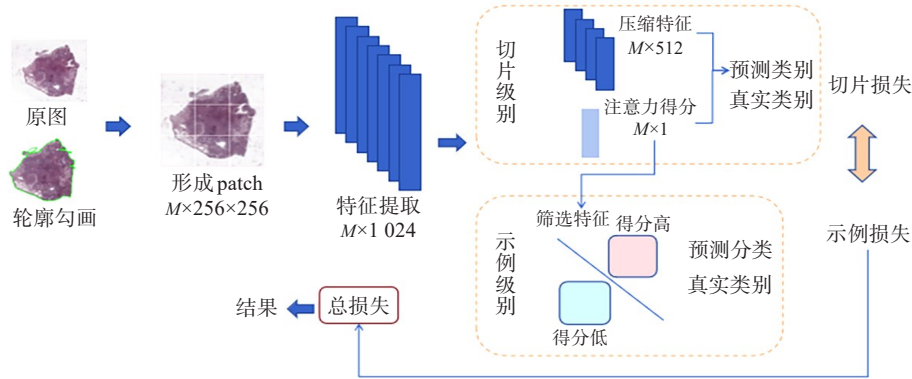


图3 基于两级注意力的多示例分类模型

Figure 3 Multi-instance classification model based on two-level attention

具体来说,首先是门注意力机制,采用全连接层 $W \in R^{512 \times 1024}$ 将 1024 维的特征压缩至 512 维的向量:

$$h_k = W_1 z_k^T \quad (1)$$

其中, $W_1 \in R^{512 \times 1024}$ 是权重矩阵, $z_k \in R^{1024 \times 1}$ 是特征输入向量, $h_k \in R^{512 \times 1}$ 表示输出向量。在压缩后的向量上采用 2 个注意力机制层作为被所有类共用的主干网络,其中 $U_a \in R^{256 \times 512}$ 以及 $V_a \in R^{256 \times 512}$, 在经过门注意力机制后将其分成 n 个平行的注意力分支 $W_{a,1}, W_{a,2}, \dots \in R^{1 \times 256}$, 同时可以得到 n 个平行的独立分类器 $W_{c,1}, W_{c,2}, \dots$ 为每一个块对应的类给出注意力得分, $a_{k,m}$ 表示为 k 个块分到 m 类的注意力得分:

$$a_{k,m} = \frac{\exp \{ W_{a,m} (\tanh(V_a h_k^T) \odot \text{sigm}(U_a h_k^T)) \}}{\sum_{j=1}^N \exp \{ W_{a,m} (\tanh(V_a h_j^T) \odot \text{sigm}(U_a h_j^T)) \}} \quad (2)$$

$h_{\text{slide},m}$ 表示一个切片对应的类的注意力得分:

$$h_{\text{slide},m} = \sum_{k=1}^N a_{k,m} h_k \quad (3)$$

再将该结果输入分类层中得到:

$$s_{\text{slide},m} = W_{c,m} h_{\text{slide},m}^T \quad (4)$$

最终,通过 softmax 得到整体切片的分数。

在注意力机制的设计过程中,通过将传统的平均池化替换为注意力分数和聚类方法,使得网络可以更好地关注于可能癌变区域。传统的网络一般采用交叉熵损失函数,通过调查发现,当数据标签有限时,对损失函数引入边界可以有效解决过拟合的问题^[7]。由于多实例的实现过程,当使用的监督标签为实例级别时,会创建很多伪标签,而这些伪标签包含

许多噪音。当为一个实例内所有的小块分配同样标签时,如果该实例为阳性,则该区域内所有块分配标签为阳性,但是实际在每一个实例中只有一小部分块有癌变,其他区域为正常组织区域,应当为阴性标签,因此引入较多噪音,如果继续使用交叉熵损失函数则不能很好学到该特征,使得模型拟合效果变差。综上所述,本研究采用二元光滑损失替代原有的交叉熵损失函数:

$$l(s, y) = \max \{ \max_{j \in \gamma \setminus \{y\}} \{ S_j + \alpha \} - S_y, 0 \} \quad (5)$$

$$J_{lr}(s, y) = \tau \log \left[\sum_{j \in \gamma} \exp \left(\frac{1}{\tau} (\alpha \prod (j \neq y) + S_j - S_y) \right) \right] \quad (6)$$

$$\text{最终得到 } L_{\text{total}} = c_1 L_{\text{slide}} + c_2 L_{\text{patch}} \circ$$

2 实验与结果

实验环境为 Windows10, 显卡型号为 P40, 内存为 24 G, 使用 Pytorch 作为深度学习框架, 使用 Pycharm 作为编译器, 在分级注意力的多示例分类模型的训练过程中, 对训练参数进行设置, 初始学习率设置为 $2e-4$, 权重衰减设置为 $1e-5$, $c_1=0.7$, $c_2=0.3$, epoch 默认设置最大为 200, 当连续出现 5 轮损失不再下降, 则提前停止训练, batch 大小设置为 8, 参数优化器选择为 Adam 优化器, ResNet 的预训练权重采用其在 ImageNet^[31] 上的预训练权重。

2.1 评价指标

使用准确率 (Accuracy)、精确度 (Precision)、F1 值和曲线下面积 (AUC) 来评估模型的性能^[32]。

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}$$

(7)

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

(8)

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

(9)

$$\text{F1} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

(10)

其中,真阳性(TP)表示正样本被正确分类的个数,真阴性(TN)表示负样本被正确分类的个数,假阳性(FP)表示负样本被错误分类为正样本的个数,假阴性(FN)表示正样本被错误分类为负样本的个数。准确率表示正确预测的样本数占总预测样本数的比值;ROC曲线是反映灵敏度和特异性连续变量的综合指标,AUC用来评估该二元分类器的可信度,其值为0~1,AUC值越大说明可信度越高。

2.2 实验结果

2.2.1 前后背景分割实验对比 表2对比了在是否做背景分割的情况下,与本文提出的基于分级注意力的病理分类模型得到的效果对比,对比后发现F1得分接近,背景分割后的AUC有一定的提升。

表2 是否分割背景结果对比

Table 2 Comparison of background segmentation or not

是否分割背景	准确率	精确度	F1值	AUC
是	0.789 5	0.842 9	0.851 8	0.843 0
否	0.752 3	0.823 1	0.833 3	0.836 9

2.2.2 特征提取网络对比 表3列举了4个较为常见的CNN特征提取模型,主要对比DenseNet和ResNet的区别,进一步比较ResNet在不同深度的情况下的表现。在对提取特征的对比实验中发现由于数据量相较于自然图像有较大区别,过深的网络效果反而有一定的下降。随着提取特征网络层数的增加,模型的数据量不足以支撑参数的训练,导致DensNet效果下降明显;ResNet中层数为50时,效果表现最为均衡。

表3 不同特征提取模型结果对比

Table 3 Comparison of results of different feature extraction models

特征提取模型	准确率	精确度	F1值	AUC
DensNet121	0.684 2	0.814 2	0.785 7	0.722 4
ResNet34	0.778 9	0.828 6	0.844 9	0.851 0
ResNet50	0.789 5	0.842 9	0.851 8	0.843 0
ResNet101	0.726 9	0.816 3	0.812 3	0.752 0

2.2.3 不同模型对比 表4列举了在提取特征后采用不同分类策略的结果。由于数据集的大小,首先考虑与传统的机器学习算法进行对比,选择机器学习中的支持向量机模型和随机森林模型进行对比,发现不同的机器学习算法在各项指标中的差异不明显;其次将两级注意力模型替换成SE注意力机制,通过机制的转变和对比发现,当引入注意力机制时可以明显对数据各方面的指标进行提升,说明注意力机制的有效性,同时通过对比SE模块和两级注意力机制发现在损失中同时考虑示例损失和切片损失可以明显提升模型的准确率,证明两级注意力机制的有效性。

表4 不同的分类策略结果对比

Table 4 Comparison of results of different classification strategies

分类模型	准确率	精确度	F1值	AUC
本文	0.789 5	0.842 9	0.851 8	0.843 0
SVM	0.736 8	0.795 6	0.812 3	0.751 2
随机森林	0.731 5	0.795 9	0.816 9	0.749 6
SE注意力机制	0.758 6	0.812 6	0.836 2	0.832 1

2.2.4 迭代次数与准确率 通过图4可以发现随着迭代次数的增加,损失越来越低,逐渐趋于稳定;相较于腺癌来说,口腔鳞癌的准确率变化更稳定,更快地趋于1,究其主要原因在于鳞癌的发病率在口腔癌中占据90%以上,因此数据集中包含的鳞癌样本更多,使得网络训练更充分。

2.2.5 公开数据集对比效果 在TCGA获得的数据集上进行实验(表5),针对其他病理类型,评估模型的效果(包括F1值和AUC)。与本研究采用的口腔癌病理数据集相比,AUC的值较为接近,但在准确率和精确度等指标上存在一定差异,这可能是由于不同部位癌症的病理特征和浸润形式的差异,以及当前数据集的有限性所致,这些结果表明模型在特定病理上的有效性,但要实现模型泛化和通用性,仍需进一步改进和验证。

3 讨论

针对多实例学习无法捕获空间和全局信息的问题,提出基于ResNet的分级注意力机制病理模型,旨在构建基于示例级别和图像块级别的分级注意力的口腔癌病理分类网络,有效考虑整体样本的阴阳性情况,同时有效利用病灶区域的信息。通过先分割再分类的方法提高病理图像分类的处理速度;在3种

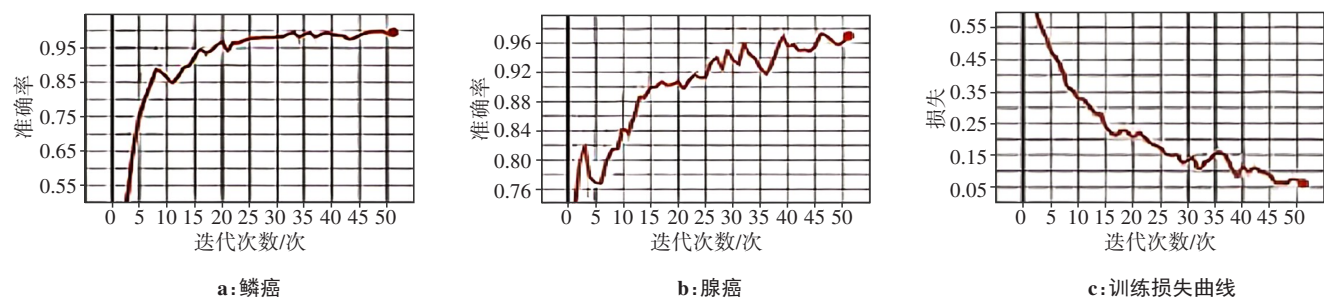


图4 训练损失值和准确率曲线图

Figure 4 Curves of training loss and accuracy

表5 公开数据集验证结果

Table 5 Dataset validation results

数据集来源	准确率	精确度	F1 值	AUC
本文	0.789 5	0.842 9	0.851 8	0.843 0
ICIA R2018	0.723 1	0.740 0	0.729 8	0.828 0
BreakHis	0.766 7	0.738 8	0.767 7	0.817 0

不同来源数据集上的实验验证了分级注意力机制的有效性,论证了有效利用图像块之间信息的传递可以显著提高病理诊断的准确性。但是该方法仍存在一些不足,首先由于数据的有限性,仅将模型应用于口腔鳞癌与腺癌的分类,对于口腔腺癌中多种腺癌的种类仍可进一步探索;其次对于模型的医学解释性的探索仍有待深入。综上所述,本研究提出的分级注意力的口腔癌病理分类网络有效提高病理图像分类精度。进一步的研究将探索模型的解释性,并将病理与其他模态融合,为病理诊断准确性的提高奠定基础。

【参考文献】

[1] Rivera C. Essentials of oral cancer[J]. Int J Clin Exp Pathol, 2015, 8 (9): 11884-11894.

[2] Pfister DG, Spencer S, Adelstein D, et al. Head and neck cancers, version 2.2020, NCCN clinical practice guidelines in oncology[J]. J Natl Compr Canc Netw, 2020, 18(7): 873-898.

[3] 李贵忠, 李洁莹, 周凯, 等. 淋巴结外扩展对口腔鳞癌患者TNM分期及生存预后的影响[J]. 中国口腔颌面外科杂志, 2023, 21(4): 377-383.

Li GZ, Li JY, Zhou K, et al. Effect of extranodal extension on TNM staging and survival in patients with oral squamous cell carcinoma[J]. China Journal of Oral and Maxillofacial Surgery, 2023, 21(4): 377-383.

[4] Amin MB, Edge SB, Greene FL, et al. AJCC cancer staging manual [M]. 8th ed. Switzerland: Springer, 2017: 202-218.

[5] 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.

[6] Zhao Y, Lin ZY, Sun K, et al. SETMIL: spatial encoding transformer-based multiple instance learning for pathological image analysis[C]// Medical Image Computing and Computer Assisted Intervention-MICCAI 2022. Cham: Springer Nature Switzerland, 2022: 66-76.

[7] Chikontwe P, Luna M, Kang M, et al. Dual attention multiple instance learning with unsupervised complementary loss for COVID-19

screening[J]. Med Image Anal, 2021, 72: 102105.

[8] Shao ZC, Bian H, Chen Y, et al. TransMIL: transformer based correlated multiple instance learning for whole slide image classification[EB/OL]. (2021-10-31). <https://arxiv.org/abs/2106.00908>.

[9] Van Der Laak J, Litjens G, Ciompi F. Deep learning in histopathology: the path to the clinic[J]. Nat Med, 2021, 27(5): 775-784.

[10] Chen Z, Tian ZQ, Zhu JH, et al. C-CAM: causal CAM for weakly supervised semantic segmentation on medical image[C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ, USA: IEEE, 2022: 11666-11675.

[11] Wang J, Xia B. Bounding box tightness prior for weakly supervised image segmentation[C]//Medical Image Computing and Computer Assisted Intervention-MICCAI 2021. Cham: Springer International Publishing, 2021: 526-536.

[12] Moccia S, De Momi E, El Hadji S, et al. Blood vessel segmentation algorithms-review of methods, datasets and evaluation metrics[J]. Comput Methods Programs Biomed, 2018, 158: 71-91.

[13] Lu C, Romo-Bucheli D, Wang XX, et al. Nuclear shape and orientation features from H&E images predict survival in early-stage estrogen receptor-positive breast cancers[J]. Lab Invest, 2018, 98(11): 1438-1448.

[14] Chan HP, Hadjiiski LM, Samala RK. Computer-aided diagnosis in the era of deep learning[J]. Med Phys, 2020, 47(5): e218-e227.

[15] Ahn J, Cho S, Kwak S. Weakly supervised learning of instance segmentation with inter-pixel relations [C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ, USA: IEEE, 2019: 2204-2213.

[16] Dietterich TG, Lathrop RH, Lozano-Pérez T. Solving the multiple instance problem with axis-parallel rectangles[J]. Artif Intell, 1997, 89(1/2): 31-71.

[17] Chen ZH, Chi ZR, Fu H, et al. Multi-instance multi-label image classification: a neural approach [J]. Neurocomputing, 2013, 99: 298-306.

[18] Srinidhi CL, Ciga O, Martel AL. Deep neural network models for computational histopathology: a survey[J]. Med Image Anal, 2021, 67: 101813.

[19] Prabhakar SK, Rajaguru H. Performance analysis of linear layer neural networks for oral cancer classification[C]//2017 6th ICT International Student Project Conference (ICT-ISPC). Piscataway, NJ, USA: IEEE, 2017: 1-4.

[20] Chen RJ, Lu MY, Weng WH, et al. Multimodal co-attention transformer for survival prediction in gigapixel whole slide images [C]//2021 IEEE/CVF International Conference on Computer Vision (ICCV). Piscataway, NJ, USA: IEEE, 2021: 3995-4005.

[21] Chen RJ, Chen CK, Li YC, et al. Scaling vision transformers to gigapixel images via hierarchical self-supervised learning[C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ, USA: IEEE, 2022: 16123-16134.

[22] Ilse M, Tomczak J, Welling M. Attention-based deep multiple instance learning [C]//Proceedings of the 35th International Conference on Machine Learning. Chia Laguna Resort, Sardinia, Italy: PMLR, 2018: 2127-2136.

[23] 王怡伟, 李晓兵, 聂生东, 等. 基于深度学习的超声多模态乳腺肿块良恶性分类[J]. 中国医学物理学杂志, 2023, 40(8): 988-995.

Wang YW, Li XB, Nie SD, et al. Deep learning-based classification

- for benign and malignant breast masses using multimodal ultrasound images[J]. Chinese Journal of Medical Physics, 2023, 40(8): 988-995.
- [24] Kwak JT, Hewitt SM. Multiview boosting digital pathology analysis of prostate cancer[J]. Comput Methods Programs Biomed, 2017, 142: 91-99.
- [25] Sharma H, Zerbe N, Heim D, et al. A multi-resolution approach for combining visual information using nuclei segmentation and classification in histopathological images[C]//Proceedings of the 10th International Conference on Computer Vision Theory and Applications. Setúbal, Portugal: SciTePress, 2015: 37-46.
- [26] 王新宇, 赵静文, 刘翔, 等. 人工智能在肺结节筛查和肺癌诊断中的应用[J]. 中国医学物理学杂志, 2023, 40(9): 1182-1188.
- Wang XY, Zhao JW, Liu X, et al. Applications of artificial intelligence in lung nodule detection and lung cancer diagnosis [J]. Chinese Journal of Medical Physics, 2023, 40(9): 1182-1188.
- [27] Kamarudin AN, Cox T, Kolamunnage-Dona R. Time-dependent ROC curve analysis in medical research: current methods and applications [J]. BMC Med Res Methodol, 2017, 17(1): 53.
- [28] Das DK, Chakraborty C, Sawaimoon S, et al. Automated identification of keratinization and keratin pearl area from *in situ* oral histological images[J]. Tissue Cell, 2015, 47(4): 349-358.
- [29] Das N, Hussain E, Mahanta LB. Automated classification of cells into multiple classes in epithelial tissue of oral squamous cell carcinoma using transfer learning and convolutional neural network[J]. Neural Netw, 2020, 128: 47-60.
- [30] Gupta RK, Kaur M, Manhas J. Tissue level based deep learning framework for early detection of dysplasia in oral squamous epithelium [J]. J Multimed Inf Syst, 2019, 6(2): 81-86.
- [31] He KM, Zhang XY, Ren SQ, et al. Deep residual learning for image recognition [C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ, USA: IEEE, 2016: 770-778.
- [32] Ghaffari Laleh N, Muti HS, Loeffler CM, et al. Benchmarking weakly-supervised deep learning pipelines for whole slide classification in computational pathology[J]. Med Image Anal, 2022, 79: 102474.
- (编辑:谭斯允)