楚胜轩, 谷雪莲, 管睿. 基于阴道镜三视图特征集成的宫颈病变诊断[J]. 智能计算机与应用,2025,15(5):52-60. DOI:10.20169/j.issn.2095-2163.25011004

基于阴道镜三视图特征集成的宫颈病变诊断

楚胜轩[1], 谷雪莲[1], 管 睿[2]
(1 上海理工大学 健康科学与工程学院, 上海 200093; 2 海军军医大学第一附属医院 妇产科, 上海 200433)

摘 要: 宫颈癌早期筛查可有效防控病情,降低发病死亡率。人工阴道镜检查虽广泛用于宫颈癌筛查,但主观性强且特异性低。因此,本研究提出一种基于阴道镜三视图特征集成的算法,以辅助临床医生筛查,提高诊断效率与准确性。该方法首先利用设计的 CNN-Transformer 网络 CTLG 对同一患者的醋酸图像、绿透镜图像和碘图像同时进行特征提取,融合后实现对宫颈病变准确诊断。为了训练和评估所提出算法的性能,基于 6 774 张临床阴道镜图像数据集进行实验。在测试集上,该算法的总体准确率达到 90.93%,召回率、精确率、特异性和 *F*1 分数分别为 95.40%、92.12%、95.82%和 93.69%。误诊率和漏诊率仅为 4.60%和 4.18%。综上所述,该算法具有较高的临床应用价值,有助于降低误诊率和漏诊率,提高宫颈癌筛查效率。

关键词: 宫颈癌; 三视野图像; 特征集成; 分类诊断
中图分类号: R737.25 **文献标志码**: A **文章编号**: 2095-2163(2025)05-0052-09

Cervical lesion diagnosis based on colposcopic three-view feature integration

CHU Shengxuan[1], GU Xuelian[1], GUAN Rui[2]
(1 School of Health Science and Engineering, University of Shanghai for Science and Technology, Shanghai 200093, China; 2 Department of Gynecology and Obstetrics, the First Affiliated Hospital of Naval Medical University, Shanghai 200433, China)

Abstract: Early screening for cervical cancer is effective in preventing and controlling the disease and reducing morbidity and mortality. Although manual colposcopy is widely used for cervical cancer screening, it is highly subjective and has low specificity. Therefore, this study proposes an algorithm based on the integration of colposcopic three-view features to assist clinicians in screening and improve diagnostic efficiency and accuracy. The method starts with simultaneous feature extraction of acetic acid images, green lens images and iodine images of the same patient using a designed CNN-Transformer network CTLG, which is fused to achieve accurate diagnosis of cervical lesions. In order to train and evaluate the performance of the proposed algorithm, experiments are conducted based on a dataset of 6 774 clinical colposcopy images. On the test set, the algorithm achieves an overall accuracy of 90.93% with recall, precision, specificity and *F*1 - Score of 95.40%, 92.12%, 95.82% and 93.69%, respectively. The misdiagnosis and underdiagnosis rates are only 4.60% and 4.18%. In conclusion, the algorithm has high clinical application value, which helps to reduce the misdiagnosis and missed diagnosis rate and improve the efficiency of cervical cancer screening.

Key words: cervical cancer; three-view images; feature integration; classification diagnosis

0 引 言

宫颈癌是最常见的女性恶性肿瘤之一,对女性健康构成了严重威胁。据国际癌症研究机构数据,2024 年全球宫颈癌新增 60 万例,死亡 34 万例[1]。宫颈癌前病变一般需 5~8 年发展为癌症,其主要特征为宫颈上皮内病变(CIN)[2]。WHO 将 CIN 分为 CIN1、CIN2 和 CIN3,其中 CIN1 属于低级别鳞状上皮内病变(LSIL),CIN2/CIN3 属于高级别鳞状上皮内病变(HSIL)[3]。LSIL 患者通常可通过自身免疫恢复,而 HSIL 患者有进一步发展为宫颈癌的趋势。因此,宫颈癌的早期筛查,尤其是 HSIL 的检测,对

基金项目: 江苏省科技计划项目(BE2023737)。
作者简介: 楚胜轩(1998—),男,硕士研究生,主要研究方向:图像处理;管 睿(1975—),女,博士,主任医师,硕士生导师,主要研究方向:卵巢囊肿,子宫肌瘤,子宫肌腺症等常见妇科疾病的微创治疗。
通信作者: 谷雪莲(1972—),女,博士,副教授,硕士生导师,主要研究方向:基于 AI 技术的医疗器械优化设计研究。Email: guxuelian@usst.edu.cn。
收稿日期: 2025-01-10

于预防宫颈病变的进一步恶化至关重要[4]。

宫颈癌筛查方法包括液基薄层细胞检测(TCT)、人类乳头状瘤病毒(HPV)检测以及阴道镜检查。TCT 和 HPV 检测较为简单,但不足以全面评估女性患宫颈癌的风险,容易导致漏诊和误诊。阴道镜检查可直接观察与宫颈癌相关的病变,同时判定病灶的严重程度,必要时对疑似病变区域进行活检,从而有效提高宫颈病变的诊断准确率。因此,阴道镜检查被认为是宫颈癌前病变检测的金标准[5]。在阴道镜检查过程中,临床医生首先将生理盐水涂抹于宫颈上皮,初步确定转化区,然后将 3%~5%的醋酸溶液涂抹宫颈,观察宫颈呈现的变化,再通过绿色透镜观察有无异常血管,最后将复方碘溶液涂抹宫颈,做进一步的观察。若发现可疑区域,临床医生会从该区域取小部分组织进行活检。然而,由于受到临床医生操作水平和经验的影响,阴道镜检查存在主观性强且准确性较低的问题。因此,利用人工智能技术辅助临床医生进行宫颈癌筛查诊断是十分必要的[6]。

人工智能技术在临床医学影像辅助分析中得到了广泛的重视和应用研究[7]。在辅助宫颈病变诊断的研究中,许多研究忽略了阴道镜检查过程中绿透镜图像的状态,仅利用醋酸图像或宫颈碘图像作为模型的输入。然而,未成熟化生的鳞状上皮、雌激素缺乏的鳞状上皮等非病理组织可能在醋酸图像和宫颈碘图像中出现假阳性反应,从而导致误诊。通过绿透镜可观察到异常阴道镜图像中的血管镶嵌或点状形态等病变特征表现。因此,同时结合醋酸图像、宫颈碘图像和绿透镜图像实现信息互补,可降低单一视图不足以提供宫颈病变特征的影响。

为了捕获更充分的宫颈病变特征,本研究提出了一种基于三视野图像特征集成的宫颈病变诊断算法 CTLG Ensemble。首先,该算法利用设计的 CNN-Transformer 特征提取主干 CTLG,从 3 种视野图像中提取病变特征。具体而言,CTLG 由卷积编码块(CEB)、转置融合块(TFB)和窗口注意力块(WAB)组成。CEB 捕获多局部尺度的特征,TFB 将捕获的特征解码为相同的比例以进行融合,再将融合后的特征输入到 WAB 以捕获全局特征。其次,利用连接层将 3 个视野的特征整合在一起,再使用构建的三状态感知卷积层来降低特征维度,并进一步提取三状态特征。最后,通过 Softmax 层得到最终的预测结果。

1　相关工作

1.1　宫颈病变辅助诊断

深度学习凭借其强大的自动特征学习能力被广泛应用在基于阴道镜图像的宫颈病变辅助诊断研究中。Zhang 等学者[8]用预训练的 DenseNet 在醋酸图像数据集上训练模型,在测试集上的准确率达到了 73.08%。Li 等学者[9]使用 ResNet 作为骨干网络,并通过增添醋酸白上皮纹理和血管外观线索构建自动诊断模型,准确率为 81.07%。Kalbhor 等学者[10]使用 ResNet-50 和 GoogLeNet 提取阴道镜图像特征,在测试评估中,准确率达到了 76.64%。

然而,上述研究仅利用单张醋酸图像,忽略了检查过程中其他状态的图像,这将损失很多特征信息。为此,Yan 等学者[11]构建基于注意力机制的特征选择模块,并利用因子化双线性汇集技术将醋酸和宫颈碘图像的特征进行有效融合,实现了 74.6%的灵敏度、85.5%的准确率。Fan 等学者[12]使用预训练的 EfficientNet-B3 作为特征提取主干,有效融合醋酸图像和宫颈碘图像的特征,在区分正常、LSIL、HSIL+测试中,准确率达到 92.70%。Chen 等学者[13]采用 EfficientNet-B0 作为特征提取主干,并使用 GRU 将原始图像、醋酸图像和宫颈碘图像进行特征融合。Xu 等学者[14]利用改进的 DenseNet 分别提取盐水图像、醋酸图像以及宫颈碘图像的特征,并将其融合,有效提升了模型性能。值得注意的是,在大部分阴道镜异常表现中,点状、镶嵌等形态的血管是辅助诊断病变级别的重要特征。然而,这些采取图像融合方法的研究忽略了可明显表现异型血管的绿透镜图像。当遇到假醋白反应或假阳性碘反应的图像时,模型可能会判断错误。为此,本研究将醋酸图像、绿透镜图像和碘图像在特征级上融合来改善这个问题。

1.2　特征提取网络

由于不同级别宫颈病变之间具有相似的特征表现,设计一个具有强大学习能力的特征提取网络至关重要。在特征提取方面,卷积神经网络(CNN)能够通过连续的卷积运算捕获输入的空间结构特征。然而,阴道镜图像中宫颈病变的特征表现较为复杂,且每个病例的病灶位置及大小并不统一。CNN 的卷积运算仅捕获局部特征,对于长距离的像素关系感知能力较弱,在处理复杂问题时存在局限性。Transformer 通过自注意力机制能够捕捉输入序列中的长距离依赖关系,从而捕获图像的全局特征。因

此,将 CNN 与 Transformer 结合,可同时捕获局部和全局特征,有助于减少对潜在病变区域的忽视,提高特征提取能力。

近年来,联合 CNN 和 Transformer 的方法在计算机视觉的诸多领域得到了广泛的应用。例如,Wang 等学者[15]提出了一种 CNN 和 Transformer 联合的网络 CCTNet,将由 CNN 得到的局部信息和由 Transformer 得到的全局信息结合在一起,实现了性能的突破。Li 等学者[16]提出了 CNN 联合 Transformer 的网络 CNN-T,在 NEU-CLS 数据集上展现出优秀的性能。Lee 等学者[17]提出了一种 Plant-CNN-ViT 模型用于植物叶片分类,在 Flavia 和 Swedish Leaf 等数据集上均实现了极高的准确率。上述研究证实了 CNN 结合 Transformer 的有效性。因此,针对病变情况复杂的阴道镜图像,设计 CNN 联合 Transformer 作为特征提取主干,以提高宫颈病变诊断性能是一种有前景的方法。

2 材料和方法

2.1 方法概述

为充分利用三视野图像信息,本研究提出一种基于三视野图像特征集成的宫颈病变诊断算法 CTLG Ensemble,其总体框架如图 1 所示。首先,将从医院采集的原始阴道镜图像数据进行预处理。其次,将预处理后的三视野图像分别输入到包含 CTLG 的 3 个特征提取分支中,提取各个视野图像的特征。然后,利用连接层将 3 个视野的特征整合在一起,再使用构建的三状态感知卷积层来降低特征维度,并学习 3 种状态特征之间的关联性和依赖关系,以进一步提炼出更有判别性的高层特征。最后,通过 Softmax 层来产生 Normal、LSIL、HSIL 和 Cancer 四个类别的概率,以实现宫颈病变诊断。

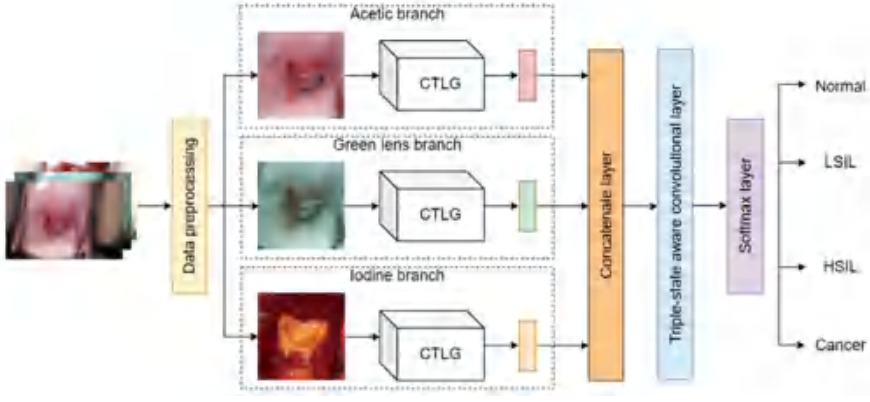


图 1 CTLG Ensemble 的总体框架

Fig. 1 The overall structure of CTLG Ensemble

2.2 数据集

本研究回顾性分析了长海医院 2 258 例阴道镜检查患者数据。每例包含 1 份检查报告单和 3 张阴道镜图像:醋酸图像、绿透镜图像和碘图像。所有图像的分辨率均为 6 000×4 000 像素,样本示例如图 2 所示。以检查报告单中的活检结果作为金标准,将所有数据分为 4 类:Normal、LSIL、HSIL 和 Cancer。其中包括 751 例 Normal,541 例 LSIL,609 例 HSIL 和 357 例 Cancer,总计 6 774 张图像。数据集的划分情况见表 1。

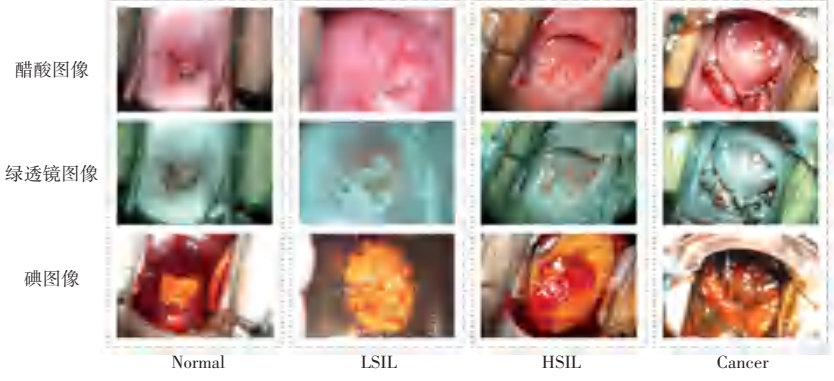


图 2 样本示例

Fig. 2 Examples of samples

表 1　数据分布

Table 1　Data distribution

数据	Normal	LSIL	HSIL	Cancer	总计
训练集	1 689	1 218	1 371	804	5 082
验证集	225	162	183	108	678
测试集	339	243	273	159	1 014
总计	2 253	1 623	1 827	1 071	6 774

2.3　数据预处理

由于采集环境差异，部分原始阴道镜图像质量不佳。实验前，首先对所有阴道镜图像执行宫颈区域提取。对质量差的图像进行图像增强或移除镜面反射。然后，为防止模型过拟合，通过随机翻转和裁剪扩增数据集。最后，将所有处理后的图像统一缩放至 224×224 像素。

2.3.1　宫颈区域的提取

为消除阴道镜图像中阴道壁、毛发和器械等信息的干扰，首先在阴道镜专家指导下用矩形框标注阴道镜图像中的宫颈区域，然后再根据大量手工标注的数据训练 Mask R-CNN[18]，最后应用训练好的模型提取阴道镜图像中的宫颈区域，这一过程的结果如图 3 所示。

图 3　宫颈区域的提取

Fig. 3　Extraction of cervical region

2.3.2　图像增强

为提高模型对目标区域的识别和分析准确性，本研究采用直方图均衡化技术[19]来增强图像对比度和细节。首先，将原始图像转换至 LAB 色彩空间。然后，对 L、A、B 3 个通道分别应用对比度受限自适应直方图均衡化（CLAHE）算法。最后，合并处理后的 3 个通道，将结果重新映射至 RGB 色彩空间。图 4 展示了图像增强的效果。

图 4　阴道镜图像增强

Fig. 4　Colposcopic image enhancement

2.3.3　移除镜面反射

镜面反射表现为宫颈表面的高亮小斑点，主要由宫颈表面黏液反射相机闪光灯而产生[20]。由于镜面反射可能会干扰模型对阴道镜图像的分析，本研究采用以下步骤处理：首先，将原图像转换为仅包含亮度信息的灰度图，再结合 Top-hat 滤波和全局阈值化方法对灰度图进行处理，生成二值图像。然后，利用二值图像作为掩码，从原图像中分离出镜面反射区域。其次，将分离出的镜面反射区域图像转换为三通道图像。最后，在每个通道上选取黑色像素，并应用形态学膨胀操作填充镜面反射区域。图 5 展示了镜面反射去除的效果。

图 5　阴道镜图像中的镜面反射去除

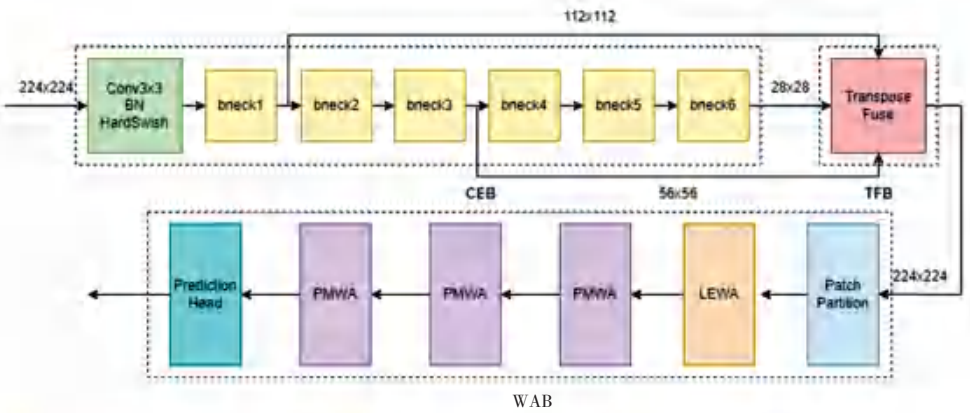
Fig. 5　Specular reflections removal in colposcopic image

2.4　特征提取主干 CTLG

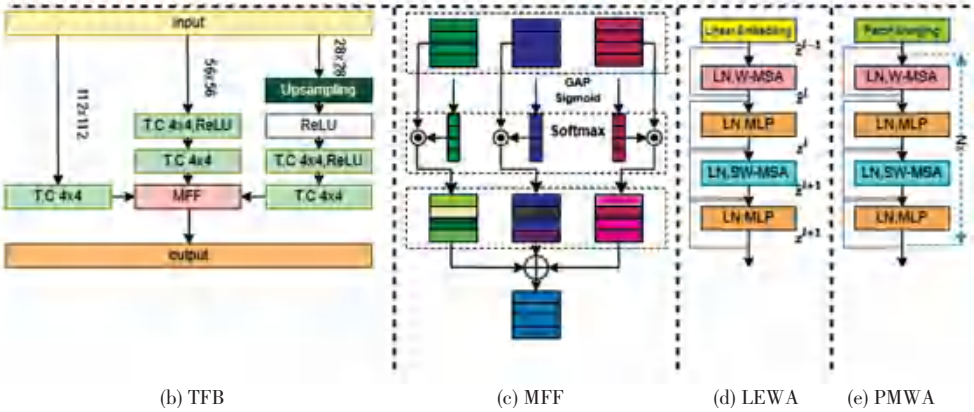
为提高宫颈病变分类的性能，本研究设计了一种强大的特征提取网络 CTLG，作为 CTLG Ensemble 的主干网络（如图 6 所示）。CTLG 由 3 个关键组件构成：卷积编码块（Convolutional Encoding Block，CEB）、转置融合块（Transpose Fusion Block，TFB）和窗口注意力块（Window Attention Block，WAB）。CEB 利用 CNN 主干网络捕获多尺度局部特征；TFB 将捕获的特征解码至相同尺度，以便进行特征融合；融合后的特征输入 WAB，以捕获全局特征信息。通过这 3 个组件的协同作用，CTLG 能够有效捕获从 28×28 到 224×224 像素尺度的多局部和全局特征。

在 CEB 中，采用了经过的 MobileNetV3[21]作为主干网络，具体选取了前 7 个子层。这一主干网络由 1 个卷积层和 6 个倒置残差块组成。当输入依次通过 bneck1、bneck3 和 bneck6 时，可分别获得 112×112、56×56 和 28×28 像素尺度的局部特征。

在 TFB 中，具体结构如图 6（b）所示。112×112 尺度的特征通过 1 个 4×4 转置卷积层，56×56 尺度的特征通过 2 个 4×4 转置卷积层和 1 个 ReLU 层，28×28 尺度的特征通过 1 个上采样层、2 个 ReLU 层和 2 个 4×4 转置卷积层。该过程可得到 3 个 shape 相同的尺度特征表示：$[X_0, X_1, X_2 \in R^{C\times H\times W}]$，采用多特征融合模块（MFF）将其进行融合，MFF 的细节如图 6（c）所示。这 3 个特征表示通过在空间维度上的全局平均池化操作获得全局特征表示，再通过 1×1 卷积建模通道间的相关性，Sigmoid 激活函数生成通道描述符：



(a) CTLG 结构



(b) TFB (c) MFF (d) LEWA (e) PMWA

图 6 CTLG 的详细结构

Fig. 6 The detailed structure of CTLG

$$X_0^{'} \in R^{C\times 1\times 1} = \text{Sigmoid}(Conv(AvgPool(X_0))) \quad (1)$$

$$X_1^{'} \in R^{C\times 1\times 1} = \text{Sigmoid}(Conv(AvgPool(X_1))) \quad (2)$$

$$X_2^{'} \in R^{C\times 1\times 1} = \text{Sigmoid}(Conv(AvgPool(X_2))) \quad (3)$$

将 $X_0^{'}$, $X_1^{'}$, $X_2^{'}$ 在第二个维度上进行拼接,得到 $X_ \in R^{C\times N\times 1}$, 然后通过 Softmax 函数获得各自的权重表示:

$$X_ = \text{Concat}(X_0^{'}, X_1^{'}, X_2^{'}) \quad (4)$$

$$W \in R^{C\times N\times 1} = \text{Softmax}(X_) \quad (5)$$

$$W[:,i,:] \in R^{C\times 1\times 1} \quad (6)$$

将生成的权重 W 与对应的输入进行逐元素乘法,并相加:

$$Y = W[:,0,:] \odot X_0 + W[:,1,:] \odot X_1 + W[:,2,:] \odot X_2 \quad (7)$$

其中 $Y \in R^{C\times H\times W}$ 是融合后的特征表示。

在 WAB 中,利用 Swin Transformer[22] 作为主干,包含 4 个阶段:Patch Partition、线性嵌入窗口注意(LEWA)块、补丁合并窗口注意(PMWA)块和 Prediction Head。WAB 将 FFB 输出的特征作为输入,捕获尺度为 224×224 的全局特征。LEWA 和 PMWA 都由 Windows Multi-head Self-Attention(W-MSA)、LayerNorm (LN)、MLP 和 Shifted Windows Multi-Head Self-Attention(SW-MSA)块组成,其中 LEWA 和 PMWA 分别利用 Linear Embedding 和 Patch Merging 处理输入,如图 6(d)和(e)所示。Prediction Head 由 2 个输出节点的线性层组成。LEWA 和 PMWA 中的计算可表示为:

$$\hat{z}^{l} = W - MSA(\text{LN}(z^{l-1})) + z^{l-1} \quad (8)$$

$$z^{l} = MLP(\text{LN}(\hat{z}^{l})) + \hat{z}^{l} \quad (9)$$

$$\hat{z}^{l+1} = SW - MSA(\text{LN}(z^{l})) + z^{l} \quad (10)$$

$$z^{l+1} = MLP(\text{LN}(\hat{z}^{l+1})) + \hat{z}^{l+1} \quad (11)$$

其中,$\hat{z}^{l}$ 和 z^{l} 分别表示第 l 个 block 的 W-MSA 和 MLP 的输出特征。

2.5 实验准备

所有实验均基于 Python3.7 和 PyTorch 1.13.1 实现,并在 NVIDIA Tesla V100 GPU(32 GB RAM)上进行训练。优化器采用随机梯度下降(SGD)算法,动量为 0.9,权重衰减为 1e-4。损失函数选择交叉熵损失函数,初始学习率设置为 1e-4,训练周期总数设置为 300,批次设置为 16。训练过程中的动态变化公式如下:

$$lr = lr_{\text{initial}} \times \left(1 - \frac{epoch}{epochs}\right)^{0.8} \quad (12)$$

其中，$lr_{initial}$ 表示初始学习率；*epoch* 表示当前训练周期；*epochs* 表示训练周期总数。

3　实验

3.1　评估指标

为了分析所提出模型的性能，选择医疗诊断领域广泛采用的 7 个评估指标，即准确率、召回率、精确率、特异性、*F*1 − *Score*、漏诊率（α），误诊率（β）。上述指标定义公式为：

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \tag{13}$$

$$Recall = \frac{TP}{TP + FN} \tag{14}$$

$$Precision = \frac{TP}{TP + FP} \tag{15}$$

$$Specificity = \frac{TN}{TN + FP} \tag{16}$$

$$F1 - Score = \frac{2 \times Precision \times Recall}{Precision + Recall} \tag{17}$$

$$\alpha = 1 - Recall = \frac{FN}{TP + FN} \tag{18}$$

$$\beta = 1 - Specificity = \frac{FP}{TN + FP} \tag{19}$$

其中，*TN*、*TP*、*FN*、*FP* 分别表示正确预测的阴性、阳性样本数和错误预测的阴性、阳性样本数，*F*1 − *Score* 表示精确率和召回率的调和平均值。

受试者工作特征曲线（*ROC*）也用于评估诊断效果。*ROC* 曲线以假阳性率为横轴、真阳性率为纵轴，通过调整阈值生成，其曲线下面积（*AUC*）越接近 1，模型性能越好。

3.2　CTLG Ensemble 的宫颈病变分类性能表现

所提出的 CTLG Ensemble 在测试集上的混淆矩阵如图 7 所示。CTLG Ensemble 表现出优异的整体分类性能，对于 Normal、LSIL、HSIL 和 Cancer 四个类别的分类，总体准确率为 90.93%，召回率、精确率、特异性和 *F*1 分数分别为 95.40%、92.12%、95.82% 和 93.69%。此外，CTLG Ensemble 的预测与实际病理诊断结果的一致性分别为 90.86%，87.66%，92.31% 和 93.71%。

3.3　CTLG Ensemble 中主干的比较

为评估所设计的 CTLG 的性能，将其与多个主流模型进行了对比实验。除 CTLG 外，还选用了 VGG16、ResNet50、MobileNetV3、DenseNet − 161 和 Swin Transformer 作为特征提取主干，并在本研究的数据集上进行了实验。实验结果见表 2。结果表明，以 CTLG 作为主干网络时，各项性能指标均优于其他模型，这充分展示了 CTLG 在学习宫颈病变特征方面的强大能力，从而有效提高了宫颈病变的分类性能。

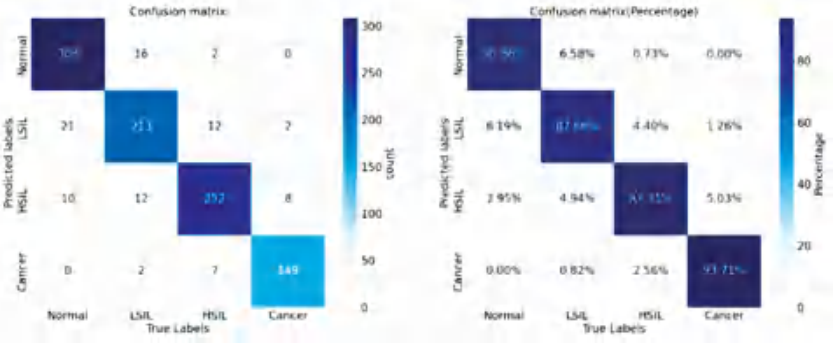


图 7　CTLG Ensemble 在测试集上的混淆矩阵

Fig. 7　Confusion matrix of the CTLG Ensemble on the test set

表 2　CTLG 与其他模型性能对比

Table 2　Performance comparison between CTLG and competing models　　%

模型	*Accuracy*	*Recall*	*Precision*	*Specificity*	*F*1 − *Score*	α	β
VGG16	82.38	89.49	88.27	89.27	88.88	10.51	10.73
ResNet−50	85.56	91.43	92.81	90.81	92.15	8.57	9.19
MobileNetV3	85.49	90.22	90.91	88.91	90.56	9.78	11.09
DenseNet−161	86.08	91.56	89.17	90.17	90.35	8.44	9.03
Swin−Transformer	86.56	90.43	88.81	93.81	89.61	9.57	6.19
CTLG	90.93	95.40	92.12	95.82	93.69	4.60	4.18

图 8 展示了 CTLG 及其他模型作为特征提取主干的 ROC 曲线。所提出的 CTLG 作为主干时，即 CTLG Ensemble，展现出最高的平均 *AUC* 值，整体性能优于所有对比模型。特别值得注意的是，对 HSIL 和 Cancer 类别表现尤为出色，这具有重要的临床意义。

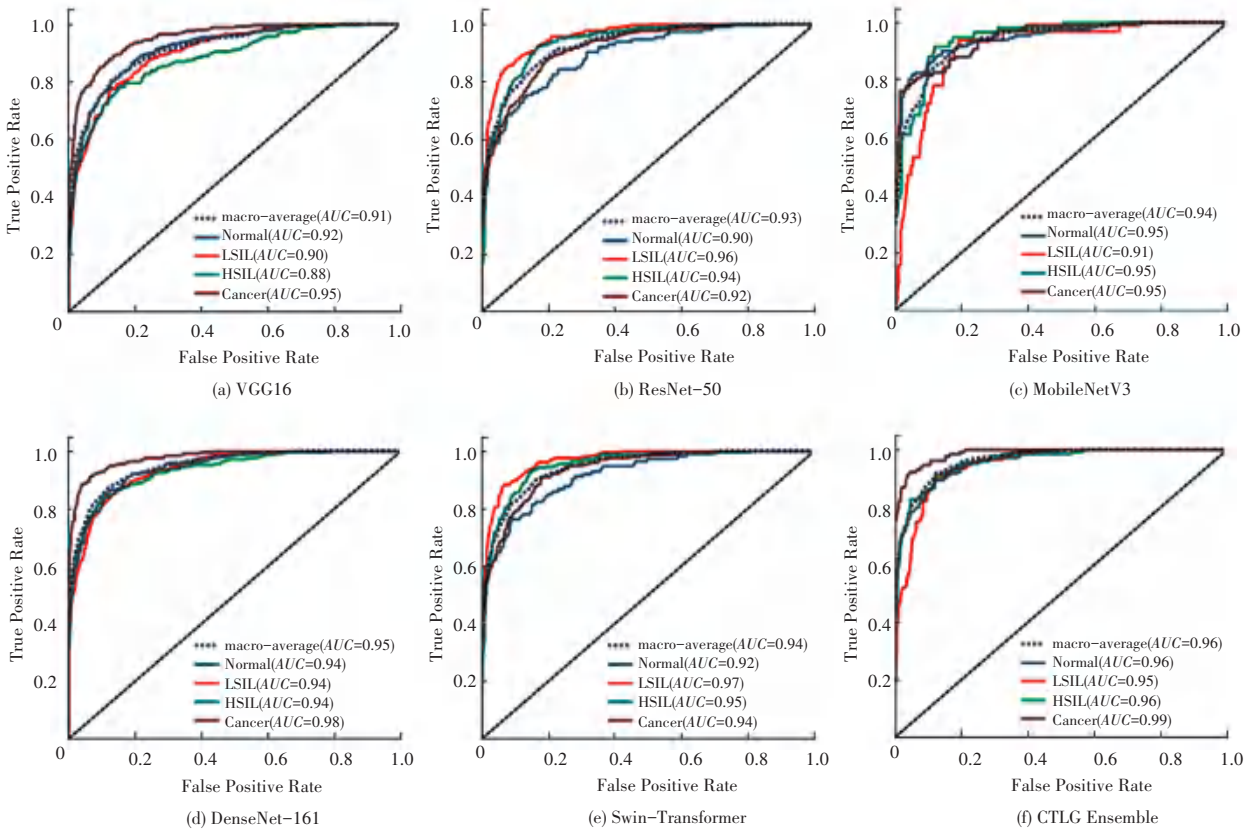


图 8 CTLG Ensemble 与其他方法的 *ROC* 曲线

Fig. 8 The *ROC* curves of CTLG Ensemble and other models

3.4 消融实验

3.4.1 数据预处理方法的作用

为验证所提出的数据预处理方法的实用性和必要性，本研究设计了一组对比实验，比较了数据预处理前后模型的分类准确率。表 3 展示了实验结果，所有模型在经过预处理的数据上均获得了更高的分类准确率。这一结果表明，经过处理的数据能够帮助模型更有效地学习关键特征。这不仅证实了所提出预处理方法的有效性，还凸显了其在提高宫颈病变分类性能中的重要作用。

表 3 数据预处理前后实验结果

Table 3 Experiments comparison before and after data preprocessing %

模型	数据	预处理的数据
VGG16	80.41	82.38
ResNet-50	84.07	85.56
MobileNetV3	83.35	85.49
DenseNet-161	84.87	86.08
Swin-Transformer	84.72	86.56
CTLG	87.34	90.93

3.4.2 CTLG 中不同块和特征捕获尺度的作用

为验证捕获多局部和全局特征能力的有效性和重要性，本研究对 CTLG 的不同块配置和特征捕获尺度进行了一系列消融实验。在不同的块配置和特征捕获尺度上进行实验，实验结果见表 4。首先，在仅使用 CEB 的情况下，采用设计的分类头进行训练，分别评估了其捕获 28×28、56×56 和 112×112 局部特征后的分类性能。结果表明，这种配置下整体性能较差。随后，添加了 TFB。通过 CEB 和 TFB 的结合，可获得 3 个局部特征融合后的进阶特征，整体性能得到了进一步提高。对于仅使用 WAB 的情况，由于其具有 Prediction Head 可直接进行训练，结果显示在只提取 224×224 全局特征后的分类性能表现不错。本研究未单独评估 TFB，因其主要用于特征解码和融合。当同时结合 CEB、TFB 和 WAB 时，模型能够捕获 28×28 到 224×224 的多尺度局部和全局特征。相比仅使用 CEB 或 WAB 的方案，这种组合表现最佳，达到了 90.93%的准确率、95.40%的召回率和 95.82%的特异性。实验结果表明，随

着捕获特征量的增加，模型的判别能力得到提升，从而提高了分类性能。这凸显了多尺度特征捕获在宫颈病变分类中的重要性。同时，给出了与其他相关方法的比较结果见表 5。

表 4　CTLG 的消融实验

Table 4　Ablation experiment of CTLG

CEB	TFB	WAB	28×28	56×56	112×112	224×224	*Accuracy*/%	*Recall*/%	*Specificity*/%
√	×	×	√	×	×	×	65.09	65.42	68.51
√	×	×	×	√	×	×	68.64	69.02	67.82
√	×	×	×	×	√	×	71.38	79.46	71.25
√	√	×	√	√	√	×	80.57	82.57	83.36
×	×	√	×	×	×	√	86.56	90.43	93.81
√	√	√	√	√	√	√	90.93	95.40	95.82

表 5　与其他相关方法的比较

Table 5　Comparison with other related methods

引文	年份	算法	数据类型	诊断类别	准确率/%
Li 等学者[9]	2021	ResNet+MR	醋酸图像	CIN1/(CIN2/3)	81.07
Yan 等学者[11]	2021	BF-CNN	醋酸图像和碘图像	LSIL-/HSIL+	85.50
Fang 等学者[23]	2022	ShuffleNet_SE	醋酸图像	Normal/LSIL/HSIL/Cancer/Neoplasm	81.38
Xu 等学者[14]	2023	CTIFI	盐水图像、醋酸图像和碘图像	Normal/LSIL/HSIL/Cancer	71.00
Chen 等学者[24]	2024	MFEM-CIN	醋酸图像	LSIL/HSIL	89.20
本文	2025	CTLG Ensemble	醋酸图像、绿透镜图像和碘图像	Normal/LSIL/HSIL/Cancer	90.93

4　结束语

在阴道镜检查中，临床医生需要综合分析醋酸图像、绿透镜图像和碘图像来做出诊断。基于这一临床实践，本研究提出了一种基于三视野图像特征集成的宫颈病变诊断算法 CTLG Ensemble。首先，对所有原始阴道镜图像预处理以排除干扰信息提高图像质量，再用 CTLG 提取三视野图像的特征，通过连接层整合后用三状态感知卷积层降维提取特征。最后经过 Softmax 实现 Normal、LSIL、HSIL 和 Cancer 四类准确诊断。

经过训练和验证，CTLG Ensemble 在测试集上展现出优异的性能：准确率达到 90.93%，召回率、精确度、特异性和 *F*1 分数分别为 95.40%、92.12%、95.82%和 93.69%。此外，本研究还对比了近年来宫颈癌辅助诊断领域的突出研究。大部分研究所使用的数据并不全面，多集中于二分类研究，对于临床医师的帮助有限。与 Xu 等学者[14]和 Fang 等学者[23]的方法相比，本研究提出的方法不仅可实现多种病变级别的诊断，而且具有更高的准确率，能在临床筛查诊断中发挥重要的辅助作用。鉴于这些出色的结果，本研究提出的算法有望成为有效的辅助诊断工具，协助临床医生进行全面而准确的诊断工作。

参考文献

[1] SIEGEL R L, GIAQUINTO A N, JEMAL A. Cancer statistics, 2024[J]. CA: A Cancer Journal for Clinicians, 2024, 74: 12-49.

[2] BUCKLEY C H, BUTLER E B, FOX H. Cervical intraepithelial neoplasia[J]. Journal of Clinical Pathology, 1982, 35: 1-13.

[3] MOSCICKI A B, SHIBOSKI S, HILLS N K, et al. Regression of low-grade squamous intra-epithelial lesions in young women[J]. The Lancet, 2004, 364: 1678-1683.

[4] SHIIRAZ A, CRAWFORD R, EGAWA N, et al. The early detection of cervical cancer. The current and changing landscape of cervical disease detection[J]. Cytopathology, 2020, 31: 258-270.

[5] REN Hongyan, JIA Mengzhe, ZHAO Shujun, et al. Factors correlated with the accuracy of colposcopy-directed biopsy: A systematic review and meta-analysis[J]. Journal of Investigative Surgery, 2022, 35: 284-292.

[6] XUE P, NG M T A, QIAO Y. The challenges of colposcopy for cervical cancer screening in LMICs and solutions by artificial intelligence[J]. BMC Medicine, 2020, 18: 169.

[7] SHEN D, WU G, SUK H I. Deep learning in medical image analysis[J]. Annual Review of Biomedical Engineering, 2017, 19: 221-248.

[8] ZHANG Tao, LUO Yanmin, LI Ping, et al. Cervical precancerous lesions classification using pre - trained densely connected convolutional networks with colposcopy images[J]. Biomedical Signal Processing and Control, 2020, 55: 101566.

[9] LI Y, LIU Z H, XUE P, et al. GRAND: A large-scale dataset and benchmark for cervical intraepithelial Neoplasia grading with fine-grained lesion description[J]. Medical Image Analysis, 2021, 70: 102006.

[10] KALBHOR M M, SHINDE S V. Comparative analysis of deep learning pre-trained models and machine learning classifiers for accurate cervical cancer diagnosis using colposcopy images[C]// Proceedings of the 2023 7th International Conference on Computing, Communication, Control and Automation (ICCUBEA). Pune, India: dblp, 2023: 1-6.

[11] YAN Ling, LI Shufeng, GUO Yi, et al. Multi-state colposcopy image fusion for cervical precancerous lesion diagnosis using BF-CNN[J]. Biomedical Signal Processing and Control, 2021, 68: 102700.

[12] FAN Yinuo, MA Huizhan, FU Yuanbin, et al. Colposcopic multimodal fusion for the classification of cervical lesions[J]. Physics in Medicine & Biology, 2022, 67: 135003.

[13] CHEN Xiaoyue, PU X, CHEN Zirou, et al. Application of EfficientNet-B0 and GRU-based deep learning on classifying the colposcopy diagnosis of precancerous cervical lesions[J]. Cancer Medicine, 2023, 12: 8690-8699.

[14] XU T, LIU P, WANG X, et al. CTIFI: Clinical-experience-guided three-vision images features integration for diagnosis of cervical lesions[J]. Biomedical Signal Processing and Control, 2023, 80: 104235.

[15] WANG H, CHEN X, ZHANG T, et al. CCTNet: Coupled CNN and transformer network for crop segmentation of remote sensing images[J]. Remote Sensing, 2022, 14: 1956.

[16] LI Shunfeng, WU Chunxue, XIONG Naixue. Hybrid architecture based on CNN and transformer for strip steel surface defect classification[J]. Electronics, 2022, 11: 1200.

[17] LEE C P, LIM K M, SONG Yuxuan, et al. Plant-CNN-ViT: Plant classification with ensemble of convolutional neural networks and vision transformer[J]. Plants, 2023, 12: 2642.

[18] WU Minghu, YUE Hanhui, WANG Juan, et al. Object detection based on RGC mask R-CNN[J]. IET Image Processing, 2020, 14: 1502-1508.

[19] ABDULLAH - Al - WADUD M, KABIR MD H, AKBER DEWAN M A, et al. A dynamic histogram equalization for image contrast enhancement[J]. IEEE Transactions on Consumer Electronics, 2007, 53: 593-600.

[20] ZIMMERMAN - MORENO G, GREENSPAN H. Automatic detection of specular reflections in uterine cervix images[J]. Proceedings of SPIE, 2006, 6144: 2037-2045.

[21] HOWARD A, SAMDLER M, CHU G, et al. Searching for MobileNetV3[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. Piscataway, NJ: IEEE, 2019: 1314-1324.

[22] LIU Ze, LIN Yutong, CAO Yue, et al. Swin transformer: Hierarchical vision transformer using shifted windows[C]// Proceedings of the IEEE/CVF International Conference on Computer Vision. Piscataway, NJ: IEEE, 2021: 10012-10022.

[23] FANG S, YANG J, WANG M, et al. An improved image classification method for cervical precancerous lesions based on ShuffleNet[J]. Computational Intelligence and Neuroscience, 2022, 2022: 9675628.

[24] CHEN P, LIU F, ZHANG J, et al. MFEM-CIN: A lightweight architecture combining CNN and transformer for the classification of pre-cancerous lesions of the cervix[J]. IEEE Open Journal of Engineering in Medicine and Biology, 2024, 5: 216-225.