

人工智能阴道镜在大规模宫颈癌 筛查异常转诊人群中的价值

王 婷,梁 慧,何雪梅,杜 慧,蒋梦雨,豆秋彦,王邯邯

(徐州市妇幼保健院宫颈疾病科,江苏 徐州 221009)

[摘要]目的 评估人工智能阴道镜在宫颈癌筛查异常转诊中的诊断价值。方法 基于深度学习算法开发人工智能阴道镜系统,选择2021年7月至2024年2月于本院收治的400例自愿参加免费宫颈癌筛查的女性为研究对象并随机分为两组,200例进行人工阴道镜活检纳入人工组,200例使用人工智能(AI)辅助诊断软件进行AI阴道镜活检纳入AI组。以病理结果为金标准进行诊断比较。通过受试者工作特征(ROC)曲线分析诊断效能,采用Kappa指数进行一致性检验。结果 两组基线特征未见显著统计学差异。AI组对HSIL+的诊断符合率显著高于人工组(89.00% vs. 82.00%, $\chi^2=3.95$, $P=0.047$),两组在其他病理分级的诊断符合率及总符合率比较,差异均无统计学意义($P>0.05$)。ROC曲线分析结果显示:人工阴道镜诊断宫颈癌的曲线下面积(AUC)值优于AI组[AUC及95%CI分别为0.942(0.833~1.000)、0.786(0.640~0.932)];人工阴道镜诊断HSIL的AUC值仍优于AI组[AUC及95%CI分别为0.850(0.793~0.906)、0.741(0.670~0.813)]。AI组展现更高特异性(84.30% vs. 76.80%, $\chi^2=4.27$, $P=0.039$)和更低假阳性率(15.70% vs. 23.20%, $\chi^2=5.84$, $P=0.016$)。LSIL及炎症诊断中AI组的灵敏度(52.10% vs. 73.80%, $\chi^2=12.45$, $P<0.001$)与特异度(68.40% vs. 82.30%, $\chi^2=9.67$, $P=0.002$)均显著低于人工组。人工组Kappa值更高(0.67 vs. 0.53),即人工组诊断一致性上优于AI,AI组在复杂病例中表现不足。结论 AI组在宫颈癌和HSIL诊断中符合率和特异性显著优于人工组,AI阴道镜在大规模数据中表现出较高特异性和较低假阳性率。随着技术优化,AI有望提升低级别病变识别和诊断一致性,二者优势互补可为异常转诊患者提供更精准的诊断支持。

[关键词]人工智能;阴道镜;宫颈癌筛查;转诊;诊断价值

Doi:10.3969/j.issn.1673-5293.2025.11.005

[中图分类号]R173

[文献标识码]A

[文章编号]1673-5293(2025)11-0029-08

The value of artificial intelligence colposcopy in abnormal referral populations during large-scale cervical cancer screening

WANG Ting, LIANG Hui, HE Xuemei, DU Hui, JIANG Mengyu, DOU Qiuyan, WANG Hanhan

(Department of Cervical Diseases, Xuzhou Maternal and Child Health Hospital, Jiangsu Xuzhou 221000, China)

[Abstract] **Objective** To evaluate the diagnostic value of artificial intelligence (AI) colposcopy in abnormal referrals during cervical cancer screening. **Methods** An AI colposcopy system was developed based on deep learning algorithms. A total of 400 women who voluntarily participated in free cervical cancer screening at our hospital between July 2021 and February 2024 were enrolled and randomly divided into two groups. The conventional colposcopy biopsy group ($n=200$) was designated as the manual group, and the AI-assisted diagnostic software colposcopy biopsy group ($n=200$) was designated as the AI-assisted group. Pathological results were used as the gold standard for diagnostic comparison. The diagnostic performance was evaluated using receiver operating characteristic (ROC) curve analysis, and consistency between methods was assessed using the Kappa statistic. **Results** No significant differences were observed in baseline characteristics between the two groups. The diagnostic concordance rate for high-grade squamous intraepithelial lesion (HSIL) + in the AI-assisted group was significantly higher than that in the manual group (89.00% vs. 82.00%, $\chi^2=3.95$, $P=0.047$). For other pathological grades and overall diagnostic concordance, no statistically significant differences were found between the two groups ($P>0.05$). ROC curve analysis showed that the area under the curve (AUC) for cervical cancer diagnosis using manual colposcopy was superior to that of the AI-assisted group [AUC (95%CI): 0.942 (0.833-1.000) and 0.786 (0.640-0.932), respectively]. Similarly, the AUC for HSIL diagnosis in the manual group remained higher than that of the AI-assisted group [AUC (95%CI): 0.850 (0.793-0.906) and 0.741 (0.670-0.813), respectively]. The AI-assisted

[收稿日期]2024-12-17

[基金项目]2021年度中国妇幼保健协会科研创新项目(ZGFYBJXH-KY2102)

[作者简介]王 婷(1986—),女,副主任医师,硕士,主要从事妇科肿瘤(宫颈疾病方向)的研究。

[通讯作者]梁 慧,主任医师。

group displayed higher specificity (84.30% vs. 76.80%, $\chi^2=4.27, P=0.039$) and a lower false-positive rate (15.70% vs. 23.20%, $\chi^2=5.84, P=0.016$). However, for low-grade squamous intraepithelial lesions (LSIL) and inflammatory lesions, both the sensitivity (52.10% vs. 73.80%, $\chi^2=12.45, P<0.001$) and specificity (68.40% vs. 82.30%, $\chi^2=9.67, P=0.002$) of the AI-assisted group were significantly lower than those of the manual group. The manual group achieved a higher weighted Kappa coefficient (0.67 vs. 0.53), indicating superior diagnostic consistency over AI, which underperformed in complex cases requiring nuanced pathological interpretation. **Conclusion** The AI-assisted group demonstrated significantly higher concordance and specificity in the diagnosis of cervical cancer and HSIL compared to the human group. AI-assisted colposcopy showed high specificity and a lower false-positive rate in large-scale data. With further optimization of the technology, AI has the potential to improve the identification of low-grade lesions and improve diagnostic consistency. The complementary advantages of both methods can provide more accurate diagnostic support for patients referred due to abnormalities.

[Key words] artificial intelligence; colposcopy; cervical cancer screening; referral; diagnostic value

宫颈癌是全球女性健康的重大威胁,在发展中国家尤为突出,其癌前病变的早期识别是防控关键,目前筛查体系依赖细胞学检测、人乳头瘤病毒(human papilloma virus, HPV)初筛及阴道镜评估三阶梯模式^[1]。阴道镜检查作为从初级筛查到确定异常病变最佳管理策略之间的重要桥梁,在组织病理学诊断中扮演关键角色^[2]。然而,阴道镜检查存在明显的操作者依赖性,其诊断一致性受限于医生的经验水平,这在医疗资源匮乏地区尤为凸显^[3-4]。有数据显示,培养合格的操作阴道镜医师需5~10年周期^[5]。宫颈病变筛查中异常转诊的女性往往需要规范随访,但目前相关阴道镜专业医生培养不足,导致这部分转诊人群的标准化管理难以有效落实。

人工智能(artificial intelligence, AI)技术的突破为医学影像分析提供了新范式。在宫颈癌筛查领域,深度学习算法已证明其在细胞学判读中的高效性^[6]。值得注意的是,AI在阴道镜图像解析方面展现出独特优势,初步研究表明其可提升低年资医师的诊断准确性(敏感度提升23%,特异度提升18%)^[7],这对于标准化阴道镜操作流程具有重要价值。但现有研究多聚焦于算法开发,缺乏在真实转诊场景中的系统性验证。

本研究通过基于深度学习算法开发的AI阴道镜系统,旨在探索:首先,在宫颈病变筛查异常需转诊的高风险人群中,AI阴道镜对高级别鳞状上皮内病变(high-grade squamous intraepithelial lesion, HSIL)及以上的识别能力可否超越传统人工判读;其次,该技术是否能在减少误诊为癌前病变的同时保持较高准确率;最后,AI与人工诊断在判断不同严重程度病变[如低级别鳞状上皮内病变(low-grade

squamous intraepithelial lesion, LSIL)、HSIL及癌]时的各自优势及联合应用价值。本研究采用前瞻性随机对照试验(randomized controlled trial, RCT)设计,对400例宫颈癌筛查高危人群进行系统性观察,通过对比诊断准确度、检测效能曲线及结果一致性等核心指标,重点验证智能系统在提升阴道镜检查规范化水平方面的作用。这些发现将为优化宫颈癌筛查流程提供重要参考价值,并为医疗智能设备的持续升级提供实践依据。

1 资料与方法

1.1 试验前期

1.1.1 AI阴道镜软件的来源及设置标准

本研究采用基于深度学习的掩码区域卷积神经网络(mask region-based convolutional neural network, Mask R-CNN)开发的阴道镜智能辅助诊断系统,依托全国多中心(含多家三甲医院)提供的标准化、病理证实图像数据集进行训练。经五折交叉验证,模型表现稳定[曲线下面积(area under the curve, AUC)=0.92, 95%CI: 0.89~0.95]。图像标注由3名副高以上专家双盲完成, Kappa一致性系数为0.85。

1.1.2 AI阴道镜辅助诊断软件的训练阶段

AI阴道镜辅助诊断软件的性能高度依赖于高质量训练数据,需通过多中心合作收集覆盖基层常见病变与三级医院疑难病例的样本。

转诊标准筛选:①液基薄层细胞学检查(thin-prep cytology test, TCT)异常,宫颈细胞学结果显示为意义不明确的非典型鳞状上皮细胞(atypical squamous cells of undetermined significance, ASC-US)

及以上级别病变,如LSIL、HSIL、鳞状细胞癌(squamous cell carcinoma, SCC)、非典型腺细胞(atypical glandular cells, AGC)等。②HPV检测异常。③临床症状及体征可疑(如接触性出血、异常阴道出血、排液,或查体见宫颈糜烂样改变、赘生物或肥大等可疑体征)者,虽细胞学或HPV检测未显异常,但其临床表现提示宫颈病变可能,相关图像数据可增强AI对病变形态的识别能力。

排除标准:①已接受宫颈癌或癌前病变手术者,因宫颈形态改变可能干扰模型学习;②合并严重全身性疾病无法耐受阴道镜活检者;③妊娠期女性,因宫颈生理变化显著可能影响模型泛化性能。

样本量估算:基于病变类型多样性、程度分布及统计效能要求,计划纳入 $\geq 4\,000$ 例样本,以确保模型充分学习特征,保障其临床应用准确性与稳定性。

数据采集与质量控制:①标准化图像采集:严格遵循阴道镜检查规范,采用高清光电一体阴道镜,系统采集醋酸试验前、试验中及碘试验后的宫颈图像。每例保证多角度、多视野图像不少于2张/类型,以确保病变特征全面呈现。严格控制成像条件,排除模糊、遮挡及非目标区域干扰。②临床-影像数据整合:完整收录患者基线资料(年龄、婚育史、月经史)、细胞学、HPV检测、阴道镜拟诊及病理结果。融合多源数据构建训练集,通过深度学习建立临床-影像-病理映射模型,增强诊断性能。③质控与数据清洗:由符合资质(从业 ≥ 5 年,国家级培训 ≥ 3 次,年接诊 $\geq 2\,000$ 例)的中高级职称医师组成质控团队,依据标准对图像进行四级质量评估,仅采纳“满意”及“基本满意”图像;病例信息经统一审核与标准化处理,保障数据可靠性。

通过以上系统全面训练人群的选择、数据收集处理和管理保障措施,经过为期2年的训练,共收集4021例,为AI阴道镜辅助诊断软件提供了高质量的训练数据。

1.2 研究对象

选择2021年7月至2024年2月于本院收治的2548例自愿参加免费宫颈癌筛查的女性作为初始研究对象。研究对象首先接受双联筛查:通过高危型HPV检测联合宫颈TCT进行初筛。筛查出阳性病例:高危HPV阳性者、ASC-US及以上,初筛阳性

病例经随机抽样后接受阴道镜检查,最终400例具有完整病理诊断的病例纳入研究。按照不同观察方法均分为两组,200例进行人工阴道镜活检纳入人工组,200例使用AI辅助诊断软件进行AI阴道镜活检纳入AI组。

纳入标准:①年龄在25~65岁之间,均行高危型HPV筛查联合TCT;②无免疫系统疾病史;③无妇科生殖系统恶性肿瘤;④既往无宫颈活检史;⑤无阴道炎症;⑥非经期及非妊娠期。排除标准:①有阴道镜检查试剂(5%醋酸溶液/复方碘溶液)过敏史或禁忌证;②有免疫系统疾病史;③既往接受过宫颈癌前病变治疗(如宫颈锥切术、环形电切术等)者;④合并其他生殖系统恶性肿瘤病史。

本研究经我院医学伦理委员会批准(批件编号:XFYJL-2022-20),受试者对本研究均知情,并签署同意书。

1.3 检查方法

1.3.1 阴道镜检查方法

人工阴道镜选择电子高清阴道镜,经具有从事阴道镜工作经验大于5年、阴道镜规范培训的主治及以上医师进行阴道镜检查、采图、诊断。AI阴道镜选择德国莱斯康光电一体阴道镜,安装上述AI阴道镜辅助诊断软件,经过深度学习系统不断得以优化,从而做为本课题AI组辅助诊断活检的必备工具。

阴道镜检查时间为非月经期,48小时内无阴道冲洗、性生活及阴道上药。阴道镜检查及活检药具包括:复方碘溶液,5%醋酸溶液,生理盐水,活检钳,棉签,宫颈管扩张钳,组织固定液,止血纱布球。对可疑部位采用复方碘溶液及5%醋酸溶液,镜下观察转化区的血管、形态、颜色、鳞柱交界等,进行阴道镜图像描述和诊断。若阴道镜下发现可疑病变,取宫颈多点(3~4点)活检,行病理学检查。一旦受试者出现急性下生殖道感染或出血,治疗后再行阴道镜检查,以免影响阴道镜检查的准确性。

1.3.2 阴道镜印象

采用2017年ASCCP阴道镜标准^[8]。①正常/良性(宫颈炎症):未见恶性病变,包括柱状上皮、原始鳞状上皮、化生鳞状上皮、异位/外翻、纳氏囊肿及妊娠期蜕膜。②LSIL:细点状血管、细镶嵌、薄的醋

白上皮、快速消退、地图样边界。③HSIL:粗点状血管、粗镶嵌、厚醋白上皮、醋白出现快/消退慢、锐利边界。④可疑Ca:非典型血管、外生病灶、表面不规则、溃疡、坏死、肿瘤或大块新生物。

1.4 统计学方法

采用 SAS 9.4 进行数据管理和统计分析。定性资料采用频数(百分比)[$n(\%)$]进行描述,组间比较采用 χ^2 检验、Fisher 确切概率法;对于服从正态分布的定量资料,采用均值 $\pm$ 标准差($\bar{x} \pm s$)进行描述,组间比较采用 t 检验、方差分析;对于不服从正态分布的定量资料,采用中位数(四分位间距)[$M(P_{25}, P_{75})$]描述,组间比较采用 Wilcoxon 秩和检验。采用受试者工作特征(receiver operating characteristic, ROC)曲线评估预测变量的准确性,计算 AUC 判断 AI 与人工进行阴道镜诊断的区分效能,采用加权 Cohen's Kappa 统计量进行一致性检验。所有统计检验均为双侧检验,以 $P < 0.05$ 为差异具有统计学意义。

2 结果

2.1 两组间基线特征的对比分析

最终共 400 例具有完整病理诊断的病例纳入研究,所有研究对象的 TCT 结果以及宫颈活检病理结果如下,见表 1。

对人工组与 AI 组进行基线特征比较,两组在年龄和宫颈活检病理比例上比较均无显著性差异,说明两组可比性较好,见表 2。

对人工组与 AI 组进行病理诊断符合率的比较, AI 组对 HSIL+ 的诊断符合率显著高于人工组 ($89.00\% \text{ vs. } 82.00\%, \chi^2 = 3.95, P = 0.047$),差异具有统计学意义。两组在其他病理分级的诊断符合率

及总符合率比较,差异均无统计学意义($P > 0.05$),见表 3。

2.2 阴道镜诊断效能的评估

2.2.1 人工组阴道镜诊断性能的评估

ROC 曲线分析显示:人工阴道镜诊断 Ca 的 AUC 值最大($AUC = 0.942, 95\% CI: 0.833 \sim 1.000; Z = 7.94, P < 0.001$),显著高于其他病变类型(HSIL: $Z = 12.16$; LSIL: $Z = 5.60$),表明其对宫颈癌具有最优诊断效能(见表 4);HSIL、Ca 和 HSIL+ 的 AUC 值均大于 0.750,说明人工阴道镜预测有无 HSIL、Ca 和 HSIL+ 的真实性较好,区分效能较高,见表 4 及图 1。

表 1 研究对象的基线特征描述 [$n(\%)$]

Table 1 Baseline characteristics of study participants [$n(\%)$]

检测项目	病例数及占比
TCT 结果	
ASC-US	230(57.50)
LSIL	80(20.00)
ASC-H	49(12.25)
HSIL	36(9.00)
AGC	5(1.25)
宫颈活检病理	
正常(炎症)	124(31.00)
LSIL	138(34.50)
HSIL	117(29.25)
Ca(SCC/腺癌)	21(5.25)

注:本文中以 Ca 代表宫颈浸润癌。

表 2 人工组与 AI 组基线特征比较 [$M(P_{25}, P_{75}), n(\%)$]

Table 2 Comparison of baseline characteristics between manual and AI-assisted groups [$M(P_{25}, P_{75}), n(\%)$]

特征指标	总体($n=400$)	人工组($n=200$)	AI 组($n=200$)	Z/χ^2	P
年龄(岁)	47.0(39.0,55.0)	46.0(37.0,54.0)	48.0(40.0,55.0)	1.61	0.107
宫颈活检病理					
正常(炎症)	124(31.00)	67(33.50)	57(28.50)	2.30	0.513
LSIL	138(34.50)	68(34.00)	70(35.00)		
HSIL	117(29.25)	53(26.50)	64(32.00)		
Ca	21(5.25)	12(6.00)	9(4.50)		

表 3 人工组与 AI 组病理诊断符合率的比较[n(%)]

诊断分类	总例数(n=400)	人工组(n=200)	AI 组(n=200)	χ^2	P
正常(炎症)	310(77.50)	155(77.50)	155(77.50)	0.00	1.000
LSIL	272(68.00)	133(66.50)	139(69.50)	0.41	0.520
HSIL	339(84.75)	163(81.50)	176(88.00)	3.27	0.071
Ca	391(97.75)	193(96.50)	198(99.00)	1.82	0.177
HSIL+	342(85.50)	164(82.00)	178(89.00)	3.95	0.047
总符合率	256(64.00)	122(61.00)	134(67.00)	1.56	0.211

注:HSIL+定义为 HSIL 及 Ca。

表 4 人工阴道镜诊断不同病变的 ROC 曲线指标

Table 4 ROC curve indicators of manual colposcopy for different lesions

诊断分类	AUC(95%CI)	Z	P	灵敏度(95%CI)	特异度(95%CI)	约登指数	准确率(%)	PPV	NPV
正常(炎症)	0.695(0.624~0.766)	5.41	<0.001	50.88%(37.90~63.86)	88.11%(82.81~93.42)	0.390	77.50	0.630	0.818
LSIL	0.693(0.625~0.760)	5.60	<0.001	68.57%(57.70~79.45)	70.00%(62.12~77.88)	0.386	69.50	0.552	0.805
HSIL	0.850(0.793~0.906)	12.16	<0.001	76.56%(66.18~86.94)	93.38%(89.20~97.56)	0.699	88.00	0.845	0.894
Ca	0.942(0.833~1.000)	7.94	<0.001	88.89%(68.36~109.42)	99.48%(98.45~100.50)	0.884	99.00	0.889	0.995
HSIL+	0.873(0.822~0.923)	14.56	<0.001	80.82%(71.79~89.85)	93.70%(89.48~97.93)	0.745	89.00	0.881	0.895

注:HSIL+定义为 HSIL 及 Ca;PPV 为阳性预测值;NPV 为阴性预测值。

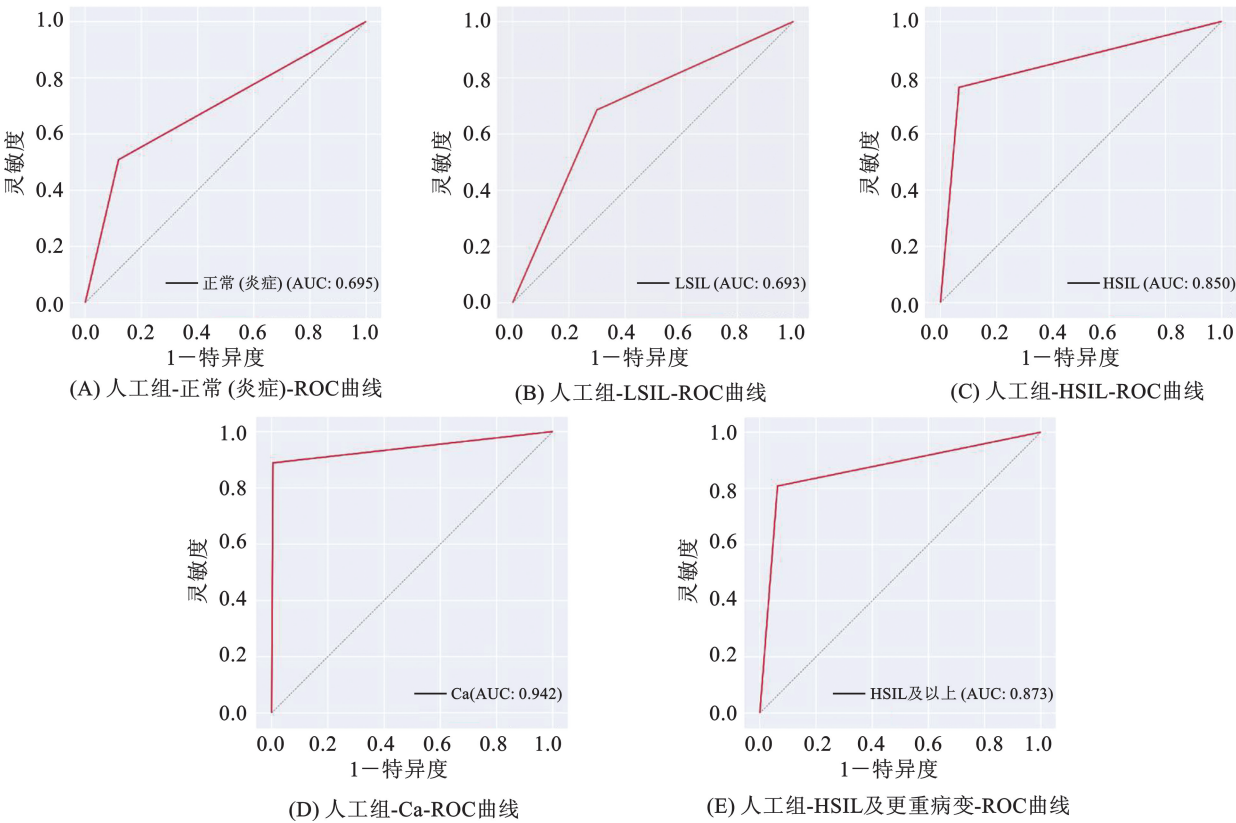


图 1 人工组相关 ROC 曲线展示

Fig.1 ROC curve related to the manual group

2.2.2 AI组阴道镜诊断性能的评估

AI阴道镜诊断效能分层显著:对宫颈癌(Ca)和HSIL+鉴别能力最优,AUC分别为0.786(95%CI:0.640~0.932)和0.775(95%CI:0.711~0.839),而LSIL诊断能力最弱(AUC=0.661)。AI特异性显著高于灵敏度:宫颈癌检测特异性达98.94%(假阳性率仅1.06%),但灵敏度仅58.33%;HSIL+综合性能最优(约登指数0.550,准确率82.00%),阴性预测值84.10%,适合高风险病例分流。HSIL+和Ca的ROC曲线左上方偏移明显,而LSIL曲线平缓,反映AI对高级别病变识别更可靠。对LSIL的灵

敏度(64.71%)和特异度(67.42%)均不足(约登指数0.321)。总体显示:AI组在高级别病变排除(特异性>90%)和癌症(特异性98.94%)中优势突出,但低级别病变诊断能力有限。见表5、图2。

AI组展现更高特异性(84.30% vs. 76.80%, $\chi^2=4.27$, $P=0.039$)和更低假阳性率(15.70% vs. 23.20%, $\chi^2=5.84$, $P=0.016$)。LSIL及炎症诊断中AI组的灵敏度(52.10% vs. 73.80%, $\chi^2=12.45$, $P<0.001$)与特异度(68.40% vs. 82.30%, $\chi^2=9.67$, $P=0.002$)均显著低于人工组。

表5 AI组阴道镜诊断不同病变的ROC曲线指标

Table 5 ROC curve indicators of AI-assisted colposcopy for diagnosing different lesions

诊断分类	AUC(95%CI)	Z	P	灵敏度(95%CI)	特异度(95%CI)	约登指数	准确率(%)	PPV	NPV
正常(炎症)	0.731(0.665~0.797)	6.86	<0.001	59.70%(47.96~71.45)	86.47%(80.65~92.28)	0.462	77.50	0.690	0.810
LSIL	0.661(0.591~0.731)	4.51	<0.001	64.71%(53.35~76.06)	67.42%(59.43~75.42)	0.321	66.50	0.506	0.788
HSIL	0.741(0.670~0.813)	6.63	<0.001	58.49%(45.22~71.76)	89.80%(84.90~94.69)	0.483	81.50	0.674	0.857
Ca	0.786(0.640~0.932)	3.85	<0.001	58.33%(30.44~86.23)	98.94%(97.47~100.40)	0.573	96.50	0.778	0.974
HSIL+	0.775(0.711~0.839)	8.46	<0.001	64.62%(52.99~76.24)	90.37%(85.39~95.35)	0.550	82.00	0.764	0.841

注:HSIL+定义为HSIL及Ca;PPV为阳性预测值;NPV为阴性预测值。

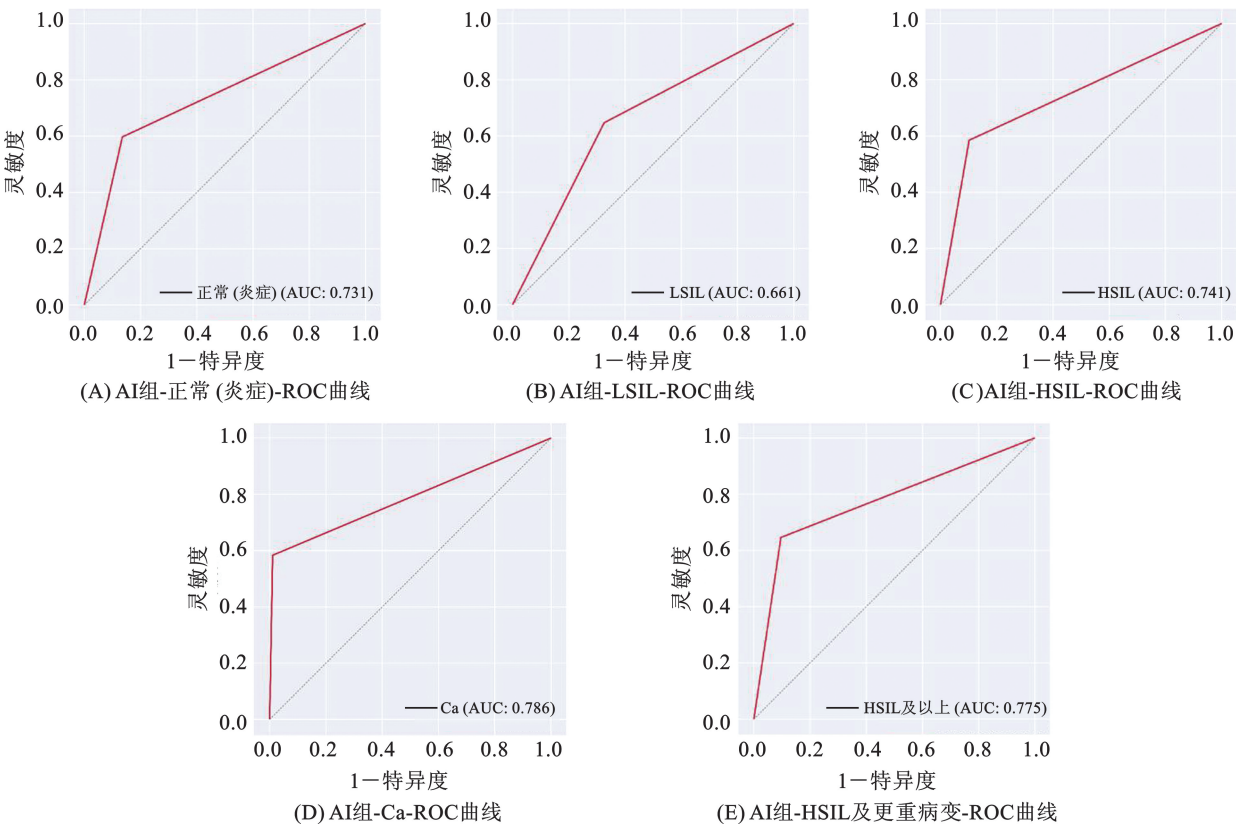


图2 AI组相关ROC曲线展示

Fig.2 ROC curves related to the AI-assisted group

2.3 诊断一致性分析

对人工与 AI 阴道镜诊断得出正常（炎症）、LSIL、HSIL 及 Ca 四种结果与宫颈病理活检的结果进行一致性检验，其中人工组的加权 Kappa 值为 0.6658，提示人工诊断与病理结果具有高度一致性；AI 组的加权 Kappa 值为 0.5270，说明一致性中等，见表 6。

表 6 人工组与 AI 组阴道镜诊断一致性分析
(加权 Kappa 系数)

Table 6 Diagnostic concordance between manual and AI-assisted groups (weighted Kappa coefficients)

组别	加权 Kappa	标准误差	95%CI	一致性强度
人工组	0.6658	0.0272	0.6124~0.7191	高度一致
AI 组	0.5270	0.0333	0.4618~0.5922	中等一致

3 讨论

宫颈癌作为全球重大公共卫生问题，其防控形势依然严峻。最新流行病学数据显示^[9]14，全球每年新增宫颈癌病例约 52.8 万例，死亡 26.6 万例，90%的疾病负担集中在发展中国家。我国作为人口大国，宫颈癌年发病率达 98.9/10 万，死亡率 30.5/10 万，且呈现持续上升趋势^[10]。尽管我国已建立多层级筛查体系，但筛查覆盖率不足（尤其是农村地区）、HPV 疫苗接种率偏低及转诊病例管理失范等问题，导致宫颈癌防控效果与发达国家存在显著差距。值得注意的是，阴道镜检查作为宫颈癌筛查的关键环节，其质量直接影响高级别宫颈上皮内瘤变（CIN2+）的早期识别率。然而基层医疗机构普遍存在阴道镜医师培养周期长（平均需要 3~5 年）、诊断水平参差不齐等现实困境，造成高级别病变漏诊率高达 15%~20%^[11]，这成为制约我国宫颈癌防控成效的瓶颈问题。

3.1 AI 辅助诊断的临床效能优势及 AI 技术的当前局限性与改进方向

本研究通过前瞻性队列研究设计，系统地比较了 AI 阴道镜与传统人工诊断在大规模宫颈癌筛查异常转诊病例中的诊断效能。在高级别病变（HSIL+）识别方面，AI 组诊断符合率达 89.00%，显著高于人工组的 82.00%（ $P=0.047$ ），这与 Wu 等^[12]报道的 AI 诊断敏感性（约 80%）与资深医师相当的结

论相印证。特别在宫颈癌诊断中，AI 组 AUC 值达 0.786，虽低于人工组的 0.942（ $P<0.01$ ），Hu 等^[6]基于 NIH 宫颈影像数据库（ $n\approx 98\ 000$ ）的研究表明，深度学习模型对 CIN2+ 的鉴别效能可达 $AUC=0.91$ ，提示扩大训练数据集规模是提升 AI 诊断性能的有效策略。值得注意的是，AI 组在特异性和假阳性率方面展现出明显优势，这对降低阴道镜过度转诊率具有重要临床意义。

但需客观认识到当前 AI 技术的局限性：在 Kappa 一致性检验中，AI 组加权 Kappa 值（0.5270）显著低于人工组（0.6658），尤其在低级别病变（LSIL）识别中，AI 的灵敏度和特异度均存在显著差距。这种差异可能源于：①AI 算法对细微黏膜改变的识别能力不足；②现有训练数据中 LSIL 样本量相对不足；③炎症、出血等干扰因素对图像特征的掩盖效应^[13]。值得关注的是，薛鹏教授团队^[14]通过纳入动态时序阴道镜图像，将 CIN2+ 诊断准确率提升至 89.6%，提示多维特征融合可能是突破现有技术瓶颈的重要方向。

3.2 AI 技术在宫颈癌筛查中的应用及未来发展方向

AI 技术的迅速发展，能够满足大规模人群的宫颈癌筛查需求，为宫颈癌早期筛查提供一定契机^[15]。从技术发展角度看，AI 阴道镜具有三大核心优势：首先，其诊断稳定性不受医师疲劳程度影响，可保持 24 小时 95% 以上的诊断一致性^[16-17]；其次，基于深度学习的特征提取能力，可识别肉眼难以察觉的微血管形态改变；最后，便携式设备与云端诊断系统的结合，能有效地解决基层病理医师短缺问题^[18-19]，但需警惕数据隐私、算法可解释性及医疗责任认定等伦理问题，在涉及少数民族等特定群体健康数据时，应建立标准化的数据匿名化处理流程。

后续研究可从三个维度深化：首先应构建覆盖多地域、多民族的阴道镜影像数据库，重点扩充低级别病变（LSIL）及罕见病理亚型的样本储备；其次需开发整合 HPV 基因型别、细胞学报告等临床信息的多维度智能诊断系统，以提升模型临床适用性；最后可探索基于分布式机器学习框架的隐私保护技术，在确保数据安全的前提下实现跨机构模型优化。通过持续技术创新和临床验证，AI 阴道镜有望成为宫颈癌“筛-诊-治”闭环管理的关键技术支撑。

综上,本研究通过系统比较 AI 阴道镜与传统人工诊断在异常转诊病例中的诊断效能,揭示了两类方法在宫颈癌筛查中的互补价值。AI 组在宫颈癌和 HSIL 诊断中符合率和特异性显著优于人工组, AI 阴道镜在大规模筛查场景中展现较高的特异性及低假阳性率,具备临床潜力。随着技术的优化, AI 阴道镜有望提升低级别病变识别和诊断一致性,进而提高筛查异常人群的诊断效率,为宫颈癌筛查异常转诊患者提供更加精准的诊断支持。

[参考文献]

- [1]陈飞,尤志学,隋龙,等.阴道镜应用的中国专家共识[J].中华妇产科杂志,2020,55(7):443-449.
- [2]Green L I, Mathews C S, Waller J, et al. Attendance at early recall and colposcopy in routine cervical screening with human papillomavirus testing [J]. Int J Cancer, 2021, 148(8):1850-1857.
- [3]Roopa H, Mittal S, Basu P. Role of colposcopy in the management of women with abnormal cytology[J]. Cytojournal, 2022, 19:40.
- [4]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 Med, 2022, 18(1):169.
- [5]薛鹏.基于人工智能的电子阴道镜辅助诊断系统构建及多中心应用评估[D].中国医学科学院北京协和医学院, 2020.
- [6]Hu L, Bell D, Antani S, et al. An observational study of deep learning and automated evaluation of cervical images for cancer screening[J]. J Natl Cancer Inst, 2019, 111(9):923-932.
- [7]陈明扬,蔡紫庭,薛鹏,等.人工智能在肿瘤研究和临床中的应用[J].基础医学与临床, 2022, 42(11):1637-1643.
- [8]薛鹏.人工智能技术在宫颈癌筛查、分流和诊断中的应用研究[D].北京:中国医学科学院, 2023.
- [9]Bray F, Ferlay J, Soerjomataram I, et al. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries[J]. CA Cancer J Clin, 2018, 68(6):394-424.
- [10]臧春宝,吴爱林,王慧妍,等.磷酸铝凝胶联合康复新液保留灌肠预防宫颈癌放疗后放射性肠炎的临床研究[J].蚌埠医学院学报, 2019, 44(8):1004-1008.
- [11]薛鹏,唐朝,乔友林,等.人工智能电子阴道镜辅助诊断系统对宫颈癌筛查的现实挑战和未来机遇[J].中国肿瘤, 2019, 28(7):483-486.
- [12]Wu A, Xue P, Abulizi G, et al. Artificial intelligence in colposcopic examination: a promising tool to assist junior colposcopists[J]. Front Med (Lausanne), 2023, 10:1060451.
- [13]陈苗,高丽,张婷,等.1 357 例宫颈上皮内瘤变临床特征及阴道镜图像的诊断意义[J].中国妇幼健康研究, 2023, 34(2):59-63.
- [14]薛鹏,乔友林,江宇.人工智能在医学内窥镜诊断中的应用[J].中华肿瘤杂志, 2018, 40(12):890-893.
- [15]何雪梅,黄辉,王婷,等.人工智能辅助宫颈细胞学筛查联合 HPV 分流在宫颈癌人群筛查中的应用分析[J].徐州医科大学学报, 2022, 42(4):273-278.
- [16]Hussain E, Mahanta L B, Das C R, et al. A comprehensive study on the multi-class cervical cancer diagnostic prediction on pap smear images using a fusion-based decision from ensemble deep convolutional neural network[J]. Tissue Cell, 2020, 65:101347.
- [17]车拴龙,刘栋,刘斯,等.人工智能技术在宫颈细胞筛查中的应用进展和挑战[J].中华临床实验室管理电子杂志, 2019, 7(4):193-198.
- [18]李艳辉,周星,徐琴,等.基于云技术和人工智能的自动化细胞学诊断平台在 703103 名女性宫颈癌筛查中的应用研究[J].中国妇产科临床杂志, 2022, 23(4):364-368.
- [19]中国妇幼保健协会,中国妇幼保健协会妇女病防治专业委员会.人工智能辅助宫颈细胞学诊断技术的应用及质量控制专家共识[J].中国妇幼健康研究, 2023, 34(12):7-13.

[专业责任编辑:安瑞芳]

[中文编辑:冯佳圆;英文编辑:杨 寅]