

# 基于多模态图像融合的早期蕈样肉芽肿识别

谢凤英<sup>1</sup>, 赵丹培<sup>1</sup>, 王 可<sup>1</sup>, 刘兆睿<sup>2,3</sup>, 王煜坤<sup>2,3</sup>, 张漪澜<sup>1</sup>, 刘 洁<sup>2,3</sup>

(1. 北京航空航天大学宇航学院图像处理中心, 北京 100191; 2. 北京协和医院皮肤科, 北京 100730; 3. 中国医学科学院北京协和医学院, 北京 100730)

**摘要:** 早期蕈样肉芽肿(Mycosis fungoid, MF)可表现为红斑鳞屑性皮炎, 很难从银屑病及慢性湿疹等良性炎症性皮肤病中鉴别出来。本文提出了一种基于多模态图像融合的早期蕈样肉芽肿识别方法。该方法基于皮肤镜图像和临床图像, 采用ResNet18网络提取单模态图像的特征; 设计跨模态的注意力模块, 实现两种模态图像的特征融合; 并且设计自注意力模块提取融合特征中的关键信息, 改善信息冗余, 从而提高蕈样肉芽肿智能识别的准确度。实验结果表明, 本文所提出的智能诊断模型优于对比算法。将本文模型应用于皮肤科医生的实际临床诊断, 通过实验组医生和对照组医生平均诊断准确率的变化证实了本文模型能够有效提升临床诊断水平。

**关键词:** 皮肤镜图像; 临床图像; 多模态图像融合; 临床应用; 辅助诊断

**中图分类号:** TP183; R751      **文献标志码:** A

## Early Mycosis Fungoides Recognition Based on Multimodal Image Fusion

XIE Fengying<sup>1</sup>, ZHAO Danpei<sup>1</sup>, WANG Ke<sup>1</sup>, LIU Zhaorui<sup>2,3</sup>, WANG Yukun<sup>2,3</sup>, ZHANG Yilan<sup>1</sup>, LIU Jie<sup>2,3</sup>

(1. Image Processing Center, School of Astronautics, Beihang University, Beijing 100191, China; 2. Department of Dermatology, Peking Union Medical College Hospital, Beijing 100730, China; 3. Peking Union Medical College, Chinese Academy of Medical Science, Beijing 100730, China)

**Abstract:** Early mycosis fungoides (MFs) may present as erythematous scaly skin lesions, which are difficult to distinguish from benign inflammatory skin diseases such as psoriasis and chronic eczema. This paper presents a new method based on multimodal image fusion for early mycosis fungoides recognition. The method adopts the ResNet18 network to extract features of single-modality images based on dermoscopic images and clinical images, designs the cross-modal attention module to achieve feature fusion of two modal images, and uses the self-attention module to extract the key information and reduce redundant information in the fusion features, thereby improving the accuracy of intelligent identification of early mycosis fungoides. Experimental results show that the proposed intelligent diagnosis model outperforms the comparison algorithms. At the same time, the proposed intelligent model is applied to the actual clinical diagnosis of dermatologists. Through the changes in the average diagnostic accuracy of the experimental group and the control group, it is confirmed that the proposed intelligent diagnostic model can

**基金项目:** 国家自然科学基金(61871011, 62071011, 61971443, 82173449); 中国医学科学院中央级公益性科研院所基本科研业务费专项资金(2019XK320024); 北京市自然科学基金(4192032)。

**收稿日期:** 2022-07-30; **修订日期:** 2022-09-10

effectively improve the clinical diagnosis level.

**Key words:** dermoscopic image; clinical image; multimodal image fusion; clinical application; computer aided diagnosis

## 引言

皮肤镜是一种在体观察皮肤表皮及真皮浅层微细结构和色素的无创性图像分析技术,是各种皮肤病临床筛查和诊断的有效工具<sup>[1]</sup>。而基于皮肤镜图像的皮肤疾病计算机辅助诊断方法也是当前皮肤影像领域发展最为完善的辅助诊断技术。早期的皮肤镜图像辅助诊断方法主要是基于传统机器学习的方法,如文献[2]提取出437个颜色纹理形状特征,并优选出18个重要特征,最后采用支持向量机进行分类,敏感性和特异性分别达到92.34%和93.33%;文献[3]从皮损区域提取了颜色、纹理和形状信息,使用Adaboost方法进行分类,获得了85%的特异度和90%的敏感度;文献[4]提取对称性、离心率、颜色、纹理和边界梯度等54种特征,用于训练随机森林分类器,其特异度和敏感度分别达到44%和97%。以上方法基于传统的机器学习方法,所提取的特征大多是低级特征,算法的泛化能力不强。

近年来,基于深度学习的皮肤镜图像自动分析方法开始流行。Bi等<sup>[5]</sup>将多个全卷积神经网络进行串联,将每一级网络的输出和原始图像作为下一级网络的输入,使用元胞自动机综合各级网络输出得到皮损边界。Yuan等<sup>[6]</sup>基于Jaccard距离的损失函数来提高分割网络的准确度,获得更精确的皮损区域。Esteva等<sup>[7]</sup>在GoogLeNet Inception-v3网络上采用迁移学习对皮肤图像进行端到端的分类。在3种皮肤疾病和9种皮肤疾病的分类任务上分别获得了72.1%和55.4%的分类精度,该结果超过了两名专业医生的诊断水平。Matsunaga等<sup>[8]</sup>将多个深度神经网络进行组合实现了黑色素瘤、色素痣和脂溢性角化3种皮肤病的多分类。Jojoa等<sup>[9]</sup>使用基于掩模和区域的卷积神经网络裁剪皮损区域,并基于深度残差网络(Deep residual network, ResNet)框架将黑色素瘤病变分类为良性或恶性,最终取得90.4%的分类准确率。

在临床诊断中,皮肤科医生通常会综合皮肤镜图像和临床图像的信息来对皮肤疾病做出诊断。皮肤镜可以揭示皮损由表皮到真皮浅层的鳞屑、色素、血管等亚宏观特征。而临床图像则是由数码相机拍摄,更关注皮损的分布模式、隆起程度、颜色和边界等大体信息。不同模态信息之间特征通常互为补充,难以相互取代。因此,为了贴近皮肤科临床诊断的实际情况,进一步提高皮肤疾病智能诊断的准确率,学者们开始研究基于多模态的皮肤病辅助诊断系统。Yap等<sup>[10]</sup>将临床信息和皮肤影像同时输入ResNet,提高了5类肿瘤性疾病的分类准确率。Bi等<sup>[11]</sup>提出一种超连接网络HcCNN用于融合临床图像和皮肤镜图像,实现皮损的良恶性分类,并对病变表面的血管、点和球等模式进行预测。Tang等<sup>[12]</sup>提出了一种用于多标签皮肤病分类的两阶段多模态算法FusionM4Net,实现了78.5%的分类准确率。

上述工作促进了皮肤疾病领域计算机辅助诊断的发展,但这些工作比较关注于智能诊断方法的研究,它们以提高算法的诊断准确率为研究目标。在一些研究工作中,为了验证所提出方法的有效性,会进行一些人机对比的实验。例如文献[7]在三分类任务和九分类任务上,将智能诊断模型与专业医生的诊断结果进行对比,验证了所提出的模型已经超过了人类医生的诊断水平。然而,智能诊断方法是一种辅助诊断方法,临床诊断水平的提高仍然依赖于医生诊断水平的提高,如何将智能诊断模型应用于临床并提升临床医生的诊断水平,是当前比较欠缺的研究,也很少引起人们的关注。

本文采用深度学习方法,将临床图像与皮肤镜图像进行融合,研究蕈样肉芽肿(Mycosis fungoides, MF)的智能诊断模型,并将智能诊断模型应用于临床来提升医生的诊断准确率。MF是最常见的皮肤T细胞淋巴瘤(Cutaneous T cell lymphoma, CTCL)。据统计,早期MF患者5年总体生存率为85.8%,

晚期 MF 患者的 5 年生存率为 41%~68.5%<sup>[13]</sup>。因此, MF 晚期预后差、治疗手段有限, 而早期识别及诊治能够改善患者预后, 具有重要临床意义。早期 MF 的皮损临床典型表现为躯干四肢散在分布、大小不一的斑片/斑块, 表面略萎缩、覆有细小鳞屑, 而其皮肤镜图像则可以观察到线状血管、精子样血管及橘黄色无结构区等特征。因此, 通过融合 MF 的皮肤镜图像和临床图像, 构建基于图像智能分析技术的诊断模型, 有望实现 MF 的早期诊断。

本文的主要贡献包括: (1) 提出了一种基于注意力机制的皮损图像融合网络, 该网络将临床图像与皮肤镜图像融合, 实现早期蕈样肉芽肿的精确识别; (2) 将所提出的智能诊断模型与医生诊断结果进行对比分析, 验证了模型的有效性; (3) 将智能模型应用于实际的临床诊断, 提高了医生的临床诊断准确性, 为人工智能方法与临床诊断的实际结合提供了借鉴。

## 1 智能诊断模型设计

临床图像和皮肤镜图像作为临床中最常见的两类皮损影像模态, 临床图像强调皮损宏观结构, 皮肤镜图像强调皮损深层纹理, 二者信息互为补充, 为皮肤科医生提供了重要的诊断依据。然而, 现有的皮损图像辅助诊断方法大多基于皮肤镜图像, 缺乏临床图像的宏观信息。此外, 由于皮肤镜图像和临床图像空间特征差异大, 现有的融合方法大多无法很好地利用两种图像间的关联互补信息, 同时也不能有效降低多模态信息引入带来的信息冗余。本文提出了一种基于注意力机制的双模态融合方法, 该方法设计跨模态注意力模块实现模态间的特征互补, 并设计自注意力模块提取融合特征的关键信息, 从而提升 MF 识别的准确度。图 1 为基于注意力机制融合网络的整体结构图。该网络由特征提取模块、模态融合模块两部分组成。特征提取模块以 ResNet18 作为骨干网络, 分别提取皮肤镜图像和临床

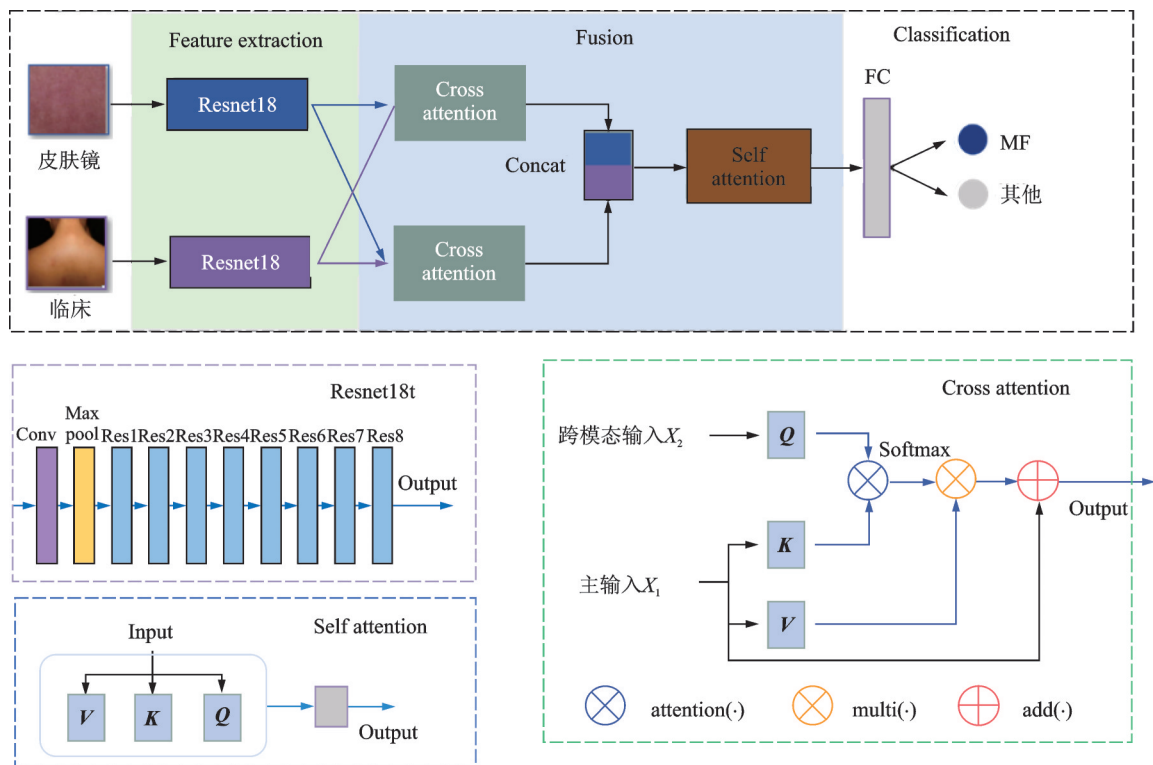


图 1 设计的多模态融合网络结构

Fig.1 Designed multi-modal fusion network structure

图像两种单模态的深层特征信息;双模态融合模块,首先利用跨模态注意力获取两种模态之间的关联信息,实现特征的互补;然后通过自注意力获得跨模态联合表征的内部相关作用,为重要程度不同的信息赋予权重系数,使模型关注更有利于分类的重要信息。

对于跨模态注意力(Cross attention),一种模态  $X_1$  作为主输入,另一种模态  $X_2$  作为跨模态输入,将其分别映射到两个不同的空间得到两个矩阵,即键矩阵  $K$  和查询矩阵  $Q$ ;然后利用点积缩放的注意力打分方式计算查询矩阵  $Q$  和键矩阵  $K$  之间的相关性,经过 softmax 获得注意力分布 attention,最终将  $X_1$  映射的值矩阵  $V$  和注意力分布 attention 相乘即可得到模态  $X_1$  关于模态  $X_2$  的跨模态表征向量,计算公式为

$$\text{attention} = \text{softmax}\left(\frac{K^T Q}{\sqrt{D_k}}\right) \quad (1)$$

$$H = V \times \text{attention} \quad (2)$$

式中: $D_k$ 为输入的维度; $H$ 为输出。

如图1所示,将皮肤镜图像特征看作  $X_1$ 、临床图像特征看作  $X_2$ ,将两者分别作为一次主输入,通过跨模态注意力,可得到两组跨模态输出  $H_{X_1 \rightarrow X_2}$  和  $H_{X_2 \rightarrow X_1}$ ,将其进行拼接操作,即可得到双模态数据的跨模态表征。

自注意力<sup>[14]</sup>可以学习输入特征的内部相关性,分配不同权重以关注关键信息,同时有效缓解多模态信息引入的信息冗余问题。类似于跨模态注意力的计算方式,但不同的是,自注意力机制中的查询矩阵  $Q$ 、键矩阵  $K$  和值矩阵  $V$  均由输入特征映射而来,得到的融合特征为

$$Z = \text{self-attention}\left(H_{X_1 \rightarrow X_2} \oplus H_{X_2 \rightarrow X_1}\right) \quad (3)$$

式中 $\oplus$ 为拼接操作。通过自注意力层,网络得到最终的双模态融合特征,经过全连接层输出分类结果。

## 2 实验数据收集

### 2.1 数据纳入对象

回顾性纳入2016年1月至2020年12月北京协和医院皮肤科门诊确诊的早期MF患者和临床表现与之相似的炎症性皮肤病(银屑病、慢性湿疹、脂溢性皮炎、玫瑰糠疹和扁平苔藓)患者。纳入标准:(1)早期MF均经组织病理学、免疫组化、T细胞受体基因重排检查等确诊,符合欧洲癌症治疗研究组织(European Organization of Research and Treatment of Cancer, EORTC)以及国际皮肤淋巴瘤学会(International Society for Cutaneous Lymphoma, ISCL)提出的早期MF诊断标准;(2)根据临床表现、皮肤镜、组织病理(部分病例)、治疗反应等由2位经过皮肤镜培训且有经验的皮肤科医师共同做出炎症性皮肤病的诊断;(3)入组前1个月内未接受系统或局部治疗;(4)均经皮肤镜检查,对于MF患者,选取早期(斑片期及斑块期)皮损进行分析。排除标准:(1)仅毛发、甲、黏膜等特殊部位受累的病例;(2)皮肤镜检查图像不清晰者。

### 2.2 图像数据采集

针对每位患者采集临床图像,同时根据皮损表面情况,采用偏振光浸润式、偏振光非浸润式或非偏振光浸润式进行皮肤镜图像采集,使用75%乙醇作为镜头与皮损之间的浸润液体,放大倍数为20、30或40倍。每位患者采集临床图像1~2张,采集皮肤镜图像若干张不等,视皮损的患病部位、分散性和采集角度而定。共纳入144个病例,包括48例早期MF患者、96例炎症性皮肤病患者(银屑病、慢性湿疹、脂溢性皮炎、玫瑰糠疹、扁平苔藓),共收集皮肤镜图像996幅、临床图像277幅。由于每个病例患者其临床图像数目少于皮肤镜图像数目,为制作多模态数据集,将单个病例内临床图像数目扩增至与皮肤

镜图像数目相同,扩增方式为在原始图像内随机复制,最终得到完成一一配对(一张皮肤镜图像对应一张临床图像)的多模态数据集。144个病例中,训练集117例,其中MF有40例、炎症性皮肤病77例(银屑病20例、慢性湿疹20例、扁平苔藓17例、脂溢性皮炎10例、玫瑰糠疹10例);测试集27例,其中MF有8例,炎症性皮肤病19例(银屑病5例、慢性湿疹5例、扁平苔藓4例、脂溢性皮炎3例、玫瑰糠疹2例)。图2给出了一组皮肤镜图像实例,其中图2(a)为早期MF,图2(b~f)为5种常见的炎症性皮肤病,可以看到MF与其他5种炎症性皮肤病具有相似的临床表现。

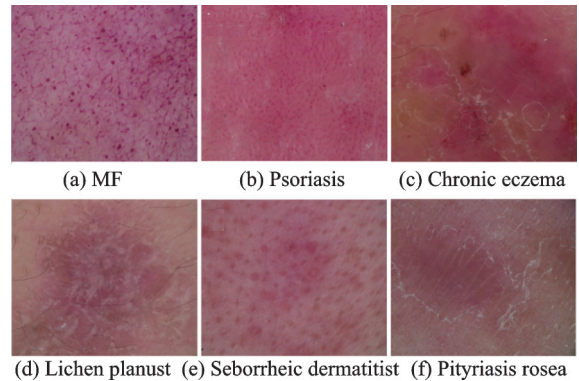


图2 早期MF及炎症性皮肤病皮肤镜图像

Fig.2 Dermoscopic images of early MF and inflammatory dermatosis

### 3 实验结果分析

本文在Pytorch环境下训练本文提出的基于多模态的智能诊断模型。网络的训练采用随机梯度下降(Stochastic gradient descent, SGD)优化器,交叉熵损失函数(按类别数量设置权重),训练中初始学习率设为0.01,并在固定批次设置学习率的衰减,批处理大小为32,权重衰减为0.001,同时采用Dropout策略来提高模型的泛化能力,Dropout的具体设定值为0.3。实验中训练轮次epoch上限为100,设置patience为30,当训练损失等待30个epoch仍不继续下降时视为网络收敛,停止训练,保存模型参数。每个训练周期中,图像按按0.5的概率随机水平、竖直翻转,旋转 $\pm 30^\circ$ 以内的角度,以增强模型对图片拍摄角度的泛化能力。

训练集有117例病例,完成一一配对后共883幅图像对。将训练集按病例重新划分5份,采用五折交叉验证法对二分类模型进行训练,报告不同数据划分下验证集上的实验结果,以均数正负标准差( $\bar{x} \pm s$ )表示,并采用交叉配对t检验报告平均 $p$ 值表明组内数据是否具有显著性差异, $p > 0.05$ 表明无显著性差异。

本文用灵敏度Sen、特异度Spe、准确度Acc、阳性预测值PPV以及阴性预测值NPV来评价算法的分类任务性能,具体公式为

$$\text{Sen} = \text{TPR} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (4)$$

$$\text{Spe} = 1 - \text{FPR} = 1 - \frac{\text{FP}}{\text{FP} + \text{TN}} = \frac{\text{TN}}{\text{FP} + \text{TN}} \quad (5)$$

$$\text{Acc} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{TN} + \text{FN}} \quad (6)$$

$$\text{PPV} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (7)$$

$$\text{NPV} = \frac{\text{TN}}{\text{TN} + \text{FN}} \quad (8)$$

式中:TPR为真正类率;FPR为负正类率;TP为实际正类而预测正类数目;FP代表实际负类而预测正类的数目;FN为实际正类而预测负类的数目;TN为实际负类而预测负类的数目。以上指标均为数值越高,算法的分类性能越好。

对于智能诊断模型的性能统计指标,分为按图像分类和按病例分类两种情况。测试集中包含

27个病例,完成一一配对后共224对图像对。按图像分类时,每一组临床图像+皮肤镜图像一个分类结果,分类性能是对224对图像对的分类结果统计。按病例分类时,对一个病例的所有皮肤镜图像和临床图像对进行分类,将这些分类结果进行投票,作为该病例的最终分类结果。测试集中共有27例病例,按病例分类是对27个结果的统计。

### 3.1 智能诊断模型的性能验证

#### 3.1.1 与临床图像融合前后的分类对比

为了验证皮肤镜图像在与临床图像融合前后对分类性能的变化,本文实验基于ResNet18的特征提取网络,在皮肤镜图像单模态、皮肤镜和临床图像双模态条件下对MF进行识别,表1给出了两种分类方式在测试集上按图像分类的统计结果( $\bar{x} \pm s$ )。可以看出,相较于单纯采用皮肤镜图像的分类,皮肤镜图像在与临床图像融合后,所有的分类指标都得到了提升,说明单纯的皮肤镜图像模态特征信息不够丰富,难以全面地描述皮损目标的特征,而双模态融合网络则可以有效地利用皮肤镜图像的皮损深层纹理和临床图像的皮损宏观结构,二者信息互为补充,从而提升了MF的识别准确率。

表1 融合皮肤镜图像前后的实验对比结果

方法	灵敏度	特异度	准确度	阳性预测值	阴性预测值
皮肤镜模态	73.71±8.33	69.43±7.79	70.44±7.03	67.47±8.12	74.30±8.21
双模态融合	78.05±7.97	89.12±4.21	83.94±4.24	86.52±4.82	82.39±5.80

#### 3.1.2 注意力机制融合的消融实验

为了验证本文所提出的注意力机制对MF识别准确率的影响,本文设计消融实验,基于ResNet18的特征提取网络,分别搭建仅跨模态注意力、仅自注意力、跨模态注意力和自注意力结合的融合网络,对比分析3种模型的各项分类指标,实验结果见表2。可以看到,本文设计的最终融合网络,相比仅跨模态注意力和仅自注意力机制的融合网络,所有指标均获得了提升,其中准确率分别提高了3.7%和5.89%,灵敏度分别提高了1.07%和2.6%,特异度分别提高了5.7%和8.4%。加入跨模态注意力可以更好地实现两种模态之间的有效交互,自注意力可以更好地学习融合模态的内部特征,确定不同数据的贡献度,关注更重要的信息,二者结合提升了整个融合网络的表现性能。

表2 不同注意力机制的实验对比结果

方法	灵敏度	特异度	准确度	阳性预测值	阴性预测值
跨模态注意力	76.98±8.45	83.42±6.86	80.24±4.95	81.56±5.37	79.67±7.33
自注意力	75.45±8.55	80.72±8.17	78.05±5.74	78.72±7.87	78.14±7.04
跨模态注意力+自注意力	78.05±7.97	89.12±4.21	83.94±4.24	86.52±4.82	82.39±5.80

#### 3.1.3 与其他融合算法的对比

本文将本文提出的基于注意力机制的融合算法与其他的多模态融合方法进行对比,包括简单的Concat方法、张量融合网络(Tensor fusion network, TFN)<sup>[15]</sup>和低秩多模态融合(Low-rank multimodal fusion, LMF)方法<sup>[16]</sup>,其中TFN对不同模态特征进行笛卡尔积计算,获取不同模态信息之间的所有交互作用,LMF是TFN方法的升级,采用低秩权重分解来减少参数数量,提升运算效率。实验结果见表3,可以看到相对于多模态融合算法,本文方法在多数指标上达到了最优,说明与文献方法相比,本文的多模态融合算法能更有效地融合皮肤镜图像和临床图像,获取其相关互补信息,提升网络对MF的识别准确率。

表3 不同融合算法的实验对比结果

Table 3 Experimental comparison results of different fusion algorithms

%

方法	灵敏度	特异度	准确度	阳性预测值	阴性预测值
Concat	76.31±8.63	83.00±7.45	79.70±5.58	80.54±6.85	79.57±7.39
TFN	74.88±8.13	82.97±7.76	79.45±6.17	79.30±6.73	79.90±4.78
LMF	75.12±9.09	<b>89.26±3.36</b>	82.66±5.01	86.40±3.16	80.12±6.37
本文方法	<b>78.05±7.97</b>	89.12±4.21	<b>83.94±4.24</b>	<b>86.52±4.82</b>	<b>82.39±5.80</b>

### 3.1.4 与其他皮损分类模型进行对比

将本文所提出的多模态融合皮损分类算法与其他4种皮肤镜图像分类算法进行对比,以验证本文方法对于皮肤镜图像蕈样肉芽肿的识别性能。实验结果见表4,其中SENet<sup>[17-18]</sup>、DenseNet<sup>[18-19]</sup>和EfficientNet-B0<sup>[20]</sup>是常用于皮肤镜图像分类的网络,ARL\_ResNet<sup>[21]</sup>是在ResNet基础上提出的改进算法。可以看到,本文方法在所有指标上优于对比算法,说明相对于其他已有的皮损分类算法,本文基于注意力机制的双模态融合算法更能有效地提取皮损目标更丰富的特征信息,具有更好的表征能力和泛化能力,提升了网络对MF皮肤病的识别能力。

表4 不同皮损分类算法的实验对比结果

Table 4 Experimental comparison results of different skin lesion classification algorithms

%

方法	灵敏度	特异度	准确度	阳性预测值	阴性预测值
SENet	57.36±8.17	81.94±5.06	71.26±5.04	70.76±6.90	71.68±5.45
DenseNet	65.88±8.31	76.40±8.66	71.98±3.16	68.76±4.01	74.94±4.50
EfficientNet-B0	72.82±5.54	80.96±6.96	77.48±3.57	75.18±6.54	79.68±3.97
ARL_ResNet	74.71±7.54	77.58±5.26	76.28±2.87	76.28±5.65	76.59±3.57
本文方法	<b>78.05±7.97</b>	<b>89.12±4.21</b>	<b>83.94±4.24</b>	<b>86.52±4.82</b>	<b>82.39±5.80</b>

## 3.2 与医生诊断结果的对比分析

本文将上述训练好的智能模型应用于皮肤科医生的临床诊断,来验证其对于临床诊断的辅助价值。参加诊断测试的皮肤科医师一共有10名,其中实验组医师和对照组医生各5名。所有参与诊断的皮肤科医师均经过皮肤镜诊断的培训且考核通过,同时具有1年以上皮肤镜判读经验。所有参与实验的医生将得到一份问卷,并被要求一次进行所有题目的回答。实验组的医师进行诊断时,针对测试集中的每个病例,随机给出1张临床图像和1张皮肤镜图像,同时给出智能模型预测为MF及其他炎症性皮肤病的概率值。医师在不知病理结果或其他临床特征情况下对该病例进行人工诊断。对照组的医生进行诊断时,针对测试集中的每个病例,给出与实验组医生相同的1张临床图像和1张皮肤镜图像,医师在不知智能诊断模型分类、病理结果或其他临床特征情况下对该病例进行人工诊断。

本文分别计算每位医生对27例病例的诊断灵敏度、特异度、准确度、阳性预测值和阴性预测值,在各自的实验组内求取各项指标的平均值并报告95%的置信区间(Confidence interval, CI)来表示该组医生的整体诊断水平。

### 3.2.1 智能模型与医生诊断结果的对比

表5给出了测试集上智能诊断模型与对照组医生诊断的结果统计,智能模型的分类统计包括按图像分类和按病例分类两种情况( $p$ 值均大于0.05,表明组内数据无显著性差异)。针对所提出的智能诊断模型,总体上按病例分类的性能要优于按图像分类的性能。这是因为患者的皮损特征局限性分布,单幅的皮肤镜图像采集存在病变特征不明显甚至漏采集的可能性。当按病例进行分类时,多幅皮

表5 智能诊断模型与对照组医生的诊断结果对比

Table 5 Comparison of diagnosis results between intelligent diagnosis model and doctors in control group

指标	对照组医生 均值(95% CI)	CNN按图像分类 均值(95% CI)	CNN按病例分类 均值(95% CI)
灵敏度/%	70.00(39.41~100.00)	79.97(75.28~84.65)	<b>80.96</b> (74.95~85.24)
特异度/%	93.68(85.16~100.00)	89.40(85.12~93.68)	<b>94.64</b> (91.72~97.56)
准确度/%	86.67(75.59~97.74)	84.37(81.78~86.97)	<b>90.13</b> (86.78~93.48)
阳性预测值/%	84.48(62.25~100.00)	<b>88.95</b> (83.96~93.95)	87.20(80.24~94.16)
阴性预测值/%	88.67(78.36~98.98)	80.33(78.09~82.58)	<b>91.36</b> (88.97~93.75)
$p$ 值	0.473	0.180	0.503

肤镜图像意味着对皮损的多点采集,减少错误分类。该实验提示,辅助诊断模型的研究应该基于采用多点采集进行判别,这与专业医师临床诊断的形式也是一致的。本文后续的实验也是采用按病例分类的方式将智能模型应用于医生的临床诊断。从表5中还可以看出,对照组医生的诊断准确率在所有指标上低于智能模型按病例分类,说明智能诊断模型优于医生诊断水平的平均值。

3.2.2 智能诊断模型辅助临床诊断前后的对比

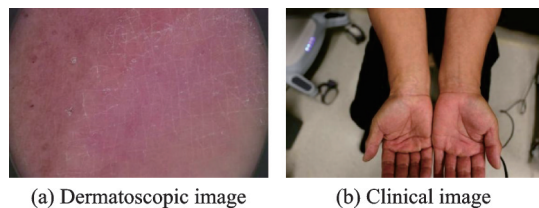
表6给出了对照组医生、实验组医生、以及CNN智能模型对测试集的结果统计( $p$ 值均大于0.05,组内数据无显著性差异)。可以看到,相比于未使用智能模型的对照组医生和智能诊断模型分类,使用智能诊断模型的实验组医生各项指标均有了显著的提升,平均值均达到90%以上水平。其中,实验组灵敏度(95.00%,86.50%~100%)相比于对照组(70.00%,39.41%~100%)提升了25%,表明对MF的漏诊减少。

表6 智能诊断模型辅助医生临床诊断前后的结果对比

Table 6 Comparison of results before and after clinical diagnosis assisted by intelligent diagnosis model

指标	对照组医生 均值(95% CI)	实验组医生 均值(95% CI)	CNN按病例分类 均值(95% CI)
灵敏度/%	70.00(39.41~100.00)	<b>95.00</b> (86.50~100.00)	80.96(74.95~85.24)
特异度/%	93.68(85.16~100.00)	<b>96.30</b> (93.05~99.55)	94.64(91.72~97.56)
准确度/%	86.67(75.59~97.74)	<b>96.30</b> (93.05~99.55)	90.13(86.78~93.48)
阳性预测值/%	84.48(62.25~100.00)	<b>93.06</b> (85.15~100.00)	87.20(80.24~94.16)
阴性预测值/%	88.67(78.36~98.98)	<b>97.95</b> (94.46~100.00)	91.36(88.97~93.75)
$p$ 值	0.473	0.489	0.503

图3给出了一个被智能诊断模型诊断正确的MF病例,未使用辅助诊断模型的对照组医生均将该病例误诊为其他炎症性皮肤病,而使用辅助诊断模型的实验组医生中有80%的医生将该病例正确诊断为MF。智能诊断模型输出的概率值分别为MF65.76%、其他炎症性皮肤病34.23%,MF的概率值较高,表明智能诊断模型可以对难分病例提供有效提示,提升医生的诊断准确率。



(a) Dermatoscopic image (b) Clinical image

图3 被对照组医生误诊的一个MF病例

此外,本文对智能诊断模型误诊的病例 Fig.3 An MF case misdiagnosed by doctors in the control group



进行分析,如图4所示,该病例为MF病例,而智能诊断模型将其错误诊断为其他炎症性皮损。针对该病例,对照组中60%的医生诊断正确,实验组中80%的医生诊断正确。实验组医生和对照组医生是不同的人,实验组在有智能诊断模型信息输出的情况下,5个人中仍然有4人坚持判别为MF,比对照组诊断准确率还

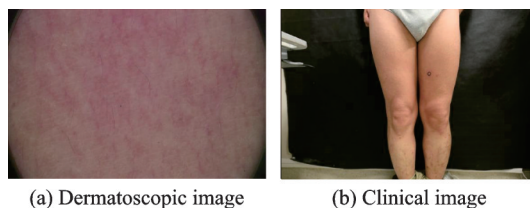


图4 被智能诊断模型误诊的一个MF病例  
Fig.4 An MF case misdiagnosed by the intelligent diagnosis model

要高,说明智能诊断模型未显示出明显干扰医生判断的情况,其原因可能为智能诊断模型输出的概率值分别为MF的44.50%和其他炎症性皮肤病的55.50%,其概率值相对接近。在该种情况下提示医生此病例智能诊断模型难以区分,需要医生仔细分辨,结合临床特征做出诊断。

#### 4 结束语

MF是最常见的皮肤T细胞淋巴瘤,晚期预后差、治疗手段有限,早期识别及诊治能够改善患者预后,具有重要临床意义。然而,早期MF可表现为红斑鳞屑性皮炎,与银屑病、慢性湿疹等良性炎症性皮肤病症状类似,难以鉴别。本文提出了一种基于注意力机制的双模态融合方法,设计跨模态注意力模块实现皮肤镜图像和临床图像间的特征互补,并且设计自注意力模块提取融合特征的关键信息,从而提升了MF智能识别的准确度。同时本文将所研究智能诊断模型的预测概率值辅助皮肤科专业医生的临床诊断,通过实验组和对照组医生的诊断准确率变化,探索了智能诊断模型在临床中的应用模式,证明了将智能模型应用于临床可以有效提高MF的临床诊断水平。但是,由于蕈样肉芽肿病例较难收集,本文所收集的数据以及参与实验的医生还不够多。下一步将收集更多的病例,设计更完备的实验方法和流程,进行更全面的分析。同时会将该研究方法在其他皮肤病病种中进行验证,为人工智能在皮肤科临床诊断中的实际应用提供经验。

#### 参考文献:

- [1] 谢凤英,刘洁,崔勇,等.皮肤镜图像计算机辅助诊断技术[J].中国医学文摘:皮肤科学,2016,33(1):45-50.  
XIE Fengying, LIU Jie, CUI Yong, et al. Computer aided diagnosis technology of dermatoscope image[J]. Chinese Medical Digest: Dermatology, 2016, 33(1): 45-50
- [2] CELEBI M E, KINGRAVI H A, UDDIN B, et al. A methodological approach to the classification of dermoscopy images[J]. Computerized Medical Imaging and Graphics, 2007, 31(6): 362-373.
- [3] CAPDEHOURAT G, COREZ A, BAZZANO A, et al. Toward a combined tool to assist dermatologists in melanoma detection from dermoscopic images of pigmented skin lesions[J]. Pattern Recognition Letters, 2011, 32(16): 2187-2196.
- [4] FERRIS L K, HARKES J A, GILBERT B, et al. Computer-aided classification of melanocytic lesions using dermoscopic images[J]. Journal of the American Academy of Dermatology, 2015, 73(5): 769-776.
- [5] BI L, KIM J, AHN E, et al. Dermoscopic image segmentation via multistage fully convolutional networks[J]. IEEE Transactions on Biomedical Engineering, 2017, 64(9): 2065-2074.
- [6] YUAN Y, CHAO M, LO Y C. Automatic skin lesion segmentation using deep fully convolutional networks with jaccard distance[J]. IEEE Transactions on Medical Imaging, 2017, 36(9): 1876-1886.
- [7] ESTEVA A, KUPREL B, NOVOA R A, et al. Dermatologist-level classification of skin cancer with deep neural networks[J]. Nature, 2017, 542(7639): 115-118.
- [8] MATSUNAGA K, HAMADA A, MINAGAWA A, et al. Image classification of melanoma, nevus and seborrheic keratosis by deep neural network ensemble[EB/OL]. (2017-03-09)[2022-06-15]. <https://arxiv.org/abs/1703.03108v1>.
- [9] JOJOA ACOSTA M F, CABALLERO TOVAR L Y, GARCIA-ZAPIRAIN M B, et al. Melanoma diagnosis using deep

- learning techniques on dermoscopic images[J]. BMC Medical Imaging, 2021, 21(1): 1-11.
- [10] YAP J, YOLLAND W, TSCHANDL P. Multimodal skin lesion classification using deep learning[J]. Experimental Dermatology, 2018, 27: 1261-1267.
- [11] BI L, FENG D D, FULHAM M, et al. Multi-label classification of multi-modality skin lesion via hyper-connected convolutional neural network[J]. Pattern Recognition, 2020, 107: 107502.
- [12] TANG P, YAN X, NAN Y, et al. FusionM4Net: A multi-stage multi-modal learning algorithm for multi-label skin lesion classification[J]. Medical Image Analysis, 2022, 76: 102307.
- [13] MOURAD A, GNIADDECKI R. Overall survival in mycosis fungoides: A systematic review and meta-analysis[J]. Journal of Invest Dermatol, 2020, 140: 495-497.
- [14] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[C]//Proceedings of the 31st International Conference on Neural Information Processing Systems. [S.l.]: ACM, 2017, 30: 6000-6010.
- [15] ZADEH A, CHEN M, PORIA S, et al. Tensor fusion network for multimodal sentiment analysis[EB/OL].(2017-0723)[2022-0615]. <https://arxiv.org/abs/1707.07250?context=cs>.
- [16] LIU Z, SHEN Y, LAKSHMINARASIMHAN V B, et al. Efficient low-rank multimodal fusion with modality-specific factors [EB/OL].(2018-05-31)[2022-0615]. <https://arxiv.org/abs/1806.00064v1>.
- [17] HU J, SHEN L, SUN G. Squeeze-and-excitation networks[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake City, UT, USA: IEEE, 2018: 7132-7141.
- [18] 刘兆睿, 张漪澜, 谢凤英, 等. 基于皮肤镜图像智能分析的早期蕈样肉芽肿诊断模型构建[J]. 协和医学杂志, 2021, 12(5): 689-697.
- LIU Zhaorui, ZHANG Yilan, XIE Fengying, et al. Early diagnosis model of mycosis fungoides based on intelligent analysis of dermoscopic images[J]. Union Medical Journal, 2021, 12 (5): 689-697
- [19] HUANG G, LIU Z, VAN DER MAATEN L, et al. Densely connected convolutional networks[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, HI, USA: IEEE, 2017: 4700-4708.
- [20] YANG Y, WANG J, XIE F, et al. A convolutional neural network trained with dermoscopic images of psoriasis performed on par with 230 dermatologists[J]. Computers in Biology and Medicine, 2021, 139: 104924.
- [21] ZHANG J, XIE Y, XIA Y, et al. Attention residual learning for skin lesion classification[J]. IEEE Transactions on Medical Imaging, 2019, 38(9): 2092-2103.

## 作者简介:



谢凤英(1972-),通信作者,女,博士,教授,博士生导师,研究方向:医学图像处理、遥感图像理解和应用、图像质量评估, E-mail: xfy\_73@buaa.edu.cn.



赵丹培(1977-),副教授,博士生导师,研究方向:医学图像分析、天基空间目标的信息处理、遥感图像处理与内容理解等, E-mail: zhaodanpei@buaa.edu.cn.



王可(2001-),女,硕士研究生,研究方向:医学图像处理、计算机视觉。



刘兆睿(1993-),女,博士,住院医师,研究方向:皮肤影像学及皮肤淋巴瘤。



王煜坤(1995-),男,博士研究生,研究方向:皮肤影像及人工智能。



张漪澜(1999-),女,硕士研究生,研究方向:医学图像处理、计算机视觉。



刘洁(1975-),女,博士,教授,主任医师,博士生导师,研究方向:皮肤影像学及皮肤淋巴瘤。