

人工智能辅助阅片与单纯人工阅片在女性阴道微生态系统形态学诊断中的比对研究

杨 华^a, 孙天舒^b, 王 瑶^c, 徐英春^c, 孙宏莉^c

(北京协和医院 a. 妇产科; b. 医学科学研究中心; c. 检验科, 北京 100730)

摘要: 目的 评估不同级别检验员对革兰染色涂片阴道微生态形态学评价的基线水平, 分析检验员在电子阅片和镜下阅片的差异, 探究使用人工智能分析系统独立进行微生态评价以及辅助检验员进行微生态评价的能力表现, 评价人工智能分析系统在临床中的应用价值。方法 该研究样本来源于北京协和医院《中国女性人群下生殖道微生态菌群基线研究项目》, 收集 2021 年 5 月~2021 年 7 月女性阴道分泌物涂片共 385 例, 经革兰染色和图像采集后, 分别进行检验员等级考核以及人工显微镜镜下阅片、人工电子阅片、人工智能 (artificial intelligence, AI) 独立阅片和 AI 辅助检验员阅片。在确定镜下阅片金标准和电子阅片金标准之后, 分析两种不同阅片方式在 AV 评分和 Nugent 评分的差异, 比较不同级别检验员、AI, 以及经 AI 辅助后, 在 AV 评分和 Nugent 评分上的能力表现。结果 镜下阅片和电子阅片在需氧菌性阴道炎 (aerobic vaginitis, AV) 和细菌性阴道病 (bacterial vaginosis, BV) (含 BV 中间型) 诊断的 Kappa 一致性分析分别为 0.91 和 0.93 ($P < 0.01$)。AI 独立阅片在 AV 和 BV (含 BV 中间型) 诊断的准确度分别为 0.85 和 0.92, 灵敏度分别为 0.86 和 0.88, Kappa 值分别为 0.62 和 0.79。初级检验员在电子阅片下的 AV 和 BV (含 BV 中间型) 诊断的准确度分别为 0.85 ± 0.02 和 0.89 ± 0.01 , 灵敏度分别为 0.64 ± 0.06 和 0.84 ± 0.07 , Kappa 值分别为 0.55 ± 0.07 和 0.72 ± 0.04 。高级检验员在电子阅片下的 AV 和 BV (含 BV 中间型) 诊断的准确度分别为 0.92 ± 0.03 和 0.91 ± 0.03 , 灵敏度分别为 0.87 ± 0.02 和 0.92 ± 0.04 , Kappa 值分别为 0.78 ± 0.07 和 0.79 ± 0.06 。经 AI 辅助诊断后, 初级检验员 AV 和 BV (含 BV 中间型) 诊断的 Kappa 值提升至 0.77 ± 0.04 和 0.78 ± 0.02 , 高级检验员 AV 和 BV (含 BV 中间型) 诊断的 Kappa 值提升至 0.82 ± 0.05 和 0.85 ± 0.01 。结论 镜下阅片和电子阅片的一致性非常高, 电子阅片或可替代镜下阅片成为一种新的阅片方式。AI 独立阅片诊断 AV 和 BV (含 BV 中间型) 的能力优于普通检验员, 比高级检验员略差。不同级别检验员经 AI 辅助诊断后, AV 和 BV (含 BV 中间型) 的诊断能力均有提升, 其中初级检验员提升明显, 能力接近高级检验员的水平, 且各检验员之间的偏差缩小明显。整体结果表明, 使用人工智能 Descartes-Image 妇科微生态辅助分析软件不仅能提升检验员诊断能力, 还能减小检验员之间的偏差, 使诊断结果不容易因为人为因素而出现较大波动, 保证了结果的稳定性和可靠性。

关键词: 阴道微生态; 形态学检测; 人工智能; 性能评估; 人工智能辅助阅片

中图分类号: R446.19 **文献标识码:** A **文章编号:** 1671-7414 (2023) 01-169-07

doi: 10.3969/j.issn.1671-7414.2023.01.032

Comparative Study of Artificial Intelligence-assisted Analysis and Manual Visual Analysis in Gynecological Microbiome Diagnosis

YANG Hua^a, SUN Tian-shu^b, WANG Yao^c, XU Ying-chun^c, SUN Hong-li^c

(a. Department of Obstetrics and Gynecology; b. Medical Science Research Center; c. Department of Laboratory Medicine, Peking Union Medical College Hospital, Beijing 100730, China)

Abstract: Objective To evaluate the accuracy of morphological evaluation of vaginal microecology on Gram-stained vaginal smears by operators of differing levels of experience, determine differences between analyses using previously captured images versus live microscope images, explore the use of an artificial intelligence (AI) analysis system to independently conduct vaginal microecological evaluation, assist operators in vaginal microecological evaluation, and determine the application value of the AI analysis system in a clinical setting. **Methods** A total of 385 cases of female vaginal secretion smears from May 2021 to July 2021 were collected. After gram dyeing and image acquisition, the inspector's grade assessment, manual microscope film reading, manual electronic film reading, AI independent film reading and AI auxiliary inspector film reading were conducted respectively.

作者简介: 杨华 (1982-), 女, 医学博士, 主治医师, 研究方向: 阴道微生态, E-mail: huasunny82@126.com。

通讯作者: 孙宏莉 (1971-), 女, 医学博士, 副研究员, 硕士生导师, 主要研究方向: 临床微生物检验, E-mail: sunh12010@sina.com。

After determining the gold standard of microscopic viewing and the gold standard of image viewing, the differences in AV score and Nugent score of two different viewing methods were analyzed, and the performance of different operators, AI, and AI-assisted performance on AV and Nugent scores were compared. **Results** The Kappa concordance analysis of microscopic viewing and image viewing in the diagnosis of AV and BV (including BV intermediate) were 0.91 and 0.93, respectively ($P < 0.01$). The accuracy of AI independent analysis in the diagnosis of AV and BV (including intermediate BV) were 0.85 and 0.92, sensitivity was 0.86 and 0.88 respectively, and the Kappa value was 0.62 and 0.79 respectively. The diagnostic accuracy of AV and BV (including intermediate BV) by junior operators using image viewing were 0.85 ± 0.02 and 0.89 ± 0.01 , the sensitivity were 0.64 ± 0.06 and 0.84 ± 0.07 , and the Kappa value were 0.55 ± 0.07 and 0.72 ± 0.04 respectively. The diagnostic accuracy of AV and BV (including intermediate BV) by senior operators using image viewing was 0.92 ± 0.03 and 0.91 ± 0.03 , the sensitivity was 0.87 ± 0.02 and 0.92 ± 0.04 respectively, and the Kappa values was 0.78 ± 0.07 and 0.79 ± 0.06 respectively. After AI-assisted diagnosis, the Kappa values of AV and BV (including intermediate BV) diagnosed by junior operators were increased to 0.77 ± 0.04 and 0.78 ± 0.02 , and the Kappa values of senior operators AV and BV (including intermediate BV) diagnosis were increased to 0.82 ± 0.05 and 0.85 ± 0.01 . **Conclusion** The consistency between microscopic viewing and image viewing was very high, suggesting that image viewing could replace microscopic viewing as a new viewing method. The ability of AI to independently analyze and diagnose AV and BV (including intermediate BV) was better than that of junior operators, and slightly inferior to senior operators. With AI-assisted analysis, the diagnostic capabilities for AV and BV (including intermediate BV) among both junior and senior operators improved. The performance of junior operators improved significantly and nearly approached the performance of senior operators, significantly reducing the performance gap between junior and senior operators. The overall results indicated that the use of the Turing Microbial Gynecology Microbiome Auxiliary Analysis System not only improved the diagnostic ability of the operators, but also reduce the deviation between different operators thereby reducing fluctuations due to human factors and improving reliability of diagnosis.

Keywords: vaginal microecology; morphological detection; artificial intelligence; performance evaluation; AI-assisted analysis

女性生殖道感染作为女性常见的妇科疾病之一,严重影响女性的健康。阴道感染性疾病是临床最常见的女性生殖道感染性疾病,主要包括细菌性阴道病(Bacterial vaginosis, BV)、阴道毛滴虫病(*Trichomonas vaginitis*, TV)、外阴阴道假丝酵母菌病(*Vulvovaginal candidiasis*, VVC)、需氧菌性阴道炎(Aerobic vaginitis, AV)及混合性阴道炎。据卫健委数据统计,我国在妇科疾病普查时20~64岁之间的女性阴道炎患病率为11%^[1]。女性生殖道感染性疾病已成为重大社会及公共卫生问题,给家庭及社会带来严重负担。

临床上对女性阴道微生态评价系统包括形态学检测和功能学检测两部分,目前以形态学检测为主要参考指标,功能学检测为辅^[2]。形态学诊断方法包括革兰染色涂片镜检法和湿片镜检法,主要采用显微镜人工形态学检验观察阴道分泌物内的微生物和细胞形态特征进行判读。但无论是形态学检测还是功能学检测都会受到诸多因素的影响导致误、漏诊率较高。在实验室检测及判读过程中高度依赖于检验员的主观判断,其诊断可能会受到个人技能和经验的影响,如检验员操作不当、经验缺乏等^[3-4]。另外对于大型综合性医院,检验员工作任务重、责任大,不仅不利于医疗安全,而且影响到医院和检验科在患者心目中的信任度^[5]。因此,临床迫切需

要一种更有效的临床辅助诊断方法帮助临床提高检验员水平,缩小各医院检验科技术人员的相关检验能力差距,提高诊断准确性和时效性。

国内外已有许多将深度学习应用于医疗领域的研究,在计算机视觉任务中表现出出色的性能。清华大学WANG等^[6]开发并优化了卷积神经网络(Convolutional Neural Networks, CNN)模型,可实现从显微镜图片中自动识别并进行Nugent评分三分类,且模型表现优于参与测试的高级检验科医生^[7]。图灵微生物妇科微生态辅助分析系统进一步将卷积神经网络(CNN)深度学习算法应用于Nugent评分、AV评分等多个指标,以量化革兰染色,从而实现对阴道微生态多个指标的自动化诊断。目前已有文章报道使用自动化分析系统对革兰染色涂片的形态学进行评价,结果表明,在BV, VVC以及其他阴道菌群指标上与人工镜检有较好的一致性^[8]。但不同级别的检验员在阴道微生态评价的水平如何,人工智能或自动化分析系统是否能替代检验员,如何使用人工智能或自动化分析系统从而达到最优效果,目前尚缺乏数据支持。因此,本研究通过分析比较6位检验员(3名初级检验员,3名高级检验员)在显微镜镜下阅片和人工电子阅片的结果,了解不同水平检验员诊断AV, BV以及BV中间型的基线水平,分析镜下阅片和电子阅片

两种方法在诊断 AV 和 BV 上是否存在显著差异。与此同时,通过使用图灵微生物妇科微生态辅助分析系统,了解人工智能(artificial intelligence, AI)独立诊断的能力以及不同级别检验员使用 AI 辅助分析系统后的能力表现,评价图灵微生物妇科微生态辅助分析系统的应用效果。

1 材料与方法

1.1 研究对象 本研究收集了 2021 年 5 ~ 7 月女性阴道分泌物涂片,并筛选出 385 例符合入组和排除标准的样本。入选标准为:年龄 ≥ 20 岁,有性生活的女性阴道分泌物涂片样本;诊断指标包括 Nugent 评分和 AV 评分。排除标准为:月经期、妊娠期女性阴道分泌物涂片样本;不符合制片要求,或信息不完全的样本;玻片或图片受损,检验员认为无法进行阅片的样本。本研究样本来源于《中国女性人群下生殖道微生物菌群基线研究项目》,并且已经通过北京协和医院伦理审查委员会审批(伦理审查批件编号:ZS-2902)。

1.2 仪器与试剂 TURING M3F 显微图像扫描仪(湖南唯知医学科技有限公司),Descartes-Image 妇科微生态辅助分析软件(湖南唯知医学科技有限公司),奥林巴斯 BX43 生物显微镜,贝索全自动染色机以及配套染色液(珠海贝索生物技术有限公司)。

1.3 方法

1.3.1 取样、涂片制备及革兰染色:根据 2016 年《阴道微生态评价的临床应用专家共识》^[3],样本提供者取膀胱截石位,将窥器以少量生理盐水润滑后放入阴道内,暴露子宫颈,以干棉签从阴道上 1/3 侧壁刮取分泌物,并在清洁载玻片上均匀涂抹,涂抹玻片后静置室温自然晾干。待涂片干燥和固定后进行革兰染色。

1.3.2 镜检阅片方法:每张分泌物涂片均进行人工显微镜镜下阅片、人工电子阅片、AI 独立阅片以及 AI 辅助阅片四种方法。其中,①人工显微镜镜

下阅片:由检验员人工镜下独立阅片,给出每张分泌物涂片的 Nugent 评分和 AV 评分。用于人工镜检的显微镜型号为奥林巴斯 BX43,1 000 倍油镜检测至少 40 个视野。②人工电子阅片:每张分泌物涂片通过 TURING M3F 显微图像扫描仪拍摄 200 张图片,并对拍摄的数字图片进行人工阅片,给出每张分泌物涂片的 Nugent 评分和 AV 评分。③ AI 独立阅片:采用与人工电子阅片相同的图片,由图灵微生物妇科微生态辅助分析软件对图片进行自动分析后,给出每张分泌物涂片的 Nugent 评分和 AV 评分。④ AI 辅助阅片:采用与人工电子阅片相同的图片,使用图灵微生物妇科微生态辅助分析软件,辅助检验员给出每张分泌物涂片的 Nugent 评分和 AV 评分。

1.3.3 判定标准:根据 2016 年《阴道微生态评价的临床应用专家共识》和《细菌性阴道病诊治指南(2021 修订版)》^[2,8],BV 和 AV 诊断标准如下:

① BV:按照 Nugent 评分标准,记录相应 Nugent 评分^[9]。其中,Nugent 评分 0 ~ 3 分为正常,4 ~ 6 分为 BV 中间型,7 ~ 10 分为 BV。本次试验将 Nugent 评分 ≥ 4 分标记为阳性病例。具体评分标准见表 1。② AV:采用基于革兰染色玻片的评分方式,AV 评分标准见表 2^[9-10]。这里将阴道分泌物显微镜下 AV 评分 ≥ 3 分标记为 AV 阳性病例。

表 1 Nugent 评分标准

评分	乳杆菌	加德纳菌及类杆菌	革兰染色不定的弯曲小杆菌
0	4+	0	-
1	3+	1+	1+ 或 2+
2	2+	2+	3+ 或 4+
3	1+	3+	-
4	0	4+	-

注:各项根据每 10 个油镜视野下观察到的每类形态细菌的平均数量进行评分;0:油镜视野未见细菌;1+: <1 个细菌/油镜视野(此为平均数);2+:1~4 个细菌/油镜视野;3+:5~30 个细菌/油镜视野;4+: >30 个细菌/油镜视野;—无此项。

表 2 AV 评分标准

AV 评分	LBG	白细胞数	背景菌落	PBC 所占比例(%)
0	I 或 II a	$\leq 10/\text{hpf}$	不明显或溶解性	无或 <1
1	II b	$> 10/\text{hpf}$ 和 $\leq 10/$ 上皮细胞	大肠埃希菌类的小杆菌	≤ 10
2	III	$> 10/$ 上皮细胞	球菌样或呈链状	> 10

注:LBG(Lactobacillary Grades)乳杆菌分级:I:多量多形性乳杆菌,无其他细菌;II a:指混合菌群,但主要为乳杆菌;II b:混合菌群,但乳杆菌比例明显减少,少于其他菌群;III:乳杆菌严重减少或缺失,其他细菌过度增长。高倍视野(Highpower field, HPF):等效 400 倍放大目视视野大小;PBC(parabasal epitheliocytes):基底旁上皮细胞。

1.3.4 检验员等级考核:从五家医院检验科邀请 10 名检验员参与本次实验。10 名检验员从事医学检验工作 3 ~ 20 年不等,从事阴道微生态评价工作的

经验也不尽相同。为了更公平的考核检验员在阴道微生态形态学评价的能力,从入选样本中随机挑选 50 张分泌物涂片,用于检验员等级考核。其中,

AV, BV 和 BV 中间型分泌物涂片样本各 10 张, AV+BV 混合感染 5 张, 菌群正常分泌物涂片样本 15 张。10 名检验员分别对 50 张分泌物涂片进行镜下阅片, 按照上述评分标准进行判定, 并参考初次完成的革兰染色微生态形态学评价结果, 对每例样本进行打分。若样本形态学评价结果正确, 记 2 分。 > 80 分, 认定为高级检验员; $61 \sim 80$ 分, 认定为初级检验员; ≤ 60 分, 则认为不合格, 将不参与后续实验。

1.3.5 金标准: 由于显微镜下阅片和电子阅片略有不同, 因此分别确定镜下阅片金标准和电子阅片金标准。金标准按照高级检验员诊断结果的多数原则来确定。如果某例样本有 2 名高级检验员的诊断结果一致, 则该结果将作为金标准; 若结果无法达成一致, 则由清华长庚医院妇产科微生态检验实验室的高级检验员在不知晓其他检验员结果的情况下判读后, 确定金标准。

1.4 统计学分析

1.4.1 样本量估计: 本研究采用灵敏度和特异度评价诊断价值, 因此在不同诊断方法的分析比较中, 通过使用 PASS v22.0.2 软件中的诊断性试验方法估算所需样本量。假设 AI 独立诊断 AV 和 BV(含 BV 中间型)的灵敏度和特异度均为 0.85, 在 $\alpha=0.05$ (单侧), $\beta=0.1$, 患病率 0.2, 采用 PASS 估算样本量。结果表明, 至少需要纳入 AV 患者 59 人, BV(含 BV 中间型)患者 59 人, 对照组 236 人。考虑初次完成的革兰染色微生态形态学评价结果存在 10% 的错误率, 因此考虑入组 AV 患者 65 人, BV(含 BV 中间型)患者 65 人, 非 AV, BV 或 BV 中间型 260 人。如果某例样本诊断结果为 AV+BV 或 AV+BV 中间型的混合感染, 则该样本将同时入组 AV 和 BV(含 BV 中间型)患者中。

1.4.2 统计学分析: 使用 R4.0.5 统计分析软件, 分析比较镜下阅片和电子阅片两种不同阅片方式的差异, 采用 Kappa 一致性检验和配对卡方检验(McNemer test)进行评估分析。当 $Kappa \geq 0.75$, 且 $P < 0.05$ 时, 说明两种方法诊断结果一致性较好。当 McNemer 检验 $P \geq 0.05$ 时, 说明两种方法差异无统计学意义。所有统计检验均采用双侧检验, $P < 0.05$ 为差异有统计学意义。

2 结果

2.1 镜下阅片和电子阅片在 AV 评分和 Nugent 评分的差异分析 根据多数原则, 确定 335 例样本镜下阅片和电子阅片的金标准, 详细分布结果见表 3 和表 4。AV 阳性一致性检验 Kappa 值为 $0.91(P < 0.01)$, McNemer 检验 $P=0.7518$, 说明镜下阅片和电子阅片两种方法在 AV 诊断上差异无统计学意义。Nugent 评分三分类一致性检验 Kappa 值为

$0.848(P < 0.01)$, Nugent 评分二分类($Nugent \geq 4$ 分)一致性检验 Kappa 值为 $0.925(P < 0.01)$, McNemer 检验 $P=0.7518$, 说明镜下阅片和电子阅片两种方法在 BV 诊断上差异无统计学意义。

表 3 AV 镜下阅片金标准和电子阅片金标准混淆矩阵

人工镜下阅片金标准	人工电子阅片金标准		合计
	阳性	阴性	
阳性	66	6	72
阴性	4	259	263
合计	70	265	335

表 4 Nugent 评分三分类镜下阅片金标准和电子阅片金标准混淆矩阵

人工镜下阅片金标准	人工电子阅片金标准			合计
	BV	BV 中间型	正常	
BV	35	7	0	42
BV 中间型	5	39	4	48
正常	0	6	239	245
合计	40	52	243	335

2.2 不同级别检验员分别采用显微镜镜下阅片和电子阅片在 AV 评分和 Nugent 评分的差异分析 3 名初级检验员和 3 名高级检验员分别对相同样本进行显微镜镜下阅片、电子阅片。通过与金标准比对分析发现, 初级检验员镜下阅片和电子阅片诊断 AV 的灵敏度分别为 0.68 ± 0.06 和 0.64 ± 0.06 , 特异度为 0.91 ± 0.02 和 0.91 ± 0.02 。与金标准的一致性检验分析表明, Kappa 值分别为 0.59 ± 0.07 , 0.55 ± 0.07 , 且 $P < 0.01$, 说明初级检验员 AV 存在较严重的漏检, 且与金标准的一致性处于中等水平。高级检验员镜下阅片和电子阅片诊断 AV 的灵敏度分别为 0.87 ± 0.04 和 0.87 ± 0.02 , 特异度分别为 0.91 ± 0.06 和 0.93 ± 0.04 , 一致性检验 Kappa 值分别为 0.74 ± 0.09 和 0.78 ± 0.07 , 高于初级检验员的 Kappa 值, 且电子阅片的平均 Kappa 值 > 0.75 , 说明高级检验员在电子阅片时的平均一致性较好。AV 诊断的详细结果见表 5。

初级检验员镜下阅片和电子阅片诊断 BV(含 BV 中间型, 即 Nugent 评分 ≥ 4 分)的灵敏度分别为 0.85 ± 0.06 和 0.84 ± 0.07 , 特异度分别为 0.91 ± 0.02 和 0.90 ± 0.02 , Kappa 值分别为 0.73 ± 0.04 和 0.72 ± 0.04 , 说明初级检验员在 BV(含 BV 中间型)的诊断存在少量的漏检, 与金标准的一致性处于中等水平。高级检验员镜下阅片和电子阅片诊断 BV(含 BV 中间型, 即 Nugent 评分 ≥ 4 分)的灵敏度分别为 0.90 ± 0.05 和 0.92 ± 0.04 , 特异度分别为 0.91 ± 0.03 和 0.91 ± 0.02 , 一致性检验 Kappa 值分别为 0.78 ± 0.08 和 0.79 ± 0.06 ,

高于初级检验员的 Kappa 值, 且 Kappa> 0.75, 说明高级检验员在 BV 和 BV 中间型诊断的平均一致性较好, 详细结果见表 6。

表 5 不同检验方法以及检验员在 AV 诊断中的表现

检验方法	分组	准确度	灵敏度	特异度	阳性预测值	阴性预测值	Kappa
镜下	初级检验员	0.86 ± 0.02	0.68 ± 0.06	0.91 ± 0.02	0.67 ± 0.06	0.91 ± 0.02	0.59 ± 0.07
	高级检验员	0.91 ± 0.04	0.87 ± 0.04	0.91 ± 0.06	0.75 ± 0.13	0.96 ± 0.01	0.74 ± 0.09
电子	初级检验员	0.85 ± 0.02	0.64 ± 0.06	0.91 ± 0.02	0.66 ± 0.05	0.90 ± 0.02	0.55 ± 0.07
	高级检验员	0.92 ± 0.03	0.87 ± 0.02	0.93 ± 0.04	0.79 ± 0.10	0.96 ± 0.01	0.78 ± 0.07
AI		0.85	0.86	0.85	0.61	0.96	0.62
AI 辅助分析	初级检验员	0.92 ± 0.02	0.87 ± 0.03	0.93 ± 0.03	0.78 ± 0.07	0.96 ± 0.01	0.77 ± 0.04
	高级检验员	0.94 ± 0.02	0.90 ± 0.03	0.94 ± 0.02	0.82 ± 0.06	0.97 ± 0.01	0.82 ± 0.05

表 6 不同检验方法以及检验员在 BV(含 BV 中间型) 诊断中的表现

检验方法	分组	准确性	灵敏度	特异度	阳性预测值	阴性预测值	Kappa
镜下	初级检验员	0.89 ± 0.01	0.85 ± 0.06	0.91 ± 0.02	0.78 ± 0.03	0.94 ± 0.02	0.73 ± 0.04
	高级检验员	0.91 ± 0.03	0.90 ± 0.05	0.91 ± 0.03	0.79 ± 0.07	0.96 ± 0.02	0.78 ± 0.08
电子	初级检验员	0.89 ± 0.01	0.84 ± 0.07	0.90 ± 0.02	0.77 ± 0.03	0.94 ± 0.02	0.72 ± 0.04
	高级检验员	0.91 ± 0.03	0.92 ± 0.04	0.91 ± 0.02	0.80 ± 0.04	0.97 ± 0.02	0.79 ± 0.06
AI		0.92	0.88	0.93	0.83	0.96	0.79
AI 辅助分析	初级检验员	0.91 ± 0.01	0.89 ± 0.03	0.92 ± 0.02	0.80 ± 0.03	0.96 ± 0.01	0.78 ± 0.02
	高级检验员	0.94 ± 0.01	0.92 ± 0.04	0.95 ± 0.02	0.87 ± 0.04	0.97 ± 0.01	0.85 ± 0.01

2.3 不同级别检验员 AI 辅助阅片在 AV 评分和 Nugent 评分的差异分析 使用图灵微生物妇科微生态辅助分析软件对相同样本进行自动分析, 详细结果见表 5 和表 6 的 AI 辅助分析。结果表明, AI 独立诊断 AV 的灵敏度为 0.86, 灵敏度高于初级检验员, 与高级检验员接近; Kappa 一致性结果为 0.62 ($P < 0.01$), 高于初级检验员的 Kappa 值, 说明 AI 与金标准的一致性高于初级检验员的平均一致性。AI 独立诊断 BV(含 BV 中间型) 的灵敏度和 Kappa 值分别为 0.88 和 0.79, 均高于初级检验员, 且与金标准的一致性较好。

检验员在使用图灵微生物妇科微生态辅助分析软件后, 诊断 AV 和 BV(含 BV 中间型) 的详细结果见表 5 和表 6 的 AI 辅助分析。结果表明, 使用 AI 辅助分析软件后, 初级检验员和高级检验员诊断 AV 和 BV(含 BV 中间型) 的能力均有所提升。其中, 初级检验员在 AV 诊断的灵敏度由 0.64 ± 0.06 提升到 0.87 ± 0.03 , Kappa 一致性由 0.55 ± 0.07 提升至 0.77 ± 0.04 , 一致性接近高级检验员人工电子阅片的水平。高级检验员 AV 诊断的 Kappa 一致性也由 0.78 ± 0.07 提升至 0.82 ± 0.05 , 除了 Kappa 均值有所提升之外, 标准偏差也在减小。而在 BV(含 BV 中间型) 的诊断中, 各个指标均有小幅度提升,

其中阳性预测值提升最为明显, 初级检验员阳性预测值由 0.77 ± 0.03 提升至 0.80 ± 0.03 , Kappa 值由 0.72 ± 0.04 提升至 0.78 ± 0.02 ; 高级检验员阳性预测值由 0.80 ± 0.04 提升至 0.87 ± 0.04 , Kappa 值由 0.79 ± 0.06 提升至 0.85 ± 0.01 。

3 讨论

基于光学显微技术的体液形态学镜检在临床微生物检验中具有重要地位, 其中白带形成分镜检为妇科生殖道感染临床诊断提供了重要的临床信息。长期以来, 形态学镜检主要采用光学显微镜目镜下人工阅片的方式, 严重依赖人员经验并极易产生疲劳。随着各种自动化检验技术和仪器在临床检验中的应用, 人工形态学镜检在实际检验工作中应用越来越少, 年轻检验员对于形态学镜检技术的掌握程度也逐渐变低。近年来, 自动化光学显微技术及人工智能技术的快速发展为临床形态学镜检这一重要的传统检验技术带来了重新广泛发挥其临床价值的可能性。利用自动扫描显微镜进行数字成像并进行高清显示, 检验人员可以摆脱光学显微镜目镜下镜检带来的疲劳, 并借助人工智能技术大大提高了形态学镜检的效率和准确度。

阴道微生态形态学评价中 AV, BV, VVC, TV 及混合感染的诊断标准中都涉及依据形态学信

息对微生物的种类和数量进行判读,其中BV、VVC和TV的形态学标准均可以使用革兰染色涂片进行判读。只有AV标准采用了DONNERS等^[9]提出的400倍相差显微镜下观察生理盐水湿片的方法,该评分体系中包括乳杆菌分级、白细胞数量、含中毒颗粒的白细胞所占比例、背景菌落及基底旁上皮细胞比例,对这5个项目分别评分,每项0~2分,总分10分;累计评分 ≥ 3 分诊断为AV。该方法有以下几点不足之处:①基于生理盐水湿片,未进行染色,放大倍数仅为400 \times 。②镜下细菌形态较小且可随介质流动,需配备相差显微镜,要求检验人员有丰富的经验。③生理盐水湿片无法保存,导致不能够回顾性阅片。为了满足阴道微生态评价BV和AV的同时评价、配合自动化图像扫描仪和人工智能辅助阅片,很多医院临床实际应用中依然采用基于革兰染色玻片的评分方式来对乳杆菌分级、白细胞数量、背景菌落及基底旁上皮细胞比例进行评价,天津医科大学总医院的薛凤霞团队并据此提出了基于革兰染色涂片结合临床症状的AV评分方法,通过对比验证了其临床可行性和独特优势^[10]。因此,利用基于革兰染色涂片的方式是一种临床实现多种女性下生殖道感染统一评价的可行方法,并十分有利于自动化显微扫描结合AI技术辅助诊断技术的开发和推广。

本研究通过比较AI以及不同级别检验员在显微镜镜下阅片、人工电子阅片和AI辅助阅片的结果,分析镜下阅片和电子阅片两种阅片方法的一致性,评估AI和检验员在阴道分泌物革兰染色微生态形态学评价中的能力表现,以及检验员在AI辅助诊断后的能力提升,评估图灵微生物妇科微生态辅助分析软件的临床应用价值。

人工镜下阅片和电子阅片诊断AV和BV(含BV中间型)的一致性分析表明,AV和BV(含BV中间型)的Kappa值分别为0.91和0.93,且均 $P < 0.01$,说明镜下阅片和电子阅片的一致性非常高。考虑到电子图片可永久存储,便于样本追溯,且可极大地降低由镜下阅片引起的视觉疲劳,电子阅片或可替代镜下阅片成为一种新的阅片方式。

比较AI以及不同级别检验员在显微镜镜下阅片、人工电子阅片和AI辅助阅片的差异,结果表明,对于AV诊断来说,AI独立诊断AV的准确性与初级检验员的准确性接近,但是AI独立诊断的灵敏度和Kappa值均高于初级检验员,且灵敏度与高级检验员接近,说明AI诊断AV的能力优于初级检验员,但比高级检验员略差。对于BV(含BV中间型)诊断来说,除了灵敏度略低于高级检验员之外,AI独立诊断的准确性和Kappa值均比初级检验员高,

且与高级检验员接近,说明AI诊断BV(含BV中间型)表现良好。检验员使用AI辅助分析软件重新进行AV评分和Nugent评分,结果表明,初级检验员的诊断水平有大幅度提升,尤其是诊断AV的灵敏度由 0.64 ± 0.06 提升至 0.87 ± 0.03 ,达到高级检验员电子阅片的水平。高级检验员使用AI辅助后各个指标均有小幅度提升,且检验员之间的标准偏差也在降低,说明使用图灵微生物妇科微生态辅助分析软件不仅能提升检验员诊断能力,还能减小检验员之间的偏差,使诊断结果不容易因为人为因素而出现较大波动,保证了结果的稳定性和可靠性。

阴道微生态评价及其检测方法的应用为女性下生殖道感染性疾病的诊治带来了显著的临床益处,并且在解决临床常见的混合性阴道炎和症状不典型的阴道感染中发挥了重要作用。但是由于该评价方法的复杂性和对检验人员经验要求高等因素,阴道微生态评价方法在广大基层医疗机构中推广效果不佳。Descartes-Image妇科微生态辅助分析软件是基于深度卷积神经网络ResNet50搭建的分类模型结合计算机视觉方法,用于对阴道分泌物革兰染色微生态形态学自动化诊断和辅助分析。基于人工智能技术解决了人工镜检进行微生态评价中的效率和可靠性问题,为阴道微生物评价方法的临床广泛应用提供了很好的技术支持。

参考文献:

- [1] 国家卫生健康委员会.中国卫生健康统计年鉴[M].北京:中国协和医科大学出版社,2020:222.
National Health Commission. China Health Statistics Yearbook[M].Beijing: China Union Medical University Press, 2020: 222.
- [2] 中华医学会妇产科学分会感染性疾病协作组.阴道微生态评价的临床应用专家共识[J].中华妇产科杂志,2016,51(10):721-723.
Cooperative Group of Infectious Disease, Chinese Society of Obstetrics and Gynecology, Chinese Medical Association. Expert consensus on the clinical application of vaginal microecology test [J]. Chinese Journal of Obstetrics and Gynecology, 2016,51(10): 721-723.
- [3] 杨广全,来忠梅.阴道分泌物检查结果的影响因素分析[J].中外女性健康研究,2016(11):92,94.
YANG Guangquan, LAI Zhongmei. Analysis of influencing factors of vaginal discharge examination results[J]. Women's Health Research, 2016(11):92, 94.
- [4] 李云,冯伟,王丕明,等.细菌性阴道病分泌物漏误诊相关因素分析[J].航空航天医学杂志,2014,25(8):1121-1122.
LI Yun, FENG Wei, WANG Peiming, et al. Analysis of factors related to missed and misdiagnosed secretion of bacterial vaginosis [J]. Journal of Aerospace Medicine, 2014, 25(8):1121-1122.

(下转第198页)

- on reducing adverse events of hemolytic transfusion[J]. *Clinical Research and Practice*, 2020, 5(5): 115-117.
- [5] 张印则, 徐华, 周华友. 红细胞血型原理与检测策略[M]. 2版. 北京: 人民卫生出版社, 2019:93-95.
- ZHANG Yinze, XU Hua, ZHOU Huayou. Principle and detection strategy of red blood cell blood group[M]. 2th ed. Beijing: People's Medical Publishing House, 2019: 93-95.
- [6] BELINGA S, NGO-SACK F, BILONG C, et al. High prevalence of anti D antibodies among women of child bearing age at Centre Pasteur of Cameroun[J]. *Afr J Reprod Health*, 2009, 13(3): 47-52.
- [7] 彭进, 梁延连, 徐华, 等. 西安地区献血人群 MNS 血型基因多态性研究[J]. *中国输血杂志*, 2017, 30(6):590-592.
- PENG Jin, LIANG Yanlian, XU Hua, et al. Analysis of genetic polymorphism of blood donors in Xi'an based on the MNS human blood group system[J]. *Chinese Journal of Blood Transfusion*, 2017, 30(6): 590-592.
- [8] 杰夫·丹尼尔斯. 人类血型[M]. 2版. 朱子严, 译. 北京: 科学出版社, 2007:171-173.
- JEFF Daniel. Human Blood Groups[M]. 2th ed. ZHU Ziyang, Translate. Beijing: Science Publishing House, 2007:171-173.
- [9] EMMANUEL A, TAWFEQ N, MARY G, et al. Anti-M-induced delayed hemolytic transfusion reaction[J]. *Laboratory Medicine*, 2020, 51(4): 426-429.
- [10] DANIELS G. Human blood groups[Z]. Oxford: Blackwell Science Ltd, 2013: 96-161.
- [11] 范亮峰, 刘曦, 沈伟, 等. 不规则抗体体内消减速率的研究[J]. *中国输血杂志*, 2016, 29(11):1229-1231.
- FAN Liangfeng, LIU Xi, SHEN Wei, et al. The persistence and evanescence of red cell irregular antibody in vivo[J]. *Chinese Journal of Blood Transfusion*, 2016, 29(11):1229-1231
- [12] 高娃, 何燕京, 白英哲, 等. 2012 ~ 2017 年某三甲医院患者 Kidd 血型系统特异性抗体回顾调查[J]. *中国输血杂志*, 2019, 32(7):682-685.
- GAO Wa, HE Yanjing, BAI Yingzhe, et al. Retrospective investigation of specific antibodies against Kidd blood group system in a single hospital patient population, from 2012 to 2017[J]. *Chinese Journal of Blood Transfusion*, 2019, 32(7):682-685.
- [13] 严凤好, 曾少丽, 曾演强, 等. 广东惠州地区献血者 Kidd 血型基因频率分布与输血风险评估[J]. *中国输血杂志*, 2018, 31(12):1398-1401.
- YAN Fenghao, ZENG Shaoli, ZENG Yanqiang, et al. Frequency distribution of Kidd blood group gene and tansfusion risk assessment of blood donors in Huizhou, Guangdong Province[J]. *Chinese Journal of Blood Transfusion*, 2018, 31(12):1398-1401.
- [14] 洪毅, 王华. 20 例非 ABO-HDFN 血型血清学分析及临床干预结果比较[J]. *现代检验医学杂志*, 2021, 36(6):120-123.
- HONG Yi, WANG Hua. Comparison of serological analysis and clinical intervention results of 20 non-ABO-HDFN serum types[J]. *Journal of Modern Laboratory Medicine*, 2021, 36(6):120-123.

收稿日期: 2022-04-11

修回日期: 2022-09-24

(上接第174页)

- [5] 陈黔. 大型综合性医院检验科门诊工作存在的问题与对策[J]. *西南军医*, 2007, 9(4): 100-101.
- CHEN Qian. Problems and countermeasures in outpatient work of laboratory department in large general hospitals[J]. *Journal of Military Surgeon in Southwest China*, 2007, 9(4):100-101.
- [6] WANG Zhongxiao, ZHANG Lei, ZHAO Min, et al. Deep neural networks offer morphologic classification and diagnosis of bacterial vaginosis zhongxiao[J]. *J Clin Microbiol*, 2021, 59(2):e02236-20.
- [7] 王瑶, 孙宏莉, 赵颖, 等. 自动化镜检结合人工智能分析系统对阴道分泌物革兰染色涂片形态学的准确性评价[J]. *协和医学杂志*, 2021, 12(4): 503-509.
- WANG Yao, SUN Hongli, ZHAO Ying, et al. Accuracy assessment of the morphological analysis system with automated microscopy and artificial intelligence for gram-stained vaginal discharge smears [J]. *Medical Journal of Peking Union Medical College Hospital*, 2021, 12(4):503-509.
- [8] 中华医学会妇产科学分会感染性疾病协作组. 细菌性阴道病诊治指南(2021 修订版)[J]. *中华妇产科杂志*, 2021, 56(1): 3-6.
- Cooperative Group of Infectious Disease, Chinese Society of Obstetrics and Gynecology, Chinese Medical Association. Guideline for diagnosis and treatment of bacterial vaginosis (2021 revised edition) [J]. *Chinese Journal of Obstetrics and Gynecology*, 2021, 56(1): 3-6.
- [9] DONDEERS G G, VERECKEN A, BOSMANS E, et al. Definition of a type of abnormal vaginal flora that is distinct from bacterial vaginosis: aerobic vaginitis[J]. *BJOG-An International Journal of Obstetrics and Gynaecology*, 2002, 109(1): 34-43.
- [10] 董梦婷, 王辰, 李会阳, 等. 基于革兰染色涂片结合临床特征的需氧菌性阴道炎联合诊断标准专家建议[J]. *中国实用妇科与产科杂志*, 2021, 37(3): 327-335.
- DONG Mengting, WANG Chen, XUE Huiyang, et al. Aerobic vaginitis diagnosis criteria combining gram stain with clinical features: an establishment and prospective validation study[J]. *Chinese Journal of Practical Gynecology and Obstetrics*, 2021, 37(3):327-335.

收稿日期: 2022-05-07

修回日期: 2022-06-07