论著·智能医学与预测模型

一种可显著提升初级超声医师诊断准确性与效率的深度学习模型:基于YOLOv13的腹部超声异常实时识别系统开发与验证

李 苇[1*],李华龙[2*],曾小科[1],何丽君[1],钟飞飞[1],刘 原[1],杨 毅[2],杨 信[1],粘永健[2],徐亚丽[1] [1]陆军军医大学(第三军医大学)第二附属医院超声科,重庆;[2]陆军军医大学(第三军医大学)生物医学工程与影像医学系数字医学教研室,重庆

[摘要] **目的** 开发并验证一种基于腹部超声影像的深度学习模型,用于17类常见腹部异常的实时识别与定位,为基层超声诊断提供可靠技术支撑。**方法** 回顾性收集陆军军医大学第二附属医院超声科2017年1月至2024年9月腹部异常患者18 947例共39 530幅B型超声图像。按异常类别分层随机抽取1 891幅图像构建独立测试集;剩余的37 639幅图像按7∶1∶2比例随机划分为训练集(26 258幅)、验证集(3 840幅)及内部测试集(7 541幅)。经标准化标注后,在内部测试集上对比RetinaNet、Faster R-CNN、RT-DETR及YOLOv13这4种主流目标检测网络的性能,筛选最优模型;评估指标包括交并比(intersection over union, IoU)、精确率(precision)、召回率(recall)、F_1值(F_1)及平均精度均值(mAP@50);在独立测试集上,进一步比较4种诊断模式的诊断效能(最优模型、3名无标注经验的初级超声医师、3名中级超声医师及模型辅助下初级医师)。**结果** 内部测试集上,YOLOv13模型表现最优,mAP@50为0.653(95%*CI*:0.608~0.669),高于RetinaNet(0.546,95%*CI*:0.511~0.563)、Faster R-CNN(0.578,95%*CI*:0.531~0.591)及RT-DETR(0.627,95%*CI*:0.585~0.643);其精确率为0.633(95%*CI*:0.624~0.643),召回率为0.715(95%*CI*:0.704~0.725),F_1值为0.671(95%*CI*:0.663~0.681);独立测试集上,IoU为0.5时,6名医师的F_1值均值为0.57(初级组:0.484,中级组:0.656);YOLOv13模型的F_1值为0.624,显著高于所有初级医师(P<0.001)。模型辅助后,初级医师的F_1值均值从0.484提升至0.644(提升33.1%,P<0.001),接近中级医师水平;平均精确率从0.529提升至0.638(提升20.6%),平均召回率从0.472提升至0.674(提升42.8%),平均诊断耗时从213.33 min缩短至161.33 min(缩短24.4%)。**结论** YOLOv13深度学习模型在17类腹部异常超声检测中表现最优,能显著提升初级超声医师的诊断效能,可作为基层及社区医院腹部超声筛查的辅助工具。

[关键词] 腹部超声;深度学习;检测算法;计算机辅助诊断

[中图法分类号] R195.1;R319;R445.1 [文献标志码] A

[基金项目] 陆军军医大学优秀人才项目(2019R038)

* 同等贡献者

[通信作者] 徐亚丽,E-mail:xuyal1976@tmmu.edu.cn

2097-0927/Copyright © 2026, 陆军军医大学(第三军医大学)。这篇开放获取文章遵循CC BY许可协议(https://creativecommons.org/licenses/by/4.0/)

引用格式:李苇,李华龙,曾小科,等. 一种可显著提升初级超声医师诊断准确性与效率的深度学习模型:基于YOLOv13的腹部超声异常实时识别系统开发与验证[J]. 陆军军医大学学报,2026,48(04):493-503. DOI:10.16016/j.2097-0927.202512098.

Citation: LI W, LI H L, ZENG X K, *et al*. A deep learning-based model for significantly enhancing diagnostic accuracy and efficiency in novice sonographers: development and validation of a YOLOv13-powered real-time abdominal ultrasound anomaly detection system[J]. J Army Med Univ, 2026, 48(04): 493-503. DOI: 10.16016/j.2097-0927.202512098.

A deep learning-based model for significantly enhancing diagnostic accuracy and efficiency in novice sonographers: development and validation of a YOLOv13-powered real-time abdominal ultrasound anomaly detection system

LI Wei[1], LI Hualong[2], ZENG Xiaoke[1], HE Lijun[1], ZHONG Feifei[1], LIU Yuan[1], YANG Yi[2], YANG Xin[1], NIAN Yongjian[2], XU Yali[1] [1]Department of Ultrasonography, Second Affiliated Hospital, Army Medical University (Third Military Medical University), Chongqing; [2]Department of Digital Medicine, College of Biomedical Engineering and Imaging Medicine, Army Medical University (Third Military Medical University), Chongqing, China

[**Abstract**] Objective To develop and validate a deep learning model for real-time identification and localization of 17 common abdominal anomalies on ultrasound images, providing reliable technical support for primary care sonography. Methods Retrospectively collected 39 530 B-mode ultrasound images from 18 947 patients with abdominal anomalies at the Second Affiliated Hospital of Army Medical University from January 2017 to September 2024). A stratified random sample of 1 891 images formed an independent testing set; the remaining 37 639 images were split 7:1:2 into training (26 258), validation (3 840), and internal test (7 541) sets. After standardized annotation, RetinaNet, Faster R-CNN, RT-DETR, and YOLOv13 were compared on the internal test set using IoU, precision, recall, F_1-score, and mAP@50. The optimal model was further evaluated against 6 sonographers (3 novice, 3 intermediate) and novice sonographers assisted by the model on the independent test set. Results YOLOv13 achieved superior performance on the internal testing set: mAP@50 0.653 (95%*CI*: 0.608 to 0.669) versus RetinaNet 0.546 (95%*CI*: 0.511 to 0.563), Faster R-CNN 0.578 (95%*CI*: 0.531 to 0.591), RT-DETR 0.627 (95%*CI*: 0.585 to 0.643); precision 0.633 (95%*CI*: 0.624 to 0.643), recall 0.715 (95%*CI*:0.704 to 0.725), F_1-score 0.671 (95%*CI*:0.663 to 0.681). On the independent testing set (IoU=0.5), YOLOv13's F_1-score (0.624) exceeded all novice sonographers' mean F_1-score (0.484; $P<0.001$). Model assistance improved novice sonographers' mean F1-score from 0.484 to 0.644 (33.1% improvement; $P<0.001$), approaching intermediate-level performance (0.656); mean precision increased 20.6% (0.529 to *o* 0.638), mean recall increased 42.8% (0.472 to 0.674), and mean interpretation time decreased 24.4% (213.33 min to 161.33 min). Conclusion The YOLOv13 model demonstrates optimal performance for detecting 17 abdominal ultrasound anomalies and significantly enhances diagnostic efficiency of novice sonographers, serving as an effective assistive tool for abdominal ultrasound screening in primary care settings.

[**Key words**] abdominal ultrasonography; deep learning; detection algorithms; computer-assisted

LI Wei and LI Hualong contributed equally to this article. Supported by the Excellent Talent Program of Army Medical University (2019R038). Corresponding author: XU Yali, E-mail: xuyal1976@tmmu.edu.cn

超声检查是一种非侵入性、无辐射的实时成像技术,已成为腹部疾病临床筛查的一线影像学手段,在肝、胆、肾等腹部器官病变诊断中发挥着不可替代的作用[1-2]。然而,临床实践中超声检查面临多重现实痛点:其一,超声实时扫描帧率通常可达25 fps以上[3],单次腹部超声扫描可产生数千帧图像,仅少数含明确声学特征的图像帧具备诊断价值,人眼易遗漏病灶;其二,超声图像易受噪声与伪影干扰[4-5],再加上医师高强度工作导致的疲劳,显著增加漏诊、误诊风险[6-7]。因此,实现多种腹部异常的自动实时识别与定位,成为降低医师工作负荷、提升诊疗效率的核心临床需求。

人工智能技术的快速发展为超声影像辅助诊断提供了新路径,已有多项研究证实其在腹部超声病灶识别中具备良好应用前景。YANG等[8]构建了基于ResNet-18网络的分类模型,用于肝脏超声局灶性病灶良恶性的鉴别,其性能优于具备15年临床经验的影像科医师;刘丽等[9]提出的随机森林模型,实现了早期肝肿瘤良恶性的高效区分;GUPTA等[10]研发的深度学习模型可自动识别超声图像中的胆囊癌病灶;SALAM

等[11]构建的混合深度学习框架,对肾积水超声影像的二分类准确率达95%;在脂肪肝超声分类研究中,深度学习模型也表现出可靠性[12-13]。然而,此类研究多聚焦单一疾病诊断,难以适配患者常合并多种腹部异常的实际场景,应用范围受限。尽管部分研究已转向多病灶分类,如实现多种肝局灶性结节、胆道系统病变的识别[14-18],但这类模型仅能输出多分类结果,无法提供病灶的具体定位信息,而定位是超声临床诊断的核心环节,信息缺失严重制约了模型的临床转化。此外,为提升模型可解释性,部分研究者尝试将热力图技术与分类模型相结合[10,14,17-18],通过高亮标注病变可疑区域辅助诊断,但热力图仅能反映影响模型分类决策的图像区域,其高亮位置与病灶真实解剖位置存在偏差,定位精准度不足[19]。因此,现有研究尚缺乏有效解决"多类腹部异常实时识别+精准定位"的核心需求,该领域仍存在明确的研究空白,因此,亟需依托大规模数据集探索适配的目标检测技术,弥补现有研究的不足。本研究通过构建大规模标准化腹部B型超声异常数据集,探究并验证深度学习目标检测模型,旨在实现17类常见腹部异常的实时识别与精确定位,为超声医师提供快速精准的辅助诊断工具。

1　资料与方法

1.1　研究设计

本研究围绕腹部超声多器官多异常实时诊断的核心临床需求,重点研究了17类常见的腹部超声异常,包括脂肪肝、肝内血管瘤、肝囊肿、肝内钙化灶、肝内其他占位、胆囊息肉、胆囊结石、胆囊其他占位、胰腺占位、脾囊肿、脾血管瘤、脾其他占位、肾结石、肾囊肿、肾其他占位、膀胱结石或膀胱肿瘤等。本研究设计的技术路线如图1所示,先构建标准化腹部超声影像数据集,随后在该数据集上分别训练RetinaNet、Faster R-CNN、RT-DETR及YOLOv13共4种主流深度学习目标检测网络,通过多维度指标系统比较各模型的检测性能,筛选出综合表现最优的深度学习模型;最后通过与6名不同年资超声医师(3名初级、3名中级)的诊断表现进行对照分析,评估最优模型的临床辅助诊断应用潜力。

本研究经陆军军医大学第二附属医院医学伦理委员会批准(2024-研第276-01)。此外,在收集之前,所有回顾性数据集都被去标识,以删除任何与患者相关的信息。

1.2　数据集的构建

本研究回顾性收集2017年1月至2024年9月陆军军医大学第二附属医院超声科的腹部B型超声图像数据,纳入标准:①年龄≥18岁;②行腹部超声检查,且经超声医师明确诊断存在17类目标腹部异常中1类或多类的患者。排除标准:①无明确目标病灶超声图像的患者;②病灶超声图像不清晰无法满足标注要求的患者。经筛选,最终纳入了18 947例患者的39 530幅B型超声图像,然后,采用按异常类别分层随机抽样的方法,抽取927例患者的1 891幅图像作为独立测试集;剩余18 020例患者的37 639幅影像为模型构建数据集,按7∶1∶2比例随机划分为训练集、验证集与内部测试集。腹部B型超声图像数据的纳入、排除流程图见图2。模型构建数据集的患者基线特征、异常类别分布见表1。

数据集中所有图像均由以下设备采集:①中国迈瑞公司的Mindray Resona 7,SC6-1U型凸阵探头;②中国开立公司的SonoScape S50/P60,3C-A/C1-6A型凸阵探头;③中国飞依诺公司的ULTIMUS 9E,S1-8C型凸阵探头;④美国飞利浦公司的EPIQ5/EPIQ7,C1-5型凸阵探头。

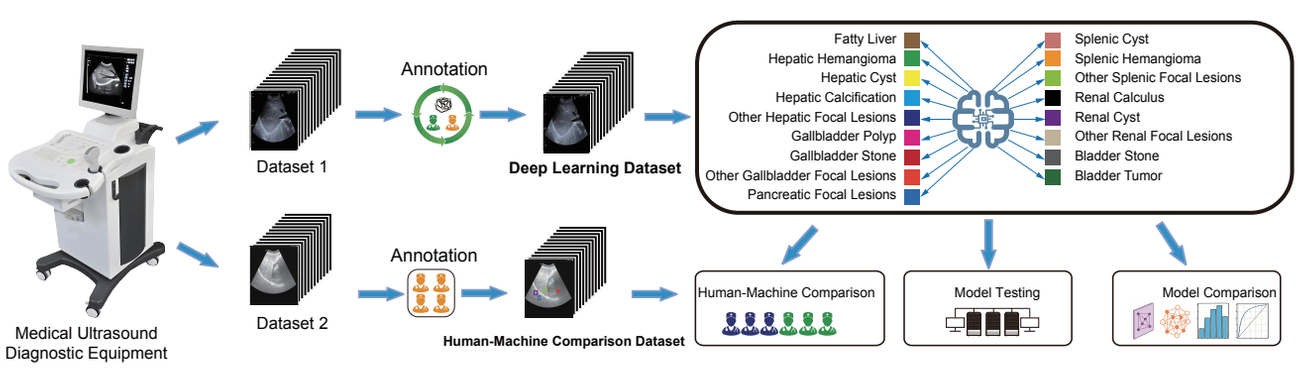


图1　研究流程

Figure 1　Research flowchart

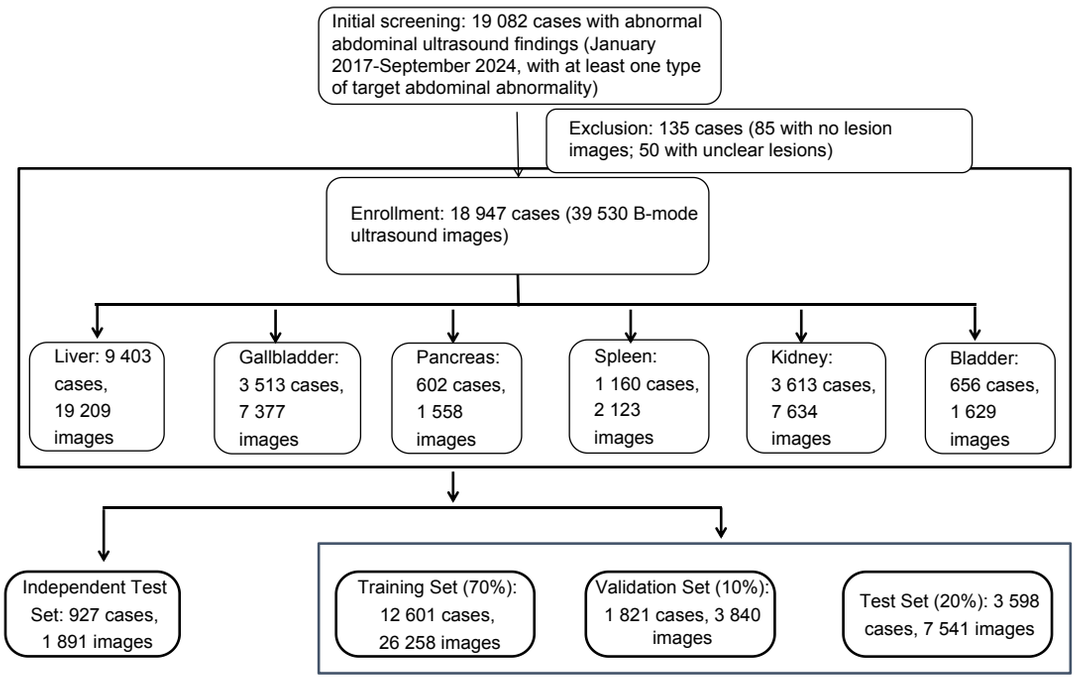


图2 腹部B型超声图像的筛选流程图

Figure 2 Screening flowchart of abdominal B-mode ultrasound images

表1 数据集中患者的一般分布及各类异常分布情况[$M(P_{25},P_{75})$,例(%)]

Table 1 Baseline distribution of patients and distribution of various types of abnormalities in the dataset [$M(P_{25},P_{75})$, n(%)]

Category	Training set (n=12 601)	Validation set (n=1 821)	Test set (n=3 598)	Date set (n=18 020)
Images	26 258	3 840	7 541	37 639
Bounding boxes	29 007	4 213	8 287	41 507
Patients	12 601	1 821	3 598	18 020
Age/years	51.00(43.00, 56.00)	51.00(43.00, 56.00)	51.00(42.00, 56.00)	51.00(43.00, 56.00)
Gender (male)	6 531(51.83)	973(53.43)	1 822(50.64)	9 326(51.75)
Fatty liver	2 586	337	681	3 604
Hepatic hemangioma	3 593	501	1 034	5 128
Hepatic cyst	1 788	219	474	2 481
Hepatic calcification	2 546	408	731	3 685
Other hepatic focal lesions	2 237	337	774	3 348
Gallbladder polyp	517	61	123	701
Gallbladder stone	3 678	570	1 119	5 367
Other gallbladder focal lesions	525	111	185	821
Pancreatic focal lesions	1 033	161	291	1 485
Splenic cyst	493	61	135	689
Splenic hemangioma	431	71	127	629
Other splenic focal lesions	523	70	155	748
Renal calculus	3 080	500	877	4 457
Renal cyst	1 424	188	352	1 964
Other renal focal lesions	710	70	176	956
Bladder stone	811	122	215	1 148
Bladder tumor	283	53	92	428
Total	26 258	3 840	7 541	37 639

1.3　图像标注

超声医生采用开源的LabelImg(v1.8.0,https://pypi.org/project/标签1mg/1.8.0)定位框标注工具,对腹部B型超声图像中各类异常的类别信息及位置坐标进行标注,每个定位框均包含异常类别标签与异常区域最小外接矩形框的坐标数据。标注标准化是保障数据集质量的核心前提[20]。所有参与标注的超声医师完成培训,熟练掌握各类异常的标注原则后,开展正式标注工作。具体标注流程如下:所有图像采用随机分配方式,由4名中级超声医师完成初次标注;随后由2名具备12年以上资质的超声专家对标注结果进行交叉审核;若审核发现标注分歧,引入第3名资深超声专家讨论后最终确定统一的标注结果,即真实标签(ground truth, GT)。图3展示了各类异常的典型标注示例。

1.4　深度学习模型的训练与性能评价

1.4.1　深度学习模型训练

本研究选取医学影像领域广泛应用的4种代表性目标检测网络进行对比分析:RetinaNet作为单阶段(single-stage)检测经典代表,其通过特征金字塔网络(Feature Pyramid Network, FPN)增强多尺度病灶检测能力,并采用焦距损失函数(focal loss)有效缓解类别不平衡问题[21],在小病灶识别中具有显著优势。Faster R-CNN是两阶段检测标杆性架构,创新性引入区域建议网络(Region Proposal Network,RPN),突破传统R-CNN系列模型因重复特征计算导致的效率瓶颈[22],在复杂病灶检测中仍具有重要参考价值。RT-DETR是首个实时端到端目标检测器[23],在保留Transformer架构高精度特性的同时,针对性优化推理速度瓶颈,为传统Transformer与实时检测器(如YOLO系列)的技术融合提供新范式,符合超声动态扫查的实时需求。YOLOv13通过持续优化网

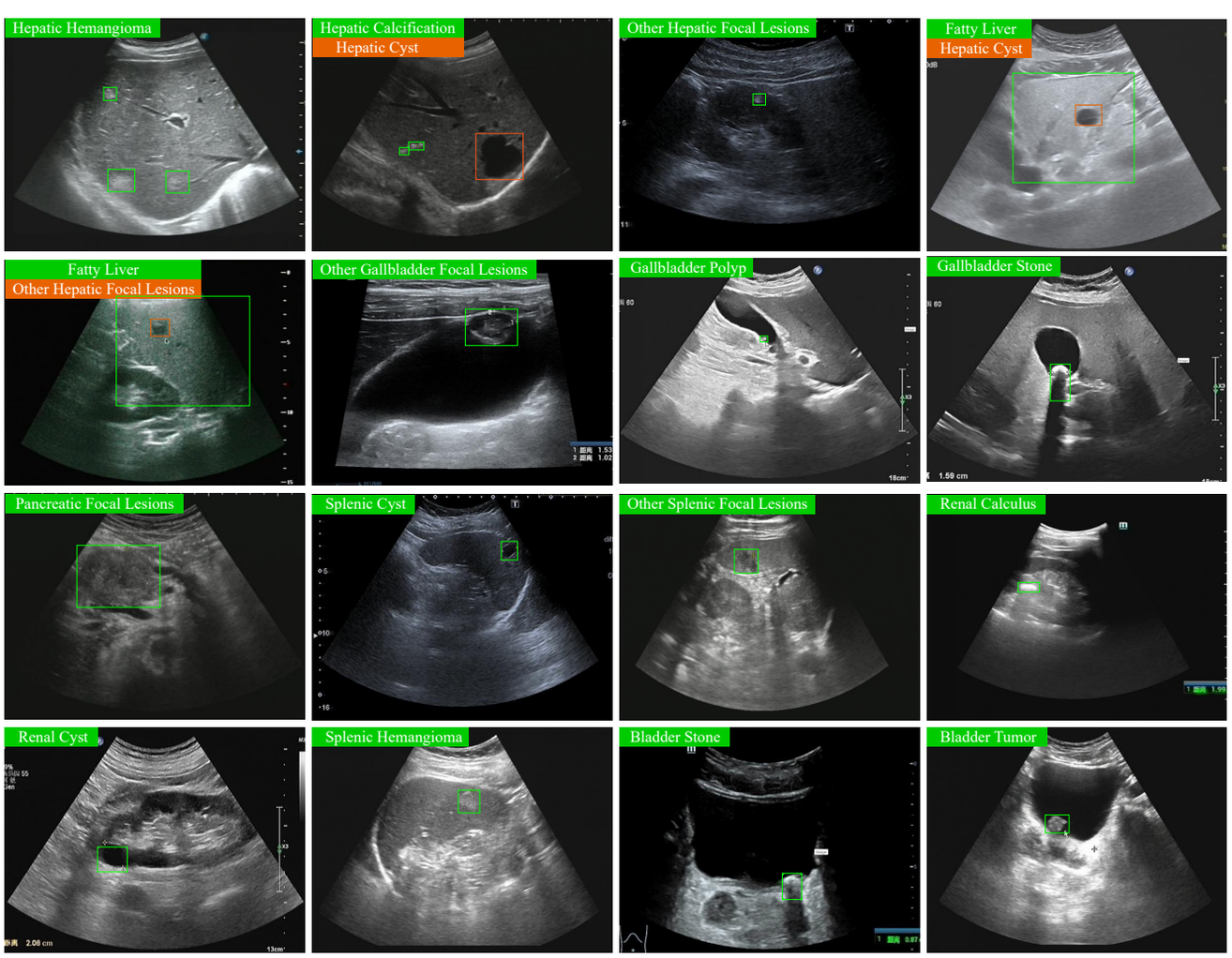


图3　腹部B型超声中17类异常的标注示例图

Figure 3　Annotated example images of 17 types of abnormalities in B-mode abdominal ultrasound

络结构、升级特征提取模块及算法迭代,在YOLO系列中综合性能领先(检测精度/速度/多目标兼容性)[24],故本研究将其作为YOLO代表模型纳入对比。

本研究首先对数据进行预处理,通过裁剪和重采样将图像尺寸固定为640×640,然后通过划分的训练集、验证集和测试集分别对4种模型分别进行训练、验证和测试。使用随机旋转、随机水平翻转、随机亮度和对比度变换进行数据增强。训练过程中使用交叉熵作为损失函数,优化器选择SGD,初始学习率设置为0.001,训练epoch设置为300,根据验证集最小损失保存最优的深度模型。本研究所有模型训练、推理及性能测试基于统一服务器环境完成。

1.4.2 模型的评价 本研究通过模型在测试集上的预测结果来评价模型性能。采用交并比(intersection over union, IoU)、精确率(precision)、召回率(recall)、F_1分数(F_1)、平均精度(average precision, AP)及平均精度均值(mean average precision, mAP)作为深度学习模型的评价指标。其中,IoU用于量化模型预测框与专家标注"金标准"的重叠度,其阈值越高,对应重叠度要求越严格。精确率用来衡量在所有被模型标记为病变的区域中,真正属于病变的比例;召回率衡量了所有真正的病变中,被模型正确检测出来的比例;F_1分数是精确率和召回率的调和平均数,F_1分数越高,说明模型在精确率和召回率之间取得了更好的平衡;针对每类病变绘制精确率-召回率(P-R)曲线,曲线下面积即为AP值,所有病变AP的均值计算为mAP,mAP@50为IoU阈值为0.5时的mAP,反映异常区域的粗粒度空间匹配能力。为验证模型在便携式超声仪等计算资源有限设备上的部署可行性,本研究从计算效率、资源占用及计算复杂度3个维度评估,采用帧率、模型参数量及浮点运算次数量化表征。

此外,本研究设计人机对比实验评估模型的诊断效能。在独立测试集中,将最优模型的检测性能与6名无前期标注经验的超声医师(初级医师3名:工作经验为2~3年;中级医师3名:工作经验为6~8年)的表现进行对比(6名超声医师均先培训,熟练掌握17类异常的标注原则)。为进一步评估模型对医师表现的提升效果,我们对上述3名初级超声医师设计了前后对照实验:在完成首次独立标注后,设置15 d的洗脱期,然后在最优模型辅助下该医师再次完成相同图像的标注(即AI辅助初级医师),并比较前后2次的表现,量化评估模型对初级医师表现的辅助增益。

通过模拟临床中不同设备的图像质量波动(如图像亮度、对比度的异常波动),评估最优模型对图像质量变化的耐受能力,间接验证模型的跨设备泛化能力。将原始图像的亮度、对比度参数值设为1.0,分别测试亮度和对比度的负范围和正范围,对亮度与对比度分别进行"负向调整"(参数值<1.0,模拟亮度不足、对比度偏低场景)与"正向调整"(参数值>1.0,模拟亮度过强、对比度过高场景),每次调整的增量或减量均为0.1,确保覆盖从低质量到过增强的全范围图像质量场景。

本研究中YOLOv13模型的程序代码已公开,可在以下网址下载:https://github.com/LiHuaGO/Ultrasound-Object-Detection。

1.5 统计学分析

采用mAP、精准率、召回率、F_1值、帧率、模型参数量及浮点运算次数来评价模型的性能。精确率与召回率的95%置信区间通过Wilson评分法估算。人机对比测试中,多组之间F_1分数均值的整体差异采用IBM SPSS Statistics 22.0软件进行方差分析以及最小显著性差异法(least significant difference,LSD)开展组间比较。所有统计分析,$P<0.05$为差异具有统计学意义,并采用双尾检验。

2 结果

2.1 YOLOv13模型显著提升腹部异常检测综合性能与推理效率,大幅降低参数量且适配多算力环境

在内部测试集上,以精确率、召回率、F_1值及mAP@50为核心指标,对比RetinaNet、Faster R-CNN、RT-DETR及YOLOv13这4种模型的腹部异常检测性能,结果如表2所示。YOLOv13为最优模型:综合指标mAP@50达0.653,较RT-DETR(0.627)、Faster R-CNN(0.578)、RetinaNet(0.546)分别提升4.1%、12.9%、19.6%;召回率0.715(95%*CI*:0.704~0.725)、精确率0.633(95%*CI*:0.624~0.643),较所对比模型最高分别提升23.7%、20.3%,实现漏诊率与误诊率降低;各指标95%*CI*跨度窄,模型稳定性优;参数量2.5×10^6(较Faster R-CNN降低98.2%),推理帧率31.55 fps(较RT-DETR提升53.9%),见表2。在便携医疗终端级CPU平台验证显示,模型单幅推理耗时106.2 ms/张,适配低算力设备,结合GPU测试结果,证实其对高低算力环境均具良好适配性(表2)。提示YOLOv13模型在腹部异常检测的综合性能、稳定性及轻量化适配性

上表现突出,可满足不同硬件场景的临床应用需求。

2.2 YOLOv13模型检测效能显著优于初级医师,召回率略高于中级医师

在独立测试集中(IoU阈值为0.5),以精确率、召回率、F_1值为指标,对比YOLOv13模型与不同年资超声医师(3名初级、3名中级)对17类腹部异常的检测性能,结果如表3、图4所示。模型召回率为0.658,不仅高于初级医师组均值(0.472),且略高于中级医师组均值(0.640);精确率为0.626,高于初级医师组均值(0.529),但低于中级医师组均值(0.705);模型F_1值为0.624,显著高于初级医师组均值(0.484,$P<0.001$),整体综合性能接近中级医师水平,见图4、图5B。提示YOLOv13模型可有效弥补初级医师诊断短板,凭借更优召回率展现出良好临床辅助应用潜力。

2.3 AI模型显著提升初级超声医师诊断效能,大幅缩小与中级超声医师差距

采用P-R曲线、精确率、召回率及F_1值等指标,分析3名初级超声医师在AI模型辅助前后的腹部异常

表2 4种主流目标检测模型在内部测试集上的性能

Table 2 Performance of 4 mainstream object detection models on the internal test set

Models	Precision (95%*CI*)	Recall (95%*CI*)	F_1 (95%*CI*)	mAP@50 (95%*CI*)	Number of parameters/ ($\times10^6$)	Floating point operations/ (GFLOP/s)	Frames per second/fps
RetinaNet	0.517 (0.508~0.527)	0.578 (0.567~0.589)	0.546 (0.538~0.554)	0.546 (0.511~0.563)	55.47	101.30	18.12
Faster R-CNN	0.426 (0.419~0.434)	0.679 (0.669~0.690)	0.524 (0.517~0.532)	0.578 (0.531~0.591)	137.02	370.10	15.80
RT-DETR	0.526 (0.516~0.537)	0.707 (0.696~0.717)	0.603 (0.593~0.613)	0.627 (0.585~0.643)	15.52	37.50	20.49
YOLOv13	0.633 (0.624~0.643)	0.715 (0.704~0.725)	0.671 (0.663~0.681)	0.653 (0.608~0.669)	2.50	6.40	31.55

表3 4种诊断模式下17类腹部超声异常的诊断效能指标对比

Table 3 Comparison of diagnostic fficacy across four modes for 17 abdominal ultrasound abnormalities

Class	Junior			Attending			AI + Junior			AI		
	Precision	Recall	F_1	Precision	Recall	F_1	Precision	Recall	F_1	Precision	Recall	F_1
Fatty liver	0.719	0.440	0.539	0.707	0.364	0.437	0.476	0.806	0.585	0.437	0.799	0.565
Hepatic hemangioma	0.714	0.660	0.686	0.795	0.745	0.764	0.813	0.835	0.824	0.775	0.857	0.814
Hepatic cyst	0.746	0.693	0.718	0.831	0.830	0.828	0.833	0.860	0.847	0.842	0.880	0.860
Hepatic calcification	0.492	0.429	0.458	0.626	0.628	0.625	0.672	0.663	0.667	0.695	0.700	0.697
Other hepatic focal lesions	0.505	0.581	0.539	0.602	0.637	0.615	0.629	0.708	0.666	0.551	0.645	0.594
Gallbladder polyp	0.386	0.393	0.389	0.584	0.600	0.590	0.562	0.521	0.539	0.583	0.262	0.361
Gallbladder stone	0.440	0.396	0.417	0.463	0.488	0.474	0.459	0.506	0.481	0.454	0.482	0.468
Other gallbladder focal lesions	0.397	0.139	0.181	0.533	0.319	0.394	0.402	0.389	0.393	0.455	0.458	0.457
Pancreatic focal lesions	0.762	0.635	0.692	0.945	0.735	0.801	0.843	0.863	0.853	0.737	0.849	0.789
Splenic cyst	0.436	0.333	0.378	0.913	0.594	0.711	0.612	0.652	0.629	0.584	0.783	0.669
Splenic hemangioma	0.484	0.542	0.506	0.810	0.896	0.842	0.625	0.854	0.720	0.660	1.000	0.795
Other splenic focal lesions	0.457	0.281	0.330	0.758	0.684	0.712	0.630	0.351	0.435	0.659	0.407	0.503
Renal calculus	0.329	0.305	0.317	0.456	0.412	0.432	0.411	0.449	0.429	0.436	0.340	0.382
Renal cyst	0.624	0.533	0.575	0.784	0.767	0.771	0.759	0.817	0.786	0.673	0.760	0.714
Other renal focal lesions	0.617	0.679	0.633	0.790	0.833	0.809	0.884	0.786	0.827	0.945	0.607	0.739
Bladder stone	0.573	0.392	0.465	0.730	0.654	0.687	0.769	0.621	0.687	0.697	0.569	0.626
Bladder tumor	0.306	0.595	0.403	0.654	0.690	0.652	0.464	0.786	0.578	0.456	0.786	0.577
Mean	0.529	0.472	0.484	0.705	0.640	0.656	0.638	0.674	0.644	0.626	0.658	0.624

诊断效能。AI辅助后,初级医师的P-R曲线均向模型最优操作点靠拢(图4);对比无模型辅助时的表现,其精确率提升20.6%(0.529 *vs* 0.638),召回率提升42.8%(0.472 *vs* 0.674),F_1值提升33.1%(0.484 *vs* 0.644,$P<0.001$),见表3;与中级医师相比,辅助后初级医师召回率反超(0.674 *vs* 0.640),精确率差距从12.9%缩小至9.5%(0.638 *vs* 0.705),F_1值差距从32.2%缩小至9.7%(0.644 *vs* 0.656,$P<0.001$),见图5B。同时,为评估模型对工作效率的改善价值,本研究记录3名初级医师AI辅助前后的平均诊断耗时,结果显示:辅助后,初级医师平均诊断耗时由213.33 min降至161.33 min,效率提升24.4%,且耗时显著低于中级医师的234.67 min,实现诊断效率的显著提升(图5A)。但在胆囊结石、胆囊其他占位及肾结石检测中,初级医师组、中级医师组、AI组及AI+初级医师组的F_1值均低于0.5(表3)。提示AI模型可有效弥补初级医师的经验短板,显著提升其诊断效能与效率,大幅缩小与中级医师的能力差距,具有重要临床辅助价值。

2.4 YOLOv13模型对非理想成像条件具备良好耐受度

通过调节图像亮度、对比度开展压力测试,评估YOLOv13模型的性能稳定性,其性能变化曲线见图6。结果显示,在图像亮度(0.4~1.5)对比度(0.5~1.6)区间波动时,模型mAP@50始终维持在0.6以上;即便处于超出临床常规的极端参数条件下,核心性能仍保持稳定,未出现断崖式下降。而临床实践中,超声医师会主动将图像参数调节至合理阅片范围,此类极端成像条件极少出现。提示YOLOv13模型对非理想成像条件的耐受能力优异,可充分适配临床实际应用场景。

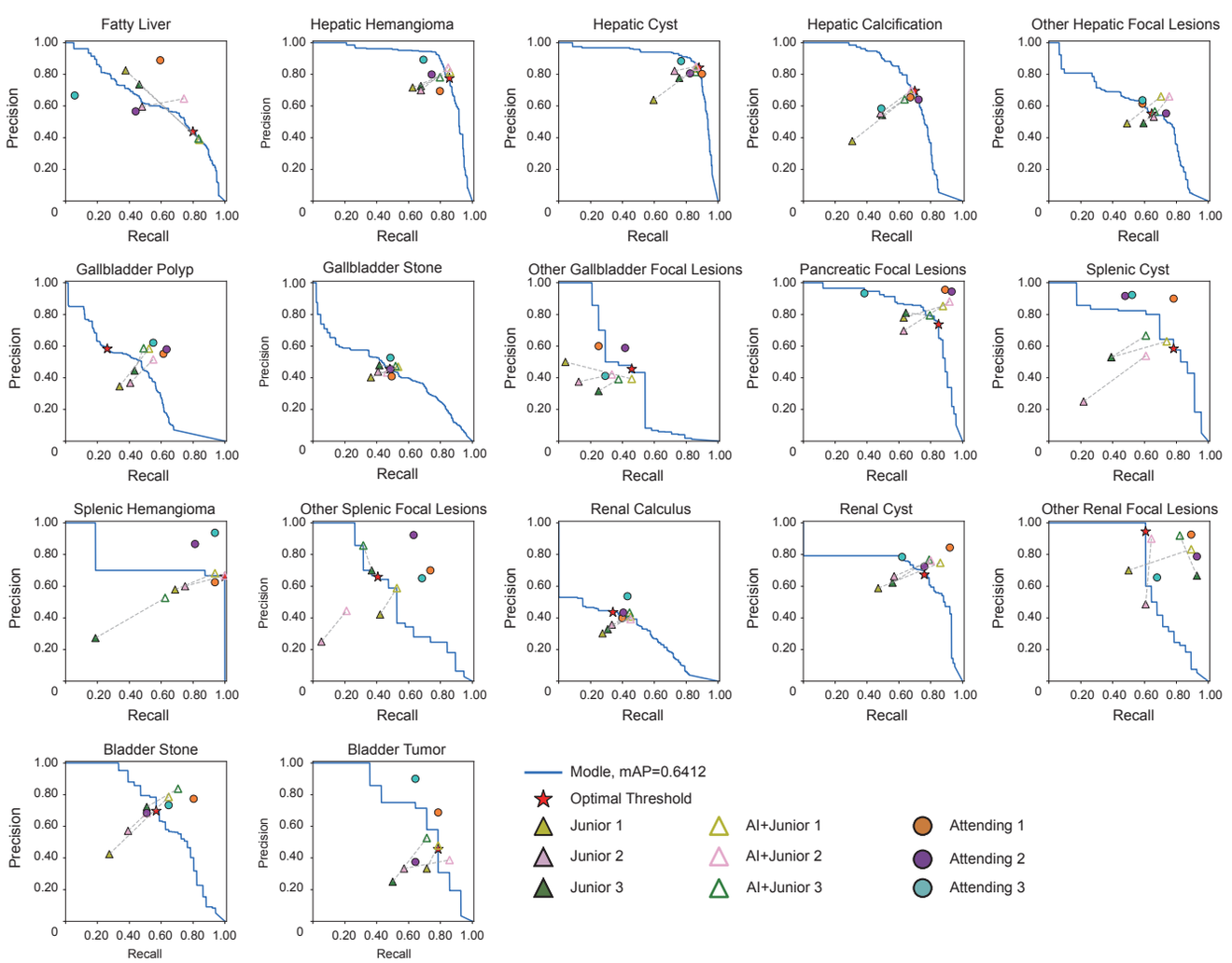
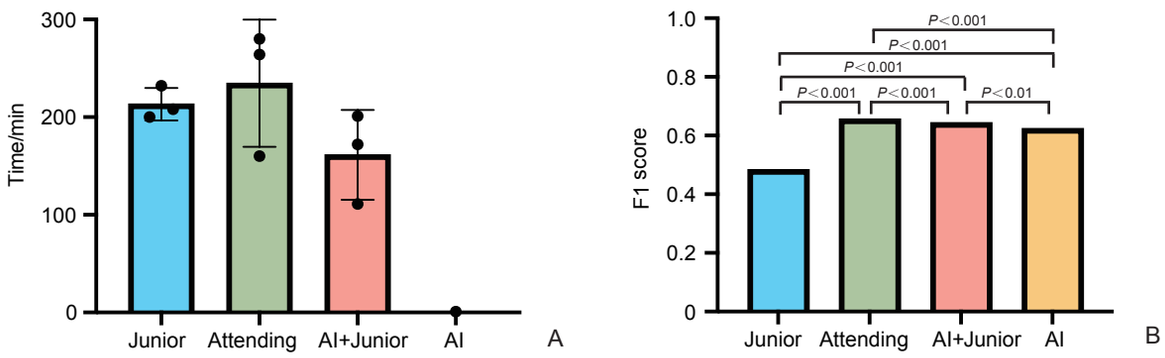


图4 4种诊断模式下17类腹部超声异常的P-R曲线特征及诊断效能评估

Figure 4 Analysis of P-R curve features and diagnostic efficacy across 4 modes for 17 abdominal ultrasound abnormalities



A: Average diagnostic time among the four diagnostic modes; B: F_1 scores and *P* values obtained by the 4 groups at an IoU threshold of 0.5

图5　4种诊断模式的腹部超声诊断平均耗时与F1值对比

Figure 5　Comparison of abdominal ultrasound diagnostic time and F_1 scores across 4 diagnostic modes

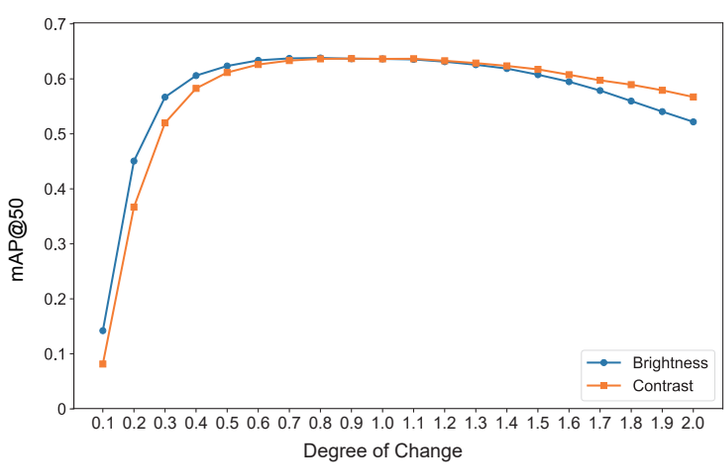


图6　亮度/对比度变化对腹部超声检测模型mAP@50的影响曲线

Figure 6　Effect curves of brightness/contrast variations on mAP@50 of the abdominal ultrasound detection model

3　讨论

3.1　腹部超声多异常检测的深度学习应用瓶颈与本研究突破

腹部多病变因病灶差异细微、累及器官范围广，其识别与检测难度远高于单病变[21]。现有深度学习研究虽已涉足该领域，但仍存在瓶颈：KARAKO等[25]仅基于1例患者的91张图像建模，样本量匮乏致模型泛化性不足，mAP@50仅0.549、推理帧频4 FPS，难以满足临床需求；DADOUN等[26]构建的DETR模型虽在肝脏局灶性病变检测中F_1值达0.8，但局限于单一器官，未覆盖多器官联合筛查场景；WU等[27]构建的YOLOR-D6模型对以肝癌为主的8类病灶mAP@50仅0.315，检测效能与临床实用标准差距较大。针对上述瓶颈，本研究锚定超声一线筛查的临床定位，聚焦肝、胆、胰、脾、肾、膀胱6大器官的17类常见异常，构建含39 530幅图像的大规模高质量数据集，实现多器官多异常的同步实时检测与定位。模型对囊肿、典型血管瘤等高确定性病变直接分类以提升诊断效率；对争议性占位病变则仅精准框定位置、不做定性诊断，既规避了模型独立决策的伦理风险，又与超声先筛查异常、后确诊的临床流程高度契合。这一设计不仅突破了现有研究样本量不足、应用场景单一的局限，更为本研究模型的性能优势奠定了基础。

3.2　YOLOv13架构设计提升腹部多异常检测性能

腹部17类异常存在病灶尺度差异悬殊、图像声学特征复杂等检测技术难点，YOLOv13架构设计有助于提升检测性能。在特征提取环节，跨阶段局部网络(CSPNet)与路径聚合网络(PANet)的级联架构，可在削减冗余计算量的同时，充分保留微小病灶的细粒度

特征,这一设计直接推动模型召回率提升至0.715(较Faster R-CNN提高5.3%),核心综合指标mAP@50达0.653(较RT-DETR提高4.1%),显著优于传统模型依赖单一尺度特征的检测模式。在定位精度层面,CIoU损失函数可适配不同病灶的形态多样性,有效降低定位偏差,进而助力模型精确率提升至0.633(较RT-DETR提高20.3%),为减少临床误诊风险提供了技术支撑。此外,YOLOv13在轻量化与高效性之间实现了精准平衡,其2.5×10^6的参数量较Faster R-CNN降低98.2%,31.55 fps的推理帧频较RT-DETR提升53.9%,这一优势使其可直接适配基层医疗单位的低算力设备,为后续临床推广扫清了硬件障碍。

3.3 YOLOv13模型的临床实用价值与基层医疗运用潜力

临床诊断中,精确率与召回率分别直接关联误诊风险与漏诊风险:精确率越高,因假阳性结果导致的误诊概率越低;召回率越高,因假阴性结果造成的病灶漏诊及治疗延误风险越小。YOLOv13模型召回率不仅高于5名参试医师的个体水平,更优于初级医师组均值,略高于中级医师组均值,凸显其在降低漏诊风险方面的显著优势;尽管模型精确率略低于中级医师组均值,但F_1值较初级医师组提升32.2%,仅与中级医师组相差3.0%,实现了漏诊与误诊风险的最优平衡。模型的临床价值体现在对初级医师的辅助赋能上。3名初级医师经AI辅助后,召回率提升42.8%,并反超中级医师组均值,精确率提升20.6%,平均诊断耗时缩短24.4%,F_1值提升33.1%($P<0.001$),综合诊断能力接近中级医师水平。即使在胆囊结石、肾结石等静态B型超声检出难度较高的病变中,AI辅助组的召回率仍高于中级医师组,进一步印证了模型的临床实用价值。此外,模型的推理速度远超临床实时检测的基准要求[28],适配超声医师边扫查、边判断的工作流程,有效规避静态离线分析模型的漏判问题;其轻量化架构便于基层低算力设备部署,降低基层医疗机构的部署门槛。

本研究存在以下局限性:首先,模型训练基于单一医疗中心的回顾性数据集,未来需通过多中心前瞻性研究进一步验证模型泛化性;其次,模型构建仅纳入B型超声灰度图像,未融合彩色多普勒血流信号及患者病史、实验室检查等临床资料,导致对占位性病变的定性诊断能力不足,难以完全满足临床精准诊疗的需求。

综上所述,本研究通过构建大规模腹部B型超声数据集,系统对比4种主流深度学习模型,证实YOLOv13在腹部多器官多异常实时检测中综合性能最优。该模型不仅凭借针对性架构设计,实现了检测效能对传统模型的超越,更在人机对比中远超初级医师水平,辅助初级医师后可使其诊断能力接近中级医师标准。同时,模型的轻量化特性、实时推理能力及非理想成像条件耐受能力,可直接适配社区医院、部队基层卫生单位等资源受限场景。本研究为缓解基层超声专业人才短缺困境、提升腹部疾病筛查效能提供了可落地的技术方案,具有重要的临床转化价值。

利益冲突 本文所有作者声明不存在利益冲突

参考文献:

[1] TOMIZAWA M, SHINOZAKI F, HASEGAWA R, *et al*. Abdominal ultrasonography for patients with abdominal pain as a first-line diagnostic imaging modality[J]. Exp Ther Med, 2017, 13(5): 1932-1936. DOI:10.3892/etm.2017.4209.

[2] 孙蓉, 葛四明, 曹晓岚. 腹部超声在健康体检中的应用分析[J]. 中国医学创新, 2017, 14(22): 56-58.
SUN R, GE S M, CAO X L. Application analysis of abdominal ultrasound in physical examination[J]. Med Innov China, 2017, 14(22): 56-58.

[3] PEI Y Y, WANG G J, CAO H W, *et al*. A deep-learning pipeline to diagnose pediatric intussusception and assess severity during ultrasound scanning: a multicenter retrospective-prospective study[J]. NPJ Digit Med, 2023, 6(1): 182. DOI:10.1038/s41746-023-00930-8.

[4] XU M L, LV P, LI M Y, *et al*. Medical image denoising by parallel non-local means[J]. Neurocomputing, 2016, 195: 117-122. DOI:10.1016/j.neucom.2015.08.117.

[5] CONTRERAS ORTIZ S H, CHIU T, FOX M D. Ultrasound image enhancement: a review[J]. Biomed Signal Process Control, 2012, 7(5): 419-428. DOI: 10.1016/j.bspc. 2012.02.002.

[6] KRUPINSKI E A, BERBAUM K S, CALDWELL R T, *et al*. Long radiology workdays reduce detection and accommodation accuracy[J]. J Am Coll Radiol, 2010, 7(9): 698-704. DOI: 10.1016/j.jacr.2010.03.004.

[7] NAM J G, KIM M, PARK J, *et al*. Development and validation of a deep learning algorithm detecting 10 common abnormalities on chest radiographs[J]. Eur Respir J, 2021, 57(5): 2003061. DOI:10.1183/13993003.03061-2020.

[8] YANG Q, WEI J W, HAO X H, *et al*. Improving B-mode ultrasound diagnostic performance for focal liver lesions using deep learning: a multicentre study[J]. EBioMedicine, 2020, 56: 102777. DOI:10.1016/j.ebiom.2020.102777.

[9] 刘丽, 谭鹰, 唐春霖, 等. 基于机器学习的早期肝肿瘤超声造影诊断模型建立及应用价值[J]. 陆军军医大学学报, 2023, 45(21): 2275-2283. DOI: 10.16016/j. 2097-0927.

202307131.
LIU L, TAN Y, TANG C L, *et al*. Establishment and application value of a diagnostic model for early hepatic tomors using contrast-enhanced ultrasound based on machine learning[J]. J Army Med Univ, 2023, 45(21): 2275-2283. DOI:10.16016/j.2097-0927.202307131.

[10] GUPTA P, BASU S M, RANA P, *et al*. Deep-learning enabled ultrasound based detection of gallbladder cancer in northern India: a prospective diagnostic study[J]. Lancet Reg Health Southeast Asia, 2024, 24: 100279. DOI: 10.1016/j.lansea.2023.100279.

[11] SALAM A, NAZNINE M, CHOWDHURY M E H, *et al*. Hybrid neural networks for precise hydronephrosis classification using deep learning[J]. Urology, 2025, 206: 17-24. DOI:10.1016/j.urology.2025.08.005.

[12] VIANNA P, CALCE S I, BOUSTROS P, *et al*. Comparison of radiologists and deep learning for US grading of hepatic steatosis[J]. Radiology, 2023, 309(1): e230659. DOI: 10.1148/radiol.230659.

[13] YANG Y, LIU J, SUN C X, *et al*. Nonalcoholic fatty liver disease (NAFLD) detection and deep learning in a Chinese community-based population[J]. Eur Radiol, 2023, 33(8): 5894-5906. DOI:10.1007/s00330-023-09515-1.

[14] SCHMAUCH B, HERENT P, JEHANNO P, *et al*. Diagnosis of focal liver lesions from ultrasound using deep learning[J]. Diagn Interv Imaging, 2019, 100(4): 227-233. DOI:10.1016/j.diii.2019.02.009.

[15] DING W Z, MENG Y Q, MA J, *et al*. Contrast-enhanced ultrasound-based AI model for multi-classification of focal liver lesions[J]. J Hepatol, 2025, 83(2): 426-439. DOI: 10.1016/j.jhep.2025.01.011.

[16] NISHIDA N, YAMAKAWA M, SHIINA T, *et al*. Artificial intelligence (AI) models for the ultrasonographic diagnosis of liver tumors and comparison of diagnostic accuracies between AI and human experts[J]. J Gastroenterol, 2022, 57(4): 309-321. DOI:10.1007/s00535-022-01849-9.

[17] INTHARAH T, WIRATCHAWA K, WANNA Y, *et al*. BiTNet: Hybrid deep convolutional model for ultrasound image analysis of human biliary tract and its applications[J]. Artif Intell Med, 2023, 139: 102539. DOI:10.1016/j.artmed.2023.102539.

[18] BASU S, GUPTA M, RANA P, *et al*. RadFormer: Transformers with global-local attention for interpretable and accurate Gallbladder Cancer detection[J]. Med Image Anal, 2023, 83: 102676. DOI: 10.1016/j.media.2022.102676.

[19] ZHAO W T, ZHANG Z, WANG Z, *et al*. ECLNet: Center localization of eye structures based on Adaptive Gaussian Ellipse Heatmap[J]. Comput Biol Med, 2023, 153: 106485. DOI:10.1016/j.compbiomed.2022.106485.

[20] LIANG W X, TADESSE G A, HO D, *et al*. Advances, challenges and opportunities in creating data for trustworthy AI[J]. Nat Mach Intell, 2022, 4(8): 669-677. DOI: 10.1038/s42256-022-00516-1.

[21] 王生生,路淑贞,曹斌.面向隐私保护联邦学习的医学影像目标检测算法[J].计算机辅助设计与图形学学报,2021,33(10):1553-1562. DOI:10.3724/SP.J.1089.2021.18416.
WANG S S, LU S Z, CAO B. Medical image object detection algorithm for privacy-preserving federated learning[J]. J Comput Aided Des Comput Graph, 2021, 33(10): 1553-1562. DOI:10.3724/SP.J.1089.2021.18416.

[22] REN S Q, HE K M, GIRSHICK R, *et al*. Faster R-CNN: towards real-time object detection with region proposal networks[J]. IEEE Trans Pattern Anal Mach Intell, 2017, 39(6): 1137-1149. DOI:10.1109/tpami.2016.2577031.

[23] ZHAO Y, LYU W, XU S, *et al*. DETRs Beat YOLOs on Real-time Object Detection[EB/OL]// 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2024: 16965-16974. (2024-06-17)[2026-01-29 11:35]. DOI: 10.1109/CVPR52733.2024.01605. Available from: https://openaccess.thecvf.com/content/CVPR2024/html/Zhao_DETRs_Beat_YOLOs_on_Real-Time_Object_Detection_CVPR_2024_paper.html

[24] JIA H J, ZHANG L J, LIANG X M, *et al*. DPDB-YOLO: a lightweight YOLOv13 cherry tomato ripeness detection method with adaptive extraction module and multi-scale feature fusion architecture[J]. Ind Crops Prod, 2025, 238: 122419. DOI:10.1016/j.indcrop.2025.122419.

[25] KARAKO K, MIHARA Y, ARITA J, *et al*. Automated liver tumor detection in abdominal ultrasonography with a modified faster region-based convolutional neural networks (Faster R-CNN) architecture[J]. Hepatobiliary Surg Nutr, 2022, 11(5): 675-683. DOI:10.21037/hbsn-21-43.

[26] DADOUN H, ROUSSEAU A L, DE KERVILER E, *et al*. Deep learning for the detection, localization, and characterization of focal liver lesions on abdominal US images[J]. Radiol Artif Intell, 2022, 4(3): e210110. DOI: 10.1148/ryai.210110.

[27] WU C H, SHEU J C, CHOU P L, *et al*. Automatic real-time detection and diagnosis of liver tumor with ultrasound[J]. J Hepatocell Carcinoma, 2025, 12: 1599-1611. DOI: 10.2147/JHC.S524311.

[28] ZHAO Z J, CHEN Z R, VOROS S, *et al*. Real-time tracking of surgical instruments based on spatio-temporal context and deep learning[J]. Comput Assist Surg, 2019, 24(sup1): 20-29. DOI:10.1080/24699322.2018.1560097.

(**收稿**:2025-12-18;**修回**:2026-01-13)
(编辑 栾 嘉)