专论

Deep learning models for automatically differentiating endometrial polyps and submucosal fibroids in conventional ultrasound

LIU Xiaotang[1], HUANG Haishan[2], MA Weibo[3], HUANG Zhixi[4], LI Jin[1], CHEN Minglang[5], LIANG Yuanhao[6], JIANG Yao[7], XU Kai[8], CHEN Yayan[9*]

(1. *Department of Ultrasonography, Shenzhen Nanshan Maternity & Child Health Care Hospital, Shenzhen* 518067, *China*; 2. *School of Software Engineering, Sun Yat-sen University, Zhuhai* 519082, *China*; 3. *School of Nursing and Health Management, Shanghai University of Medicine & Health Sciences, Shanghai* 201318, *China*; 4. *Department of Neonatology*, 7. *Department of Ultrasonography, Shenzhen Maternity & Child Health Care Hospital Affiliated to Southern Medical University, Shenzhen* 518028, *China*; 5. *Guangxi Key Laboratory of Machine Vision and Intelligent Control, Wuzhou University, Wuzhou* 543002, *China*; 6. *Department of Ultrasonography, Dongguan Maternal and Child Health Care Hospital, Dongguan* 523000, *China*; 8. *School of Software, Yunnan University, Kunming* 650500, *China*; 9. *Department of Ultrasonography, Shenzhen Longhua District Maternal & Child Health Hospital, Shenzhen* 518000, *China*)

[**Abstract**] **Objective** To explore the value of deep learning (DL) models for automatically differentiating endometrial polyps and submucosal fibroids in conventional ultrasound. **Methods** A total of 1 291 individuals who underwent transvaginal ultrasound examination were retrospectively enrolled and randomly divided into training set (n=1 166, including 572 cases of endometrial polyps, 215 cases of submucosal uterine fibroids and 379 normal controls) and validation set (n=125, including 59 cases of endometrial polyps, 23 cases of submucosal uterine fibroids and 43 normal controls) at the ratio of 9∶1. Based on conventional ultrasonic images, 8 convolutional neural networks, including DenseNet-121, DenseNet-201, EfficientNet-B0, VGG-16, MobileNetV2, ResNet-18, ResNet-34 and Inception-v3 were used to construct DL models for automatic identification of endometrial polyps, submucosal uterine fibroids and normal uterus, and the classification performance of each DL model in validation set was evaluated. **Results** The mean precision, recall and F1 scores of 8 models for differentiating endometrial polyps, submucous fibroids and normal uterus were all 73.00%—86.00%, among which the mean precision (85.22%), recall (80.69%) and F1 score (82.46%) of Inception-v3 model were all relative higher. Gradient-weighted class activation mapping analysis showed that the regions of interest identified of Inception-v3 model during feature extraction highly overlapped with the key regions of interest in clinical diagnosis. The visualization results of t-distributed stochastic neighbor embedding indicated that Inception-v3 model had a strong feature representation ability in three-classification task and best overall classification performance. **Conclusion** DL models could be utilized to differentiate endometrial polyps and submucosal fibroids in conventional ultrasound, among which Inception-v3 model demonstrated the best performance.

[**Keywords**] endometrium; polyps; leiomyoma; ultrasonography; artificial intelligence

DOI: 10.13929/j.issn.1003-3289.2025.12.004

[基金项目] 广东省基础与应用基础研究基金(2025A1515011404)。

[第一作者] 刘晓棠(1988—),男,广东汕尾人,本科,主治医师。研究方向:妇产科超声。E-mail: liuxiaotang12345@outlook.com

[通信作者] 陈亚岩,深圳市龙华区妇幼保健院超声科,518000。E-mail: 905108774@qq.com

[收稿日期] 2025-05-29　[修回日期] 2025-12-14

深度学习模型用于自动鉴别常规超声所示子宫内膜息肉与黏膜下肌瘤

刘晓棠[1]，黄海珊[2]，马微波[3]，黄稚熙[4]，厉　进[1]，陈明朗[5]，梁元豪[6]，江　瑶[7]，许　凯[8]，陈亚岩[9*]

（1. 深圳市南山区妇幼保健院超声科，广东 深圳　518067；2. 中山大学软件工程学院，广东 珠海　519082；3. 上海健康医学院护理与健康管理学院，上海　201318；4. 南方医科大学附属深圳妇幼保健院新生儿科，7. 超声科，广东 深圳　518028；5. 梧州学院广西机器视觉与智能控制重点实验室，广西 梧州　543002；6. 东莞市妇幼保健院超声科，广东 东莞　523000；8. 云南大学软件学院，云南 昆明　650500；9. 深圳市龙华区妇幼保健院超声科，广东 深圳　518000）

［**摘　要**］ **目的**　探索基于深度学习(DL)模型自动鉴别常规超声所示子宫内膜息肉与黏膜下肌瘤的价值。**方法**　回顾性收集接受经阴道超声检查的 1 291 人，按 9∶1 比例随机划分训练集(n=1 166，含 572 例子宫内膜息、215 例子宫黏膜下肌瘤及 379 名正常对照)与验证集(n=125，含 59 例子宫内膜息、23 例子宫黏膜下肌瘤及 43 名正常对照)。基于常规超声图像分别以 DenseNet-121、DenseNet-201、EfficientNet-B0、VGG-16、MobileNetV2、ResNet-18 及 ResNet-34 和 Inception-v3 共 8 种卷积神经网络构建 DL 模型以自动鉴别子宫内膜息肉、子宫黏膜下肌瘤与正常子宫；以验证集评估各 DL 模型分类效能。**结果**　8 种模型鉴别子宫内膜息肉、子宫黏膜下肌瘤及正常子宫的平均精确率、召回率、F1 分数均为 73.00%～86.00%；其中，Inception-v3 模型的平均精确率(85.22%)、召回率(80.69%)及 F1 分数(82.46%)均较高。基于梯度加权类激活映射分析结果显示，Inception-v3 模型提取特征过程关注的区域与临床诊断关注的关键区域高度重叠；t-分布随机邻域嵌入可视化结果显示，Inception-v3 模型在三分类任务中表征特征能力均较强，整体分类效能最佳。**结论**　DL 模型可用于鉴别常规超声所示子宫内膜息肉与黏膜下肌瘤，尤以 Inception-v3 模型效能最佳。

［**关键词**］ 子宫内膜；息肉；平滑肌瘤；超声检查；人工智能

［**中图分类号**］ R711.74；R445.1　［**文献标识码**］ A　［**文章编号**］ 1003-3289(2025)12-1954-06

子宫内膜息肉与子宫黏膜下肌瘤均为妇科常见疾病[1]，超声表现高度相似而临床处理方式差异显著；准确鉴别二者对改善预后至关重要[2-3]。相关共识[2]建议对直径＞1.5 cm 子宫内膜息肉进行手术治疗，而对无症状子宫黏膜下肌瘤可采取保守观察[3]。超声为影像学检查子宫内膜息肉及子宫黏膜下肌瘤的首选方式[2,4]，但常需结合宫腔声学造影、宫腔镜等[2-6]，而上述检查方式具有侵入性，可引发出血、感染甚至子宫穿孔、空气栓塞等严重并发症[7]。目前人工智能(artificial intelligence，AI)技术已在妇产科超声领域取得多项研究进展[8-12]。本研究探索基于深度学习(deep learning，DL)模型自动鉴别常规超声所示子宫内膜息肉与子宫黏膜下肌瘤的价值。

1　资料与方法

1.1　研究对象　回顾性收集 2021 年 1 月—2024 年 12 月于深圳市南山区妇幼保健院接受经阴道超声检查的 1 291 人，均为女性，年龄 17～51 岁、平均(38.2±7.5)岁；其中 631 例子宫内膜息肉(均为单发病灶)、238 例子宫黏膜下肌瘤(均为单发病灶)、422 名正常对照。按 9∶1 比例随机划分训练集(n=1 166，含 572 例子宫内膜息肉、215 例子宫黏膜下肌瘤及 379 名正常对照)与验证集(n=125，含 59 例子宫内膜息肉、23 例子宫黏膜下肌瘤及 43 名正常对照)。纳入标准：①子宫内膜息肉或黏膜下肌瘤经阴道超声诊断并经宫腔镜或病理证实；②接受常规妇科检查，或监测排卵而接受经阴道超声的正常女性；③超声图像质量良好，完整显示关键解剖结构。排除合并其他宫腔病变及处于妊娠、哺乳期者。本研究获医院伦理委员会批准(NSFYEC-LW-2025009)；检查前患者均知情同意。

1.2　仪器与方法　采用 Canon Aplio i800、Phillips EPIQ 7C、GE Voluson E10、Samsung XW80A 或 Mindray Nuewa R9Q 超声仪，以频率 7～10 MHz 放入经阴道超声探头进行扫查，范围包括宫腔、子宫内膜及肌层，重点观察子宫内膜厚度、形态、回声及肌层内有无占位性病变，必要时行多切面扫查，存储子宫矢状面及横切面二维静态图像。见图 1。

1.3　构建 DL 模型

1.3.1　图像预处理　对每人选取 1 幅能明显反映宫腔的超声图像进行医学信息去标识化处理。基于训练集灰度直方图对每幅图像行直方图均衡化处理，以凸显 2 种病变边界的回声差异。将图像分辨率统一标准化至 224×224 像素，于训练阶段进行随机裁剪、上下

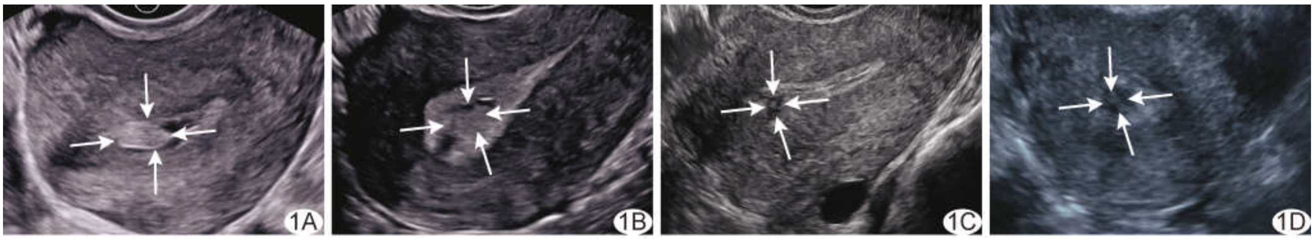


图 1　经阴道超声图示子宫内膜息肉与黏膜下肌瘤　A.患者女,35 岁,子宫内膜息肉;B.患者女,36 岁,子宫内膜息肉;C.患者女,30 岁,子宫黏膜下肌瘤;D.患者女,41 岁,子宫黏膜下肌瘤　(箭示病灶)

翻转、轻度旋转(±10°)及随机仿射变换等数据增强操作,以构建丰富的训练样本分布,降低模型对特定投影角度或操作习惯的依赖;于验证和测试阶段对图像进行中心裁剪和等比例缩放,以确保评估结果的可重复性。之后将像素值线性映射至[0,1]区间,并按ImageNet 预训练模型的通用均值(0.485,0.456,0.406)与标准差(0.229,0.224,0.225)对各通道进行归一化,转换为 PyTorch 张量后,按批次=16 输入后续网络。

1.3.2　设计网络架构　采用 Python 3.8.20 软件,分别以 DenseNet-121、DenseNet-201、EfficientNet-B0、VGG-16、MobileNetV2、ResNet-18、ResNet-34 及 Inception-v3 共 8 种卷积神经网络构建 DL 模型,将其用于自动鉴别超声图像中的子宫内膜息肉、子宫黏膜下肌瘤与正常子宫的三分类任务。先将各网络架构的原有全连接或分类器层替换为输出维度为 3 的线性层,即对 DenseNet 系列的 model.classifier 替换为新线性层,对 EfficientNet-B0 替换分类器层 classifier [1],对 VGG-16 及 MobileNetV2 修改分类器层最后一层 classifier [-1],对 ResNet 系列则采用 nn.Linear(in_features, 3) 替换 model.fc,对 Inception-v3 同时修改主分类层 model.fc 及辅助分类层 AuxLogits.fc;随后在每个网络结构的末级卷积特征图中附加全局平均池化与全局最大池化并行输出以增强其特征表达能力,并于拼接后送入分类头,以有效保留不同尺度下的关键信息。

1.3.3　训练策略及优化超参数　在配置为 NVIDIA A6000(单卡 48G 显存)及 Intel Xeon Platinum CPU 的服务器上通过 PyTorch 实现分布式数据并行,以最大化显存利用率。选择交叉熵损失处理多分类任务,优化器采用带动量(momentum=0.9)和权重衰减(weight_decay=0.0001)的随机梯度下降(stochastic gradient descent, SGD)算法,设置初始学习率为0.001,探索以更高的学习率(0.01)加速模型收敛。经综合对比余弦退火与阶梯下降两种学习率后,最终采用如下组合策略:在前 5 个 epoch 中实施线性预热(warm-up),之后启用余弦退火策略,使学习率在第 10 个 epoch 时平滑衰减至预设最小值。

1.4　评估模型效能及统计学分析　利用 scikit-learn 工具包(v1.0.2)自动计算精确率、召回率及 F1 分数,绘制混淆矩阵,于验证集评估各模型的鉴别效能,分析其误判类型。基于梯度加权类激活映射(gradient-weighted class activation mapping, Grad-CAM)方法生成解剖级热力图,将其叠加于原始超声图像中,突出模型重点关注的边界和阴影特征,以定位模型判别依赖的关键信息区域。利用 t-分布随机邻域嵌入(t-distributed stochastic neighbor embedding, t-SNE)降维,将高维特征降至二维空间,并按实际标签分色作图,直观展示各类别在高维特征空间中的聚类与分布关系,以反映模型对特征的表征能力。

2　结果

2.1　模型分类效能　8 种模型鉴别子宫内膜息肉、子宫黏膜下肌瘤及正常子宫的平均精确率、召回率、F1 分数均为 73.00%～86.00%。平均精确率以 Inception-v3 模型最高(85.22%)、DenseNet-121 模型次之(81.13%)而 VGG-16 模型最低(74.40%);平均召回率以 Inception-v3 模型最高(80.69%),其余模型均低于 80.00%;平均 F1 分数亦以 Inception-v3 模型最高(82.46%),ResNet-18 模型为 79.10%、DenseNet-121 模型为 78.30%,其余模型均低于 78.00%。

混淆矩阵显示,8 种模型的误分类主要针对子宫内膜息肉与正常子宫,DenseNet-121 模型误判数目最多(9 例),ResNet-18 及 Inception-v3 模型最少(5 例);其次是针对子宫黏膜下肌瘤与子宫内膜息肉,Inception-v3 模型误分类最少(仅 4 例)。8 种模型对子宫黏膜下肌瘤与正常子宫的分类效能均较好,Inception-v3 模型识别子宫黏膜下肌瘤精确率达

100％。见图 2。

2.2　Grad-CAM 分析　Grad-CAM 热力图显示，8 种模型中，Inception-v3 模型提取特征过程关注的区域与临床诊断时所关注的关键区域高度重叠。见图 3。

2.3　模型特征表征能力　t-SNE 算法可视化分析结果显示，针对验证集三分类任务，8 种模型中，Inception-v3 模型特征表征能力较强，虽然不同类别分布较为分散，但分离边界较清晰，以鉴别子宫黏膜下肌瘤的效能最佳，整体分类效能最佳；而 VGG-16 模型特征表征能力相对不足，表现为不同类别分布松散且存在显著重叠，整体分类效果较差，尤其鉴别子宫内息肉和黏膜下肌瘤的效能较弱。见图 4。

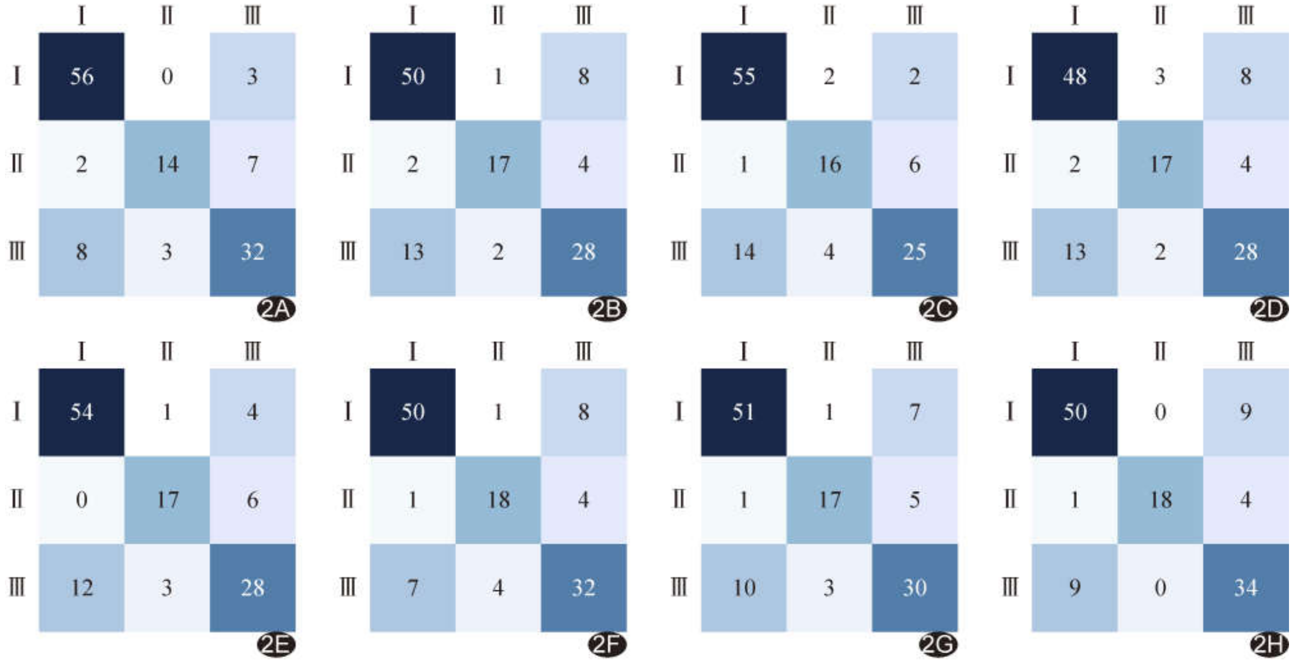


图 2　8 种模型区分验证集子宫内膜息肉、子宫黏膜下肌瘤及正常子宫的混淆矩阵图　A. DenseNet-121 模型；B. DenseNet-201 模型；C. EfficientNet-B0模型；D. VGG-16 模型；E. MobileNetV2 模型；F. ResNet-18 模型；G. ResNet-34 模型；H. Inception-v3 模型　（Ⅰ代表正常子宫；Ⅱ代表子宫黏膜下肌瘤；Ⅲ代表子宫内膜息肉）

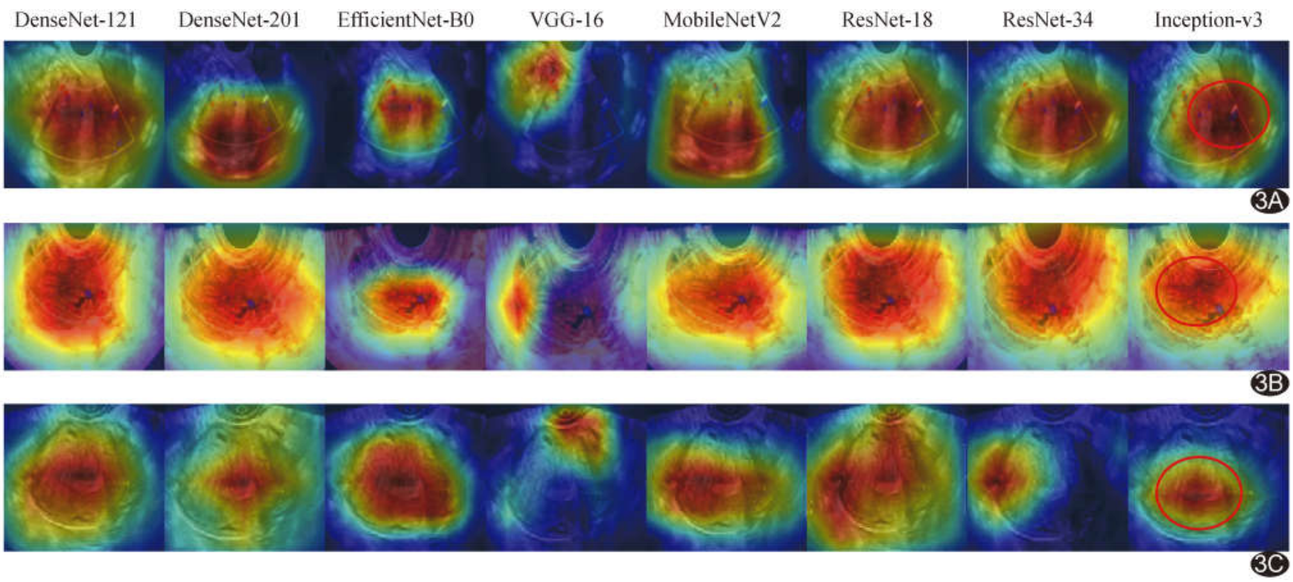
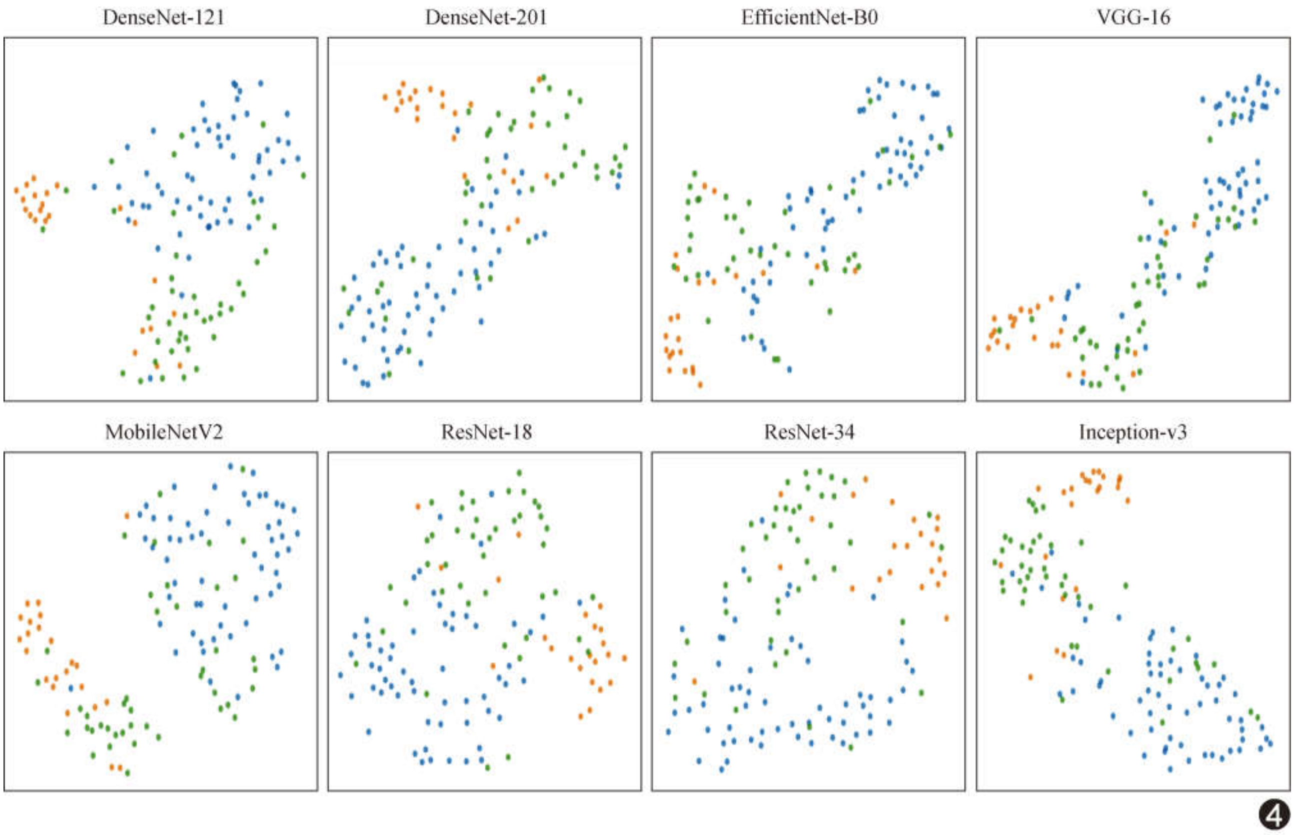


图 3　8 种模型于分类任务超声图像的关注区域的 Grad-CAM 热力图　A. 子宫内膜息肉；B. 子宫黏膜下肌瘤；C. 正常对照　（红色区域代表最高激活值，是模型最关注的区域）



4　8 种模型针对验证集学习超声图像特征的 t-SNE 可视化效果图　(蓝色代表正常子宫;橙色代表子宫黏膜下肌瘤;绿色代表子宫内膜息肉)

3　讨论

宫腔镜为诊断子宫内膜息肉和黏膜下肌瘤的金标准[5],但有创且可能导致严重并发症[8],部分病例因恐惧侵入性操作而延误诊治,或因基础疾病、凝血功能异常等禁忌证而无法进行宫腔镜检查。

本研究分别以 DenseNet-121、DenseNet-201、EfficientNet-B0、VGG-16、MobileNetV2、ResNet-18、ResNet-34 和 Inception-v3 共 8 种卷积神经网络构建 DL 模型用于鉴别超声所示子宫内膜息肉与黏膜下肌瘤,结果显示,8 种模型的平均精确率、召回率、F1 分数均为 73.00%～86.00%,尤以 Inception-v3 模型表现最优,其平均精确率、召回率及 F1 分数分别为 85.22%、80.69%及 82.46%,且识别子宫黏膜下肌瘤的精确率为 100%。这是由于子宫黏膜下肌瘤在声像图中的表现较为典型,子宫内膜息肉则超声表现多样,可呈子宫内膜局部增厚(病灶小)或囊性改变或出血(病灶大)、识别难度大。Inception-v3 的多尺度卷积架构能同时提取病灶局部细节和整体形态特征以适应这种复杂表现,而其他模型或因参数量大易过拟合(如 VGG-16)及特征冗余(如 DenseNet-121、DenseNet201),或因轻量化设计而牺牲了精度(如 MobileNetV2),难以充分适配三分类任务中病灶的多尺度形态特征。

本研究 Grad-CAM 分析结果显示,5 种 DL 模型,包括 DenseNet-121、DenseNet-201、MobileNetV2、ResNet-18 及 Inception-v3,提取病灶特征过程均主要关注病灶-子宫内膜交界、内部回声和基底部形态,其注意力模式与临床诊断中强调的超声评估要点相符;而 EfficientNet-B0、VGG-16 及 ResNet-34 模型的显著性区域分布相对分散,或部分聚焦于病灶以外的非特异性组织区域,其决策依据与临床超声诊断强调的特征的关联性较弱,可能影响其诊断准确性。t-SNE 可视化分析结果表明,Inception-v3 模型针对验证集三分类任务中的特征表征能力较强,整体分类效能最佳,尤其对于区分正常子宫与子宫黏膜下肌瘤。

综上,DL 模型可用于鉴别常规超声所示子宫内膜息肉与黏膜下肌瘤,尤以 Inception-v3 模型效能最佳,适合临床初筛和随访。本研究的局限性在于数据来自单中心而以不同设备采集,且未涉及动态超声图像或罕见的子宫内膜息肉恶变及黏膜下肌瘤变性等;有待未来通过多中心、大样本设计,联合彩色多普勒和三维超声随机对照等进一步观察。

利益冲突：全体作者声明无利益冲突。

作者贡献：刘晓棠研究设计和实施、查阅文献、数据分析、撰写文章；黄海珊、马微波和陈明朗研究实施；黄稚熙数据分析；厉进指导、修改文章；梁元豪和江瑶图像处理；许凯数据分析；陈亚岩研究设计、经费支持。

［参考文献］

[1] 杨岳州，叶旭萍，钱春蓓，等. 经阴道超声及彩色多普勒能量图对子宫黏膜下肌瘤和子宫内膜息肉的对照分析[J]. 中国医学影像技术，2001，17(3)：257-258.

[2] 田文艳，张慧英，薛凤霞. 子宫内膜息肉诊治中国专家共识(2022年版)解读[J]. 实用妇产科杂志，2023，39(1)：29-33.

[3] 子宫肌瘤的诊治中国专家共识专家组. 子宫肌瘤的诊治中国专家共识[J]. 中华妇产科杂志，2017，52(12)：793-800.

[4] 何卫东. 经阴道彩色多普勒超声诊断子宫内膜病变[J]. 中国医学影像技术，2010，26(10)：1937-1939.

[5] BITTENCOURT C A, DOS SANTOS SIMÕES R, BERNARDO W M, et al. Accuracy of saline contrast sonohysterography in detection of endometrial polyps and submucosal leiomyomas in women of reproductive age with abnormal uterine bleeding: Systematic review and meta-analysis [J]. Ultrasound Obstet Gynecol, 2017, 50(1): 32-39.

[6] YUAN X, SARAVELOS S H, WANG Q, et al. Endometrial thickness as a predictor of pregnancy outcomes in 10 787 fresh IVF-ICSI cycles [J]. Reprod Biomed Online, 2016, 33(2): 197-205.

[7] 孙宇婷，冯力民. 子宫内膜息肉的手术治疗新思考[J]. 中国计划生育和妇产科，2021，13(7)：23-24.

[8] LI X, TAN Y, LIANG B, et al. TKR-FSOD: Fetal anatomical structure few-shot detection utilizing topological knowledge reasoning [J]. IEEE J Biomed Health Inform, 2025, 29(1): 547-557.

[9] LU Y, TAN G, PU B, et al. SKGC: A general semantic-level knowledge guided classification framework for fetal congenital heart disease[J]. IEEE J Biomed Health Inform, 2024, 28(10): 6105-6116.

[10] PU B, WANG L, YANG J, et al. M3-UDA: A new benchmark for unsupervised domain adaptive fetal cardiac structure detection[C]. 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2024: 11621-11630.

[11] LIANG B, PENG F, LUO D, et al. Automatic segmentation of 15 critical anatomical labels and measurements of cardiac axis and cardiothoracic ratio in fetal four chambers using nnU-NetV2[J]. BMC Med Inform Decis Mak, 2024, 24(1): 128.

[12] 郭文佳，罗丹丹，徐荟，等. 新型智能产前超声技术 SFA 自动识别与获取胎儿标准切面[J]. 中国医学影像技术，2023，39(1)：65-69.