

基于不同深度学习架构建立结肠镜质量控制的人工智能辅助系统

陈健¹, 张子豪², 王甘红³, 王珍妮¹, 夏开建⁴, 徐晓丹¹

1. 常熟市第一人民医院/苏州大学附属常熟医院消化内科, 江苏 苏州 215500; 2. 上海豪兄教育科技有限公司, 上海 200434; 3. 常熟市中医院消化内科, 江苏 苏州 215500; 4. 常熟市医学人工智能与大数据重点实验室, 江苏 苏州 215500

【摘要】目的:利用不同深度学习架构模型构建结肠镜质量控制的深度学习模型,并深入探索其决策机制。**方法:**基于HyperKvasir和苏州大学附属常熟医院的数据集,筛选结肠镜图像,涵盖不同清洁度的肠道、息肉及盲肠。图像经过预处理和增强后,采用基于卷积神经网络(CNN)和Transformer的预训练模型进行迁移学习。模型训练采用交叉熵损失函数,使用Adam优化器,并实施学习率调度。为提高模型透明度,进行深入的可解释性分析,包括梯度加权分类激活映射、指导式梯度加权分类激活映射和沙普利加性解释等技术。最后,模型被转换为开放神经网络交换格式(ONNX)并部署到多种设备终端,以实现结肠镜质量的实时控制。**结果:**在3831张结肠镜内窥镜图像中,EfficientNet模型在测试集上表现最佳,准确率达到0.992,超过其他CNN(DenseNet121、ResNet50、VGG19)和Transformer(ViT、Swin、CvT)架构模型,其精确率、召回率和F1值分别为0.991、0.989和0.990。在358张外部测试集图像上,EfficientNet模型的平均AUC、精确率和召回率分别为0.996、0.948和0.952。尽管模型整体表现出色,但仍存在一些误判情况。模型可解释性分析揭示其决策中所依赖的图像区域。此外,模型已成功转换为ONNX格式并在多种平台和设备上部署,实现每秒超过60帧的平均推理速度,确保结肠镜检查的实时质量控制。**结论:**本研究为结肠镜质量控制开发7种基于CNN与Transformer的模型,EfficientNet在各类别中展现出卓越性能,并已在多终端实现实时预测,为患者提供更高水平的医疗服务。

【关键词】深度学习; Transformer; 结肠镜质量控制; 结肠镜; 模型部署

【中图分类号】R574.6

【文献标志码】A

【文章编号】1005-202X(2024)11-1443-10

Constructing an artificial intelligence assisted system for colonoscopy quality control based on various deep learning architectures

CHEN Jian¹, ZHANG Zihao², WANG Ganhong³, WANG Zhenni¹, XIA Kaijian⁴, XU Xiaodan¹

1. Department of Gastroenterology, Changshu No. 1 People's Hospital/Changshu Hospital Affiliated to Soochow University, Suzhou 215500, China; 2. Shanghai Haoxiong Education Technology Co., Ltd., Shanghai 200434, China; 3. Department of Gastroenterology, Changshu Hospital of Traditional Chinese Medicine, Suzhou 215500, China; 4. Changshu Key Laboratory of Medical Artificial Intelligence and Big Data, Suzhou 215500, China

Abstract: Objective To develop deep learning models for colonoscopy quality control using various deep learning architectures, and to delve into the decision-making mechanisms. **Methods** The colonoscopy images were selected from two datasets separately constructed by the HyperKvasir and Changshu Hospital Affiliated to Soochow University, encompassing intestines of varying degrees of cleanliness, polyps, and cecums. After image preprocessing and enhancement, transfer learning was carried out using the pre-trained models based on convolutional neural network (CNN) and Transformer. The model training adopted cross-entropy loss functions and Adam optimizer, and simultaneously implemented learning rate scheduling. To enhance model transparency, a thorough interpretability analysis was conducted using Grad-CAM, Guided Grad-CAM, and SHAP. The final model was converted to ONNX format and deployed on various equipment terminals to achieve real-time colonoscopy quality control. **Results** In a dataset of 3 831 colonoscopy images, EfficientNet model outperformed the other models on the test set, achieving an accuracy of 0.992 which was higher than those of the other models based on CNN (DenseNet121, ResNet50, VGG19) and Transformer (ViT, Swin, CvT), with a precision, recall rate, and F1 score of 0.991, 0.989, and 0.990. On an external test set of 358 images,

【收稿日期】2024-06-19

【基金项目】苏州市科技发展计划(临床机构能力提升)项目(SLT2023006);常熟市医学人工智能与大数据重点实验室能力提升项目(CYZ202301);常熟市科技发展计划项目(CS202019, CSWS202316);常熟市科技计划(社会发展)项目(CS202452)

【作者简介】陈健,副主任医师,研究方向:消化内镜人工智能, E-mail: szcsdoctor@gmail.com

【通信作者】徐晓丹,主任医师,研究方向:医学人工智能、机器学习等, E-mail: xxddoctor@gmail.com

EfficientNet model had an average AUC, precision, and recall rate of 0.996, 0.948, and 0.952, respectively. Although EfficientNet model is high-performing, some misjudgments still occurred. Interpretability analysis highlighted key image areas affecting decision-making. In addition, EfficientNet model was successfully converted to ONNX format and deployed on multiple platforms and devices, and it ensured real-time colonoscopy quality control with an inference speed of over 60 frames per second. **Conclusion** Among the 7 models developed for colonoscopy quality control based on CNN and Transformer, EfficientNet demonstrated exemplary performance across all categories and is deployed for real-time predictions on multiple terminals, aiming to provide patients with better medical care.

Keywords: deep learning; Transformer; colonoscopy quality control; colonoscopy; model deployment

前言

随着消化内镜技术的不断进步,结肠镜检查已成为筛查早期肠癌和诊断结肠病变的首选,识别并及时切除腺瘤性息肉可显著降低结直肠癌的风险^[1-3]。然而,肠道的清洁度对检查结果至关重要,不佳的肠道准备可能导致病变的漏诊^[4-5]。完整的盲肠插管也是影响结肠镜检查质量的关键因素^[6-7]。近年来,欧洲胃肠内镜学会、中华医学会消化内镜学分会和美国消化内镜学会发布关于结肠镜筛查的质量控制声明,其中突出了腺瘤性息肉检出率、肠道准备质量和盲肠插管率等关键指标^[8-10]。深度学习凭借其卓越的特征提取和数据处理能力,为结肠镜质量控制提供智能化方案,特别是在息肉检测、肠道准备质量评估以及盲肠到达率的评估等方面^[11-13]。卷积神经网络(CNN)主要处理固定形状的数据,而基于自注意力的Transformer在自然语言处理中已树立新标准,并向计算机视觉领域拓展^[14]。本研究采用深度学习架构,目的是智能评估结肠镜的关键质量指标,为医生提供实时反馈,并为进一步培训提供数据,从而提升结肠镜的诊断和治疗效果。

1 资料与方法

1.1 数据集

本研究基于两个数据集进行:数据集1(HyperKvasir)用于模型的开发与测试,而数据集2(苏州大学附属常熟医院)作为模型的外部测试集。收集的结肠镜图像涵盖不同清洁度的肠道(采用波士顿评分BBPS)、息肉及盲肠。HyperKvasir(数据集1)是最大的胃肠道内窥镜数据集(<https://datasets.simula.no/hyper-kvasir/>),它包含超过110 079张图像和374个视频^[15]。这些数据是在挪威Bærum医院真实的胃肠和结肠镜检查中收集的,部分由经验丰富的胃肠道内窥镜医生标记,该数据集代表解剖学上的标志物以及病理学和正常的发现。本研究从中筛选1 009张盲肠图像、1 028张息肉图像以及1 794张不同清洁度的肠道图像进行分析。回顾性收集苏州大学附属常熟医院内镜中心358张结肠镜图像,作为外部测试集(数据集2),相关图像示例见图1。为提高模型的泛化性,收集的肠镜图像采用多种图像增强内镜技术,如窄带成像、蓝光成像、柔性光谱成像。

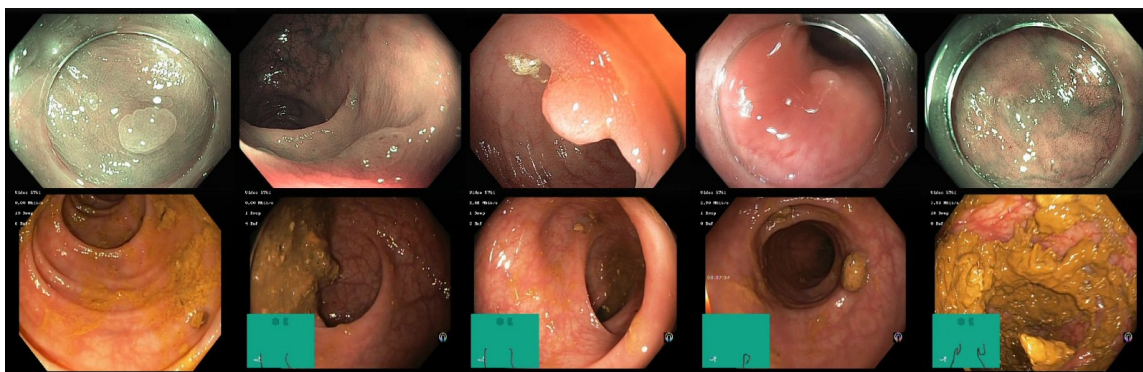


图1 数据集2中的图像示例

Figure 1 Image examples from the dataset 2

本研究排除以下情况的患者:炎症性肠病、结肠炎活动期、凝血功能障碍、家族性息肉病、急诊结肠

镜检查以及诊治信息不完整的患者。图2a展示两个数据集的图像尺寸分布,红色表示该尺寸的同类图

像数量较集中,而蓝色则表示数量较少。数据集包含多种尺寸的图像,其中尺寸为 622×529 像素和 633×532 像素的图像合计占比超过 50%。在训练集、测试集和外部测试集中,各类别图像的分布情况详

见图 2b。本研究经苏州大学附属常熟医院伦理委员会审批通过(批准号:L20230930),由于研究的非干预性回顾性设计,患者同意被豁免。

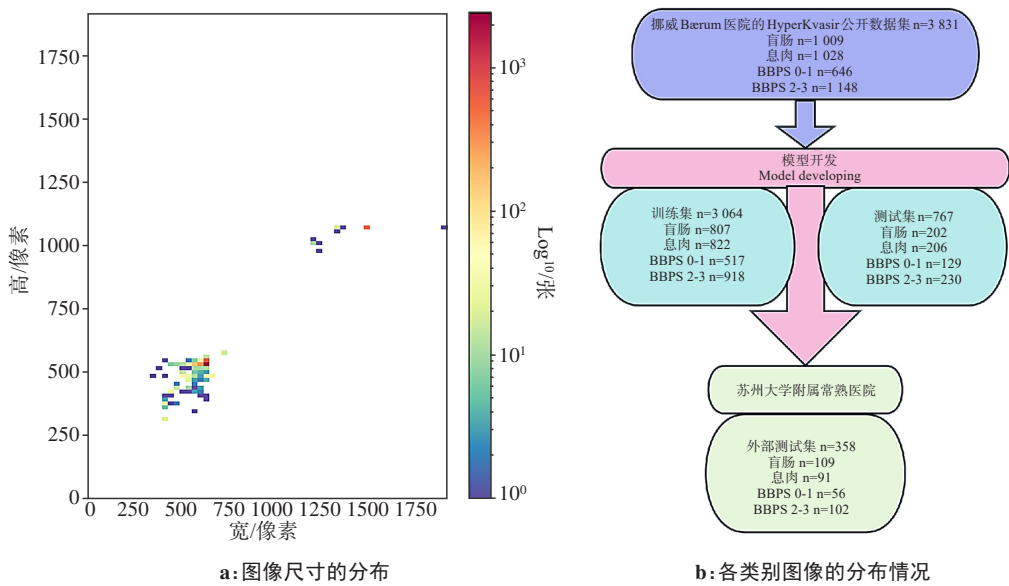


图2 数据集图像特征分析
Figure 2 Analysis of image features in the dataset

1.2 深度学习网络

1.2.1 图像预处理 为确保模型具有更好的泛化性,对图像数据进行一系列预处理和增强操作。对于训练集,首先随机调整图像大小并裁剪至 224×224 的尺寸。为了增加数据的多样性,进行随机的水平翻转。将图像从 PIL Image 或 numpy.ndarray 格式转换为 PyTorch Tensor,并将其范围归一化到[0, 1]。最后,对图像的 RGB 通道进行标准化,使用的均值和标准偏差分别为 [0.485, 0.456, 0.406] 和 [0.229, 0.224, 0.225]。对于测试集,采取不同的策略,首先调整图像的短边至 256,然后从中心裁剪出 224×224 的尺寸。后续的转变和归一化步骤与训练集相同,同时也采用相应的 RGB 通道标准化参数。所有这些步骤都是利用 PyTorch 的 torchvision 库来实现。这种处理方式在图像分类任务中已被证实可以有效提升模型表现。

1.2.2 模型训练配置 为实现图像分类,采用基于 CNN 和 Transformer 架构的预训练模型进行迁移学习。在 CNN 架构中,选择 DenseNet121、EfficientNet^[16]、ResNet50^[17]和 VGG19^[18]模型。而在 Transformer 架构中,采用 ViT(Vision Transformer)^[19]、Swin (Shifted Window Transformer)^[20]和 CvT (Convolutional Vision Transformer)^[21]模型。这些

CNN 模型均包含卷积层、平均池化层和带 ReLU 激活的全连接层。为了更好地适配数据集,在每个预训练模型上添加两个带 ReLU 激活的致密层,以及一个带 Softmax 激活的输出层进行分类。输出层的特征数被设定为 4,以适应本研究的分类任务。模型采用交叉熵作为损失函数,并使用 Adam 优化器进行 30 个训练周期的训练。同时实施学习率调度,每 5 个周期减半学习率。Transformer 模型在处理输入图像时,首先进行随机裁剪、水平翻转和最多 15° 的旋转。接着,模型将图像分为多个固定大小的块,为每块加入位置编码。这些块被输入到 Transformer 编码器中,通过编码器捕捉块之间的关联关系,最后仅利用第一个块的输出进行 4 类分类。所有操作均在 PyTorch 框架下执行。相关的神经网络架构见图 3。

1.2.3 模型可解释性分析 尽管先进的计算机视觉技术已广泛应用于医学成像,其在医学界的普及仍受高计算成本、数据限制和深度学习的黑盒特性影响。为提高透明度,可解释性人工智能被引入,旨在揭示深度学习模型的内部机制和决策过程。为了克服“黑盒效应”,对基于 CNN 和 Transformer 的高性能模型进行深入的可解释性分析,采用梯度加权分类激活映射(Grad-CAM)、指导式梯度加权分类激活映射(Guided Grad-CAM)和沙普利加性解释(SHAP)等技

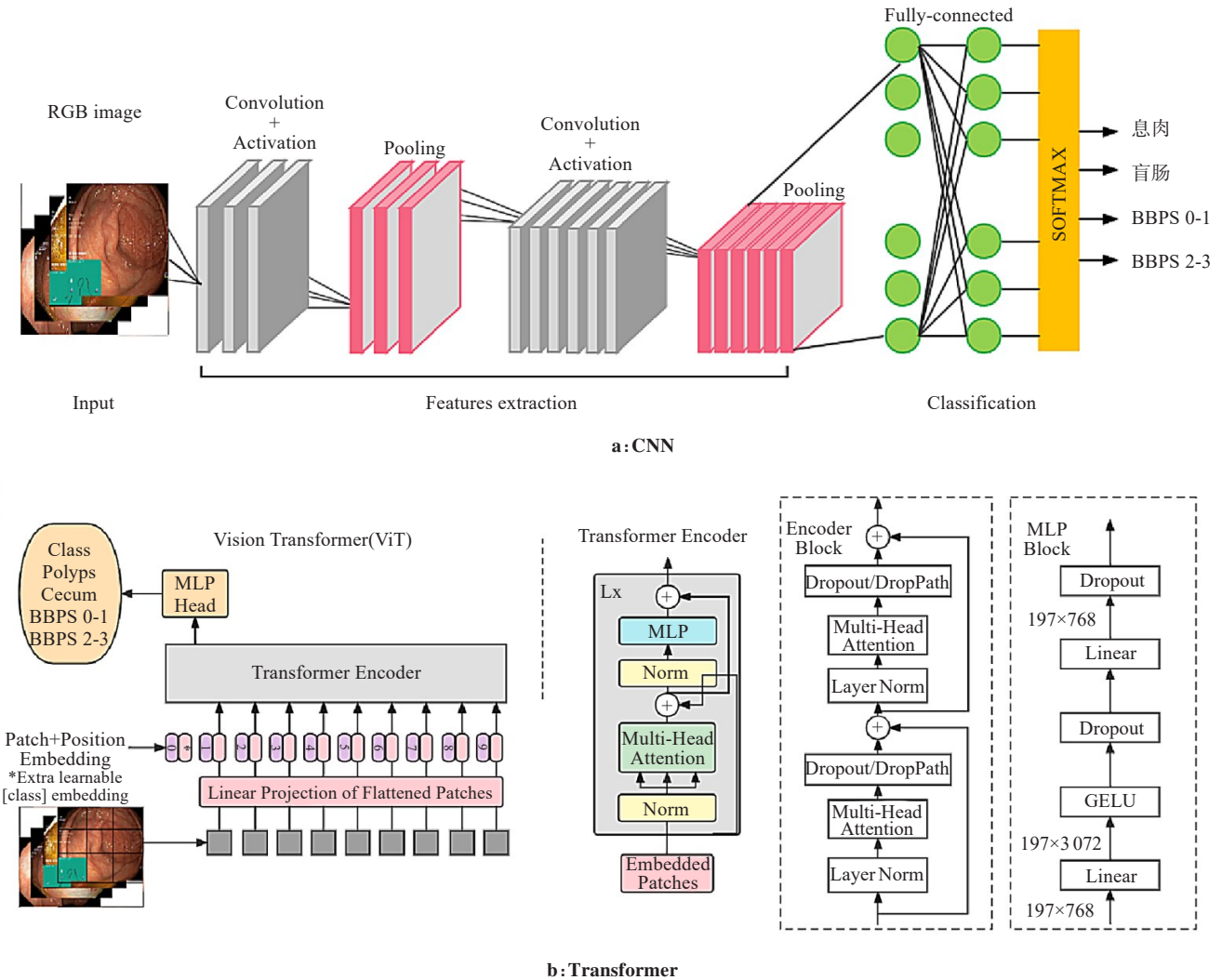


图 3 CNN 和 Transformer 的神经网络架构示意图

Figure 3 Schematic diagrams of the neural network architectures for CNN and Transformer

术^[22-24]。其中,Grad-CAM 生成热图以揭示模型决策中的关键图像区域;Guided Grad-CAM 结合 Grad-CAM 和 Guided Backpropagation,提供一个更精细的视角来理解像素如何影响决策;SHAP 则为图像分类中的每个像素分配重要性,明确其在决策中的角色。这些技术共同帮助深入理解模型如何解读不同的结肠镜图像。

1.2.4 多设备终端部署 为了系统地提升结肠镜检查的质量,本研究开发一个深度学习模型,并将其部署到多种设备终端,如内镜中心的台式电脑、笔记本电脑及浏览器。该模型旨在实时提供结肠镜质量控制,无论是检查期间还是检查后。首先,通过迁移学习得到专门针对本研究需求的 Pytorch 深度学习模型。随后,为了确保模型的跨平台部署,将其转换为开放神经网络交换格式(ONNX)。ONNX 是一个开放的模型表示标准,可以确保模型的互操作性和广泛的部署选项。通过利用 ONNX Runtime,该模型可

以在各种操作系统(例如 Linux、Windows、MacOS)上高效运行,并针对不同硬件(如 CPU、GPU)进行优化。ONNX 作为一个开放的深度学习模型标准,不仅提供模型的互操作性,还提供广泛的部署选择,从而确保结肠镜检查的准确性和效率^[25]。模型开发部署流程见图 4。

1.3 实验平台

本研究使用了配备 RTX 3060 显卡(12 GB 显存)、CPU 为 5×E5-2680 v4 以及拥有 350 GB 磁盘空间的计算机设备。借助于 TensorFlow(2.7.0)、Keras(2.7.0)以及 OpenCV(4.5.4.60)的 Python 工具库,成功地完成深度学习模型的搭建、培训及图像的处理任务。为了数据的整理、分析及可视化,采用 Pandas(1.3.4)、NumPy(1.21.4)、Matplotlib(3.5.0)和 Plotly(5.4.0)。此外,模型优化工作是通过 PyTorch(1.10.0+cu113)来实现的,模型的保存和载入则依赖于 H5py(3.6.0)。为确保模型鲁棒性和泛化能力,在

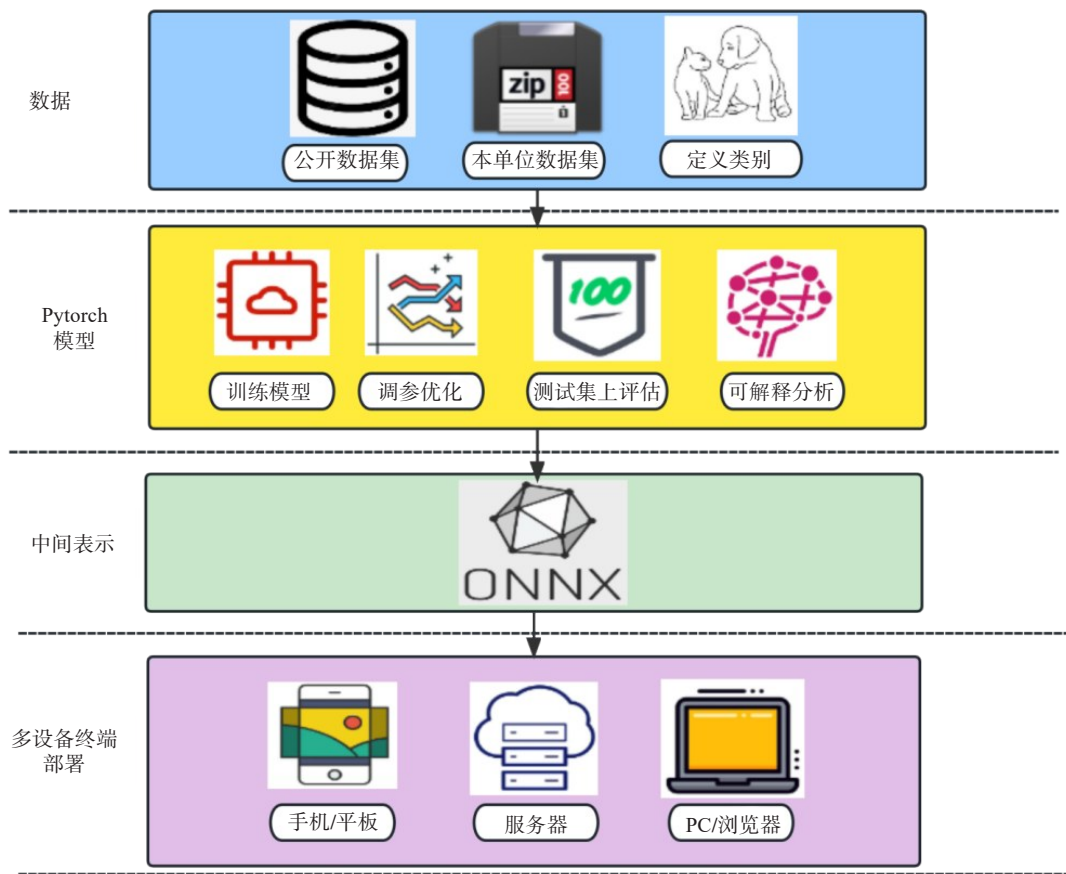


图4 模型开发部署流程

Figure 4 Model development and deployment process

独立测试集上评估其性能,评估指标包括召回率、特异性、精确率、准确率和F1值,并使用混淆矩阵进行分析。通过ROC曲线及其AUC值评估模型在各阈值下的表现。

2 结果

2.1 不同深度学习模型在结肠镜质量控制中的性能比较

本研究共采用3 831张结肠内窥镜图像,3 064张图像分配至训练集和767张图像分配至测试集。基于CNN与Transformer的预训练模型进行迁移学习微调。在CNN架构中,选择DenseNet121、EfficientNet、ResNet50和VGG19模型,而在Transformer架构中,选择ViT、Swin和CvT模型。为适应结肠镜质量控制的4类标签,对这些模型的输出层进行调整。在模型训练中,采用Adam优化器以确保效率。这些模型在测试集上的性能如图5所示。在结肠镜质量控制任务中,EfficientNet的性能表现最为出色,在测试集上准确率达到0.992,领先于VGG19(0.848)和DenseNet121(0.808)。此外,EfficientNet在精确率、召回率和F1值上均表现

卓越,分别为0.991、0.989和0.990,证明其在结肠镜质量控制任务中的稳定性与准确性。而在Transformer架构中,ViT模型准确率为0.948,成为该架构下的佼佼者。尽管ViT的准确率不如EfficientNet,但其性能同样令人瞩目。

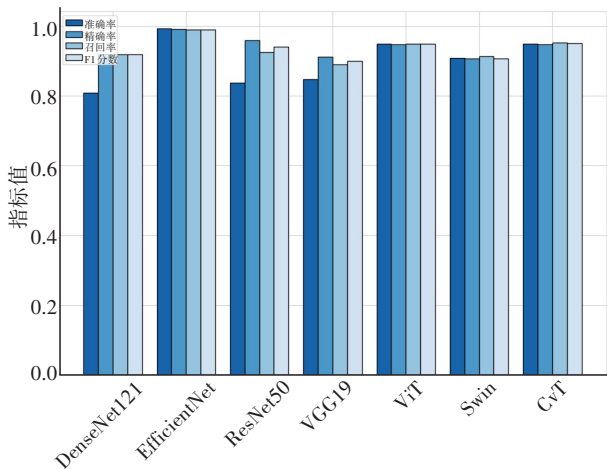


图5 基于深度学习的计算机视觉模型在测试集中的表现

Figure 5 Performance of the deep learning-based computer vision model on the test set

2.2 模型训练与评估结果

图 6a 展示了性能最佳深度学习模型 (EfficientNet) 在训练中损失值的演变。可见,随着迭代次数的增加,损失值持续下降并逐步稳定,说明模型正在向收敛方向进展。图 6b 揭示了该模型训练集准确率的生长轨迹,其中准确率稳步上升至一个较高水平,这证明模型已有效地捕获训练数据中的关

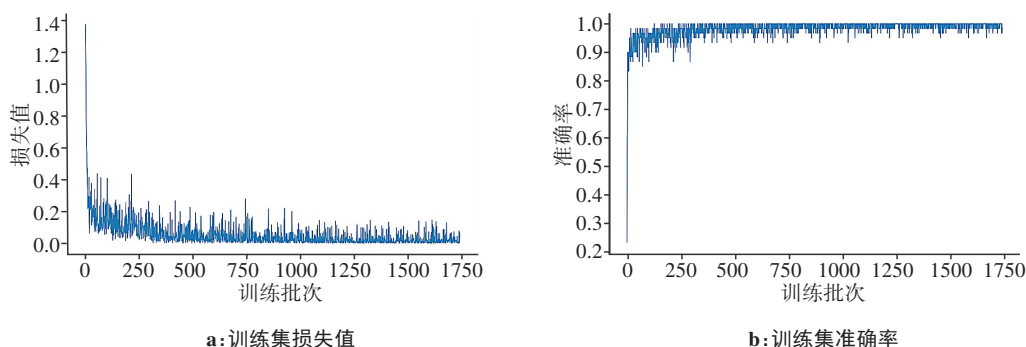


图 6 EfficientNet 模型在训练过程中性能指标的变化

Figure 6 Variations of performance metrics of EfficientNet model during the training

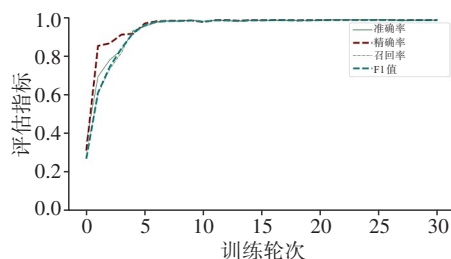


图 7 最佳模型 EfficientNet 的测试集分类评估指标

Figure 7 Evaluation metrics of EfficientNet model on the test set

2.3 外部测试集的模型预测性能

为确保模型的泛化性能,选用苏州大学附属医院的 358 张结肠镜图像作为完全独立的外部测试集进行验证。使用此独立测试集的优势在于能更准确地评估模型在实际应用中的预测效果,并确保没有过拟合的问题。EfficientNet 深度学习模型在结肠镜图像分类上展现卓越的性能,特别是在 BBPS (0~1 分,表示肠道清洁度较好) 和 BBPS (2~3 分,表示肠道清洁度较差) 类别中, AUC 值分别达到 0.997 和 0.999,显示模型的高度鉴别能力。盲肠类别的 AUC 值为 0.996,而息肉类别的 AUC 值 0.993。整体上,模型的平均 AUC、精确率和召回率值分别为 0.996、0.948 和 0.952,均表现出色,见图 8a。如图 8b 所示,对于 BBPS (0~1 分) 和 BBPS (2~3 分) 类别,模型在精确率和召回率上都展现了非常好的表现,几乎接近理想状态。盲肠类别的性能也相当高,尽管与前两

键特征并实现良好的拟合。本研究评估了 EfficientNet 深度学习分类模型在测试集上的性能。经过 30 个训练周期后,该模型在测试集上的损失值为 0.067,准确率为 0.992,证明其出色的分类能力。图 7 是模型的准确率、精确率、召回率和 F1 值的曲线变化,可以更直观地观察到模型的性能趋势。

个类别略有差距。对于息肉类别,虽然精确率仍然很高 (0.976),但召回率有所降低 (0.890),这可能意味着模型在某些情况下可能错过某些阳性样本。然而,每个类别的平均准确率都超过 0.97,这进一步证明模型在不同阈值下的鲁棒性和一致性。

对模型的分类效果进行混淆矩阵分析,进一步验证其在各个类别上的分类准确性和鲁棒性,详细结果如图 9a 所示。虽然模型整体上表现出色,但仍然存在一些误判的情况。图 9b 展示了两种典型的误判情况:一张真实标签为盲肠的图像被模型错误预测为息肉;一张标记为息肉的图像被误判为盲肠。这些误判可能是由于图像中的某些特征与息肉类似,导致模型产生混淆。

2.4 模型可解释分析

为了深入理解结肠镜质量控制模型的决策机制,采用 torchcam 算法库,结合 Grad-CAM 方法进行可视化分析。图 10a 展示了原始内镜图像,图 10b 呈现了模型基于 EfficientNet 特征提取的像素激活热图。这些激活显著地标注了模型在决策中所依赖的图像区域。图 10c 将激活热图叠加在原始图像上,其中黄绿色区域指示了模型识别息肉的关键部分。为了更细致地展现模型的关注点,图 10d 采用 Guided Grad-CAM 技术,结合 Grad-CAM 和 Guided Backpropagation,生成既具有类别判别性又具有细粒度的热力图,突显模型在分类时所依赖的细节特征。

为深入解析模型预测,采用 SHAP 方法,如图 11

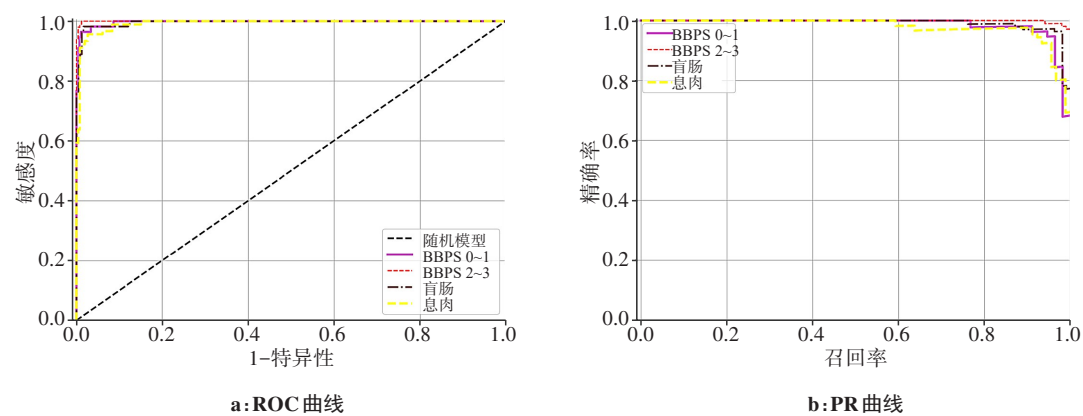


图8 EfficientNet 模型在外部测试集上的预测表现

Figure 8 Predictive performance of EfficientNet model on an external test set

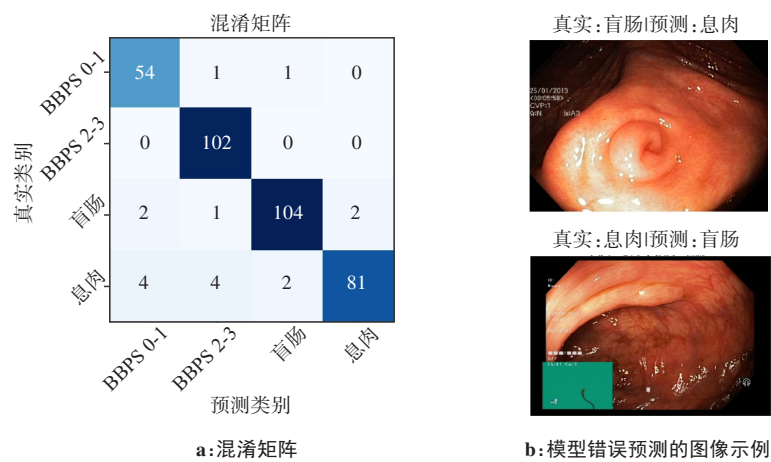


图9 模型在外部测试集上的预测表现

Figure 9 Predictive performance of the model on an external test set

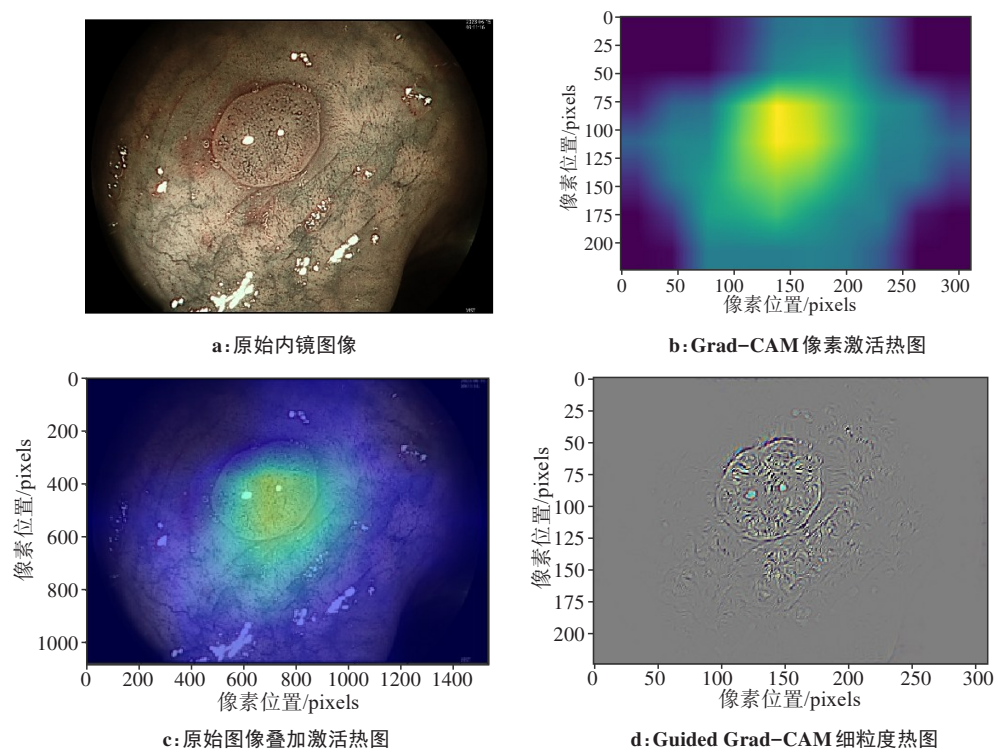


图10 结肠镜质量控制模型的可解释分析

Figure 10 Interpretability analysis of the colonoscopy quality control model

所示。图 11a 和图 11b 的实际类别分别是息肉和 BBPS(0~1 分)。图中每个像素的颜色深浅代表其对预测的影响大小(红色表示正向影响,蓝色则表示负向影响)。在图 11a 中,相对于盲肠和两个 BBPS 类

别,息肉类别的红色区域占据显著优势,从而被模型准确预测为息肉。图 11b 明确地被识别为 BBPS(0~1 分)。

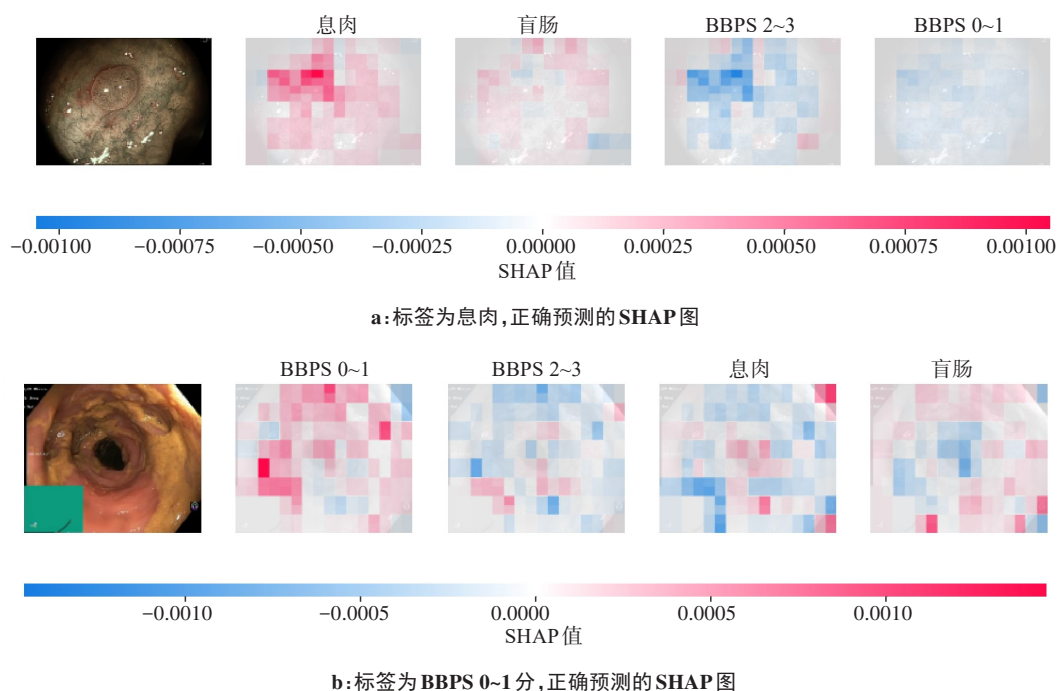


图 11 SHAP 解释性分析

Figure 11 SHAP interpretability analysis

2.5 模型对视频预测及多终端部署

为了在推理引擎上实现便捷部署,采用迁移学习方法训练 PyTorch 深度学习模型,并将其转换为 ONNX 格式,ONNX 为模型提供了一个标准化的表示,使研究者能够轻松地在多种平台和设备上进行部署并分享研究成果。这种转换使得能在本地电脑、Web 前端等环境中部署 EfficientNet 模型,从而实现随时随地的结肠镜质量实时质控。利用 OpenCV 库,从视频源中实时捕获每帧画面,并将其逐帧送入 ONNX 格式模型进行推理。图 12a 展示了模型对单帧图像的预测,其真实标签分别为盲肠和 BBPS(0~1 分)图像。在画面左上角,红色字体显示了模型对前两名类别的预测及其相应的置信度。图 12b 通过柱状图展示了不同分类的预测置信度。图 12c 通过二维码展示了模型在视频序列上的实时预测效果,用户可以扫描二维码以观看模型在本地电脑上的部署效果和使用本地电脑摄像头进行的实时视频预测。

3 讨论

本研究为结肠镜质量控制开发 7 种基于不同深

度学习架构的人工智能模型:其中 4 种采用 CNN 架构,3 种使用 Transformer 架构。从当前最大的胃肠道内窥镜数据集(HyperKvasir)中,筛选 3 831 张具有盲肠、肠道 BBPS 评分和息肉特征的图像进行模型训练。经外部测试集验证,EfficientNet 模型表现最为出色。此模型已成功部署于多个终端,实现实时视频预测。本研究是首次对 CNN 和 Transformer 在结肠镜质控中的性能进行比较,并确定最佳模型。

在一项前后对照试验中,研究团队利用 CNN 设计了 Endo.Adm 内窥镜质量控制系统,显著提升腺瘤和胃癌前病变的检测效率^[26]。这一研究突显了深度学习在结肠镜质量控制中的潜在价值。尽管如此,Endo.Adm 系统未采纳 Transformer 架构,其决策“黑盒”特性尚需进一步研究。在本研究中,在 Transformer 架构上采用了 ViT、Swin 和 CvT 模型。经过对比,ViT 模型在测试集上表现最佳,准确率达到 0.948。虽然 ViT 的表现未能超越 EfficientNet,但其表现同样出色。在本研究中,对性能最优的模型进行深入的可解释性分析,采用 Grad-CAM、Guided Grad-CAM 和 SHAP 等多种技术。利用 torchcam 算

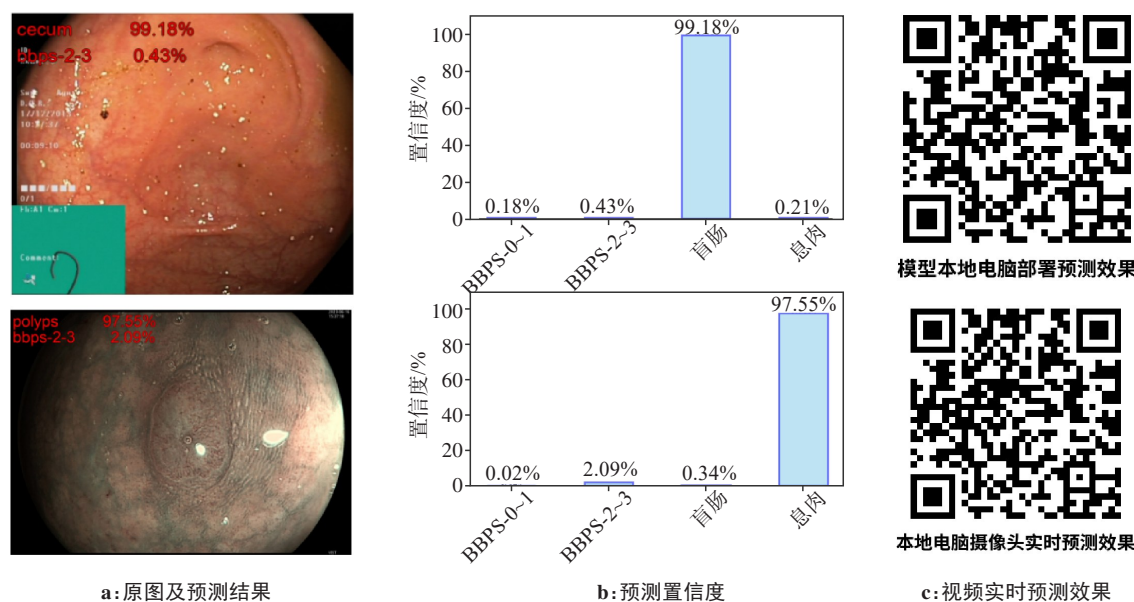


图 12 模型部署后对单帧图像和视频的预测及其置信度

Figure 12 Predictions and their confidence levels for individual frames and videos following model deployment

法库和 Grad-CAM 方法,对结肠镜质量控制模型的决策机制进行可视化探索。基于 EfficientNet 的激活热图以及其与原图的叠加效果,揭示模型在决策中关注的核心区域。为进一步细化模型的关注焦点,引入 Guided Grad-CAM 技术,生成既有类别区分性又有细粒度的热图。同时,通过 SHAP 方法分析每个像素对预测的贡献程度。这些综合分析为大家提供了对模型决策机制的深入理解。

针对结肠镜质量控制任务,本研究开发的 EfficientNet 深度学习模型在测试集及外部验证集上均得到充分验证。模型在各类别上表现优异,特别是在 BBPS(2~3 分)和盲肠类别,其 F1 值达到或超过 93%,这一结果揭示模型在这些类别上既具有高准确率又有高召回率,实现良好的性能平衡。进一步地将 EfficientNet 模型转换为 ONNX 格式并部署到多种终端设备上,在实时预测摄像头输入和视频文件时,模型不仅展现优越的分类准确性,还维持超过 60 FPS 的高效率,确保实时应用中的即时反馈。EfficientNet 通过均衡地扩展网络的深度、宽度和分辨率,实现了高性能,在确保高准确性的同时,也具备快速的推理速度。

目前,国际上广泛接受的肠道准备评分量表为 BBPS 和渥太华肠道准备量表,而国内主要采用 BBPS 评分。在临床实践中,持续使用这些评分量表可能会对内镜医护人员构成挑战,特别是在评价方法的不一致性和主观性偏见方面,此外医护人员在评估肠道清洁度时可能存在观点差异。因此,建立一个实用且方便的肠道准备质量评估方法显得至关

重要。本研究开发的 EfficientNet 深度学习模型在 BBPS 的肠道清洁度识别上表现卓越,特别是在 BBPS(0~1 分)和 BBPS(2~3 分)类别,其在外部验证集的 AUC 值分别为 0.997 和 0.999,显示高度的鉴别能力。

本研究成功开发 EfficientNet 深度学习模型,专为结肠镜质量控制设计,在 BBPS 肠道清洁度、息肉检测和盲肠识别等核心类别上,模型均展现出优异的表现。EfficientNet 在确保高准确性的同时,实现迅速的数据推理,满足实时应用需求。鉴于结直肠癌早期诊断的重要性,此模型为医护人员提供有力的技术支持,有望进一步满足国家卫健委的结肠镜质量标准,为患者提供更优质的服务。尽管本研究对深度学习在结肠镜质量控制任务上提供了新的见解,但挑战仍然存在。在未来研究中,建议结合结肠镜图像与临床结构化数据,如患者基线特征和实验室检查结果,以增强模型的预测能力。这种多模态数据融合可以为模型提供更全面的信息,提高其解释性和鲁棒性。考虑到结直肠癌早期诊断的重要性,结合深度学习和传统机器学习技术,计划开发一个高效的数据融合框架,为临床决策提供更准确的参考。

【参考文献】

[1] 国家癌症中心中国结直肠癌筛查与早诊早治指南制定专家组. 中国结直肠癌筛查与早诊早治指南(2020, 北京)[J]. 中华肿瘤杂志, 2021, 43(1): 16-38.
National Cancer Center, China, Expert Group of the Development of China Guideline for the Screening, Early Detection and Early Treatment of Colorectal Cancer. China guideline for the screening,

- early detection and early treatment of colorectal cancer (2020, Beijing) [J]. Chinese Journal of Oncology, 2021, 43(1): 16-38.
- [2] Pop OL, Vodnar DC, Diaconescu Z, et al. An overview of gut microbiota and colon diseases with a focus on adenomatous colon polyps[J]. Int J Mol Sci, 2020, 21(19): 7359.
- [3] Song MY, Emilsson L, Bozorg SR, et al. Risk of colorectal cancer incidence and mortality after polypectomy: a Swedish record-linkage study[J]. Lancet Gastroenterol Hepatol, 2020, 5(6): 537-547.
- [4] Su H, Lao YQ, Wu J, et al. Personal instruction for patients before colonoscopies could improve bowel preparation quality and increase detection of colorectal adenomas[J]. Ann Palliat Med, 2020, 9(2): 420-427.
- [5] Gómez-Reyes E, Tepox-Padrón A, Cano-Manrique G, et al. A low-residue diet before colonoscopy tends to improve tolerability by patients with no differences in preparation quality: a randomized trial [J]. Surg Endosc, 2020, 34(7): 3037-3042.
- [6] Belderbos TD, Grobbee EJ, van Oijen MG, et al. Comparison of cecal intubation and adenoma detection between hospitals can provide incentives to improve quality of colonoscopy[J]. Endoscopy, 2015, 47(8): 703-709.
- [7] Zhang QM, Dong ZY, Jiang YX, et al. The impact of sedation on adenoma detection rate and cecal intubation rate in colonoscopy[J]. Gastroenterol Res Pract, 2020, 2020: 3089094.
- [8] 国家癌症中心, 国家肿瘤质控中心结直肠癌质控专家委员会. 中国原发性结直肠癌规范诊疗质量控制指标(2022版)[J]. 中华肿瘤杂志, 2022, 44(7): 623-627.
National Cancer Center, Colorectal Cancer Expert Committee of National Cancer Quality Control Center. Quality control index for standardized diagnosis and treatment of colorectal cancer in China (2022 edition)[J]. Chinese Journal of Oncology, 2022, 44(7): 623-627.
- [9] Rembacken B, Hassan C, Riemann JF, et al. Quality in screening colonoscopy: position statement of the European Society of Gastrointestinal Endoscopy (ESGE)[J]. Endoscopy, 2012, 44(10): 957-968.
- [10] Rex DK, Schoenfeld PS, Cohen J, et al. Quality indicators for colonoscopy[J]. Am J Gastroenterol, 2015, 110(1): 72-90.
- [11] Zhao SB, Yang W, Wang SL, et al. Establishment and validation of a computer-assisted colonic polyp localization system based on deep learning[J]. World J Gastroenterol, 2021, 27(31): 5232-5246.
- [12] Zhou W, Yao LW, Wu HL, et al. Multi-step validation of a deep learning-based system for the quantification of bowel preparation: a prospective, observational study[J]. Lancet Digit Health, 2021, 3(11): e697-e706.
- [13] 阿依木克地斯·亚力孔, 庄惠军, 蔡世伦, 等. 基于深度学习人工智能在结肠镜检查中应用研究[J]. 中国实用外科杂志, 2020, 40(3): 353-357.
Ayimukedisi YL, Zhuang HJ, Cai SL, et al. Application of artificial intelligence based on deep learning in colonoscopy [J]. Chinese Journal of Practical Surgery, 2020, 40(3): 353-357.
- [14] Liu ZS, Lv QJ, Yang ZD, et al. Recent progress in transformer-based medical image analysis[J]. Comput Biol Med, 2023, 164: 107268.
- [15] Borgli H, Thambawita V, Smedsrud PH, et al. HyperKvasir, a comprehensive multi-class image and video dataset for gastrointestinal endoscopy[J]. Sci Data, 2020, 7(1): 283.
- [16] Chen LC, Zhu YK, Papandreou G, et al. Encoder-decoder with atrous separable convolution for semantic image segmentation[C]//Computer Vision-ECCV 2018. Cham: Springer International Publishing, 2018: 833-851.
- [17] 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.
- [18] Bansal M, Kumar M, Sachdeva M, et al. Transfer learning for image classification using VGG19: Caltech-101 image data set[J]. J Ambient Intell Humaniz Comput, 2023, 14(4): 3609-3620.
- [19] Yao T, Li YH, Pan YW, et al. Dual vision transformer[J]. IEEE Trans Pattern Anal Mach Intell, 2023, 45(9): 10870-10882.
- [20] Liu Z, Lin YT, Cao Y, et al. Swin transformer: hierarchical vision transformer using shifted windows[C]//2021 IEEE/CVF International Conference on Computer Vision (ICCV). Piscataway, NJ, USA: IEEE, 2021: 9992-10002.
- [21] Wu HP, Xiao B, Codella N, et al. CvT: introducing convolutions to vision transformers[C]//2021 IEEE/CVF International Conference on Computer Vision (ICCV). Piscataway, NJ, USA: IEEE, 2021: 22-31.
- [22] Zhang YY, Hong D, McClement D, et al. Grad-CAM helps interpret the deep learning models trained to classify multiple sclerosis types using clinical brain magnetic resonance imaging [J]. J Neurosci Methods, 2021, 353: 109098.
- [23] Kikutsuji T, Mori Y, Okazaki KI, et al. Explaining reaction coordinates of alanine dipeptide isomerization obtained from deep neural networks using explainable artificial intelligence (XAI)[J]. J Chem Phys, 2022, 156(15): 154108.
- [24] Ye TY, Li S, Zhang Y. Genomic pan-cancer classification using image-based deep learning [J]. Comput Struct Biotechnol J, 2021, 19: 835-846.
- [25] Li PS, Wang XH, Huang KZ, et al. Multi-model running latency optimization in an edge computing paradigm[J]. Sensors (Basel), 2022, 22(16): 6097.
- [26] Yao LW, Liu J, Wu LL, et al. A gastrointestinal endoscopy quality control system incorporated with deep learning improved endoscopist performance in a pretest and post-test trial [J]. Clin Transl Gastroenterol, 2021, 12(6): e00366.

(编辑:陈丽霞)