

引文格式:辛家隆,张雪松,强彦,等. 基于改进焦点损失与图像特征融合的CT图像分类方法[J]. 太原理工大学学报, 2026, 57(2): 378-386.

XIN Jialong, ZHANG Xuesong, QIANG Yan, et al. CT image classification method based on improved focus loss and image feature fusion[J]. Journal of Taiyuan University of Technology, 2026, 57(2): 378-386.

## 基于改进焦点损失与图像特征融合的 CT图像分类方法

辛家隆<sup>1</sup>, 张雪松<sup>1</sup>, 强彦<sup>1,2</sup>, 赵涓涓<sup>1,3</sup>, 冯对平<sup>4</sup>, 高龙<sup>4</sup>

(1. 太原理工大学 计算机科学与技术学院(大数据学院), 山西 太原; 2. 中北大学 软件学院, 山西 太原;

3. 太原理工大学 软件学院, 山西 太原; 4. 山西医科大学第一医院 山西放射治疗(介入放射学)

临床医学研究中心, 山西 太原)

**摘要:**【目的】针对医学图像临床应用场景下样本类别失衡问题, 提出一种以混合交叉注意力为核心的双分支特征融合网络。【方法】在网络输入端对输入图像进行特征融合与重要性评分, 促使模型聚焦提取感兴趣区域, 缓解图像前景与背景的失衡, 并且提出一种用于二分类任务的改进焦点损失函数, 该损失函数利用三次幂函数的中心对称特性, 通过改进焦点损失加权因子, 优化错误分类惩罚项的同时减少超参数的引入; 其次, 针对样本数量不足问题, 采用针对CT图像的数据增强方法与样本替换策略, 有效提升可用样本数。【结果】在公开数据集与真实医学临床数据集上的实验结果表明, 改进的焦点损失提高了分类模型稳定性, 所提方法能够在门静脉高压患者术前筛选任务中取得超过82%的正确率, 在多个公共数据集中也可稳定提升任务表现。

**关键词:** 类不平衡; 特征融合; 改进焦点损失; 医学图像; 数据增强

**中图分类号:** TP391 **文献标识码:** A

**DOI:** 10.16355/j.tyut.1007-9432.20240755

**文章编号:** 1007-9432(2026)02-0378-09

## CT Image Classification Method Based on Improved Focus Loss and Image Feature Fusion

XIN Jialong<sup>1</sup>, ZHANG Xuesong<sup>1</sup>, QIANG Yan<sup>1,2</sup>, ZHAO Juanjuan<sup>1,3</sup>,  
FENG Duiping<sup>4</sup>, GAO Long<sup>4</sup>

(1. College of Computer Science and Technology (College of Data Science), Taiyuan University of Technology, Taiyuan, Shanxi, China; 2. School of Software, North University of China, Taiyuan, Shanxi, China; 3. School of Software, Taiyuan University of Technology, Taiyuan, Shanxi, China; 4. Shanxi Radiation Therapy (Interventional Radiology) Clinical Medical Research Center, First Hospital of Shanxi Medical University, Taiyuan, Shanxi, China)

**Abstract:** 【Purposes】Aiming at the problem of sample category imbalance in clinical application scenarios of medical images, a dual-branch feature fusion network with mixed cross-attention as the core is proposed. 【Methods】The input image was subjected to feature fusion and importance score at the input end of the network, which promotes the model to focus on extracting the region of interest and alleviates the imbalance between the foreground and background of the image. An improved focal

收稿日期: 2024-10-10; 修回日期: 2024-11-13

基金项目: 国家自然科学基金项目资助(U21A20469, 62376183)

第一作者: 辛家隆, 硕士研究生, (E-mail) 343925807@qq.com

通信作者: 强彦, 教授, 博士生导师, 主要从事医学图像处理、高级人工智能研究, (E-mail) qiangyan@tyut.edu.cn

loss function for the two-classification task was proposed. The centrosymmetric characteristics of cubic power function was used in the proposed loss function. The penalty term of the misclassification was smoothed by improving the focus loss weighting factor and the introduction of fewer hyperparameters. Then the class imbalance factor of the focus loss was theoretically derived. The reference range was given to reduce the parameter search space. Second, aiming at the problem of insufficient sample size, the data enhancement method and sample replacement strategy for CT images were adopted to effectively increase the number of available samples. **【Results】** The experimental results on public datasets and real medical clinical datasets show that the improved focus loss improves the stability of the classification model. The proposed method can achieve more than 82% accuracy in the preoperative screening task of patients with portal hypertension, and also steadily improve task performance in multiple public datasets.

**Keywords:** class imbalance; feature fusion; improved focal loss; medical images; data enhancement

随着计算机技术与医学成像技术的发展,医学图像处理与分析已经成为现代医疗的关键技术之一,对于疾病早期筛查诊断、治疗规划和预后效果评估具有不可替代的作用<sup>[1]</sup>。传统的医学图像处理技术通过统计学分析方法或影像组学方法手动构建数学模型解决特定问题,存在效率低、准确度较低的问题<sup>[2]</sup>。随着人工智能技术的发展,端到端的深度学习方法可以自动从图像中提取高度抽象的信息<sup>[3]</sup>,发掘大量人类无法直观感知的潜在信息,在各种医疗相关任务中表现出色<sup>[4-5]</sup>。然而,出于对患者隐私保护的考虑以及患病样本远远少于正常样本的客观事实,数据集规模有限和类别不平衡的存在使得深度模型在训练时难以收敛或引发过拟合的问题,导致模型对少数类样本失去识别能力,深度网络失去实际意义,进而影响临床决策的准确性<sup>[6]</sup>。针对上述问题的解决思路主要有两个:一是使用采样策略,对少数类过采样或对多数类降采样从而达到平衡数据集的目的,医学数据集规模受限的特性使得对少数类过采样成为首选手段,但是传统的过采样方法可能会导致过拟合,而基于特征空间分布的SMOTE等方法存在真实性和医学伦理层面的风险;第二个解决方法被称为成本敏感型方法,即从损失函数设计层面对不同的分类错误施加不同的惩罚,这不可避免地引入超参数来调节惩罚力度,超参数庞大的搜索空间增加了此类方法的应用成本。

此外,图像特征信息融合也是医学领域应用人工智能方法时需要考虑的一项挑战,深度网络在应用于医学临床实践中时,分析单一信息往往不足以全面反映患者的状况,难以达到很好的效果。特别是医学图像,不同模态不同视角下的图像之间存在

大量冗余信息,造成信息熵急剧减小,有效信息被淹没,深度网络难以学到可分性强的特征,造成模型表现不佳。针对这一问题,采取有效的图像特征信息融合方法,在不丢失关键信息的基础上有效融合多模态多视角的图像特征信息将极大地提升模型的表现<sup>[7]</sup>。

综上所述,本文提出了一种基于改进焦点损失与图像特征融合的医学图像分类方法,通过在焦点损失的基础上引入非参数幂函数优化损失平滑性与收敛稳定性应对类别不平衡挑战;其次设计一种双分支图像信息融合网络,弥补单一模态单一视角表征能力不足的问题,同时在模型中加入了一种混合交叉注意力机制,降低背景前景不平衡对特征提取带来的干扰。主要贡献如下:

1)针对CT图像无关背景占比大,感兴趣区域占比小,有效前景与无关背景严重失衡问题,提出一种新的高灵活性混合交叉注意力模块为核心的双分支图像信息融合网络,有效减少背景干扰,提高特征提取效率与质量。

2)针对CT扫描序列相邻相似性强的问题,提出了基于样本分布的CT随机数据增强方法和基于概率的批量训练输入策略,可以有效地增加样本量,且不存在医学伦理方面的担忧。

3)针对医学图像存在的类别不平衡问题,提出一种改进的焦点损失,利用非参数中心对称幂函数优化加权因子。改进的焦点损失仅引入一项超参数,且通过理论推导证明和实验提供参数的搜索范围,大大降低了超参数搜索的难度和成本。

## 1 相关工作

类别不平衡问题的存在使得传统损失函数,如

交叉熵损失等无法有效约束训练模型以识别所有类别,倾向于更多关注多数类样本。为解决这一问题,研究者通常会在损失函数中引入一系列超参数,以实现对难易样本以及多数样本和少数样本的不同效果是解决问题的主要思路之一<sup>[8]</sup>。在目标检测与分割领域,前景和背景之间的严重不平衡影响了检测器的性能。KAIMIN<sup>[9]</sup>团队率先引入焦点损失来提高探测器的性能,取得了非常好的结果。该方法通过为具有挑战性或容易被错误分类的样本分配更高的权重,同时降低简单样本的权重来解决类不平衡问题。RIDNIK et al<sup>[10]</sup>将这种方法扩展到多标签分类任务,设计了一种新的不对称损失为正负样本分配不同的权重,并采用不同的损失计算机制。文献[6]从训练过程入手,提出了一种新的多加权损失函数和端到端累积学习策略,降低了异常样本对训练的影响。此外,WANG et al<sup>[10]</sup>将质量因素嵌入到焦点损失中,使网络能够专注于学习高质量的样本,而不是难以分类的样本。以上方法增强了焦点损失的灵活性,提升了性能表现,但都不同程度地引入额外的超参数,显著扩大了参数搜索空间。本文通过引入非参数的幂函数优化焦点损失,仅保留类别因子一项超参数,从源头降低参数搜索工作的负担,此外本文通过理论和实验推导出唯一超参数的搜索范围与搜索方法,进一步降低所提方法的应用成本。

## 2 方法

### 2.1 双分支图像特征融合网络

在医学图像处理领域,图像特征融合是一项关键技术。它旨在整合不同模态、不同视角或不同参数下的图像数据,弥补单一图像信息量不足的缺点,为医生和研究人员提供更加全面的信息,从而提升诊断的准确性与可靠性<sup>[11]</sup>。然而大量特征信息的融合意味着大量的数据输入,导致计算成本过高。此外,CT图像不含有效信息的黑色背景占比较大,而与病变相关的感兴趣区域占比极小。背景与前景的不平衡使得特征信息融合带来的信息冗余进一步加剧。本文提出的双分支图像特征融合网络如图1所示,网络输入包括两个部分: $S_L$ 和 $S_M$ ,它们来自同一患者的不同CT图像。输入图像首先经过混合交叉注意力模块提取感兴趣的区域,注意力模块的输出特征图与原始图像叠加。这一步的目的是在强化被关注的区域,实现对背景区域的弱化,缓解背景前景不平衡的同时,保证原始信息的

不丢失,避免表达能力的下降。然后进入特征提取阶段,两个分支的编码器结构遵循相同的设计但由于处理信息存在显著区别,因此并不共享权重。多层特征提取网络的头部为两层 $7\times 7$ 的大尺寸卷积网络实现降采样,使用含有残差连接的数个 $3\times 3$ 与 $1\times 1$ 的卷积层堆积作为特征提取层,编码器的层数可根据实际应用任务的复杂程度进行调整。经多层编码器构成的特征提取网络后,使用全连接层,将特征图转换成一维特征向量的形式。利用向量拼接的方法将两个分支的特征进行融合。选用特征向量拼接进行特征融合的原因是该方法对输入向量的数量、尺寸要求并不严格,当不同分支输入尺寸不一致,或根据实际需要增加更多支路时也可以轻松应对,显著增强了网络的灵活性。融合后形成的新特征向量根据下游分类任务的需要,利用若干全连接层灵活调整输出尺寸,输出最终的分类结果。

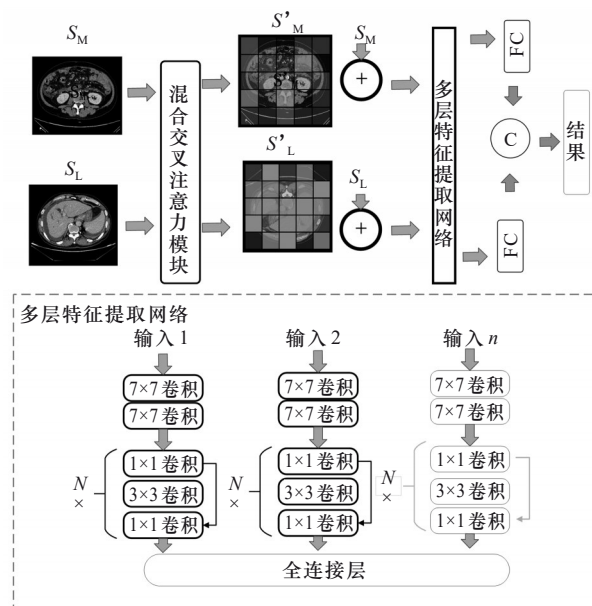


图1 双分支图像特征融合网络结构图

Fig.1 Dual-branch image feature fusion network structure

### 2.2 混合交叉注意力模块

混合交叉注意力模块结构示意图如图2所示,该模块要求两个输入参数,即两个分支输入的通道数。卷积运算对输入大小没有要求但是有通道数的限制,因此在模块的头部和尾部添加一个 $1\times 1$ 卷积层以匹配卷积通道,保证输入图像与输出特征图尺寸完全一致,从而实现即插即用的效果。输入序列 $S_L$ 和 $S_M$ 通过带残差结构的多级编码器生成 $B\times C_{in}\times H\times W$ 的特征图。在逐级解码阶段,特征向量按照层次逐步混合向后传递,混合方式为对应值相

加。在模块的末尾,混合特征映射被拉伸化为一维特征向量,通过Softmax函数激活后,调整尺寸进行恢复,达到2D平面上的Softmax的效果,对整个平面所有像素点进行评分。所有像素点都被赋予了0~1的一个分值,且所有像素点的分值和为1,这意味着在模型训练过程中,注意力机制不仅需要找到重要的区域,同时对找到的感兴趣区域进行取舍,而像素级别的评分显著优于基于池化降采样的注意力机制,拥有更小的细粒度。评分结果与原始图像对应位置点乘,评分低区域的像素值被压缩到非常低的水平,而重点感兴趣区域的像素值有所保留。在后续卷积网络对其进行卷积运算时,低像素值对卷积结果的影响非常小,从而达到避免无关信息干扰,专注于有价值感兴趣区域的作用。

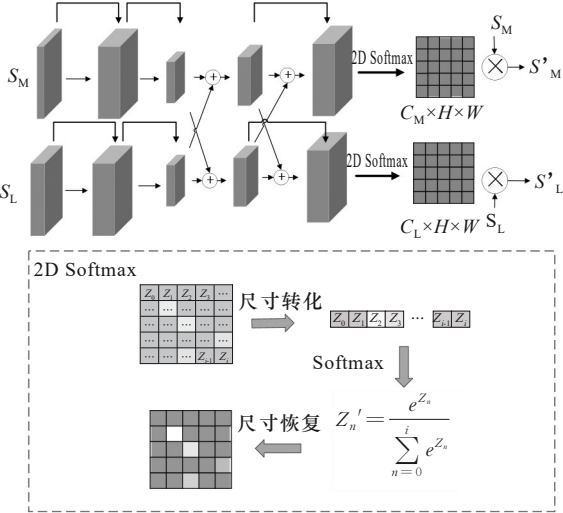


图2 混合交叉注意力模块结构图

Fig.2 Hybrid cross-attention module structure

### 2.3 CT序列图像增强与样本替换

CT图像是连续切片的结果,每幅图像都是连续体素的二维表示,因此在相邻图像之间存在结构和纹理方面的高度连续性。特别是CT薄层扫描,层间距非常小,相邻图像展示的解剖结构变化小,显著增加了特征提取难度,增加了过拟合风险。针对医学CT序列影像相邻相似性强的特点,设计了专门的数据增强方法。首先,将序列进行分区,形成多个子序列,分区依据可以引入先验知识也可以均等分区。根据子序列长度的比例关系,确定从中抽取图像数量的比例。第二步随机选择,即从子序列中以第一步中确定的比例进行随机抽样。第三步,随机排序。抽取图像的序列被随机变换以擦除潜在的空间位置信息。第四步,将各个子序列中产生的图像随机组合并随机挑选,应用一种几何转换

(水平翻转、选择、随机裁剪等)来形成输入数据对。

通过上述方法,一个CT序列可以有效地扩展数十倍,数据的连续相似性被显著降低,从数据输入层面降低了过拟合的风险。在第三步中去除的空间位置信息有助于增强特征提取网络鲁棒性。扩充样本来源于真实的医学图像,且未修改任何像素值,新产生的样本虽然经过了随机抽取与组合,但是数据操作范围始终隶属于同一病例,因此可以有效避免医学伦理相关的风险。此外,这一过程对不同长度的序列进行了标准化处理转换成相同大小的数据而无需额外操作。

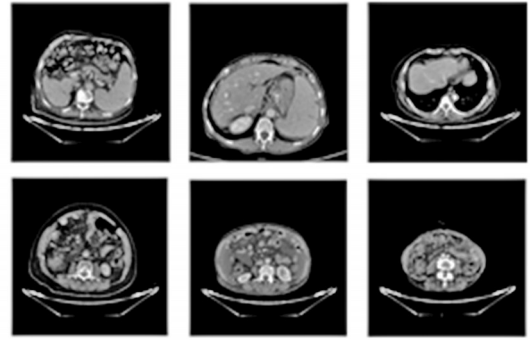


图3 部分CT影像展示

Fig.3 Part of CT images

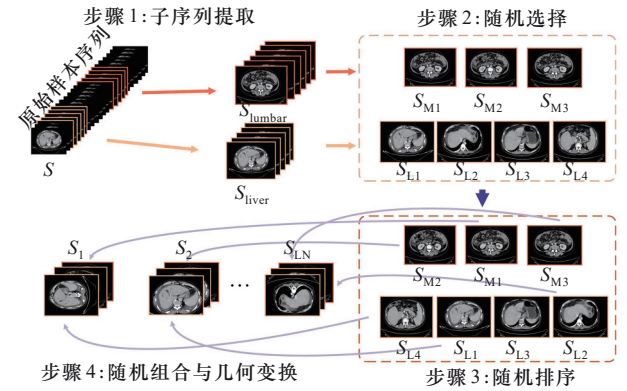


图4 CT扫描数据增强步骤示意图

Fig.4 Schematic diagram of CT scanning data enhancement steps

### 2.4 基于概率的样本替换策略

批次样本中的类别分布反映了数据集中整体的样本分布情况,当数据集类不平衡时,模型学习将变得困难。为了解决这个问题,提出一种基于概率的样本替换策略,即部分多数类被少数类取代,确保批内的类分布接近均匀分布。数学表达式为:

$$P_s B = P_L B = \frac{B}{C}. \quad (1)$$

$$P_s R = P_L. \quad (2)$$

$$p P_L B + P_s B = P_L B - p P_L B. \quad (3)$$

$B$ 表示批次大小, $C$ 为分类任务中的类别数, $P_s$ 和 $P_L$

分别表示少数类和多数类样本出现的频率,  $R$  为多数类与少数类的比值, 为一常数。多数类样本以概率  $p$  被少数类取代, 可以得到  $p$  的取值为:

$$p = \frac{RP_s - P_s}{2RP_s}. \quad (4)$$

使用概率而不是精确值作为样本替换依据的原因是每一批次随机输入的样本的比例是动态的, 使用精确值需要额外的计算成本, 并且动态的标签分布有利于提高模型的鲁棒性与训练的稳定性。

## 2.5 改进焦点损失函数

从最基本的交叉熵函数出发, 二元交叉熵损失可以表示为式(5)和式(6)。它对每个样本都是公平的, 在处理类不平衡数据时的这一特性成为缺点。简单样本( $p_i > 0.5$ )的损失没有限制在一个很低的水平, 多数类的损失轻易支配了模型。

$$CE(p, y) = CE(p_i) = -\log(p_i). \quad (5)$$

$$p_i = \begin{cases} p, y = 1 \\ 1 - p, y \neq 1 \end{cases}. \quad (6)$$

焦点损失在其基础上增加了调制项 $-(1-p_i)^\gamma$ , 数学表达式如式(7)所示, 其中 $\gamma \geq 0$ 。随着 $p_i$ 的增加, 简单样本的惩罚权重被削弱, 反之, 当 $p_i$ 减小, 调制项逐渐增大。

$$FL(p_i) = -(1-p_i)^\gamma \log(p_i). \quad (7)$$

在焦点的基础上, 引入了类平衡因子 $\alpha$ , 其维度与分类任务中的类别数量一致, 如式(8)所示。

$$FL_{CB}(p_i) = -\alpha_i \sum_{i=1}^C (1-p_i)^\gamma \log(p_i). \quad (8)$$

$\alpha$  的设计通常是基于正负样本比例的, 常见的有以下两种<sup>[13]</sup>:

$$\alpha = \frac{1}{N}. \quad (9)$$

$$\alpha = \frac{1-\beta}{1-\beta^N}. \quad (10)$$

定性分析类别平衡因子 $\alpha$ 和焦点因子 $\gamma$ 共同调节传统的交叉熵损失。然而,  $\alpha$  的增加和 $\gamma$  的减少都会导致函数值的增加, 这进一步增加了参数搜索过程的复杂性。统一或简化超参数是降低参数搜索难度的有效方法之一。本研究引入幂函数项来改善焦点损失。

$$PF_{CB} = (0.5 - p_i)^3 (\log(p_i))^2. \quad (11)$$

利用三次幂函数作为加权因子核心的原因在于其单调性和中心对称特性, 它的导数特性使得离对称中心近的区域变化缓慢, 距离对称中心远的区域变化率大, 因此, 选择三次幂函数并将其对称中

心移动至 0.5 压缩了  $p_i > 0.5$  的值, 从而增大  $p_i < 0.5$  的惩罚。原焦点损失的对数项调整为平方项, 以确保函数的正负性依赖于  $p_i$ 。然而, 这种修改引入了一个新问题: 函数的非负性无法保证。因此, 引入一个校正项, 如式(12)所示。

$$PF_{CB} = (0.5 - p_i)^3 (\log(p_i))^2 + (1 - p_i)^2. \quad (12)$$

然后, 向它添加一个类平衡项。在确保少数类得到高度重视的同时, 应该保证合适的程度, 以免表现出过分的少数类倾向。需要满足以下条件:

$$L_{FS} < L_{FL}. \quad (13)$$

$$Bp_L L_{FS} + Bp_S L_{TS} > Bp_S L_{FL} + Bp_L L_{TL}. \quad (14)$$

$$L_{FL} = \alpha L_{FS}. \quad (15)$$

$L_{FS}$  和  $L_{FL}$  分别表示错误少数类和错误多数类情况下的损失,  $L_{TS}$  和  $L_{TL}$  表示正确区分少数类和多数类情况下的损失。设多数类的权值为 1, 少数类的权值为  $\alpha$ , 可以推出:

$$1 < \alpha < R + \frac{L_{TS} - L_{TL}}{L_{FS}} < R + 1. \quad (16)$$

当多数类的权值为 1 时, 式(9)中  $\alpha$  的值为两类样本数量之比, 即  $\alpha = R$ 。式(10)可以证明  $\alpha \in [1, R]$ , 这里给出一个简单的证明。同上, 定义多数类权重基为 1, 则少数类权重为  $\frac{1-\beta^{RN_s}}{1-\beta^{N_s}}$ , 其导数如下:

$$\begin{aligned} & \frac{\partial}{\partial \beta} \left( \frac{1-\beta^{RN_s}}{1-\beta^{N_s}} \right) \\ &= \frac{-RN_s \beta^{(RN_s-1)} (1-\beta^{N_s}) + (1-\beta^{RN_s}) N_s \beta^{(N_s-1)}}{(1-\beta^{N_s})^2} \\ &+ \frac{N_s \beta^{(N_s-1)} (1-\beta^{(RN_s-N_s)})}{(1-\beta^{N_s})^2}. \end{aligned} \quad (17)$$

显然, 这是一个单调递增函数, 因此  $\alpha$  下限为 1, 上限是  $\beta \rightarrow 1$  处的极限值。

$$\lim_{\beta \rightarrow 1} \frac{1-\beta^{RN_s}}{1-\beta^{N_s}} = \lim_{\beta \rightarrow 1} \frac{-RN_s \beta^{(RN_s-1)}}{-N_s \beta^{(N_s-1)}} = R. \quad (18)$$

因此, 式(10)提出的平衡因子搜索空间  $\alpha \in [1, R]$ , 与本文的研究是一致的。这进一步验证了本文工作的科学性, 在理论层面上确定了类平衡因子搜索空间。最终, 本文提出的类平衡幂焦点损失的完整表达如下:

$$PF_{CB} = \sum_{k=1}^C \alpha ((0.5 - p_i)^3 (\log(p_i))^2 + (1 - p_i)^2). \quad (19)$$

图 5 展示了改进焦点损失函数的函数图像。与其他常见损失函数相比, 样本出现严重分类错误( $p_i < 0.2$ )时损失值较大, 随着  $p_i$  的增加, 损失以较快速率衰减, 这一特点保证训练初期损失值可以迅速收

敛。而当 $p_i$ 继续增大时,函数曲线变得平滑,保证模型训练中后期的稳定性。当 $p_i$ 处于0.5附近时,曲线较小的下降速度保证了处于分类临界区域样本损失值的稳定性,可以有效避免震荡,这对小样本情况下的分类任务具有显著的作用。当 $p_i > 0.5$ 时损失值被限制在很低的水平,这意味着改进的损失函数对分类正确率的关注远大于对置信度的关注程度。此外,改进的焦点损失函数需要对每个样本计算一次,计算过程中没有嵌套循环或复杂的数学运算,算法复杂度为 $O(n)$ ,与焦点损失保持一致。因此在未增加算法复杂度的情况下改进焦点损失在稳定性和收敛性上更加具有优势。

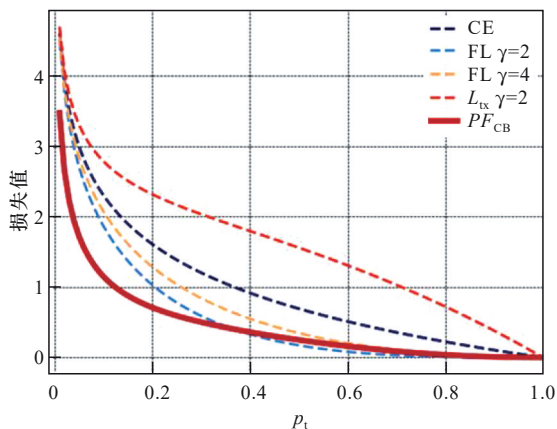


图5 损失函数图像

Fig.5 Loss function

### 3 实验结果

#### 3.1 实验数据集

本文以门静脉高压患者术前筛选任务为例,对提出方法的有效性进行评估。门静脉高压症是由肝硬化引起的严重病症。发作时经常引起各种并发症,严重时危及生命<sup>[12]</sup>。治疗门静脉高压的一种方法是通过经颈静脉肝内门静脉系统分流术(Transjugular Intrahepatic Portosystemic Shunt, TIPS)。约10%~55%患者在TIPS植入后发生肝性脑病(Hepatic encephalopathy, HE),严重影响生活质量且无有效治疗方法<sup>[13]</sup>。因此,进行患者术前筛查工作具有重要意义。门静脉高压患者术前筛选任务数据集为私有数据集,本研究在2019年7月至2022年4月在山西医科大学第一医院共收集了190例首次接受TIPS手术的患者,所有受试者均给予书面知情同意,并经山西医科大学第一医院研究伦理委员会(K-K231)批准。

此外,三个公开数据集也被用来评估本文方法的先进性。第一个数据集是2019年B-ALL白细胞

显微镜图像中正常细胞和恶性细胞(C-NMC)的分类。该挑战是在2019年IEEE生物医学成像国际研讨会(ISBI)上组织的<sup>[14]</sup>。第二个数据集为SARS-COV-2CT扫描数据集,包含来自巴西圣保罗州某医院1252例新冠阳性患者和1230例非新冠肺炎患者的CT图像。第三个公共数据集是使用CIFAR-10数据集,随机选择了两种类型的样本,自行创建了一个不平衡的数据集。训练集包括多数类5000个和少数类1000个,验证集包含1000例。

#### 3.2 双分支图像信息融合网络性能评估

为了验证多分支图像信息融合网络性能,在私有数据集与SARS-COV-2数据集上进行实验。两项实验均采用双分支信息融合网络。在私有数据集的实验中,分支图像分别为原始图像肝脏序列与原始图像胸椎序列,这是因为门静脉高压患者术前筛选任务中医生重点关注这两个部分。第一部分是包含肝静脉、门静脉和肝实质的图像。这是临床上确定TIPS植入位置最关键的部位<sup>[15]</sup>。第二部分是包含第三腰椎骨骼肌的图像,它是反映患者营养状况的有效指标之一<sup>[16]</sup>,依据两个子序列长度,在数据增强阶段抽取比例确定为4:1。在公共数据集中,分支图像分别为原始图像与Prewitt算子处理后的边缘纹理图像。

表1展示了多种人工智能方法在本研究中的表现,当模型预测输出全部为多数类时实验结果在表中由短横线代替。受类不平衡及小数据集的影响,多种先进的通用网络模型出现了严重的多数类倾向,对比其他先进方法和其他小样本学习方法,本文方法在面向类不平衡数据集时准确率方面具有显著优势。医学研究中,常使用ROC曲线来展示筛选方法的效果,图6展示了传统医学方法与本文方法在门脉高压患者术前筛选任务中的ROC曲线结果,本文方法远远优于其他方法,AUC达到0.7444。

#### 3.3 超参数 $\alpha$ 搜索

改进焦点损失的唯一超参数 $\alpha$ 的理论范围为 $[1, R+1]$ 。在训练过程中,为减少过拟合的影响,需要在有限轮次内选择收敛速度较快的超参数。私有数据集的 $R$ 约为2, $\alpha$ 的理论值为 $[1, R+1]$ ,则从 $R$ 开始搜索,步长为1。实验结果如表2所示,数据以平均数±标准差的形式表示。当 $\alpha$ 值从2增加到4时,模型性能指标呈上升趋势,当 $\alpha=4$ 时效果最佳,这略大于理论值。这是因为离群样本点不应被视为有效样本。但是,从原始图像中直接确定其在特征空间中的分布是很困难的。定性地说,这种

| 表 1 网络模型性能评估                                    |              |                |                |                |                |
|-------------------------------------------------|--------------|----------------|----------------|----------------|----------------|
| Table 1 Performance evaluation of network model |              |                |                |                |                |
| 数据集                                             | 方法           | 准确率            | 精确率            | 召回率            | ROC            |
| SARS-COV-2                                      | xDNN         | 0.973 8        | <b>0.991 6</b> | 0.955 3        | 0.973 1        |
|                                                 | Xception     | 0.889 8        | 0.905 5        | 0.867 0        | 0.886 1        |
|                                                 | Mvitv2       | 0.927 1        | 0.922 3        | 0.925 8        | 0.924 1        |
|                                                 | Metaformer   | 0.949 6        | 0.932 8        | 0.962 6        | 0.947 5        |
|                                                 | 本文方法         | <b>0.981 2</b> | <b>0.985 5</b> | <b>0.983 1</b> | <b>0.984 2</b> |
| 私有数据集                                           | Xception     | —              | —              | —              | —              |
|                                                 | Mvitv2       | —              | —              | —              | —              |
|                                                 | Metaformer   | —              | —              | —              | —              |
|                                                 | AMCL-FSIC    | 0.712 9        | 0.510 8        | 0.736 6        | 0.713 1        |
|                                                 | Mixture loss | 0.792 1        | 0.612 9        | 0.778 4        | 0.728 7        |
|                                                 | 3D-CNN       | 0.805 2        | 0.592 3        | <b>0.782 0</b> | 0.712 2        |
|                                                 | 本文方法         | <b>0.829 7</b> | <b>0.638 1</b> | 0.745 0        | <b>0.744 4</b> |

现象往往在少数类中更为显著,因此 $\alpha$ 的真实值应当大于数据分布表现出的 $\alpha$ 。因此,建议 $\alpha$ 的参数搜索空间应从理论上限附近开始,逐步向上搜索,以

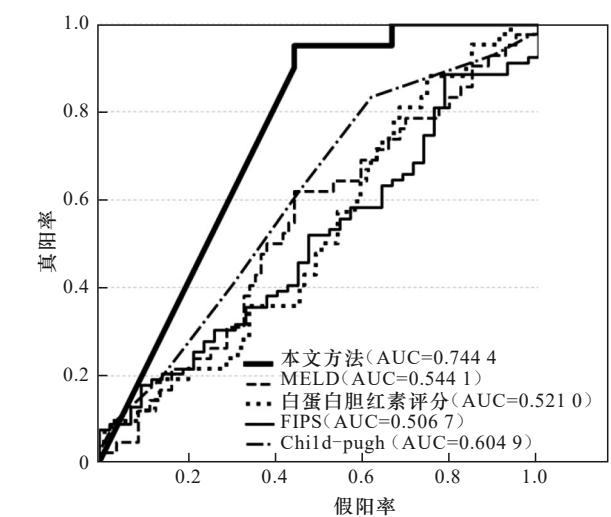


图 6 传统医学方法与本文预测方法的 ROC 曲线  
Fig.6 ROC curves of the traditional medical method and the proposed prediction method

获得最佳结果。当各项指标出现明显下降时,可能意味着搜索已经超出了最优区间。

| 表 2 超参数 $\alpha$ 搜索                    |                                    |                                    |                                    |                                    |                                    |
|----------------------------------------|------------------------------------|------------------------------------|------------------------------------|------------------------------------|------------------------------------|
| Table 2 Hyperparameter $\alpha$ search |                                    |                                    |                                    |                                    |                                    |
| $\alpha$                               | 准确率                                | 精确率                                | 召回率                                | F1                                 | ROC                                |
| 2                                      | 0.772 2 $\pm$ 0.02                 | 0.433 5 $\pm$ 0.04                 | 0.603 4 $\pm$ 0.11                 | 0.502 7 $\pm$ 0.05                 | 0.681 2 $\pm$ 0.09                 |
| 3                                      | 0.794 4 $\pm$ 0.04                 | 0.616 1 $\pm$ 0.08                 | 0.709 5 $\pm$ 0.06                 | 0.643 7 $\pm$ 0.07                 | 0.720 3 $\pm$ 0.10                 |
| 4                                      | <b>0.829 7<math>\pm</math>0.02</b> | <b>0.638 1<math>\pm</math>0.07</b> | <b>0.745 0<math>\pm</math>0.06</b> | <b>0.686 6<math>\pm</math>0.06</b> | <b>0.726 3<math>\pm</math>0.06</b> |
| 5                                      | 0.749 6 $\pm$ 0.06                 | 0.577 0 $\pm$ 0.12                 | 0.682 2 $\pm$ 0.09                 | 0.609 8 $\pm$ 0.10                 | 0.682 6 $\pm$ 0.09                 |
| 6                                      | 0.713 1 $\pm$ 0.03                 | 0.413 2 $\pm$ 0.07                 | 0.744 3 $\pm$ 0.12                 | 0.529 7 $\pm$ 0.08                 | 0.699 4 $\pm$ 0.11                 |
| 7                                      | 0.742 6 $\pm$ 0.04                 | 0.327 0 $\pm$ 0.13                 | 0.587 6 $\pm$ 0.20                 | 0.419 7 $\pm$ 0.14                 | 0.688 8 $\pm$ 0.10                 |
| 8                                      | 0.681 7 $\pm$ 0.04                 | 0.589 3 $\pm$ 0.06                 | 0.331 6 $\pm$ 0.17                 | 0.428 5 $\pm$ 0.09                 | 0.651 3 $\pm$ 0.14                 |

3.4 损失函数性能比较

为了验证改进焦点损失函数的优异性能,在三个数据集上进行测试。公共数据集使用 ResNet-50 作为主干网络。实验结果表明,传统的交叉熵损失很容易出现多数类的倾向,所有输入都被判断为多数类。在记录数据时,此类无意义的结果记录为一条短横线。在私有数据集上,本文方法的优势非常显著,准确率提高了 9% 以上,精确度达到了两倍。在两个公共数据集上,优势也较为明显。在高质量的公共数据集中,样本特征可分性高,赋予难分类的少数类样本较高权重即可轻松应对,但面对质量相对较低的私有数据集时,样本类间差距小,类内差距大,单纯聚焦难分类样本容易造成被离群样本干扰,并不能很好地完成任务,此时改进焦点损失函数稳定性强的优势更为凸显。在精度、召回率和 F1 值性能方面,尽管在某些情况下它可能不是最好的,但本文方法搜索成本小,可以迅速接近最佳性能,表现出更好的稳定性。

3.5 消融实验

表 4 为消融实验的结果。样本替换策略依赖数据增强中产生的新样本,因此二者作为整体进行实验,表中统称为样本策略。结果证明了 3 个模块的有效性,但是单个模块对基础模型的性能提升是有限的,特别是混合交叉注意力模块,该模块被用来改善特征提取能力,对样本失衡和过拟合几乎不起作用,单独使用注意力模块并不能得到有意义的结果。然而在 3 个模块的共同作用下,模型达到了最佳效果。值得注意的是改进焦点损失和混合交叉注意力模块的联合使用显著提高了模型的性能。这是由于损失函数对模型特征学习的约束只关注结果,而混合交叉注意力模块在模型训练阶段中起到提取感兴趣区域的作用,这迫使模型聚焦学习疾病本身的特征,而不是可分性强但没有实际意义的特征。因此,这两个模块相当于在端到端任务中间添加流程监督,在任务中体现了相互促进的效果,表现出了总体大于各部分之和的协同效应。

表 3 不同损失函数在三个数据集上的表现  
Table 3 Performances of different loss functions on three datasets

| 数据集                         | 损失函数  | 准确率                                | 精确率                                | 召回率                                | F1                                 |
|-----------------------------|-------|------------------------------------|------------------------------------|------------------------------------|------------------------------------|
| 本文数据集<br>( $R\approx 2$ )   | 交叉熵   | -                                  | -                                  | -                                  | -                                  |
|                             | Focal | 0.731 8 $\pm$ 0.07                 | 0.340 0 $\pm$ 0.10                 | 0.660 8 $\pm$ 0.11                 | 0.447 2 $\pm$ 0.11                 |
|                             | CBFL  | 0.745 8 $\pm$ 0.03                 | 0.355 6 $\pm$ 0.10                 | 0.660 8 $\pm$ 0.11                 | 0.460 2 $\pm$ 0.10                 |
|                             | IFL   | 0.752 7 $\pm$ 0.02                 | 0.431 7 $\pm$ 0.09                 | 0.683 3 $\pm$ 0.10                 | 0.542 8 $\pm$ 0.10                 |
|                             | 本文方法  | <b>0.829 7<math>\pm</math>0.02</b> | <b>0.638 1<math>\pm</math>0.07</b> | <b>0.745 0<math>\pm</math>0.06</b> | <b>0.686 6<math>\pm</math>0.06</b> |
| C-NMC<br>( $R\approx 2.5$ ) | 交叉熵   | 0.807 7 $\pm$ 0.02                 | 0.782 1 $\pm$ 0.04                 | 0.963 2 $\pm$ 0.03                 | 0.858 6 $\pm$ 0.04                 |
|                             | Focal | 0.818 0 $\pm$ 0.04                 | <b>0.791 1<math>\pm</math>0.06</b> | 0.958 9 $\pm$ 0.04                 | <b>0.862 5<math>\pm</math>0.06</b> |
|                             | CBFL  | 0.816 3 $\pm$ 0.02                 | 0.786 8 $\pm$ 0.05                 | 0.963 2 $\pm$ 0.03                 | 0.867 3 $\pm$ 0.06                 |
|                             | IFL   | 0.819 6 $\pm$ 0.03                 | 0.790 4 $\pm$ 0.07                 | 0.941 4 $\pm$ 0.04                 | 0.856 8 $\pm$ 0.06                 |
|                             | 本文方法  | <b>0.820 1<math>\pm</math>0.03</b> | 0.790 1 $\pm$ 0.02                 | <b>0.966 9<math>\pm</math>0.02</b> | 0.860 1 $\pm$ 0.03                 |
| CIFAR-10<br>( $R=5$ )       | 交叉熵   | -                                  | -                                  | -                                  | -                                  |
|                             | Focal | 0.531 2 $\pm$ 0.01                 | 0.063 $\pm$ 0.03                   | <b>0.984 3<math>\pm</math>0.02</b> | 0.118 4 $\pm$ 0.02                 |
|                             | CBFL  | 0.542 3 $\pm$ 0.02                 | 0.089 $\pm$ 0.02                   | 0.946 8 $\pm$ 0.01                 | 0.162 7 $\pm$ 0.02                 |
|                             | IFL   | 0.537 0 $\pm$ 0.02                 | 0.079 $\pm$ 0.02                   | 0.969 6 $\pm$ 0.03                 | 0.171 7 $\pm$ 0.03                 |
|                             | 本文方法  | <b>0.562 0<math>\pm</math>0.02</b> | <b>0.127 0<math>\pm</math>0.01</b> | 0.976 9 $\pm$ 0.02                 | <b>0.224 7<math>\pm</math>0.01</b> |

表 4 消融实验结果  
Table 4 Results of ablation experiment

| 样本策略 | 混合交叉注意力 | 改进焦点损失 | 准确率                                | 精确率                                | 召回率                                | F1                                 |
|------|---------|--------|------------------------------------|------------------------------------|------------------------------------|------------------------------------|
| ✓    |         |        | 0.689 6 $\pm$ 0.03                 | 0.363 6 $\pm$ 0.07                 | 0.666 7 $\pm$ 0.11                 | 0.470 6 $\pm$ 0.08                 |
|      | ✓       |        | -                                  | -                                  | -                                  | -                                  |
|      |         | ✓      | 0.705 8 $\pm$ 0.04                 | 0.383 7 $\pm$ 0.07                 | 0.640 0 $\pm$ 0.04                 | 0.463 1 $\pm$ 0.05                 |
| ✓    |         | ✓      | 0.745 0 $\pm$ 0.03                 | 0.403 5 $\pm$ 0.08                 | 0.677 8 $\pm$ 0.03                 | 0.519 4 $\pm$ 0.06                 |
| ✓    | ✓       |        | 0.716 2 $\pm$ 0.04                 | 0.452 7 $\pm$ 0.05                 | 0.659 3 $\pm$ 0.04                 | 0.540 1 $\pm$ 0.05                 |
|      | ✓       | ✓      | 0.793 1 $\pm$ 0.02                 | 0.543 5 $\pm$ 0.06                 | 0.703 4 $\pm$ 0.06                 | 0.615 5 $\pm$ 0.06                 |
| ✓    | ✓       | ✓      | <b>0.829 7<math>\pm</math>0.02</b> | <b>0.638 1<math>\pm</math>0.07</b> | <b>0.745 0<math>\pm</math>0.06</b> | <b>0.686 6<math>\pm</math>0.06</b> |

3.6 可视化结果

图 7 为使用 CAM 算法可视化的混合交叉注意力机制输出特征映射的类激活图。A1 和 B1 列显示第三腰椎和肝脏相关切片的 CT 原始图像。A2 和 B2 列表示混合交叉注意力机制输出特征的图像。A3 列和 B3 列展示了二者叠加的效果。A3 列显示第三腰椎感兴趣的区域。显然,脊柱及其附近的肌肉是关注的焦点,这与已有的医学研究结论一致:第三腰椎骨骼肌对 TIPS 后 HE 的预测具有重要意义<sup>[20]</sup>。B3 列所示的图像清晰地显示出模型中最感兴趣的区域是从左下到右上的狭长区域,与肝脏的位置高度吻合,与医学临床实践经验一致。

4 结束语

本文针对医学图像分析领域内存在的类不平衡小样本问题,提出一种易于扩展的多分支信息融合网络,在保留信息融合优势的情况下利用混合交叉注意力模块应对前景背景的不平衡现象,促使模型学习可分性更强的特征,并且提出了一种改进焦点损失,利用非参数幂函数的中心对称特点提高了

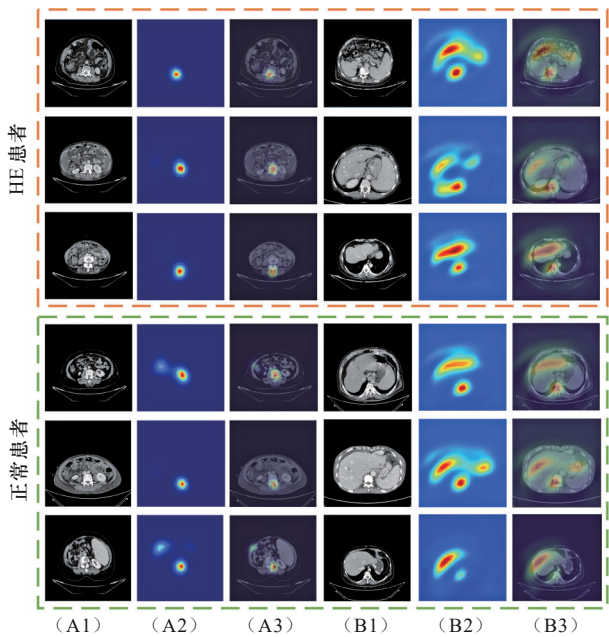


图 7 混合交叉注意力机制输出类激活图

Fig.7 Output class activation maps from mixed cross-attention mechanism

损失函数的稳定性。该方法仅引入一个超参数,通过理论推导给出取值范围并且通过实验证明了其

低搜索成本的优势。此外,提出了针对 CT 图像的数据增强方法,该方法可以有效扩展样本数量且没有合理性方面的担忧,结合基于样本分布概率的批次内样本替换策略,最终实现了高准确度的门静脉高压患者 TIPS 术前筛选及其他医学图像分析分类

任务。本文所提方法在真实的临床数据中性能显著优于现有方法。在未来的工作中,将尝试进一步降低超参数搜索带来的负担,自适应的超参数搜索方法是未来研究的方向之一。

#### 参考文献:

- [1] RAZZAK M I, NAZ S, ZAIB A. Deep learning for medical image processing: Overview, challenges and the future[J]. Classification in BioApps: Automation of decision making, 2017: 323-350.
- [2] FÜRSCHUB L, RAINER F, EFFENBERGER M, et al. A novel score predicts mortality after transjugular intrahepatic portosystemic shunt: MOTS-Modified TIPS score[J]. Liver International, 2022, 42(8): 1849-1860.
- [3] WU J, FU R, FANG H, et al. Medsegdiff: Medical image segmentation with diffusion probabilistic model[C]//Medical Imaging with Deep Learning. PMLR, 2024: 1623-1639.
- [4] 赵紫娟, 任雪婷, 宋恺, 等. 基于检索增强的中医处方生成模型[J]. 太原理工大学学报, 2025, 56(1): 114-126.  
ZHAO Z J, REN X T, SONG K, et al. Traditional Chinese medicine prescription generation model based on search enhancement[J]. Journal of Taiyuan University of Technology, 2025, 56(1): 114-126.
- [5] 李青, 李润睿, 强彦, 等. 人工智能在医学 CT 图像重建中的研究进展[J]. 太原理工大学学报, 2023, 54(1): 1-16.  
LI Q, LI R R, QIANG Y, et al. Research and progress of artificial intelligence in medical CT image reconstruction[J]. Journal of Taiyuan University of Technology, 2023, 54(1): 1-16.
- [6] YAO P, SHEN S, XU M, et al. Single model deep learning on imbalanced small datasets for skin lesion classification[J]. IEEE transactions on medical imaging, 2022, 41(5): 1242-1254.
- [7] AZAM M A, KHAN K B SALAHUDDINS, et al. A review on multimodal medical image fusion: Compendious analysis of medical modalities, multimodal databases, fusion techniques and quality metrics[J]. Computers in Biology and Medicine, 2022, 144: 105253.
- [8] 张睿, 王梓祺, 李阳, 等. 任务感知的多尺度小样本 SAR 图像分类方法[J]. 计算机科学, 2024, 51(8): 160-167.  
ZHANG R, WANG Z Q, LI Y, et al. Task-aware few-shot SAR image classification method based on multi-scale attention mechanism[J]. Computer Science, 2024, 51(8): 160-167.
- [9] LIN T Y, GOYAL P, GIRSHICK R, et al. Focal loss for dense object detection[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2017, 99: 2999-3007.
- [10] WANG Z, XIE X, YANG J, et al. Soft focal loss: Evaluating sample quality for dense object detection[J]. Neurocomputing, 2022, 480: 271-280.
- [11] JOSHI K, KUMAR M, TRIPATHI A, et al. Latest trends in multi-modality medical image fusion: A generic review[J]. Rising Threats in Expert Applications and Solutions, 2022, 2022: 663-671.
- [12] DE FRANCHIS R, BOSCH J, GARCIA-TSAO G, et al. Baveno VII - renewing consensus in portal hypertension[J]. Journal of hepatology, 2022, 76(4): 959-974.
- [13] LIN X, GAO F, WU X, et al. Efficacy of albumin - bilirubin score to predict hepatic encephalopathy in patients underwent transjugular intrahepatic portosystemic shunt[J]. European Journal of Gastroenterology & Hepatology, 2021, 33(6): 862-871.
- [14] GUPTA A, GUPTA R. ISBI 2019 C-NMC challenge: Classification in cancer cell imaging. Select Proceedings[J]. ISBI 2019 C-NMC Challenge: Classification in Cancer Cell Imaging, 2019. DOI:10.1007/978-98115-07984.
- [15] CHEN X, WANG T, JI Z, et al. 3D automatic liver and spleen assessment in predicting overt hepatic encephalopathy before TIPS: a multi-center study[J]. Hepatology International, 2023, 17(6): 1545-1556.
- [16] CAI W, LIN H, QI R, et al. Psoas muscle density predicts occurrences of hepatic encephalopathy in patients receiving transjugular intrahepatic portosystemic shunts within 1 year[J]. CardioVascular and Interventional Radiology, 2022, 45(1): 93-101.

(编辑:贾丽红)