

人工智能辅助宫颈细胞学假阳性原因分析

赫晓朋* 刘明磊 刘洪越 李晓欣 苏策

(石家庄市栾城人民医院病理科, 河北 石家庄 051430)

摘要: 目的是为了评估子宫颈细胞学计算机辅助诊断系统在宫颈病变筛查中的应用价值, 分析假阳性结果的产生机制, 并提出优化策略。方法上, 收集 2018 年至 2023 年石家庄市栾城人民医院 11594 例宫颈液基薄层细胞片, 应用湖南品信生物工程有限公司开发的子宫颈细胞学计算机辅助诊断系统 CR-SA3, 以 ASCUS 及以上为阳性结果。对比 AI 系统与细胞病理医师的诊断结果, 计算符合率、灵敏度、特异度等指标, 并以组织病理结果为金标准进行验证。结果显示, 在 8793 例合格切片中, AI 筛查提示阳性 2330 例; 与医师诊断对比, 真阳性 1547 例, 假阳性 783 例, 假阴性 0 例。AI 系统总符合率为 91.10%, 灵敏度达 100%, 特异度为 89.19%, 阳性预测值 66.39%, 阴性预测值 100%。AI 阅片效率显著高于人工, 平均每张切片仅需 70 秒, 节省 708.32 小时。AI 对炎症等良性病变的特异性较低。假阳性主要源于细胞制片差异、染色不均、炎症干扰、ASC-US 判读主观性强及算法局限性等因素。结论为, 人工智能辅助系统在宫颈细胞学初筛中表现出高灵敏度与高效率, 能有效降低漏诊风险并显著减轻医师工作负担。系统在细胞病变精确分类方面仍有不足, 假阳性率较高, 降低假阳性率需要多流程优化策略。

关键词: 人工智能; 假阳性; 液基薄层细胞学

宫颈癌严重威胁女性健康, 早期筛查是降低宫颈癌发病率和死亡率的有效途径。

传统的宫颈细胞学检查主要依赖细胞病理医师的显微镜下观察与主观判断, 存在诸多局限性。例如人工阅片效率低下, 每张涂片平均需要 5-10 分钟, 医师易产生视觉疲劳, 影响诊断一致性; 诊断标准主观性强, 不同医师间诊断差异显著, 不同等级医院宫颈细胞液基制片优良率及诊断正确率存在差异^[1]。

尤其在 ASC-US 等边界性病变的判断上; 细胞病理医师数量严重不足, 特别是在基层医疗机构, 难以满足大规模筛查需求。这些问题制约了宫颈癌筛查的普及与质量提升。

随着数字病理技术与人工智能的发展, AI 辅助宫颈细胞学诊断系统应运而生。AI 辅助宫颈细胞学诊断可以克服人工阅片的时间限制和专业技术人员缺乏的问题, 避免由主观因素引起的偏见, 有助于在资源匮乏地区实施宫颈癌筛查^[2]。人工智能辅助子宫颈细胞学诊断能够显著提高阅片速度和工作效率^{[3][4]}。然而, AI 系统在实际临床应用中 also 面临挑战, 其中假阳性问题尤为突出。假阳性结果可能导致不必要的阴道镜检查、患者焦虑和医疗资源浪费。关于 AI 系统假阳性原因的系统分析尚不充分, 缺乏针对性的优化策略。

本研究基于 11594 例宫颈液基薄层细胞学标本, 评估 AI 辅助诊断系统的性能, 分析假

基金项目: 石家庄市科技计划, 项目编号 221200353, TCT 和 HPV 分型检测对宫颈病变的诊断价值及结果分析

作者简介: 赫晓朋(1987-), 女, 主治医师, 本科, 研究方向为病理学与病理生理学。

刘明磊(1979-), 女, 副高职称。

刘洪越(1991-), 女, 初级职业医师。

李晓欣, 女。

苏策, 女。

通讯作者: 赫晓朋

阳性的产生原因，并从技术、流程和管理多维度提出优化方案，旨在为 AI 系统在宫颈癌筛查中的精准应用提供理论依据和实践指导。

1. 资料与方法

1.1 研究对象与纳入标准

本研究采用回顾性研究方法，收集 2018 年 1 月至 2023 年 12 月石家庄市栾城人民医院 11594 例宫颈液基薄层细胞学切片，最终纳入分析的合格切片共 8793 例其中细胞学阳性（ASC-US 及以上）2119 例，有组织病理学结果 1073 例。

1.2 研究方法

1.2.1 细胞学诊断标准与质控

所有涂片由两名经过宫颈细胞学诊断规范化培训的初级医师独立阅片，中级以上职称医师审核并出具最终诊断结果。诊断标准采用 2014 版 TBS 报告系统^[5]，将诊断结果分为：①未见上皮内病变或恶性细胞；②非典型鳞状细胞-意义不明确；③非典型鳞状细胞-不排除高度病变；④低度鳞状上皮内病变；⑤高度鳞状上皮内病变；⑥鳞状细胞癌；⑦非典型腺细胞等。以 ASC-US 及以上为阳性判断标准。

标本满意度评估依据 TBS 系统要求：液基制片中保存完好的鳞状上皮细胞数不少于 5000 个；无明显细胞遮盖（遮盖率<75%）；细胞结构清晰可辨。不符合上述标准视为不满意标本，排除出研究。

1.2.2 人工智能辅助诊断系统

采用湖南品信生物工程有限公司开发的子宫颈细胞学计算机辅助诊断系统 CR-SA3。该系统基于深度卷积神经网络架构，采用改进的 YOLOv3 目标检测算法作为基线模型，增加图像级分类分支和检测分支，集成软尺度锚点匹配真值分配策略。具体工作流程：①图像采集：使用玻片扫描分析系统（湖南品信生物工程有限公司）对 TCT 切片进行全玻片扫描，生成 4096×2816 像素的高分辨率数字图像，平均扫描时间约 1 分钟/张。②细胞检测与分类：系统自动识别切片中的细胞区域，标注可疑异常细胞，提取形态学特征参数（包括细胞核大小、形状、核质比、染色质分布等），按照 TBS 标准对每个可疑细胞进行分类，并给出相应的置信度。③结果生成：系统综合全片分析结果，生成诊断报告，包括阳性/阴性判断、TBS 分级及可疑细胞的定位信息。

1.2.3 组织病理学诊断标准

对细胞学阳性或临床可疑的病例行阴道镜下宫颈活检，以组织病理学结果为金标准。病理诊断依据 WHO 女性生殖器官肿瘤分类标准，由两名中级以上职称病理医师独立诊断，分歧病例由第三名高级职称医师仲裁。将组织病理结果分为：①正常/炎症；②低级别上皮内病变；③高级别上皮内病变；④宫颈癌。

1.2.4 评价指标

比较 AI 系统与细胞病理医师的诊断结果，计算符合率、灵敏度、特异度、阳性预测值、阴性预测值、假阳性率、假阴性率。

1.2.5 统计学分析

采用 SPSS 25.0 软件进行统计分析。计数资料以频数和百分比表示，组间比较采用 χ^2 检验或 Fisher 精确检验。组间比较采用 t 检验， $P<0.05$ 为差异有统计学意义。

2. 结果

2.1 基本资料分析

合格切片 8793 例，细胞病理医师诊断阳性 1547 例，阴性 7246 例。阳性病例中，ASC-US 1324 例，LSIL 64 例，ASC-H 89 例，HSIL/SCC 70 例。

2.2 AI 系统与人工诊断一致性分析

AI 系统筛查结果显示阳性 2330 例，阴性 6463 例。与细胞病理医师诊断对比，两者均为阳性 1547 例，均为阴性 6463 例，AI 假阳性 783 例，假阴性 0 例（表 1）。总符合率 91.10%、灵敏度 100%、特异度 89.19%、阳性预测值 66.39%、阴性预测值 100%、假阳性率 10.81%、假阴性率 0%。

表 1 AI 系统与细胞病理医师诊断结果对比

	医师阳性	医师阴性	合计
AI 阳性	1547	783	2330
AI 阴性	0	6463	6463
合计	1547	7246	8793

2.3 TBS 分级诊断一致性

比较 AI 系统与细胞病理医师在各 TBS 分级中的诊断一致性（表 2）。结果显示，AI 在 NILM 分类中与医师的一致性最高（89.19%），在 ASC-US 和 LSIL 分类中一致性较低，分别为 65.0%和 85.3%。值得注意的是，AI 系统将 2038 例判读为 ASC-US，明显高于医师的 1324 例，显示 AI 对 ASC-US 的判读标准较为宽松。

表 2 AI 系统与细胞病理医师 TBS 分级诊断一致性

TBS 分级	AI 诊断例数	医师诊断例数	符合率(%)
NILM	6463	7246	89.19
ASC-US	2038	1324	64.96
LSIL	75	64	85.33
ASC-H	112	89	79.46
HSIL/SCC	105	70	66.66
合计	8793	8793	

2.4 以组织病理为金标准的诊断效能

在 1073 例有组织病理结果的病例中，病理诊断分布为：炎症 507 例，CIN I 140 例，CINII 145 例，CINI 210 例，癌 71 例。在炎症病例中 AI 检出 272 例，医师检出 462 例，AI 系统的符合率仅为 53.64%，显著低于医师的 91.12%。

2.5 阅片效率比较

AI 系统平均阅片时间为 70 秒/张，总计 170.98 小时；细胞病理医师平均阅片时间为 6 分钟/张，总计 879.3 小时。AI 系统节省了 708.32 小时，效率提升约 5.14 倍。此外，AI 系统可实现 24 小时不间断工作，不受疲劳因素影响，稳定性高。

2.6 假阳性病例特征

对 783 例假阳性病例进行分析：

细胞形态学：假阳性病例中，炎性细胞团占 38.69%，萎缩细胞群占 22.35%，修复性改变细胞占 15.58%，化生细胞占 12.26%，其他（包括黏液丝、污染物等）占 10.98%。

TBS 分布：假阳性病例中，AI 判读为 ASC-US597 例（76.24%），LSIL105 例（13.41%），ASC-H56 例（7.15%），HSIL25 例（3.19%）。

3. 讨论

本研究证实了 AI 辅助系统在宫颈细胞学筛查中的重要价值。AI 系统具有高灵敏度（100%）和阴性预测值（100%），能有效避免漏诊，这一优势对于大规模筛查至关重要。AI 系统显著提高了筛查效率，阅片速度是人工的 5 倍以上，能有效缓解细胞病理医师不足的压力。全玻片分析能全面评估涂片中的每一个细胞，提高病变检出率。

尽管 AI 系统表现出色，但 10.81% 的假阳性率仍是其临床应用的主要障碍。假阳性产生原因如下：

算法局限性：当前算法对某些特殊形态细胞的区分能力有限。例如，炎性细胞团、萎缩细胞群等良性病变与 ASC-H 在形态上有相似之处，易导致误判。此外，算法对细胞核的轻微增大、染色质轻度增粗等细微变化过于敏感，导致 ASC-US 的过度诊断。

训练数据不足：宫颈细胞学领域缺乏大规模、多中心、标准化的标注数据集，特别是 ASC-H、HSIL 和腺上皮病变等相对少见的病例样本不足。

制片质量：制片方式影响细胞分布、重叠度和背景清晰度。染色过深导致细胞核细节不清，染色过浅使细胞特征不明显，均影响 AI 识别。不同的制片方式对阅片也存在一定程度的影响^[6]，干封易产生气泡，导致细胞皱缩变形；湿封能更好地保持细胞立体结构，封片胶过多或过少也会影响图像质量。细胞重叠现象仍难以完全避免。重叠细胞导致形态特征改变，核质比评估不准确，增加误判风险。

ASC-US 的主观性：ASC-US 作为 TBS 系统中最主观的诊断类别，其诊断标准本身就存在模糊性。细胞核轻微增大、染色质轻度增粗等改变既可能源于良性反应，也可能是病变的早期表现。AI 系统在学习这种模糊分类时，往往采取保守策略，倾向于将不确定细胞判为阳性，导致假阳性增加。

良性病变模仿恶性特征：某些良性病变，如修复性改变、萎缩性变化、炎性反应等，在细胞形态上可模仿上皮内病变特征。反应性变化、挖空细胞/LSIL 和 HSIL 之间存在形态重叠，主观性无法消除^{[7][8]}。

图像伪影干扰：玻片上的灰尘、指纹、气泡等伪影可能被 AI 误判为细胞结构异常，特别是当这些伪影与细胞重叠时，干扰更为明显。

降低假阳性率需要优化策略，提高分类准确性。制定统一的制片和染色标准操作程序，应用自动化制片染色设备，减少人为因素影响^{[9][10]}；建立染色质控标准，定期评估染色效果。制定数字病理扫描标准，统一扫描参数和质量要求。建立 AI 与人工协同诊断流程，整合 HPV 检测结果，对 ASC-US 病例进行分层管理。整合患者年龄、临床表现、妇科检查等信息，为 AI 诊断提供临床背景，提高诊断准确性。从标本采集、制片、染色、扫描到 AI 分析，建立全流程质量控制体系，确保每个环节符合标准。建立 AI 系统性能监测机制，定期评估其灵敏度、特异度和符合率。加强对细胞学技师和医师的培训，特别是 AI 系统原理、局限性和结果解读方面的培训，提高人机协作效果。

本研究存在以下局限性：单中心回顾性研究；ASC-H 和 HSIL 等高级别病变样本量相对不足，可能影响对这些病变的统计分析；未纳入腺上皮病变和感染性病变，未能全面评估 AI 系统对所有类型宫颈病变的诊断能力；研究使用的 AI 系统为单一品牌，不同系统间的

性能差异有待进一步比较。

综上所述本研究通过大样本数据系统评估了 AI 辅助系统在宫颈细胞学筛查中的性能,证实了其在提高筛查效率、降低漏诊率方面的显著优势。同时,分析了假阳性的产生原因,包括技术局限性、制片染色差异、细胞学特征复杂性等多方面因素。针对这些问题,提出了从技术优化、流程标准化、多模态数据融合到质量控制的全方位解决方案。

AI 辅助宫颈细胞学诊断前在细胞精确分类方面仍有不足,但其应用前景广阔。未来,我们期待 AI 技术能够适应各种染色方案、设备型号和制片标准,辅助细胞学医师准确识别各类异常细胞,同时自动识别感染因素,提高筛查效率。

参考文献:

- [1] 张兰兰,杨巧,杨莉,等.陕西省 39 家医疗机构宫颈细胞学液基制片质量与诊断水平现状分析[J].现代肿瘤医学,2024,32(19):3661-3664.
- [2] 张小松,毕蕙,董燕,等.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].中国妇幼保健研究,2023,34(12):7-13.
- [3] HongPing T,De C,YanQing K, et al.Cervical cytology screening facilitated by an artificial intelligence microscope: A preliminary study. [J].Cancer cytopathology,2021,129(9):693-700.
- [4] Heling B,Hui B,Xiaosong Z, et al.Artificial Intelligence-Assisted Cytology for Detection of Cervical Intraepithelial Neoplasia or Invasive Cancer: A Multicenter, Clinical-Based, Observational Study[J].Obstetrical & Gynecological Survey,2020,75(12):741-742.
- [5] C D W,R N.Bethesda 2014: improving on a paradigm shift. [J].Cytopathology: official journal of the British Society for Clinical Cytology,2015,26(6):339-42.
- [6] H X Z,M X L,L W Z, et al.[Application of artificial intelligence-assisted diagnosis for cervical liquid-based thin-layer cytology]. [J].Zhonghua bing li xue za zhi = Chinese journal of pathology,2021,50(4):333-338.
- [7] James M T .A Practical Approach to Squamous Abnormalities on Cervical Cytology: Overview of Interpretive Criteria and Guidance for Altering Thresholds in Response to Quality Assurance Findings. [J].Acta cytologica,2023,67(2):11-14.
- [8] Ruchika G,Neeta K,Shivani B, et al.Artificial Intelligence-driven Digital Cytology-based Cervical Cancer Screening: Is the Time Ripe to Adopt This Disruptive Technology in Resource-constrained Settings? A Literature Review. [J].Journal of digital imaging,2023,36(4):1643-1652.
- [9] 刘毓玲,袁伟,狄媛媛.宫颈细胞学人工智能分析系统在室间质量评价中的应用[J].实用检验医师杂志,2023,15(01):103-106.
- [10] 苏红丹,高丽丽,李青.全自动液基薄层细胞制片机的使用体会[J].诊断病理学杂志,2024,31(07):680-681+684.

Analysis of False-Positive Causes in AI-Assisted Cervical Cytology

HE Xiaopeng*, LIU Minglei , LIU Hongyue, LI Xiaoxin, SU Ce

(Department of Pathology, Luancheng People's Hospital of Shijiazhuang, Shijiazhuang, Hebei 051430, China)

Abstract: Objective, To evaluate the application value of a computer-aided diagnostic (CAD) system for cervical cytology in screening for cervical lesions, to analyze the mechanisms underlying false-positive results, and to propose optimization strategies. Methods, A total of 11,594 liquid-based thin-layer cervical cytology slides collected at Luancheng People's Hospital of Shijiazhuang from 2018 to 2023 were analyzed using the CR-SA3 cervical cytology CAD system developed by Hunan Pinxin Bioengineering Co., Ltd. A result of atypical squamous cells of undetermined significance (ASC-US) or higher was considered positive. Diagnostic outputs from the AI system were compared with cytopathologist interpretations to calculate concordance, sensitivity, specificity, and predictive values; histopathology served as the gold standard for validation. Results, Among 8,793 qualified slides, the AI system flagged 2,330 as positive. Compared with pathologist diagnoses, there were 1,547 true positives, 783 false positives, and 0 false negatives. The overall concordance of the AI system was 91.10%, with a sensitivity of 100%, specificity of 89.19%, positive predictive value of 66.39%, and negative predictive value of 100%. AI-assisted reading was markedly more efficient than manual review, requiring an average of 70 seconds per slide and saving 708.32 hours in total. The system demonstrated relatively low specificity for benign conditions such as inflammation. Principal causes of false positives included slide preparation variability, uneven staining, inflammatory interference, the inherently subjective nature of ASC-US interpretation, and algorithmic limitations. Conclusions, The AI-assisted system exhibits high sensitivity and efficiency for primary screening in cervical cytology, effectively reducing the risk of missed diagnoses and substantially alleviating pathologists' workload. However, limitations remain in the precise classification of cellular lesions, with a comparatively high false-positive rate. Reducing false positives will require coordinated, multi-step optimization across the diagnostic workflow.

Keywords: artificial intelligence; false positive; liquid-based thin-layer cytology