基于超声融合图像的深度学习模型评估女性膀胱膨出

冉诗懿[1]，鲁蓉[2]，李沐宸[1]，屈灿[1]

(1. 中南大学湘雅医院超声影像科，长沙 410008；2. 中南大学湘雅医院妇科，长沙 410008)

[摘 要] 目的：膀胱膨出是女性盆腔器官脱垂的常见类型，超声是诊断膀胱膨出的首选影像学方法，但其操作烦琐、依赖医师经验，不同医师间诊断结果具有主观差异。本研究拟构建基于二维超声图像、三维超声图像及二维与三维超声融合图像的深度学习模型，用于女性膀胱膨出的诊断及其超声分度评估，并对比3种模型的诊断性能。**方法：**回顾性收集2022年1月至2024年12月于中南大学湘雅医院接受经会阴盆底超声检查的女性患者资料，符合纳排标准的共625例。根据临床和超声诊断结果，最终入组膀胱膨出467例(轻度膨出167例、明显膨出300例)，无膀胱膨出158例。使用GE Voluson E10采集膀胱在静息与最大Valsalva状态下二维正中矢状切面对比图像，以及最大Valsalva状态下三维容积渲染成像模式肛提肌裂孔重建图像。将全部数据按7∶3比例随机分为训练集(437例)和测试集(188例)。通过通道堆叠融合二维与三维图像信息，组成融合图像集。以ResNet50为骨干网络构建3种模型：1)二维图像模型，以原始二维双屏图像为输入；2)三维图像模型，以三维肛提肌裂孔重建图像为输入；3)融合图像模型，以二维与三维融合图像为输入。各模型诊断性能在测试集上进行评估，评估指标包括准确率、精确率、召回率、F1分数及受试者操作特征曲线下面积(area under the curve，AUC)，模型间准确率比较采用McNemar检验。**结果：**二维图像模型总体准确率为75.5%，宏平均的精确率、召回率、F1分数、AUC分别为69.4%、69.8%、69.2%、0.878。三维图像模型总体准确率为76.1%，宏平均的精确率、召回率、F1分数、AUC分别为78.7%、71.7%、71.3%、0.908。融合图像模型总体准确率为82.3%，宏平均的精确率、召回率、F1分数、AUC分别为80.5%、78.7%、79.3%、0.930，在无膀胱膨出、轻度膨出、明显膨出3个类别上的F1分数分别为80.4%、68.4%、89.0%。融合图像模型准确率显著高于二维图像模型(82.3% vs 75.5%，P=0.031)；与三维图像模型比较，差异无统计学意义(82.3% vs 76.1%，P=0.082)。**结论：**基于超声图像构建的深度学习模型，实现了膀胱膨出的智能化诊断与分度评估。二维与三维超声图像在膀胱膨出评估中各具优势且互为补充，二者融合有助于提升诊断的准确性与均衡性；融合模型诊断准确率显著优于二维模型，在智能化诊断与分度评估中展现出应用潜力。

[关键词] 膀胱膨出；深度学习；人工智能；超声检查；图像融合；盆腔器官脱垂；多模态成像；神经网络

Deep learning models based on fused ultrasound images for the assessment of cystocele in women

RAN Shiyi[1], LU Rong[2], LI Muchen[1], QU Can[1]

(1. Department of Ultrasound Imaging, Xiangya Hospital, Central South University, Changsha 410008;
2. Department of Gynecology, Xiangya Hospital, Central South University, Changsha 410008, China)

收稿日期(Date of reception)：2026-03-19

第一作者(First author)：冉诗懿，Email: rsydoc0011@163.com, ORCID: 0009-0006-8327-1747

通信作者(Corresponding author)：鲁蓉，副教授，Email: 1752020522@qq.com, ORCID: 0000-0002-6454-386X

基金项目(Foundation item)：湖南省自然科学基金(2023JJ30920)。This work was supported by the Natural Science Foundation of Hunan Province, China (2023JJ30920).

开放获取(Open access)：本文遵循知识共享许可协议，允许第三方用户按照署名-非商业性使用-禁止演绎4.0(CC BY-NC-ND 4.0)的方式，在任何媒介以任何形式复制、传播本作品(https://creativecommons.org/licenses/by-nc-nd/4.0/)。

©Journal of Central South University (Medical Science). All rights reserved.

ABSTRACT

Objective: Cystocele is a common type of pelvic organ prolapse in women. Ultrasound is the preferred imaging modality for the diagnosis of cystocele; however, the examination procedure is relatively complicated and dependent on the operator's experience, resulting in subjective interobserver differences in diagnosis. This study aims to develop deep learning models based on two-dimensional (2D) ultrasound images, three-dimensional (3D) ultrasound images, and fused 2D and 3D ultrasound images for the diagnosis and ultrasonographic grading of cystocele in women, and to compare the diagnostic performance of the 3 models.

Methods: Data were retrospectively collected from women who underwent transperineal pelvic floor ultrasonography at Xiangya Hospital, Central South University, between January 2022 and December 2024. A total of 625 patients met the inclusion and exclusion criteria. According to the clinical and ultrasonographic diagnoses, 467 patients with cystocele, including 167 with mild cystocele and 300 with significant cystocele, and 158 patients without cystocele were ultimately included. A GE Voluson E10 ultrasound system was used to acquire paired 2D midsagittal images of the bladder at rest and during maximal Valsalva maneuver, as well as reconstructed images of the levator hiatus in 3D volume-rendering mode during maximal Valsalva maneuver. All data were randomly divided at a ratio of 7∶3 into a training set comprising 437 patients and a test set comprising 188 patients. The 2D and 3D image information was integrated through channel stacking to create a fused image dataset. Three models were constructed using ResNet50 as the backbone network: 1) a 2D image model using the original dual-screen 2D images as input; 2) a 3D image model using reconstructed 3D images of the levator hiatus as input; and 3) a fused image model using fused 2D and 3D images as input. The diagnostic performance of each model was evaluated in the test set using accuracy, precision, recall, F1 score, and the area under the receiver operating characteristic curve (AUC). Differences in accuracy among the models were compared using the McNemar test.

Results: The 2D image model achieved an overall accuracy of 75.5%, with macro-averaged precision, recall, F1 score, and AUC values of 69.4%, 69.8%, 69.2%, and 0.878, respectively. The 3D image model achieved an overall accuracy of 76.1%, with macro-averaged precision, recall, F1 score, and AUC values of 78.7%, 71.7%, 71.3%, and 0.908, respectively. The fused image model achieved an overall accuracy of 82.3%, with macro-averaged precision, recall, F1 score, and AUC values of 80.5%, 78.7%, 79.3%, and 0.930, respectively. Its F1 scores for the 3 categories of no cystocele, mild cystocele, and significant cystocele were 80.4%, 68.4%, and 89.0%, respectively. The accuracy of the fused image model was significantly higher than that of the 2D image model (82.3% vs 75.5%, P=0.031). However, the difference between the fused image model and the 3D image model was not statistically significant (82.3% vs 76.1%, P=0.082).

Conclusion: Deep learning models based on ultrasound images enabled automated diagnosis and grading of cystocele. The 2D and 3D ultrasound images each have distinct and complementary advantages in the assessment of cystocele, and their fusion may improve diagnostic accuracy and balanced performance across categories. The fused image model demonstrated significantly higher diagnostic accuracy than the 2D image model and showed potential for automated diagnosis and grading of cystocele.

©Journal of Central South University (Medical Science). All rights reserved.

KEY WORDS cystocele; deep learning; artificial intelligence; ultrasonography; image fusion; pelvic organ prolapse; multimodal imaging; neural networks

女性膀胱膨出是盆腔器官脱垂(pelvic organ prolapse，POP)中最常见的类型[1-2]，约占症状型POP的87%[3]，主要表现为阴道前壁脱垂并伴有膀胱壁脱垂。患者常伴阴道坠胀感、尿失禁及排尿困难等症状，严重影响患者生活质量与心理健康。

盆腔器官脱垂定量分度法(Pelvic Organ Prolapse Quantification，POP-Q)是当前国际应用最广泛的POP标准化评估体系，也是膀胱膨出临床评估中公认的权威分度工具[4]。但该体系无法明确具体膨出物，且操作复杂，受医师主观性影响，诊断结果可能出现假阳性或假阴性。影像学检查为POP诊断的重要补充。超声因具有实时、可重复操作、无创无辐射等优势，被国际泌尿妇科学会推荐为POP的首选影像学检查方法[5]。二维超声能直接观察盆底结构并进行测量，了解盆腔脏器在不同状态下的变化；而三维容积渲染成像模式能重建二维难以显示的冠状面及盆底整体结构，获取如肛提肌裂孔(levator hiatus，LH)等关键结构。二维与三维图像结合分析有助于全面评估盆底解剖异常。然而，现有超声评估均依赖医师手动勾画与测量，过程烦琐、耗时且主观性强，对操作者经验要求高，极大地限制了该技术在基层医疗中的普及应用。因此，亟待通过一套智能化的超声分析模型或软件来提高临床对女性膀胱膨出筛查的效率、准确性及规范性，从而提高诊治水平、改善预后。

人工智能(artificial intelligence，AI)在医学影像分析中的迅速发展为女性膀胱膨出的智能诊断提供了机遇。作为AI的前沿分支，深度学习(deep learning，DL)以多层神经网络为基础，能够自动从大量数据中学习具有判别力的层次化特征，具有强大的特征提取与模式识别能力[6]。在医学影像分析中，DL的应用显著提升了病灶检测、区域分割和疾病分类等任务的性能[7-9]。在POP领域，DL已逐步应用于多种诊断任务，例如前盆腔脱垂的自动识别[10]、POP严重程度的智能化判断[11]及POP动态视频分析与参数自动测量[12]等。然而，现有POP研究多聚焦单一成像平面，未能充分利用二维与三维超声的图像互补信息。图像融合技术作为医学影像分析的重要手段，已在多种临床场景中展现出重要价值。该技术通过整合不同来源或视角的影像，可构建信息更丰富的输入数据，从而增强模型的特征提取能力与分类性能。例如，有研究[13]采用磁共振3D-SPACE序列结合图像融合技术，显著提升了垂体大腺瘤术前评估中对颅底血管神经侵袭程度的显示效果。在DL模型的构建中，图像融合有助于增强模型的特征提取能力与分类性能，进一步提升诊断准确率与鲁棒性。然而，迄今为止，图像融合技术在女性膀胱膨出中的研究甚少。基于此，本研究拟将AI引入女性膀胱膨出的研究中，用超声图像融合技术，以残差网络(residual network，ResNet)这一常见的DL网络为基础架构，通过算法自动提取并融合二维正中矢状切面图与三维容积渲染模式轴平面中的关键解剖特征，构建评估女性膀胱膨出的超声分度模型，该模型旨在模拟超声医师综合多平面图像进行诊断的思维过程，为临床提供高效、智能的超声诊断工具，以期提升诊断效率与准确性。

1 对象与方法

1.1 伦理声明

本研究已获得中南大学湘雅医院伦理委员会批准(审批号：2025060937)。本研究为回顾性分析，患者知情同意豁免。

1.2 对象

回顾性收集2022年1月至2024年12月在中南大学湘雅医院接受盆底超声检查的女性患者的临床与影像资料。共2 464例患者被初步筛查，最终625例患者纳入本研究，根据临床诊断和超声检查结果，其中467例诊断为膀胱膨出，158例诊断为无膀胱膨出且盆底结构正常。纳入标准：1)具有完整的病史记录及系统完善的体格检查结果；2)能有效执行最大Valsalva动作(即Valsalva动作时盆腔器官向背尾侧移动、LH面积增大，且持续时间≥6 s)；3)具有质量清晰的二维及三维超声图像。排除标准：1)合并中腔室及后腔室脱垂；2)有盆腔手术史；3)存在盆腔包块(最大径≥5 cm)、泌尿生殖系统畸形及急性下尿路或生殖道感染[5]。

1.3 膀胱膨出诊断标准

膀胱膨出POP-Q诊断标准：阴道前壁最低点(Ba

©Journal of Central South University (Medical Science). All rights reserved.

点)达到或超过处女膜缘水平，脱垂分度参照国际标准(0~IV度)，其中POP-Q II度及以上作为区分前盆腔脱垂严重程度的关键临床分界点[5]。膀胱膨出超声诊断标准：最大Valsalva动作时，膀胱最低点达参考线及以下。参考线指经耻骨联合后下缘的水平线。超声分度参考Dietz等[14]基于高加索人群的研究及Wu等[15]基于亚洲人群的研究，膀胱最低点下降至参考线以下10 mm是与脱垂症状显著相关的临界值，对应POP-Q II度及以上。基于此，本研究将膀胱膨出分为：1)轻度膨出，膀胱最低点位于参考线水平/1~<10 mm；2)明显膨出，膀胱最低点位于参考线以下≥10 mm。

1.4 图像获取

采用GE Voluson E10超声诊断仪，配备RM6C型腹部容积探头，频率4~8 MHz，二维扫描角度90°，容积获得角度85°，配备超声测量分析软件。受检者适度充盈膀胱后取屈膝仰卧位，检查者将探头置于受检者会阴正中处，指示点朝向其腹侧，声束方向与人体矢状面平行，获取盆底标准二维正中矢状面图像。随后检查者选择双屏成像模式，分别采集静息状态与最大Valsalva状态时的膀胱、尿道位置的对比图像(图1，记为$I_{contrast}$)，保存于超声影像工作站中。随后恢复至正中矢状切面，启动三维容积渲染模式，调整容积框宽度为0.5~2.0 cm，激活正中矢状切面，旋转Z轴，使取样框位于耻骨联合后下缘表面到耻骨直肠肌环中心前面，连续获得LH三维重建平面，嘱患者做Valsalva动作，并采集最大Valsalva动作时的LH平面(图2)。最后，将二维图像和容积渲染模式图像经影像存档与通信系统以JPG格式导出。超声图像采集者为2位具有5年以上盆底超声诊断经验的高年资超声医师，在未知临床诊断的情况下，按照国际泌尿妇科学会的实践规范[5]进行盆底超声扫查。为确保Valsalva动作的有效性与一致性，所有受检者均在检查前接受统一动作指导：嘱受检者深吸气后屏气，并持续向下用力，如排便动作。检查者通过实时超声图像监测盆底器官向背尾侧移动及LH扩张情况，确认达到最大脱垂状态且持续时间≥6 s时采集图像。对于动作不达标的受检者，休息2 min后重新指导，重复上述步骤直至获取满意图像。

所有图像均由上述2位高年资超声医师根据标准进行诊断与分度。如遇分歧，2位医师通过讨论达成一致。随机抽取50例患者由2位医师进行评估，Cohen’s kappa为0.84，表明一致性极佳。最终评估结果作为模型训练的真实标签。

1.5 图像融合与数据集构建

1.5.1 数据集划分

将625例数据按7∶3比例随机分为训练集和测试集。

1.5.2 二维图像数据集

将原始的双屏对比图像($I_{contrast}$)直接作为二维图像模型的输入。

1.5.3 三维图像数据集

将容积渲染模式下形成的LH三维重建图像(记作I_{render})，作为三维图像模型的输入。

1.5.4 融合图像数据集

为构建基于融合图像的DL模型，本研究开发了一套基于像素级通道堆叠的图像融合流程，以整合二维正中矢状切面(静息与最大Valsalva状态)与容积渲染模式下LH三维重建平面的信息。从$I_{contrast}$中，沿垂直中线进行分割，分别获得独立的静息状态(I_{rest})与最大Valsalva状态($I_{valsalva}$)的二维正中矢状面图像。使用双三次插值算法，将尺寸较小的I_{rest}和$I_{valsalva}$放大至与I_{render}相同尺寸。最后，将所有图像统一转换为灰度图像，保证输入一致性。采用通道堆叠的融合策略，将三幅单通道图像在通道维度进行堆叠，形成一张三通道图像，其中通道1为I_{rest}，通道2为$I_{valsalva}$，通道3为I_{render}。最终以三维数组(224×224×3)的形式保存为Numpy二进制文件(npy格式)。

1.6 数据预处理与增强

在所有模型训练阶段，对训练集数据采用增强策略，包括随机翻转、颜色调整、高斯模糊等增强方法。对增强后的训练集与未增强的测试集进行归一化处理，将所有图像的像素值从[0，255]区间线性缩放至[0，1]。计算公式为：

$$I_{norm} = \frac{I}{255.0}$$

其中，I对于训练集为增强后的图像，对于测试集为原始图像。此操作确保所有输入模型的数值分布一致。

1.7 模型构建

为系统探索DL模型在女性膀胱膨出严重程度评估中的价值，本研究构建了3个模型。1)二维图像模型：该模型以$I_{contrast}$作为唯一输入，采用ResNet50作为主干网络。通过全局平均池化聚合深度特征后，连接一个输出维度为3的全连接层，用于输出图像无膀胱膨出、轻度膨出或明显膨出的概率分布。该模型作为评估二维图像在严重程度分级任务上性能的基准，其优化目标为最小化三分类交叉熵损失函数。2)三维图像模型：该模型以I_{render}作为唯一输入，同样

©Journal of Central South University (Medical Science). All rights reserved.

采用ResNet50作为主干网络，通过全局平均池化聚合深度特征，并连接一个输出维度为3的全连接层进行严重程度分级。该模型旨在评估单独使用三维渲染图像对膀胱膨出分度的能力，优化目标为最小化三分类交叉熵损失函数。3)融合图像模型：该模型以经通道堆叠融合后的I_{fused}作为输入，采用相同的ResNet50架构进行特征编码和分类，连接一个输出维度为3的全连接层。该模型旨在探讨融合二维与三维超声信息后，能否进一步提升其对女性膀胱膨出严重程度的分度诊断性能，优化目标为最小化三分类交叉熵损失函数。

1.8 模型训练与评估

1.8.1 训练设置

所有模型均在PyTorch框架下进行训练。训练参数设置为：使用Adam优化器，初始学习率为0.000 1，权重衰减为0.05，训练轮次为100轮，批次大小为32。二维图像模型与融合图像模型均使用三分类交叉熵损失函数。

1.8.2 评估指标

所有模型的诊断性能均通过在测试集上的表现进行评估。统一报告包括总体准确率及各类别与宏平均的精确率、召回率、F1分数和受试者操作特征曲线下面积(area under the curve，AUC)在内的评估指标。

1.9 统计学处理

采用Python 3.10软件进行数据处理。正态分布的计量资料以均数±标准差表示，组间比较采用独立样本t检验；非正态分布的计量资料以中位数(第1四分位数，第3四分位数)表示，组间比较采用Mann-Whitney U检验。计数资料以频数(构成比)表示，组间比较采用χ^2检验。两两模型间准确率比较采用McNemar检验。P<0.05为差异有统计学意义。

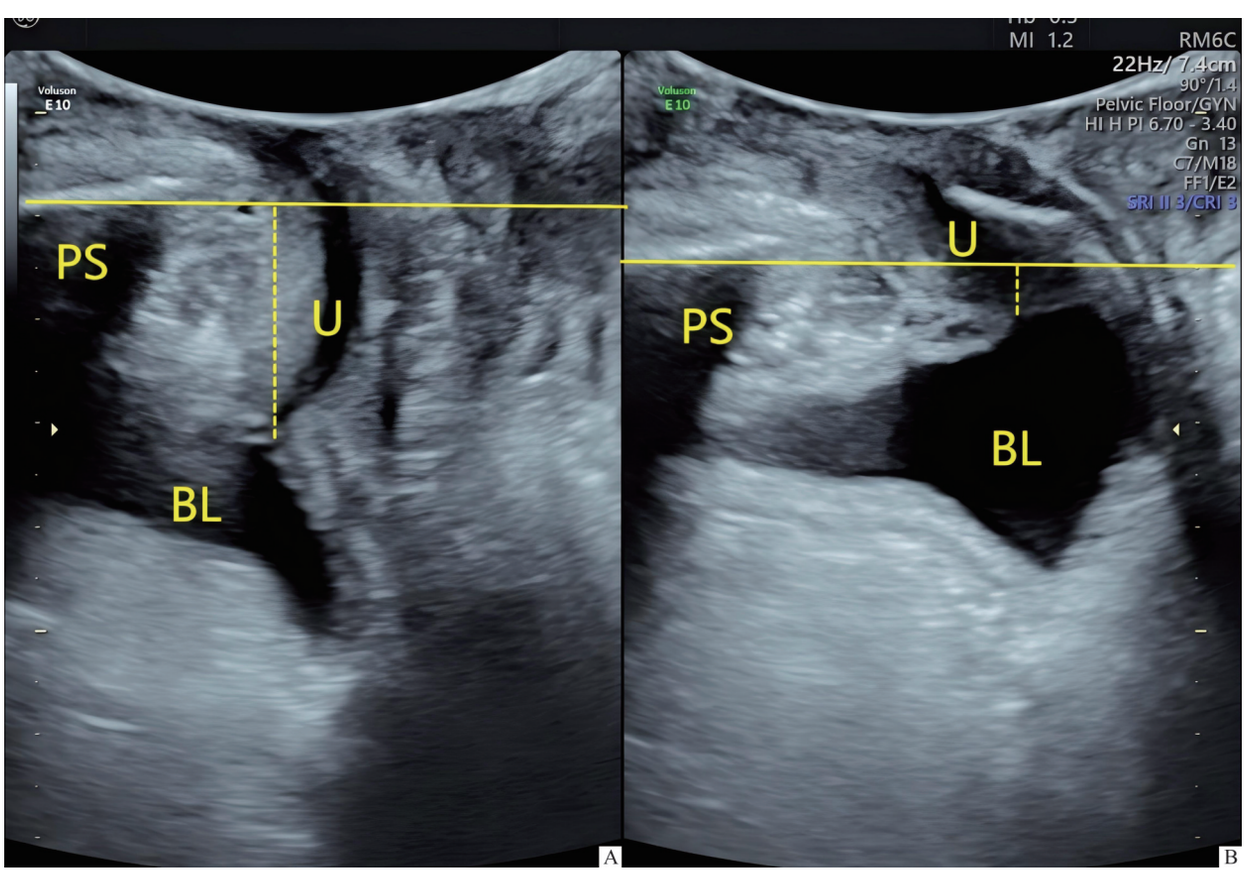


图1 静息状态与最大Valsalva状态下，膀胱、尿道位置的对比图像

Figure 1 Split-screen display images of the bladder and urethra at rest and during maximum Valsalva maneuver

A: At rest, the bladder neck is located above the reference line. B: In the same subject during maximum Valsalva maneuver, the bladder neck remains above the reference line. The solid yellow line is the reference line, and the dashed yellow line indicates the distance from the lowest point of the bladder to the reference line. The reference line is the horizontal line passing through the posterior inferior margin of the pubic symphysis. PS: Pubic symphysis; BL: Bladder; U: Urethra.

©Journal of Central South University (Medical Science). All rights reserved.

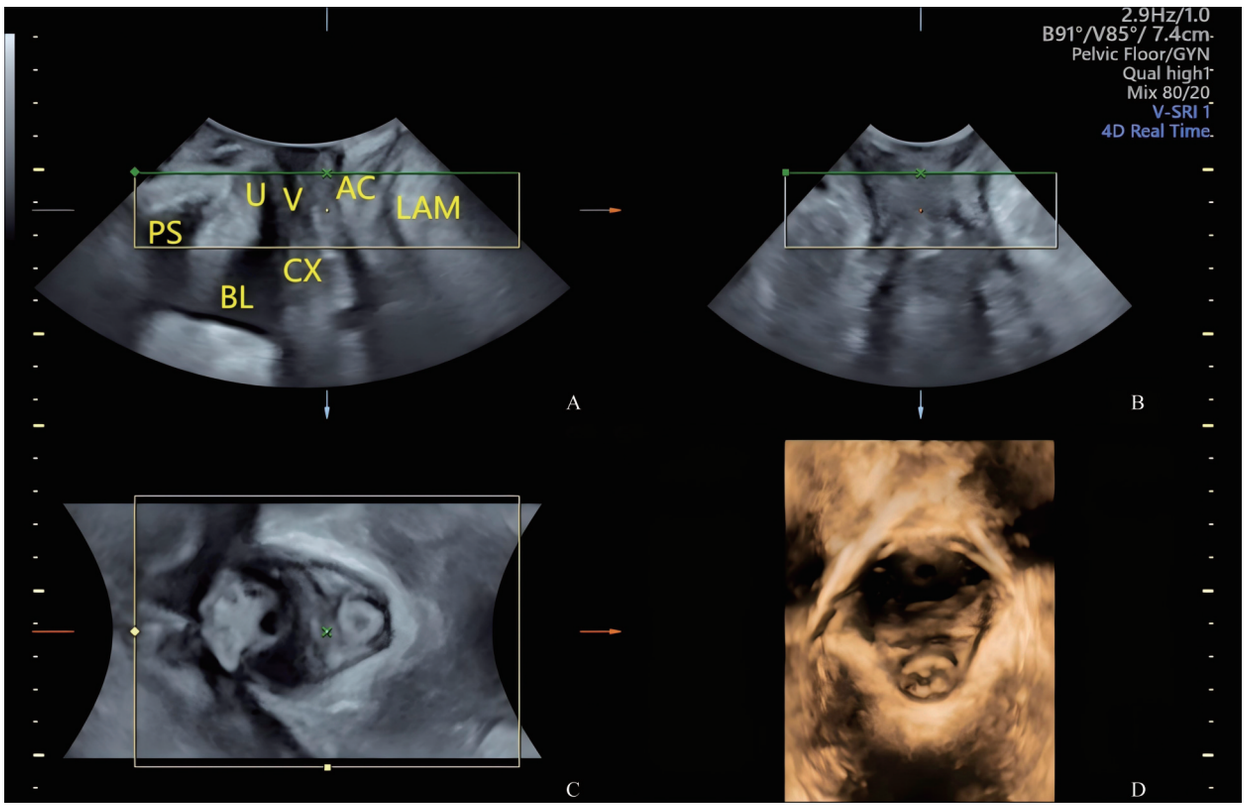


图2　最大Valsalva状态下，三维容积渲染成像模式，盆底冠状面三维重建图像

Figure 2　Three-dimensional pelvic floor coronal reconstructed image under 3D volume-rendering mode during maximum Valsalva maneuver

A: Midsagittal plane, showing the pubic symphysis (PS), bladder (BL), urethra (U), vagina (V), cervix (CX), and anorectal junction (AC), serving as the reference for 3D reconstruction. B: Coronal plane, displaying the symmetric structure of the levator ani muscle (LAM). C: Axial plane, also known as the levator hiatus plane, clearly showing the morphology of the levator hiatus, which is the core section for assessing pelvic floor integrity. D: Three-dimensional reconstructed image, rendered using volume rendering to display the overall pelvic floor structure and spatial relationships.

2 结　果

所有研究对象按7∶3比例被随机分为训练集(n=437)和测试集(n=188)，训练集与测试集在年龄、体重指数 (body mass index，BMI)、产次、阴道分娩史、绝经情况等方面差异均无统计学意义(均P>0.05，表1)。

二维图像模型以静息与最大Valsalva状态的双屏对比图像为输入，总体准确率为75.5%(表2)。该模型对膀胱明显膨出的识别能力最佳，召回率达89.7%，F1分数为88.8%，AUC为0.941；无膀胱膨出的F1分数为72.5%，精确率与召回率尚可；轻度膨出的F1分数仅为46.2%，精确率与召回率均较低，AUC为0.763，提示模型对轻度膨出的判别能力明显不足。

三维图像模型以容积渲染模式下LH三维重建图像作为输入，总体准确率为76.1%(表3)。与二维模型相比，该模型对轻度膨出的召回率升至79.6%，F1分数升至61.4%，检出能力明显改善；但对无膀胱膨出的召回率降至46.8%，虽精确率高达95.7%，F1分数仍仅为62.9%，表明模型对无膀胱膨出的检出能力欠佳，易将无膀胱膨出误判为有膀胱膨出。对于明显膨出患者，三维模型的召回率为88.7%，F1分数为89.6%，与二维模型(88.8%)基本持平。以上结果表明，三维平面信息有效增强了轻度膨出的识别能力，但牺牲了无膨出患者的召回率，对明显膨出患者的敏感性并未减弱。

融合图像模型整合二维正中矢状切面与三维LH

©Journal of Central South University (Medical Science). All rights reserved.

平面重建信息，总体准确率达82.3%(表4)，较二维图像模型的总体准确率(75.5%)提升6.8个百分点，差异具有统计学意义(χ^2=4.645，P=0.031)；较三维图像模型的总体准确率(76.1%)提升6.2个百分点，但差异无统计学意义(χ^2=3.025，P=0.082)。

融合模型在各类别上实现了性能互补与平衡。对于无膀胱膨出的患者，融合模型的F1分数为80.4%，在3个模型中最高；精确率78.0%虽低于三维模型的95.7%，但召回率达83.0%，较大幅度弥补了三维模型召回率偏低(仅46.8%)的缺陷，综合平衡性最佳。对于轻度膨出的患者，融合模型的F1分数为68.4%，高于二维模型的46.2%和三维模型的61.4%；其精确率77.1%优于三维模型的50.0%，有效减少了轻度膨出的误诊；召回率61.4%虽略低于三维模型，但整体分类效能更稳定。对于明显膨出的患者，融合模型的F1分数为89.0%，召回率高达91.8%，在3个模型中均居首位。各模型区分三类别的受试者操作特征曲线见图3~5。

表1 训练集与测试集的基本特征比较

Table 1 Comparison of baseline characteristics between the training set and the test set

组别	n	年龄/岁	BMI/(kg·m^{-2})	产次/[例(%)]			阴道分娩史/[例(%)]		绝经/[例(%)]	
				0~1	2~3	≥4	无	有	否	是
训练集	437	51.0(45.0, 57.0)	22.10(19.90, 23.94)	170 (38.90)	230 (52.63)	37(8.47)	37(8.47)	400 (91.53)	199(45.54)	238 (54.46)
测试集	188	52.0(46.0, 59.0)	22.10(20.00, 24.02)	67(35.64)	109 (57.98)	12(6.38)	24(12.77)	164 (87.23)	82(43.62)	106 (56.38)
统计量		Z=−0.765	Z=−0.111		χ^2=1.790		χ^2=2.292		χ^2=0.126	
P		0.444	0.912		0.409		0.130		0.723	

非正态分布的计量资料以中位数(第1四分位数，第3四分位数)表示，组间比较采用Mann-Whitney U检验。计数资料以频数(构成比)表示，组间比较采用χ^2检验。BMI：体重指数。

表2 二维图像模型在测试集上的诊断性能

Table 2 Diagnostic performance of the two-dimensional image model on the test set

类别	精确率/%	召回率/%	F1分数/%	AUC	总体准确率/%
无膀胱膨出	67.3	78.7	72.5	0.930	
膀胱轻度膨出	52.9	40.9	46.2	0.763	
膀胱明显膨出	87.9	89.7	88.8	0.941	
宏平均	69.4	69.8	69.2	0.878	75.5

AUC：曲线下面积。

表3 三维图像模型在测试集上的诊断性能

Table 3 Diagnostic performance of the three-dimensional image model on the test set

类别	精确率/%	召回率/%	F1分数/%	AUC	总体准确率/%
无膀胱膨出	95.7	46.8	62.9	0.932	
膀胱轻度膨出	50.0	79.6	61.4	0.850	
膀胱明显膨出	90.5	88.7	89.6	0.942	
宏平均	78.7	71.7	71.3	0.908	76.1

©Journal of Central South University (Medical Science). All rights reserved.

表4　融合图像模型在测试集上的诊断性能

Table 4　Diagnostic performance of the fused image model on the test set

类别	精确率/%	召回率/%	F1分数/%	AUC	总体准确率/%
无膀胱膨出	78.0	83.0	80.4	0.964	
膀胱轻度膨出	77.1	61.4	68.4	0.883	
膀胱明显膨出	86.4	91.8	89.0	0.944	
宏平均	80.5	78.7	79.3	0.930	82.3

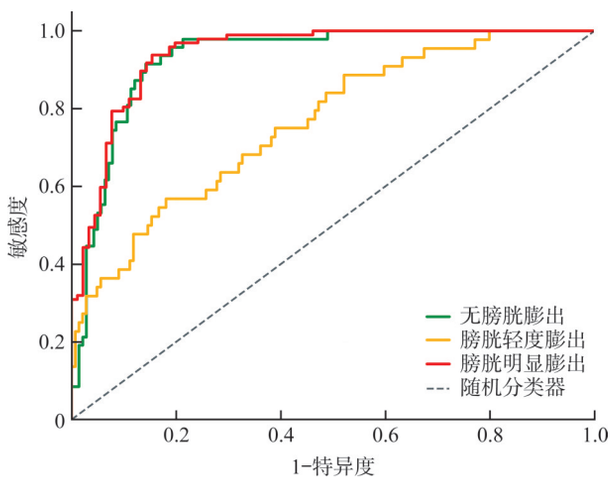


图3　二维图像模型在测试集区分无膀胱膨出、膀胱轻度膨出与膀胱明显膨出的ROC曲线

Figure 3　ROC curves of the two-dimensional image model on the test set for distinguishing no, mild, and obvious bladder prolapse

ROC: Receiver operating characteristic.

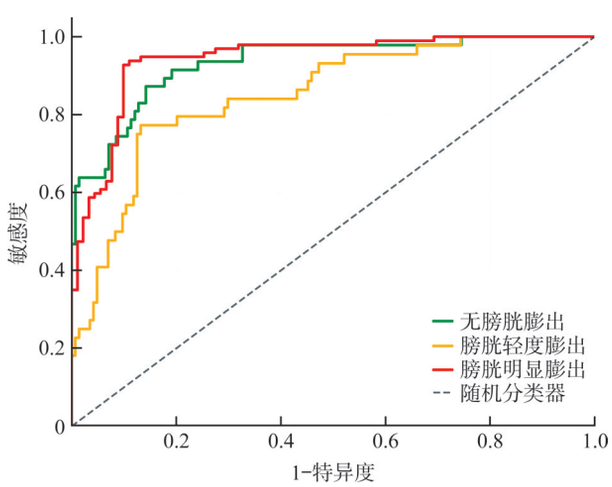


图4　三维图像模型在测试集区分无膀胱膨出、膀胱轻度膨出与膀胱明显膨出的ROC曲线

Figure 4　ROC curves of the three-dimensional image model on the test set for distinguishing no, mild, and obvious bladder prolapse

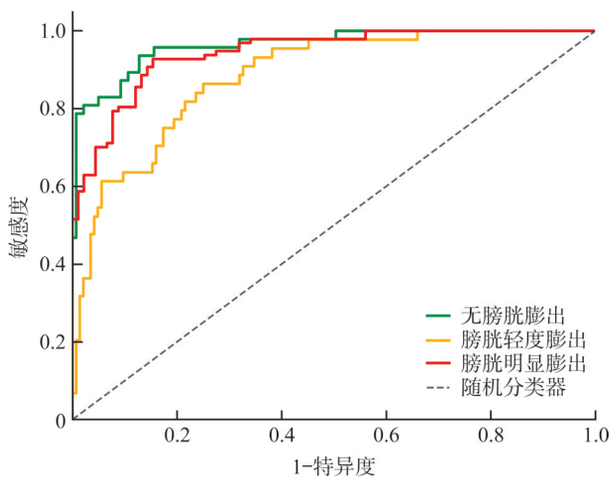


图5　融合图像模型在测试集区分无膀胱膨出、膀胱轻度膨出与膀胱明显膨出的ROC曲线

Figure 5　ROC curves of the fused image model on the test set for distinguishing no, mild, and obvious bladder prolapse

3 讨　论

超声是评估女性膀胱膨出的首选影像学方法，但其诊断过程依赖医师对多个平面图像的综合判读及手动测量，存在效率低和主观性强的问题。利用AI辅助诊断是提升诊断效率与一致性的有效方法。然而，现有关POP的AI研究多集中于单一成像模式[10-11, 16-17]，尚未探讨整合不同成像模式中的信息对诊断性能的提升作用。

二维正中矢状切面是诊断膀胱膨出的基础切面，可直接显示膀胱最低点与参考线的相对位置，是超声分度的核心依据。本研究中，二维图像模型总体准确率为75.5%，对膀胱明显膨出的识别表现出色，召回率达89.7%，表明DL模型能够有效捕捉膀胱位置的显著下移。然而，该模型对轻度膨出的F1分数仅46.2%，精确率与召回率均不足，提示其区分无膀胱膨出与轻度膨出的能力不足。这一现象可能与以

©Journal of Central South University (Medical Science). All rights reserved.

下因素有关：对于膀胱下降距离较小的轻度患者，其超声表现与正常者的差异细微，膀胱最低点移动幅度小，模型可能难以在双屏对比图像中准确定位膀胱最低点所在帧或精确识别参考线，导致分类边界模糊。

三维容积渲染模式下的LH重建图像提供了不同于二维切面的盆底支持结构信息。妊娠、分娩导致的肛提肌损伤是POP的重要诱因，静息及Valsalva状态下LH面积的测量可作为评估脱垂风险及严重程度的重要指标[18-19]。本研究中，三维图像模型总体准确率为76.1%，与二维模型的75.5%相近。在轻度膨出类别上，三维模型的召回率达79.6%，优于二维模型的40.9%，F1分数也由二维模型的46.2%提升至61.4%。这一差异可能与2类图像所承载的诊断信息类型不同有关：二维正中矢状面提供的是膀胱最低点相对于参考线的距离信息，在膀胱下降幅度较小时，该距离的组间重叠较大，判别边界模糊；三维LH平面则直接呈现盆底支持结构的形态特征，DL模型可能从中提取到与早期盆底支持结构改变相关的像素级信息，这在病变程度较轻时或具有额外的判别价值。然而，三维模型对无膀胱膨出的召回率降至46.8%，精确率虽高达95.7%，但整体平衡性欠佳。对于明显膨出的患者，三维模型的F1分数为89.6%，与二维模型的88.8%接近，表明二者在明显膨出患者的总体识别效能上相当。结合二维模型在该类别上的高召回率，可以看出二维正中矢状面在捕捉膀胱位置显著下移方面具有直接优势，三维平面信息对明显膨出的判断能力并未超越二维图像。

基于上述2类图像的互补特性，本研究采用通道堆叠策略融合二维与三维信息，构建的融合模型总体准确率达82.3%，宏平均的F1分数达79.3%，均高于2个单一模态模型。与二维模型相比，融合模型的准确率提升6.8个百分点，差异具有统计学意义(P=0.031)，验证了整合多平面信息的有效性。与三维模型相比，融合模型的准确率提升6.2个百分点，但差异无统计学意义(P=0.082)，提示在现有样本量下，融合策略相较于三维图像的单独使用，其额外增益尚需更大规模样本验证。从分类别指标来看，融合模型在3个类别上均实现了F1分数的提升。在无膀胱膨出类别中，融合模型的F1分数为80.4%，召回率达83.0%，大幅弥补了三维模型召回偏低的缺陷，同时精确率亦保持在可接受水平，实现了二者的较好平衡。在轻度膨出类别中，融合模型的F1分数为68.4%，精确率达77.1%，较三维模型的50.0%有明显改进，减少了假阳性，使分类效能更加稳定。在明显膨出类别中，融合模型的召回率达91.8%，高于二维模型和三维模型，表明联合2种图像信息有助于最大程度减少重度患者的漏诊。上述结果表明，DL模型能够有效利用二维图像中的器官位置特征与三维图像中的盆底结构特征，实现比单一模态更均衡、稳健的综合判断。

在图像融合策略上，本研究选择了通道堆叠的方式，主要基于以下考虑：通道堆叠属于像素级早期融合策略，已有综述[20]将其归类为多模态DL中最基础的融合范式之一。该方法将二维静息图像、二维Valsalva图像与三维LH平面图像直接叠加为三通道输入，使网络在首个卷积层即可同时访问所有模态的原始像素信息，从输入层开始即自主学习和发现跨通道的联合特征。相较于特征级融合或决策级融合需要在网络深层分别处理各模态后再进行整合，通道堆叠最大限度地保留了各模态原始图像的空间完整性与细粒度信息，避免因模态间特征提取路径分离而造成的信息损耗[21]。与小波变换融合相比，通道堆叠不依赖于手工设计的融合规则，而是将多模态信息的整合完全交由网络优化，避免了人工干预可能引入的偏差，同时简化了预处理流程，更利于图像复现。在网络架构选择上，ResNet50通过跳跃连接有效缓解了深层网络的梯度消失问题[22]，其适中的网络深度在保证特征提取能力的同时降低了过拟合风险，适用于本研究的样本规模。

在POP的AI诊断领域，现有研究多聚焦于单一成像模式。Yang等[10]基于二维超声静态图像，对比了AlexNet、VGG-16、ResNet-18和ResNet-50在前盆腔脱垂自动识别中的性能，其中ResNet-18最优达93.53%，但其研究为二分类任务(即诊断是否有脱垂)，未涉及脱垂程度的评估。Wu等[11]基于ResNet-34开发了POP严重程度自动评估系统，将脱垂分为轻、中和重度，其中前盆腔脱垂的准确率为86.9%，但其输入仍限于单一二维平面，未探讨不同成像模态对诊断性能的贡献。董梦等[23]基于尿道全长动态参数构建支持向量机模型进行膀胱膨出二分类诊断，发现模型对正常前盆腔结构的正确识别率仅63.3%。本研究二维模型无膀胱膨出召回率偏低的结果与该研究趋势一致。

本研究构建的融合模型准确率为82.3%，略低于Yang等[10]与Wu等[11]的研究结果。需要指出的是，本研究目的并非追求单个模型的最高准确率，而是探讨在二维图像基础上融入三维LH平面信息后，模型能否获得比单一模态模型更佳的性能。此外，本研究将无膀胱膨出作为独立类别纳入三分类分度，鉴于无膀胱膨出与轻度膨出的鉴别难度较大，也可能影响总体准确率。尽管如此，本研究仍有其不同之

©Journal of Central South University (Medical Science). All rights reserved.

处：第一，在输入模态上，本研究尝试将二维正中矢状切面与三维LH重建平面图像进行像素级通道堆叠融合，用于膀胱膨出的分度；第二，在研究设计上，通过系统对比二维模型、三维模型和融合模型三者的诊断效能，初步揭示了不同成像平面的互补特性——二维图像直接显示膀胱最低点位置变化，三维图像呈现LH等盆底支持结构的形态，二者融合有助于提升分类的均衡性。本研究结果表明，多平面信息联合在POP的超声智能化分度中具有一定的探索价值，可为后续研究提供参考。

本研究存在以下局限性：1)样本量有限且为单中心回顾性研究，融合模型与三维模型的差异无统计学意义，可能受样本量限制，模型泛化能力需通过多中心前瞻性研究进一步验证。2)本研究仅纳入了膀胱膨出患者，模型在复杂多腔室脱垂中的适用性尚不明确。3)基于静态图像的分析可能丢失Valsalva动作过程中的动态变化信息。4)DL模型的“黑箱”特性导致其决策过程缺乏可解释性，难以被临床医师完全信任。5)本研究以超声图像上的人工测量与判读结果作为金标准，未纳入POP-Q临床分度或患者症状等临床指标，结论的临床适用性有待进一步验证。

综上所述，本研究将二维正中矢状切面与三维LH重建平面图像进行像素级通道堆叠融合，用于女性膀胱膨出的超声分度评估，通过系统对比二维、三维及融合模型的诊断性能，发现不同成像平面具有互补性：二维图像直接反映膀胱最低点位置变化，三维图像则呈现盆底支持结构的形态信息，二者融合可提升多分类的均衡性。尽管本研究存在单中心小样本、仅纳入膀胱膨出患者(未纳入复杂多腔室脱垂患者)及缺乏动态信息等不足，但验证了多平面信息联合在POP超声智能分度中的可行性与潜在价值。未来需结合动态视频、临床危险因素及可解释性方法，并通过多中心前瞻性研究进一步验证模型的泛化能力。

作者贡献声明：冉诗懿　实验构思与设计，研究实施，数据采集与分析，论文撰写；鲁蓉　实验构思与设计，图像采集，论文审阅与修订；李沐宸、屈灿　研究实施。所有作者阅读并同意最终的文本。

利益冲突声明：作者声称无任何利益冲突。

参考文献

[1] Yin W, Ma Q, Xie W, et al. Three-dimensional ultrasound assessment of risk factors for cystocele and Green classification in primipara[J]. Front Med, 2022, 9: 979989. https://doi.org/10.3389/fmed.2022.979989.

[2] Weintraub AY, Glinter H, Marcus-Braun N. Narrative review of the epidemiology, diagnosis and pathophysiology of pelvic organ prolapse[J]. Int Braz J Urol, 2020, 46(1): 5-14. https://doi.org/10.1590/S1677-5538.IBJU.2018.0581.

[3] Pang H, Zhang L, Han S, et al. A nationwide population-based survey on the prevalence and risk factors of symptomatic pelvic organ prolapse in adult women in China—a pelvic organ prolapse quantification system-based study[J]. BJOG, 2021, 128(8): 1313-1323. https://doi.org/10.1111/1471-0528.16675.

[4] Barbier H, Carberry CL, Karjalainen PK, et al. International Urogynecology consultation chapter 2 committee 3: the clinical evaluation of pelvic organ prolapse including investigations into associated morbidity/pelvic floor dysfunction[J]. Int Urogynecol J, 2023, 34(11): 2657-2688. https://doi.org/10.1007/s00192-023-05629-8.

[5] AIUM/IUGA practice parameter for the performance of urogynecological ultrasound examinations: developed in collaboration with the ACR, the AUGS, the AUA, and the SRU [J]. J Ultrasound Med, 2019, 38(4): 851-864. https://doi.org/10.1002/jum.14953.

[6] Chan HP, Samala RK, Hadjiiski LM, et al. Deep learning in medical image analysis[M]//Deep Learning in Medical Image Analysis. Cham: Springer, 2020: 3-21. https://doi.org/10.1007/978-3-030-33128-3_1.

[7] 陈皓中, 刘军, 邓凯, 等. 基于多期相增强CT的肾细胞癌病理分级深度学习算法[J]. 中南大学学报(医学版), 2025, 50(4): 651-663. https://doi.org/10.11817/j.issn.1672-7347.2025.250082.
CHEN Haozhong, LIU Jun, DENG Kai, et al. Deep learning algorithm for pathological grading of renal cell carcinoma based on multi-phase enhanced CT[J]. Journal of Central South University. Medical Science, 2025, 50(4): 651-663. https://doi.org/10.11817/j.issn.1672-7347.2025.250082.

[8] 申珊珊, 李春泉, 范耀华, 等. 区分胶质母细胞瘤和单发性脑转移瘤的多模态融合深度学习模型的开发和验证[J]. 中南大学学报(医学版), 2024, 49(1): 58-67. https://doi.org/10.11817/j.issn.1672-7347.2024.230248.
SHEN Shanshan, LI Chunquan, FAN Yaohua, et al. Development and validation of a multi-modality fusion deep learning model for differentiating glioblastoma from solitary brain metastases[J]. Journal of Central South University. Medical Science, 2024, 49(1): 58-67. https://doi.org/10.11817/j.issn.1672-7347.2024.230248.

[9] 兰梓涵, 彭玉兰. 基于乳腺超声图像的人工智能诊断[J]. 中南大学学报(医学版), 2022, 47(8): 1009-1015. https://doi.org/10.11817/j.issn.1672-7347.2022.220110.
LAN Zihan, PENG Yulan. Artificial intelligence diagnosis based on breast ultrasound imaging[J]. Journal of Central South University. Medical Science, 2022, 47(8): 1009-1015. https://doi.org/10.11817/j.issn.1672-7347.2022.220110.

[10] Yang F, Hu R, Wu H, et al. Combining pelvic floor ultrasonography with deep learning to diagnose anterior compartment organ prolapse[J]. Quant Imaging Med Surg,

©Journal of Central South University (Medical Science). All rights reserved.

2025, 15(2): 1265-1274. https://doi.org/10.21037/qims-24-772.

[11] Wu S, Wu J, Xu Y, et al. Automated evaluation of female pelvic organ descent on transperineal ultrasound: model development and validation[J]. Int Urogynecol J, 2025, 36(8): 1693-1700. https://doi.org/10.1007/s00192-025-06211-0.

[12] Wei N, Li X, Lu X, et al. Study on the consistency between automatic measurement based on convolutional neural network technology and manual visual evaluation in intracavitary ultrasonic cine of anterior pelvic[J]. J Ultrasound Med, 2024, 43(4): 671-681. https://doi.org/10.1002/jum.16392.

[13] 聂吉林, 孟莉, 李臻琰, 等. 磁共振3D-SPACE序列结合图像融合技术在垂体大腺瘤术前评估中的应用价值[J]. 中南大学学报(医学版), 2020, 45(8): 980-987. https://doi.org/10.11817/j.issn.1672-7347.2020.200438.
NIE Jilin, MENG Li, LI Zhenyan, et al. Application of magnetic resonance 3D-SPACE sequence combined image fusion technique in preoperative evaluation of pituitary macroadenoma[J]. Journal of Central South University. Medical Science, 2020, 45(8): 980-987. https://doi.org/10.11817/j.issn.1672-7347.2020.200438.

[14] Dietz HP, Mann KP. What is clinically relevant prolapse? An attempt at defining cutoffs for the clinical assessment of pelvic organ descent[J]. Int Urogynecol J, 2014, 25(4): 451-455. https://doi.org/10.1007/s00192-013-2307-4.

[15] Wu M, Wang X, Zheng Z, et al. Ultrasound assessment of bladder descent and its correlation with prolapse severity in Chinese women: a prospective multicenter study[J]. Int Urogynecol J, 2022, 33(10): 2887-2893. https://doi.org/10.1007/s00192-022-05100-0.

[16] García-Mejido JA, Ramos-Vega Z, Fernández-Palacín A, et al. Predictive model for the diagnosis of uterine prolapse based on transperineal ultrasound[J]. Tomography, 2022, 8(4): 1716-1725. https://doi.org/10.3390/tomography8040144.

[17] García-Mejido JA, Galán-Paez J, Solis-Martín D, et al. Ultrasound diagnosis of pelvic organ prolapse using artificial intelligence[J]. J Clin Med, 2025, 14(11): 3634. https://doi.org/10.3390/jcm14113634.

[18] Nishibayashi M, Okagaki R. Ultrasonographic evaluation of pelvic floor structure at antepartum and postpartum periods using three-dimensional transperineal ultrasound[J]. J Med Ultrason (2001), 2021, 48(3): 345-351. https://doi.org/10.1007/s10396-021-01100-7.

[19] 李多, 鲁蓉. 盆底超声评估女性肛提肌形态及功能的研究进展[J]. 中南大学学报(医学版), 2023, 48(8): 1267-1273. https://doi.org/10.11817/j.issn.1672-7347.2023.220577.
LI Duo, LU Rong. Research progress in pelvic floor ultrasound for assessing the morphology and function of levator ani muscle in women[J]. Journal of Central South University. Medical Science, 2023, 48(8): 1267-1273. https://doi.org/10.11817/j.issn.1672-7347.2023.220577.

[20] Teoh JR, Dong J, Zuo X, et al. Advancing healthcare through multimodal data fusion: a comprehensive review of techniques and applications[J/OL]. Peerj Comput Sci, 2024, 10: e2298 [2026-01-21]. https://doi.org/10.7717/peerj-cs.2298.

[21] Dadzie AK, Iddir SP, Abtahi M, et al. Colour fusion effect on deep learning classification of uveal melanoma[J]. Eye, 2024, 38(14): 2781-2787. https://doi.org/10.1038/s41433-024-03148-4.

[22] Xu G, Wang X, Wu X, et al. Development of residual learning in deep neural networks for computer vision: a survey[J]. Eng Appl Artif Intell, 2025, 142: 109890. https://doi.org/10.1016/j.engappai.2024.109890.

[23] 董梦, 王鑫璐, 朱光宇, 等. 人工智能辅助盆底超声诊断膀胱脱膨出的效能[J/OL]. 中华妇幼临床医学杂志(电子版), 2025, 21(3): 296-303[2026-01-22]. https://doi.org/10.3877/cma.j.issn.1673-5250.2025.03.007.
DONG Meng, WANG Xinlu, ZHU Guangyu, et al. Efficacy of artificial intelligence-assisted pelvic floor ultrasound in diagnosing cystocele[J/OL]. Chinese Journal of Obstetrics & Gynecology and Pediatrics. Electronic Edition, 2025, 21(3): 296-303[2026-01-22]. https://doi.org/10.3877/cma.j.issn.1673-5250.2025.03.007.

（责任编辑 彭敏宁）

©Journal of Central South University (Medical Science). All rights reserved.