

·临床研究·

## 融合临床与超声图像多模态数据的深度学习 模型评估脂肪肝分级

蒙文仪<sup>1</sup>, 姚欣蕊<sup>2</sup>, 李任杰<sup>1</sup>, 彭键清<sup>3</sup>, 徐作峰<sup>1</sup>

(1. 中山大学附属第七医院超声科, 广东 深圳 518107; 2. 中山大学电子与通信工程学院, 广东 深圳 518107;  
3. 中山大学智能工程学院, 广东 深圳 518107)

**摘要:**【目的】本研究旨在融合多种易获取的临床数据与二维常规超声图像, 构建用于评估脂肪性肝病(SLD)不同严重程度的深度学习神经网络模型。【方法】回顾性收集649例行超声衰减成像(ATI)检查患者的临床数据和超声图像, 以ATI作为参考标准, 将患者分为正常(S0)及轻(S1)、中(S2)、重度(S3)脂肪肝4组, 按8:2比例随机划分为训练集和验证集。采用对比语言-图像预训练(CLIP)模型提取临床与图像多模态特征, 并分别训练随机森林(RF)和多层感知器(MLP)模型。在验证集中对比多模态模型与单一模态的诊断性能。【结果】与RF模型相比, MLP模型表现更优[AUC(95%CI): S0=0.96(0.93~0.99), S1=0.99(0.96~1.00), S2=0.75(0.64~0.82), S3=0.88(0.81~0.93)]。相较于单模态模型[仅临床数据AUC(95%CI): 0.89(0.81~0.95)、0.69(0.52~0.79)、0.64(0.43~0.67)、0.80(0.73~0.90); 仅图像数据AUC(95%CI): 0.91(0.86~0.96)、0.89(0.67~0.89)、0.69(0.58~0.78)、0.86(0.82~0.94)], 图像-临床数据融合模型在各分级上均表现出显著更优的性能。精确率、召回率、F1分数、F2分数、混淆矩阵与损失函数学习曲线进一步证实, 多模态数据融合可显著提升整体训练与预测效果。【结论】本研究构建的基于CLIP的MLP多模态分类模型可有效实现脂肪肝严重程度的自动分级, 证实了融合临床与超声数据在脂肪肝精准分级的显著优势, 为慢性脂肪性肝病的临床管理提供准确、可靠的辅助评估工具。

**关键词:** 脂肪肝; 超声图像; 临床数据; 深度学习; 衰减成像

中图分类号: R445.1 文献标志码: A 文章编号: 1672-3554(2026)03-0528-11

DOI: 10.11714/jysu.med.YX20260026

### A Multimodal Deep Learning Model Integrating Clinical and Ultrasound Imaging Data for grading steatotic liver disease

MENG Wenyi<sup>1</sup>, YAO Xinrui<sup>2</sup>, LI Renjie<sup>1</sup>, PENG Jianqing<sup>3</sup>, XU Zuofeng<sup>1</sup>

(1. Department of Ultrasound, the Seventh Affiliated Hospital, Sun Yat-sen University, Shenzhen, 518107, China;  
2. School of Electronic and Communication Engineering, Shenzhen Campus of Sun Yat-sen University, Shenzhen  
518107, China; 3. School of Intelligent Systems Engineering, Shenzhen Campus of Sun Yat-sen University, Shenzhen  
518107, China)

Correspondence to: PENG Jianqing, E-mail: pengjq7@mail.sysu.edu.cn; XU Zuofeng, E-mail: xuzf@mail.sysu.edu.cn

**Abstract:**【Objective】This study aimed to construct a neural network model integrating readily available clinical data and two-dimensional (2D) conventional ultrasound images to evaluate the severity of steatotic liver disease (SLD).【Methods】Clinical data and ultrasound images were retrospectively collected from 649 patients who underwent ultrasound attenuation imaging (ATI). Using ATI as the reference standard, patients were divided into four groups: normal (S0),

收稿日期: 2026-02-08

录用日期: 2026-04-24

基金项目: 广东省基础与应用基础研究基金(2024A1515012545); 临床医学研究“735计划”项目(ZSQY735202210); 深圳市科技计划项目(JCYJ20220530144804009)

作者简介: 蒙文仪, 第一作者, 研究方向: 腹部超声诊断, E-mail: mengwy3@mail2.sysu.edu.cn; 姚欣蕊, 并列第一作者, 研究方向: 医疗机器人多模态大模型, E-mail: yaoxr5@mail2.sysu.edu.cn; 彭键清, 通信作者, 副教授, 研究方向: 医疗机器人多模态大模型, E-mail: pengjq7@mail.sysu.edu.cn; 徐作峰, 并列通信作者, 主任医师, 研究方向: 腹部超声诊断, E-mail: xuzf@mail.sysu.edu.cn

mild (S1), moderate (S2), and severe (S3) SLD. Data were randomly divided into training and validation sets at an 8:2 ratio. A contrastive language-image pre-training (CLIP) model was utilized to extract features from both clinical and imaging data, which were then used to train random forest (RF) and multilayer perceptron (MLP) models. The diagnostic performance of these multimodal models was compared their single-modal counterparts in the validation set.【Results】The MLP model outperformed the RF model (AUC [95%CI]: S0 = 0.96 [0.93-0.99], S1 = 0.99 [0.96-1.00], S2 = 0.75 [0.64-0.82], S3 = 0.88 [0.81-0.93]). Furthermore, the multimodal fusion model demonstrated significantly superior performance compared to single-modal models (clinical data only: AUC [95%CI] of 0.89 [0.81-0.95], 0.69 [0.52-0.79], 0.64 [0.43-0.67] and 0.80 [0.73-0.90] for S0-S3, respectively; image data only: AUC [95%CI] of 0.91 [0.86-0.96], 0.89 [0.67-0.89], 0.69 [0.58-0.78] and 0.86 [0.82-0.94]). Metrics including precision, recall, F1-score, F2-score, confusion matrices, and loss function learning curves further confirmed that multimodal data fusion significantly improved the predictive capability.【Conclusion】The proposed CLIP-based MLP multimodal model can effectively and automatically grade SLD severity. This demonstrates the significant advantage of integrating clinical and ultrasound data, providing an accurate and reliable adjunctive tool for the clinical management of chronic SLD.

**Key words:** steatotic liver disease; ultrasound images; clinical data; deep learning; attenuation imaging

[J SUN Yat-sen Univ (Med Sci), 2026, 47(3): 528-538]

脂肪肝作为全球最常见的慢性肝脏疾病之一,向肝纤维化、肝硬化甚至肝细胞癌进展的潜在风险已成为全球公共卫生领域的重大挑战<sup>[1-2]</sup>。根据病因主要分为代谢功能障碍相关脂肪性肝病 (metabolic dysfunction-associated steatotic liver disease, MASLD)、酒精相关肝病 (alcohol-associated liver disease, ALD) 和其他罕见原因<sup>[3-4]</sup>, 它们共同的病理学特征是肝脏脂肪过度沉积<sup>[5]</sup>。尽管不同病因脂肪肝的临床管理路径不同,但其早期诊断都依赖于对肝脏脂肪变性的准确评估。然而,目前的金标准,如肝活检的有创性,或磁共振成像质子密度脂肪分数 (magnetic resonance imaging proton density fat fraction, MRI-PDFF) 等低成本或操作复杂的影像技术,给基层医疗机构的脂肪肝筛查和分级带来了巨大挑战<sup>[6-8]</sup>。超声技术的发展,推动了对肝脏脂肪变性定量评估的准确性与可重复性。其中,超声衰减成像 (attenuated imaging, ATI) 作为一种新兴的定量工具,在一众肝脏脂肪定量评估方法中脱颖而出<sup>[9]</sup>。ATI通过定量检测肝实质的超声衰减系数,能够准确反映肝脏脂肪含量,与MRI-PDFF结果以及肝脏脂肪变性的组织学分级高度一致<sup>[10-11]</sup>。在区分轻度与中重度脂肪肝时,ATI显示出更高的敏感性和特异性,优于基于传统灰阶超声的主观评估。与此同时,在多模态人工智能领域,OpenAI提出的对比语言-图像预训练模型 (contrastive language-image pre-trained model, CLIP) 通过自然语言监督将图像和文本映射到共享的嵌入空间,推动了图像-文本匹配、描述

和生成等跨模态任务的显著进展<sup>[12-13]</sup>。CLIP实现的视觉-语言对齐能力,为超声图像等医学图像的自动解读、报告生成以及多模态交互分析提供了新的技术连接点。鉴于这些技术进步,本研究旨在利用CLIP从超声图像、超声衍生指标、临床信息和血清学标志物等多模态数据中提取特征,然后进行基于随机森林 (random forest, RF) 和多层感知器 (multilayer perceptron, MLP) 的分类,从而开发一种用于脂肪肝的分类模型,为其分级提供一种无创、准确且临床实用的解决方案。

## 1 材料与方法

### 1.1 研究对象

回顾性收集2022年1月至2025年5月我院649例行ATI检查患者的临床数据和常规超声图像(共3245幅图像,附图1)。研究对象的纳入标准:具备二维超声图像和ATI检查数据,包括超声衍生指标、临床信息和血清学标志物。排除标准:年龄小于18岁的患者;标准超声图像、临床数据不完整。

本研究遵循《赫尔辛基宣言》原则,研究方案经中山大学附属第七医院伦理委员会批准(批号:KY-2025-366),并在中国临床试验注册中心注册(注册号:ChiCTR2500106945),由于本研究属回顾性研究,患者知情同意被豁免。

### 1.2 脂肪肝分级的评估

本研究首次在深度学习模型中采用ATI测量的衰减系数作为脂肪肝分级的参考标准,使用以下

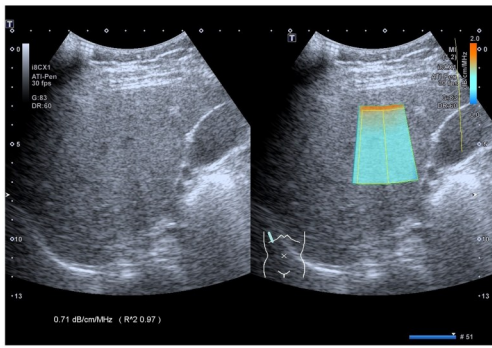


图1 超声衰减成像检查

Fig. 1 Attenuation imaging examination

阈值: S1:  $\geq 0.67$  dB/(cm·MHz), S2:  $\geq 0.72$  dB/(cm·MHz), S3:  $\geq 0.86$  dB/(cm·MHz)。这些阈值由佳能公司推荐,并在 *Radiology* 杂志中得到验证<sup>[14]</sup>。本研究以 ATI 取代病理作为分级金标准,显著降低了数据采集的难度,从而保证了数据集具备充足的样本量,满足了深度神经网络训练对大数据量的需求。

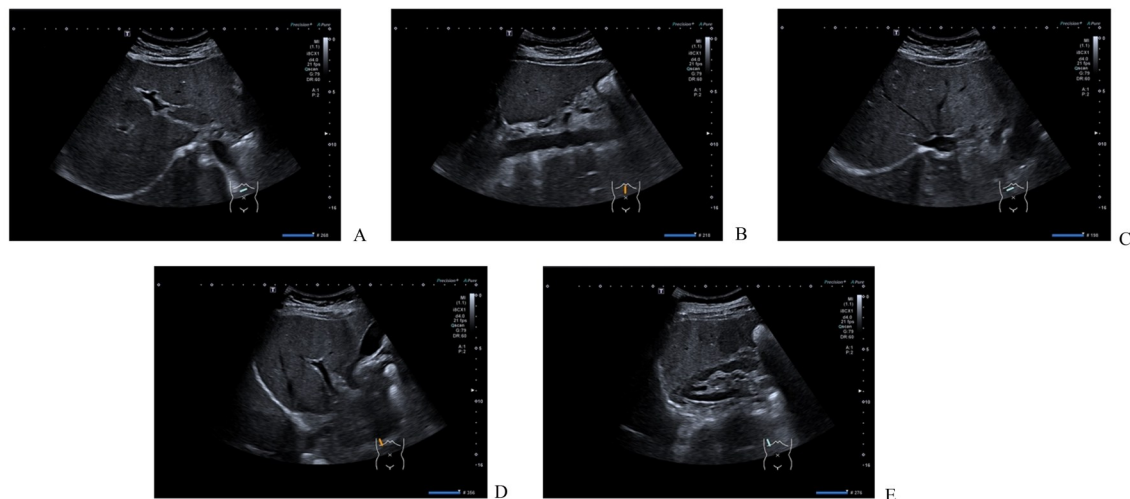
ATI 扫描具体操作步骤如下:操作者避开肝脏内的钙化灶、伪影及大血管结构。在肝脏右肋间切面确定理想扫描平面。从常规 B 模式切换到 ATI 模式,将取样框上缘置于肝包膜下 1 cm 处。取样框大小设置为 4 cm × 3 cm。冻结图像后,将黄色取样框置于绿色取样框中央区域以显示衰减系数。显示衰减系数的可靠性通过  $R^2$  (拟合优度系数)确

定:  $R^2 > 90\%$  表明可靠性极佳,  $R^2$  在 80%~90% 之间表明可靠性良好,  $R^2 < 80\%$  表明可靠性差。本研究中仅纳入  $R^2 \geq 90\%$  且显示为白色数值的衰减系数(图1)。

### 1.3 超声图像、临床数据的收集

1.3.1 标准二维超声图像选择标准 获取并存储至少 5 组静态常规超声图像用于分析。如图 2 所示,由 2 名经验丰富的超声医师收集以下标准切面图像。二维成像模式超声参数:选择 Abd-Fat i8CX1 图像质量条件,深度 16 cm, Frequency 设为 Pen, Focus 76%, DR 60 dB;增益调节以能清晰显示肝实质为宜,使肝实质呈中等强度的细小点状回声(等回声),保证肝静脉、门静脉、肝管及胆囊等结构清晰显示。

1.3.2 临床数据 ①临床信息:年龄、性别、糖尿病史、乙型肝炎史、高血压史、吸烟史、饮酒史、体质量指数(body mass index, BMI)和腰围。②血清学标志物:包括总胆固醇(total cholesterol, TC)、甘油三酯(triglycerides, TG)、低密度脂蛋白胆固醇(low-density lipoprotein cholesterol, LDL-C)、高密度脂蛋白胆固醇(high-density lipoprotein cholesterol, HDL-C)、丙氨酸氨基转移酶(alanine aminotransferase, ALT)、天冬氨酸氨基转移酶(aspartate aminotransferase, AST)、 $\gamma$ -谷氨酰转移酶( $\gamma$ -glutamyl transferase, GGT)、碱性磷酸酶(alkaline phosphatase, ALP)、空腹血糖(fasting blood glucose,



A: Subxiphoid transverse view of the left lobe showing the sagittal segment of the left portal vein. B: Left lobe longitudinal view centered on the long axis of the abdominal aorta. C: Right subcostal oblique view demonstrating the second porta hepatis (junction of middle and right hepatic veins with IVC). D: Right intercostal view targeting the main right portal vein branch. E: Hepatorenal view showing the interface between the inferior border of the right hepatic lobe and upper pole of the right kidney.

图2 标准二维超声图像选择标准

Fig. 2 Standard two-dimensional images selection criteria

FPG)、白蛋白、球蛋白、尿酸(uric acid, UA)、肌酐(creatinine, Cr)、尿素。③超声衍生指标:皮下脂肪最大厚度( $S_{max}$ )、腹壁脂肪指数(abdominal wall fat index, AFI)和颈动脉内膜中层厚度(carotid intima-media thickness, C-IMT)。AFI是用于评估内脏脂肪堆积的超声指标,能更准确地反映腹部脂肪分布<sup>[15]</sup>。具体测量方法<sup>[16]</sup>为:受试者取仰卧位。探头沿腹中线垂直于腹壁皮肤放置,自剑突至脐部进行纵向扫描。保持轻柔接触,避免压迫脂肪层。测量肝表面前方腹膜前脂肪的最大厚度( $P_{max}$ )以及沿中线皮下脂肪的最小厚度( $S_{min}$ )。AFI计算算式为 $P_{max} / S_{min}$ 。C-IMT:探头自颈总动脉起始部沿其长轴

进行纵向扫描。在血管后壁测量3个位置的内膜腔面至中膜外表面的垂直距离:颈动脉分叉近端1 cm处、分叉处本身以及沿颈内动脉分叉远端1 cm处。

#### 1.4 临床数据缺失值的插补

为提高数据质量和稳健性,采用多重插补法处理缺失值,在数据随机缺失的假设下创建多个插补数据集以考虑数据的不确定性。使用预测均值匹配法生成这些数据集。每个数据集都用于模型训练,结果合并后生成汇总估计,以减少偏倚并提高推断的可靠性和泛化性。插补缺失数据的分布及插补基本保持一致(附表1)。

表1 随机森林模型性能对比:(A)未进行图像增强前;(B)进行图像增强后

Table 1 Random forest performance comparison: (A) before and (B) after image enhancement

(A)						
Class/parameter	AUC (95%CI)	Precision	Recall	F1-score	F2-score	Accuracy
S0	0.91 (0.87, 0.96)	0.625 0	0.853 7	0.721 6	0.795 5	0.515 4
S1	0.52 (0.47, 0.75)	0.000 0	0.000 0	0.000 0	0.000 0	
S2	0.68 (0.57, 0.76)	0.365 9	0.428 6	0.394 7	0.414 4	
S3	0.79 (0.75, 0.91)	0.548 4	0.500 0	0.523 1	0.509 0	
(B)						
Class/parameter	AUC (95%CI)	Precision	Recall	F1-score	F2-score	Accuracy
S0	0.94 (0.88, 0.97)	0.587 3	0.902 4	0.711 5	0.815 0	0.553 8
S1	0.96 (0.92, 0.98)	0.750 0	0.300 0	0.428 6	0.340 9	
S2	0.74 (0.64, 0.82)	0.441 2	0.428 6	0.434 8	0.431 0	
S3	0.83 (0.78, 0.92)	0.560 0	0.411 8	0.474 6	0.434 8	

#### 1.5 图像后处理方法

为将肝脏图像标准化以适应深度学习,每例患者的标准切面调整大小为224×224像素,进行中心裁剪,使用CLIP统计量进行归一化,并堆叠为(1, 15, 224, 224)的张量输入。对于样本量较少的S1类别,应用了针对性数据增强——包括随机旋转、仿射变换(旋转±15°、平移幅度≤10%、缩放0.8~1.2倍)和颜色抖动(亮度、对比度、饱和度、色调±20%)。增强后的图像在转换为张量前,会重新调整大小并进行中心裁剪至224×224像素。

#### 1.6 图像特征提取

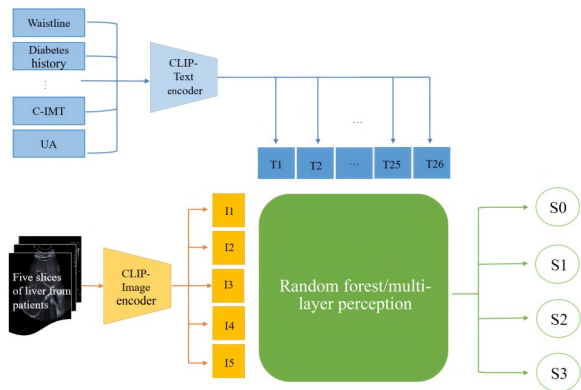
CLIP(对比语言-图像预训练)通过大规模图像-文本对比学习进行训练,提供了强大的跨模态对齐和泛化能力。它被证明在医学成像中尤其有效,即使在有限或未标记的数据下也能很好地泛

化,并且在对抗模糊和形态变异等混杂因素方面优于传统CNN,能够提取更高级别的特征。

我们使用CLIP将预处理后的图像张量编码为视觉特征,并将转换为文本描述的结构化医疗数据编码为语义特征。然后将这些特征与原始数值指标拼接,形成统一的特征向量——融合了图像、文本和结构化数据——作为分类器的输入(图4)。在本研究中,每位患者关联5张超声图像。因此,以CLIP提取的临床+超声图像多模态特征为输入,训练RF和MLP模型。由于每张RGB CT图像包含3个颜色通道,CLIP模型原始的3通道输入被扩展到15个通道以适应输入。

#### 1.7 资料分析

连续变量以均值±标准差(正态分布)或中位数[四分位距](非正态分布)表示。分类变量使用卡



The framework integrates structured clinical data (26 features, e.g., waistline, diabetes history) and liver ultrasound images (5 slices/patient) for steatosis grading (S0–S3). Clinical features are encoded into textual embeddings ( $T_1$ – $T_{26}$ ) via CLIP text encoder; images are processed into visual embeddings ( $I_1$ – $I_5$ ) by CLIP image encoder. Concatenated embeddings are classified using either RF or MLP model.

图3 基于CLIP的MLP分类模型架构

Fig. 3 Architecture of the CLIP-based MLP classification model

方检验进行分析。连续变量的组间比较采用方差分析(正态分布)或Kruskal-Wallis  $H$  检验(非正态分布),统计学显著性定义为  $P < 0.05$ 。

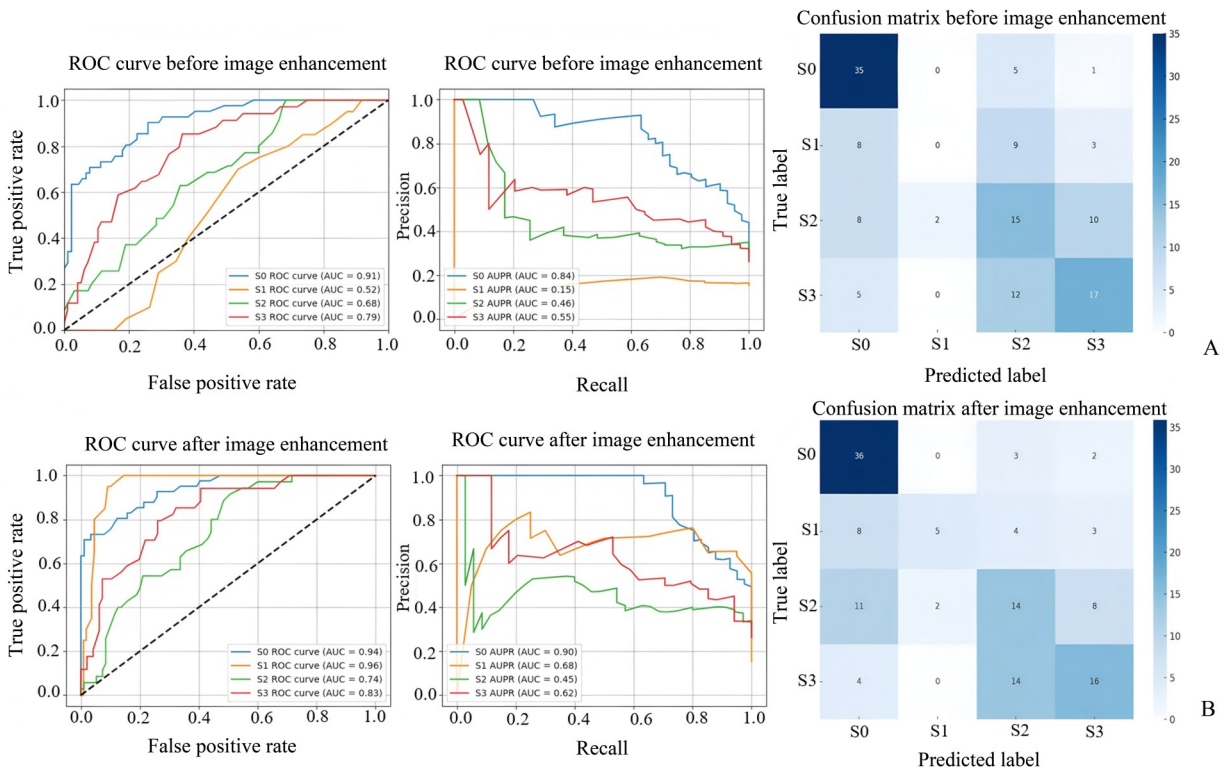
通过多种指标评估模型在脂肪肝分期中的性能,包括准确率、精确率、召回率、F1分数、F2分数、受试者工作特征(receiver operating character, ROC)曲线及其曲线下面积(area under curve, AUC)、精确率-召回率(precision-recall, PR)曲线及其曲线下面积(area under PR, AUPRC)以及混淆矩阵,其中AUC的95%置信区间采用Bootstrap法估计。通过逐轮次的损失和AUC曲线监控训练动态,以评估收敛性和稳定性。

所有统计分析均使用Python 3.11.4版和IBM SPSS Statistics 25.0进行。

## 2 结果

### 2.1 基线资料

本研究共纳入649名参与者,其中519名分配至训练集,130名分配至验证集。包括临床信息、血清学标志物、超声衍生指标,对以上资料进行了比较分析显示,结果显示数据集之间无显著差异(所有  $P > 0.05$ ;附表2)。年龄、BMI、腰围属正态性定量资料,在表中表示为均数加减标准差,其他指标均属非正态性定量资料,在表中表示为中位数及



Before (A) and after (B) image enhancement.

图4 ROC曲线、PRC曲线及混淆矩阵比较

Fig. 4 ROC curves, PRC curves and confusion matrix comparison

四分位差,定性资料表示为数量及占比(附表3-4)。在训练集中,大多数基线特征在S0~S3分期之间存在统计学显著差异( $P<0.05$ ),尽管部分指标(例如

乙型肝炎病毒血清阳性、TC)未表现出统计学上的分期差异,但根据其在先前文献中已确立的统计学显著性或潜在的临床意义而被保留用于建模。

表2 不同模型的性能:(A)随机森林模型(B)多层感知器模型  
Table 2 Performance of different models: (A) random forest model (B) multilayer perceptron model

(A)						
Class/parameter	AUC (95%CI)	Precision	Recall	F1-score	F2-score	Accuracy
S0	0.94 (0.87, 0.97)	0.606 6	0.902 4	0.725 5	0.822 2	0.592 3
S1	0.97 (0.93, 0.99)	0.900 0	0.450 0	0.600 0	0.500 0	
S2	0.75 (0.65, 0.84)	0.451 6	0.400 0	0.424 2	0.409 4	
S3	0.86 (0.76, 0.92)	0.607 1	0.500 0	0.548 4	0.518 3	
(B)						
Class/parameter	AUC (95%CI)	Precision	Recall	F1-score	F2-score	Accuracy
S0	0.96 (0.93, 0.99)	0.791 7	0.926 8	0.853 9	0.896 2	0.700 0
S1	0.99 (0.96, 1.00)	0.944 4	0.850 0	0.894 7	0.867 3	
S2	0.75 (0.64, 0.82)	0.486 5	0.514 3	0.500 0	0.508 5	
S3	0.88 (0.81, 0.93)	0.666 7	0.529 4	0.590 2	0.552 1	

表3 不同特征提取方法的性能:(A)仅临床数据特征(B)仅图像特征  
Table 3 Performance of different feature extraction methods: (A) only clinical data features (B) only image features

(A)						
Class/parameter	AUC (95%CI)	Precision	Recall	F1-score	F2-score	Accuracy
S0	0.89 (0.81, 0.95)	0.705 9	0.878 0	0.782 6	0.837 2	0.561 5
S1	0.69 (0.52, 0.79)	0.300 0	0.150 0	0.200 0	0.166 7	
S2	0.64 (0.43, 0.67)	0.392 9	0.314 3	0.349 2	0.327 4	
S3	0.80 (0.73, 0.90)	0.561 0	0.676 5	0.613 3	0.649 7	
(B)						
Class/parameter	AUC (95%CI)	Precision	Recall	F1-score	F2-score	Accuracy
S0	0.91 (0.86, 0.96)	0.700 0	0.853 7	0.769 2	0.817 8	0.546 2
S1	0.89 (0.67, 0.89)	0.600 0	0.450 0	0.514 3	0.473 7	
S2	0.69 (0.58, 0.78)	0.365 4	0.542 9	0.436 8	0.495 1	
S3	0.86 (0.82, 0.94)	0.615 4	0.235 3	0.340 4	0.268 7	

2.2 轻度脂肪肝类别中图像增强前后性能的分析

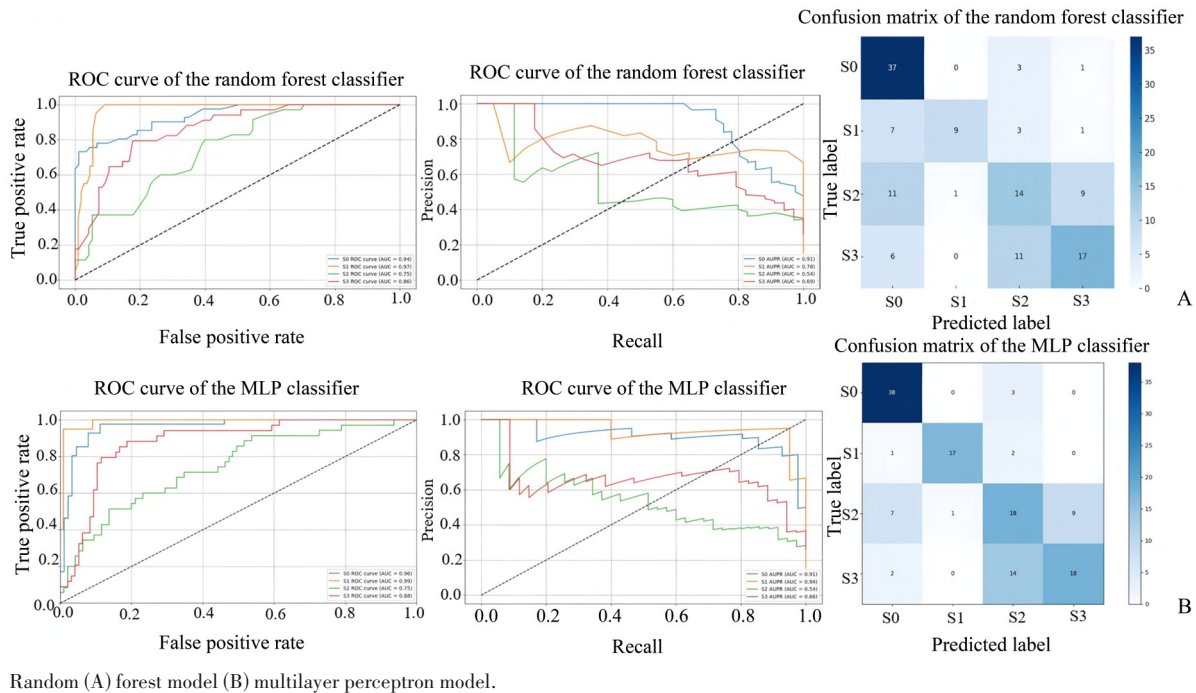
在本研究涉及的四分类脂肪肝任务中,患者数据存在显著的类别不平衡,其中S1类别样本相对较少(86例)。这种不平衡严重影响了模型学习,极大地限制了其对S1类别的识别能力。以RF模型为例,其对S1类别的召回率近乎为0,表明该类别几乎被完全忽略,从而降低了宏观平均精确度和F1分数。为缓解此问题,实施了一项专门针对S1类

别的图像增强策略。此举增加了S1类别在训练中的权重,并提升了模型的整体性能(表1;图4)。

2.3 模型比较分析:随机森林和多层感知机

对S1类别的图像进行增强后,我们训练并评估了2种分类模型:RF和MLP(表2;图5)。

结果显示,MLP模型的表现优于RF模型,尤其是在S1类别上。尽管RF模型在增强后有所改善,S1的AUC达到0.97,AUPR达到0.78,但其召回率



Random (A) forest model (B) multilayer perceptron model.

图5 不同模型的接收者操作特征曲线、精确度召回率曲线以及混淆矩阵

Fig. 5 The receiver operating characteristic curves, precision recall curves and confusion matrix of different models

仍然相对较低(0.45),F1分数保持在0.60的中等水平,表明其在该类别的识别上效果有限。相比之下,MLP模型在S1上实现了AUC 0.99和AUPR 0.94,精确率为0.94,召回率为0.85,F1分数为0.89,反映出识别准确性和整体性能的明显提升。

在整体性能方面,ROC曲线表明MLP模型表现优于RF模型。MLP在S2和S3上的AUC分别为0.75和0.88,RF则为0.75和0.86,表明MLP在中度和重度脂肪肝分级上具有更强的判别力。但是,AUPR曲线在S3上显示出相反的趋势,MLP的AUPR为0.66,而RF为0.69。

混淆矩阵分析证实了这些发现。尽管RF模型在增强后识别出了一些S1样本,但仍存在大量假阴性。MLP模型准确分类了17个S1样本,仅误判1个,表现出更高的精确率和稳定性。虽然2个模型在S0上表现良好,但MLP在S2和S3上错误更少,显示出明显改进。

#### 2.4 特征类型对模型性能影响的分析

为研究不同类型特征如何影响分类性能,我们构建了3种不同输入设置的MLP模型:①仅临床数据特征(表3A;图6A),②仅图像特征(表3B;图6B),③融合特征(表2B;图5B)。比较结果表明,融合特征模型在所有关键指标上都优于其他2个模型,特别是在更具挑战性的S1和S2类别

上(表3)。

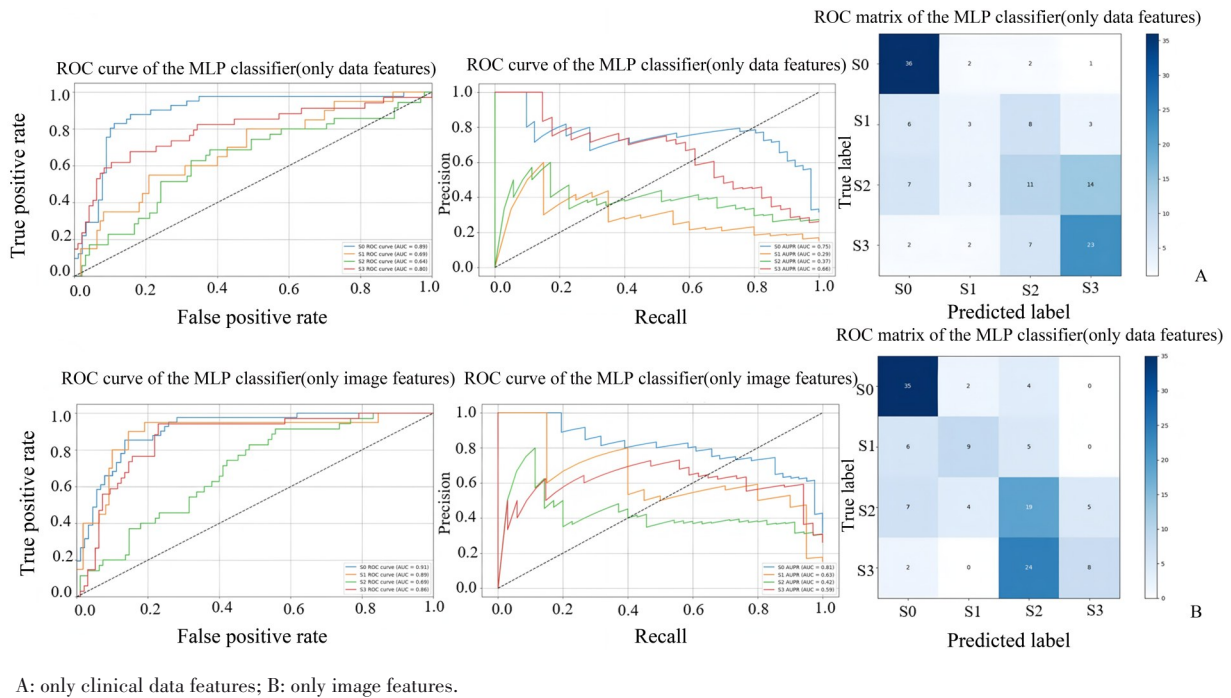
ROC曲线分析显示,融合特征模型在S0、S1、S2和S3类别上的AUC值分别为0.96、0.99、0.75和0.88。这显著优于仅临床数据模型(AUC:0.89、0.69、0.64、0.80)和仅图像模型(AUC:0.91、0.89、0.69、0.86),表明融合方法在所有类别上都具有更强的判别力。PR曲线也显示出相似的模式。融合特征模型在S1上的AUPR达到0.94,远高于仅临床数据模型(0.29)和仅图像模型(0.63),表明其在召回率和精确率之间取得了更好的平衡。

混淆矩阵证实了这些发现。仅临床数据模型在S1上表现不佳,仅识别出少量正确样本。仅图像模型略有改善,但仍存在许多误分类,特别是混淆了S2和S3。相比之下,融合特征模型对S2和S3的预测更为稳定,假阳性和假阴性显著减少。

#### 2.5 训练过程的可视化分析

为了分析不同特征类型对模型训练过程的影响,我们绘制了MLP分类器在3种输入设置下的学习曲线,即逐轮次的AUC曲线和相应的损失曲线:仅临床数据特征(图7A)、仅图像特征(图7B)、融合图像和数据特征(图7C)。

我们可以看出:使用临床数据特征训练的模型收敛迅速,AUC在前10个轮次内从0.65上升至0.90以上,并在第30个轮次左右稳定在1.0附近。



A: only clinical data features; B: only image features.

图6 不同特征提取方法的接收者操作特征曲线、精确度召回率曲线以及混淆矩阵

Fig. 6 The receiver operating characteristic curves, precision recall curves and confusion matrix of different feature extraction methods

而仅使用图像特征训练的模型收敛较慢,并且在初期表现出显著的波动。其AUC在大约80个轮次后逐渐达到0.82,而损失曲线则显示出大幅振荡(特别是在第50至70轮之间)和较慢的下降。

通过融合两种特征类型,多模态融合模型在20个轮次内迅速达到AUC 0.95,并在后续训练中保持在0.98以上。损失持续下降并稳定在0.1以下。

### 3 讨论

#### 3.1 主要发现

脂肪肝是全球慢性肝病的主要原因之一,可以进展为肝炎、纤维化、肝硬化和肝细胞癌<sup>[17-18]</sup>。目前诊断主要依赖于对二维超声图像的主观评估,无法准确分级<sup>[19-21]</sup>。这可能导致对轻度病例的过度治疗或对中重度病例的干预延迟。为解决此问题,我们开发了一种基于CLIP的MLP模型,该模型融合常规超声图像和临床数据,用于自动、客观的脂肪肝浸润(S0~S3)分级,以支持临床决策和预后判断。

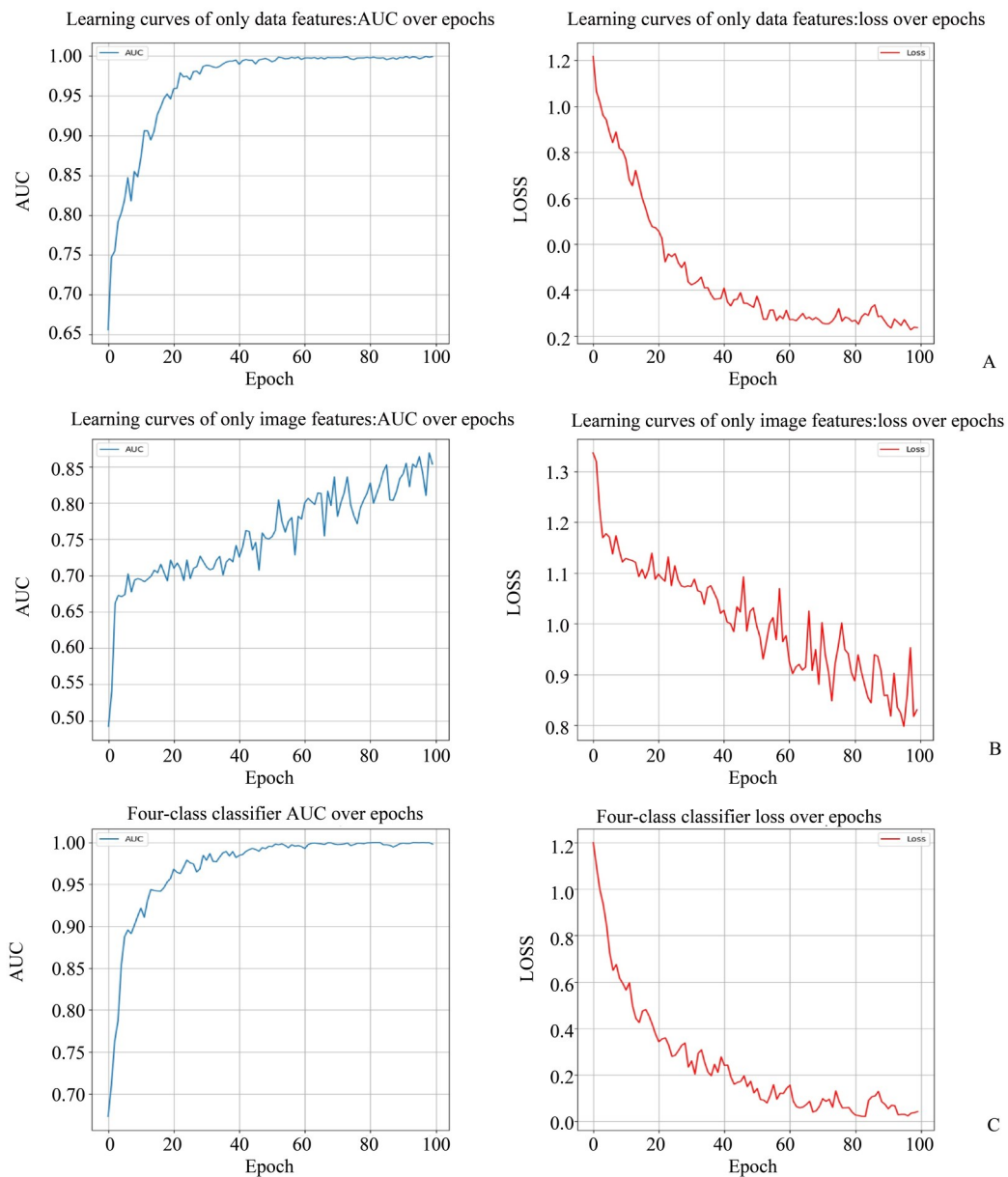
近期用于脂肪肝诊断的人工智能研究进展显示出利用多样化数据模式的趋势。Santoro等<sup>[22]</sup>

(2024)建立了一种基于人工智能的肝肾指数,显示出与MRI-PDFP标准具有很强的相关性。在定量指标的基础上,Kaffa等<sup>[23]</sup>(2025)证明,即使是常规超声图像也可被深度学习模型用于高精度地分类脂肪肝。在进一步推进分类复杂性方面,Guo等<sup>[24]</sup>(2024)利用专家标注和集成学习在四分类MASLD分层中实现了91.34%的准确率,凸显了复杂建模方法的潜力。

然而,现有方法面临3个主要局限性:数据类型单一;仅关注肝肾切面;难以获取的金标准而导致样本量受限,尤其是无及轻度脂肪肝患者的标本难以获得,所以部分研究会使用专家分级作为参考来改善样本量,但也引入了主观性。

本研究中应用临床数据促进了特征学习的加速,而图像特征尽管需要更长的训练时间,但提供了互补信息。融合临床数据和图像两种模态显著提升了训练效率(收敛速度比仅图像模型快33%)、稳定性(损失波动降低72%)和整体性能(最大AUC提升0.14),这表明多模态融合为分级任务提供了最优解决方案。

同时,通过采用ATI作为参考标准,在样本量和主观性之间取得了平衡——在减少偏倚的同时扩大了样本规模——从而能够开发出更稳健、更精



A: only data features; B: only image features; C: combined image and data features.

图7 不同特征提取方法的学习曲线

Fig. 7 Learning curves of different feature extraction methods

确的四分类模型。相较于临床另一常用的超声无创脂肪定量的受控衰减参数技术 (controlled attenuation parameter, CAP), ATI 在肝脂肪含量评估中展现出更优的诊断性能与应用优势。多项以  $^1\text{H-MRS}$ 、肝活检、MRI-PDFP 为金标准的研究证实, ATI 测得的衰减系数与肝脂肪分数相关性显著高于 CAP, 检测  $\geq 10\%$  肝脂肪变性时 AUC 达 0.939, 远优于 CAP 的 0.788<sup>[25]</sup>; 其在重度脂肪变性诊断及各分级评估中 AUC 均  $\geq 0.83$ , 性能稳定<sup>[26-27]</sup>。同时 ATI 观察者内一致性极佳 (ICC=0.931), 仅需 4 次测

量即可获可靠结果, 操作更便捷<sup>[25]</sup>。最新前瞻性多中心研究更直接证实, ATI 在所有肝脂肪变性分级中的诊断效能均优于 CAP<sup>[28]</sup>, 在脂肪变性精确分级中优势显著。值得注意的是, 尽管可能由于 ATI 截断值较窄<sup>[10, 29-30]</sup>, 本研究中 S1 组样本量偏小, 但通过针对性的图像增强与特征提取策略, 模型对该少数类别的识别能力得到显著改善, 最终在轻度脂肪肝识别中实现了 0.99 的 AUC, 体现出极高的模型稳健性——这一表现尤为可贵, 因传统超声对轻度脂肪肝的诊断敏感性仅为 17.7%~65.1%, 且结果高度

依赖操作者经验<sup>[31-32]</sup>。

本研究另一个重要的技术亮点为建立了常规二维超声图像的标准化采集与预处理流程,有效解决了临床超声图像异质性大的问题。研究严格规范了5个标准切面的采集要求,并对所有图像进行统一的尺寸归一化、裁剪及针对性数据增强,将标准化后的多切面图像整合为统一张量输入模型,保证了视觉特征提取的一致性和可靠性。超声图像的标准化为CLIP模型高效提取有效特征提供了基础,是多模态融合模型实现精准脂肪肝分级的重要前提,也为后续超声影像AI模型的临床转化提供了可复制的标准化方案。

该模型还支持在患者随访期间进行连续评估。值得注意的是,在合并纤维化的病例中,脂肪肝分级的降低可能并不表示改善——特别是如果肝脏硬度增加,则提示可能进展<sup>[33]</sup>。

我们的深度学习模型使得在活检、MRI和配备ATI中可行性较低的初级医疗的系统中,可以使准确分级得到广泛应用。

### 3.2 局限性

首先,研究数据均来自单中心佳能i900超声系统,模型泛化性受限,未来需开展多设备、多中心验证以确认结果普适性。其次,本研究尚未开展外部验证,后续将依托前瞻性队列进一步评估模型临床

应用价值。再次,模型复杂架构导致可解释性不足,计划整合LIME、SHAP等工具并开发可视化平台,以清晰呈现模型决策过程。最后,S2分级性能偏低(AUC=0.75),混淆矩阵(如图7)显示其在两种单模态下准确率仅为39.3%及36.5%,误判集中于相邻的S3级别,这与S2作为轻中度中间阶段的“特征模糊性”及样本学习偏差相关,可尝试合并中重度组以提升整体性能<sup>[23,34]</sup>。

尽管CLIP在视觉语言任务中表现优异,但直接迁移至医学影像时仍面临领域偏移问题<sup>[35]</sup>:一方面,原始CLIP侧重全局特征对齐,而脂肪肝诊断依赖局部微小病变特征;另一方面,预训练数据以自然图像为主,难以适配医学影像的细粒度病理分类需求,易导致类间相似性过高。因此,未来将引入MedCLIP或BiomedCLIP等专用医学视觉编码器,以消除领域偏移,进一步提升模型在医疗场景的泛化性与鲁棒性。

### 3.3 结论

本研究已经证实了基于多模态融合策略的深度学习模型能够有效克服传统分度方法的局限,为脂肪肝风险分层与预后管理提供可靠工具,并有望重塑其临床管理路径。



附表和附图  
Appendix table  
and figure

### 参考文献

- [1] Israelsen M, Francque S, Tsochatzis EA, et al. Steatotic liver disease[J]. *The Lancet*, 2024, 404(10464): 1761-1778.
- [2] Russo FP, Francque SM, Shawcross DL, et al. Advocating for the implementation of the new nomenclature for steatotic liver disease: a call to action for the national associations [J]. *J Hepatol*, 2024, 80(3): 384-386.
- [3] 张月霞, 王宇, 尤丛蕾, 等. 非酒精性脂肪肝患者肝功能异常相关因素分析[J]. *临床军医杂志*, 2025, 53(5): 522-524; +528.  
Zhang YX, Wang Y, You CL, et al. Analysis of factors associated with abnormal liver function in patients with non-alcoholic fatty liver disease [J]. *Clin J Med Officers*, 2025, 53(5): 522-524; +528.
- [4] Rinella ME, Lazarus JV, Ratziu V, et al. A multisociety Delphi consensus statement on new fatty liver disease nomenclature [J]. *J Hepatol*, 2023, 79(6): 1542-1556.
- [5] 张黄燕, 江杰, 胡宗强. 脂肪变性供肝缺血-再灌注损伤的发生机制及修复策略 [J]. *器官移植*, 2025, 16(4): 620-625.  
Zhang HY, Jiang J, Hu ZQ, et al. Mechanisms and repair strategies of ischemia-reperfusion injury in steatotic donor livers [J]. *Organ Transp*, 2025, 16(4): 620-625.
- [6] Chalasani N, Younossi Z, Lavine JE, et al. The diagnosis and management of nonalcoholic fatty liver disease: practice guidance from the American Association for the Study of Liver Diseases [J]. *Hepatology (Baltimore, Md.)*, 2018, 67(1): 328-357.
- [7] Pirmoazen AM, Khurana A, El Kaffas A, et al. Quantitative ultrasound approaches for diagnosis and monitoring hepatic steatosis in nonalcoholic fatty liver disease [J]. *Theranostics*, 2020, 10(9): 4277-4289.
- [8] Rinella ME, Neuschwander-Tetri BA, Siddiqui MS, et al. AASLD practice guidance on the clinical assessment and management of nonalcoholic fatty liver disease [J]. *Hepatology*, 2023, 77(5): 1797-1835.
- [9] Ferraioli G, Barr RG, Berzigotti A, et al. WFUMB guidelines/guidance on liver multiparametric ultrasound. Part 2: guidance on liver fat quantification [J]. *Ultras Med Biol*, 2024, 50(8): 1088-1098.
- [10] Pirmoazen AM, Khurana A, Loening AM, et al. Diagnostic

- performance of 9 quantitative ultrasound parameters for detection and classification of hepatic steatosis in nonalcoholic fatty liver disease [J]. *Invest Radiol*, 2021, 57(1): 23–32.
- [11] Wu S, Pan J, Song M, et al. Performance of magnetic resonance imaging and ultrasound for identifying the different degrees of hepatic steatosis: a systematic review and meta-analysis [J]. *Acad Radiol*, 2025, 32(11): 6528–6540.
- [12] Chen Z, Liu G, Zhang BW, et al. AltCLIP: altering the language encoder in CLIP for extended language capabilities [EB/OL]. (2023-12-31) [2026-02-04]. <https://aclanthology.org/2023.findings-acl.552/>.
- [13] Radford A, Kim JW, Hallacy C, et al. Learning transferable visual models from natural language supervision [EB/OL]. (2025-08-12) [2026-02-04]. <https://proceedings.mlr.press/v139/radford21a.html>.
- [14] Sugimoto K, Moriyasu F, Oshiro H, et al. The role of multiparametric US of the liver for the evaluation of nonalcoholic steatohepatitis [J]. *Radiology*, 2020, 296(3): 532–540.
- [15] Çetin N, Güneş Tatar İ, Yücege M, et al. Ultrasonographic evaluation of abdominal wall fat index, carotid intima-media thickness and plaque score in obstructive sleep apnea syndrome [J]. *Med Ultrasonogr*, 2019, 21(4): 422.
- [16] Kondo T, Abe M, Ueyama J, et al. Use of waist circumference and ultrasonographic assessment of abdominal fat distribution in predicting metabolic risk factors in healthy Japanese adults [J]. *J Physiol Anthropol*, 2009, 28(1): 7–14.
- [17] Leung PB, Davis AM, Kumar S. Diagnosis and management of nonalcoholic fatty liver disease [J]. *JAMA*, 2023, 330(17): 1687.
- [18] Thursz M, Lingford-Hughes A. Advances in the understanding and management of alcohol-related liver disease [J]. *BMJ*, 2023, 383: e077090.
- [19] de Moura Almeida A, Cotrim HP, Barbosa DBV, et al. Fatty liver disease in severe obese patients: diagnostic value of abdominal ultrasound [J]. *World J Gastroenterol*, 2008, 14(9): 1415.
- [20] Bril F, Ortiz-Lopez C, Lomonaco R, et al. Clinical value of liver ultrasound for the diagnosis of nonalcoholic fatty liver disease in overweight and obese patients [J]. *Liver Int*, 2015, 35(9): 2139–2146.
- [21] 徐海, 董帆. 棕色脂肪治疗代谢相关性脂肪肝病及影像学评估研究进展 [J]. *新医学*, 2022, 53(10): 723–726.
- Xu H, Dong Z. Research progress on brown adipose tissue in treatment of metabolic-associated fatty liver disease and imaging evaluation [J]. *J New Med*, 2022, 53(10): 723–726.
- [22] Santoro S, Khalil M, Abdallah H, et al. Early and accurate diagnosis of steatotic liver by artificial intelligence (AI)-supported ultrasonography [J]. *Eur J Int Med*, 2024, 125: 57–66.
- [23] Kaffas AE, Bhatraju KC, Vo-Phamhi JM, et al. Development of a deep learning model for classification of hepatic steatosis from clinical standard ultrasound [J]. *Ultras Med Biol*, 2025, 51(2): 242–249.
- [24] Guo L, Shi L, Wang W, et al. Neural network classification algorithm based on self-attention mechanism and ensemble learning for MASLD ultrasound images [J]. *Ultras Med Biol*, 2024, 50(9): 1361–1371.
- [25] Bao J, Lv Y, Wang K, et al. A comparative study of ultrasound attenuation imaging, controlled attenuation parameters, and magnetic resonance spectroscopy for the detection of hepatic steatosis [J]. *J Ultras Med*, 2022, 42(7): 1481–1489.
- [26] Podrug K, Sporea I, Lupusoru R, et al. Diagnostic performance of 2-D shear-wave elastography with propagation maps and attenuation imaging in patients with non-alcoholic fatty liver disease [J]. *Ultras Med Biol*, 2021, 47(8): 2128–2137.
- [27] Kuroda H, Abe T, Fujiwara Y, et al. Diagnostic accuracy of ultrasound-guided attenuation parameter as a noninvasive test for steatosis in non-alcoholic fatty liver disease [J]. *J Med Ultras*, 2021, 48(4): 471–480.
- [28] Nishimura T, Tada T, Akita T, et al. Diagnostic performance of attenuation imaging versus controlled attenuation parameter for hepatic steatosis with MRI-based proton density fat fraction as the reference standard: a prospective multicenter study [J]. *J Gastroenterol*, 2025, 60(6): 727–737.
- [29] Hobeika C, Ronot M, Guui B, et al. Ultrasound-based steatosis grading system using 2D-attenuation imaging: an individual patient data meta-analysis with external validation [J]. *Hepatology*, 2024, 81(1): 212–227.
- [30] Kwon H, Kim MG, Oh S, et al. Application of quantitative ultrasonography and artificial intelligence for assessing severity of fatty liver: a pilot study [J]. *Diagnostics*, 2024, 14(12): 1237.
- [31] Xu L, Lu W, Li P, et al. A comparison of hepatic steatosis index, controlled attenuation parameter and ultrasound as noninvasive diagnostic tools for steatosis in chronic hepatitis B [J]. *Digest Liver Dis*, 2017, 49(8): 910–917.
- [32] Kromrey ML, Ittermann T, Berning M, et al. Accuracy of ultrasonography in the assessment of liver fat compared with MRI [J]. *Clin Radiol*, 2019, 74(7): 539–546.
- [33] Tamaki N, Ajmera V, Loomba R. Non-invasive methods for imaging hepatic steatosis and their clinical importance in NAFLD [J]. *Nat Rev Endocrinol*, 2021, 18(1): 55–66.
- [34] Liu Y, Yu W, Wang P, et al. Deep learning with ultrasound images enhance the diagnosis of nonalcoholic fatty liver [J]. *Ultras Med Biol*, 2024, 50(11): 1724–1730.
- [35] Zhao Z, Liu Y, Wu H, et al. CLIP in medical imaging: a survey [J]. *Med Ima Ana*, 2025, 102: 103551.