·医学工程·

面向LR-M肝癌精准诊断的超声造影时空感知深度学习系统开发

林雅丹[1] 阮思敏[2] 陈立达[2] 何云[1*]

（1. 广西医科大学第一附属医院超声医学科，南宁 530021；2. 中山大学附属第一医院超声医学科，广州 510080）

【摘要】 目的 利用深度学习（deep learning, DL）结合超声造影（contrast-enhanced ultrasound, CEUS）视频开发一个时空混合感知的DL模型，从肝脏影像报告与数据系统（liver imaging reporting and data system, LI-RADS）LR-M类肝结节中无创诊断肝细胞癌（hepatocellular carcinoma, HCC）。**方法** 对来自中山大学附属第一医院和广西医科大学第一附属医院的经CEUS诊断为LR-M类的患者进行回顾性分析，以病理和综合诊断为标准，分为HCC组和非HCC组。使用R(2+1)D-18卷积神经网络对CEUS视频进行多尺度时空分时相并行特征提取，构建HCC诊断模型，并通过帧采样策略（15帧、30帧、45帧）平衡信息完整性与计算效率，辅以梯度加权分类激活映射技术（gradient-weighted class activation mapping, Grad-CAM）可视化解释决策依据。使用准确率、敏感度、特异度、阳性预测值、阴性预测值和受试者工作特征（receiver operating characteristic, ROC）曲线来评估模型的诊断效能。**结果** 来自中山大学附属第一医院的176例LR-M患者按7∶3划分训练集和验证集；来自广西医科大学第一附属医院的46例LR-M患者为测试集。当以45帧视频作为输入时，模型诊断效能最佳，在验证集和测试集中，敏感度为91.67%和84.21%，特异度为96.43%和81.48%，曲线下面积（area under curve, AUC）为0.940和0.828。与甲胎蛋白（alpha-fetoprotein, AFP）指标相比，模型对HCC的检出率更高（验证集：50.00%，测试集：31.58%；P值均<0.05），联合使用DL模型与AFP可以优化整体诊断效能。**结论** 基于CEUS时空感知的深度学习模型可以精准诊断LR-M类中的HCC，诊断效能显著优于AFP，两者联合诊断有望减少不必要的肝穿刺活检。

【关键词】 超声造影；肝脏影像报告与数据系统；肝细胞癌；深度学习；诊断效能

【中图分类号】 TH77；R445.1 **【文献标识码】** A **【DOI】** 10.11876/mimt202503001

Development of a spatiotemporal perception deep learning system based on contrast-enhanced ultrasound for precise diagnosis of LR-M hepatocellular carcinoma

Lin Yadan[1], Ruan Simin[2], Chen Lida[2], He Yun[1]*

(*1. Department of Medical Ultrasound, the First Affiliated Hospital of Guangxi Medical University, Nanning 530021, China; 2. Department of Medical Ultrasound, the First Affiliated Hospital of Sun Yat-sen University, Guangzhou 510080, China*)

【Abstract】 Objective To develop a spatiotemporal hybrid perception deep learning (DL) model using contrast-enhanced ultrasound (CEUS) videos for diagnosis of hepatocellular carcinoma (HCC) in liver imaging reporting and data system (LI-RADS) category LR-M liver nodules. **Methods** This study conducted a retrospective analysis of patients diagnosed with LR-M nodules by CEUS from the First Affiliated Hospital of Sun Yat-sen University and the First Affiliated Hospital of Guangxi Medical University. Using pathological and comprehensive diagnostic results as the gold standard, the patients were categorized into HCC and non-HCC. A multi-scale spatiotemporal split-phase parallel feature extraction of CEUS video using R(2+1)D-18 convolutional neural network is used to construct a diagnostic model for HCC, and the information integrity and computational efficiency are balanced by a frame

基金项目：国家自然科学基金（82272076）

第一作者：林雅丹，女，硕士研究生，研究方向：医学影像与人工智能

通信作者：何云，女，硕士研究生，主任医师，研究方向：医学影像与人工智能

sampling strategy (15/30/45 frames), supplemented by gradient–weighted class activation mapping (Grad–CAM) visualisation to explain the decision basis. The diagnostic efficacy of the model is evaluated using accuracy, sensitivity, specificity, positive predictive value, negative predictive value and receiver operating characteristic (ROC) curve. **Results** A total of 176 LR–M patients from the First Affiliated Hospital of Sun Yat–sen University were divided into training and validation sets in a 7:3 ratio, while 46 LR–M patients from the First Affiliated Hospital of Guangxi Medical University served as the test set. When using 45–frame videos as input, with sensitivities of 91.67% and 84.21%, specificities of 96.43% and 81.48%, and area under curves (AUCs) of 0.940 and 0.828 in the validation and test sets. The model had a higher detection rate of HCC compared with the alpha–fetoprotein (AFP) index (validation set: 50.00%, test set: 31.58%; both P value < 0.05), and the combined use of the DL model with AFP could optimize the overall diagnostic efficacy. **Conclusion** The DL model with CEUS spatiotemporal perception enables accurate diagnosis of HCC in LR–M category liver lesions, demonstrating significantly superior performance over AFP. Moreover, the combined diagnosis of the both can reduce unnecessary liver biopsy.

【**Key words**】 Contrast–enhanced ultrasound; Liver imaging reporting and data system; Hepatocellular carcinoma; Deep learning; Diagnostic performance

肝癌是全球第6大常见癌症和第3大癌症死亡原因[1]。在我国，肝癌的发病率位列第4，致死率高居第2位[2]。肝细胞癌（hepatocellular carcinoma, HCC）是原发性肝癌的常见病理类型。肝癌的恶性程度高且预后差，因此早筛、早诊、早治对于改善患者预后至关重要[3–4]。

在肝癌高危人群中，超声造影（contrast–enhanced ultrasound, CEUS）检查因其实时、动态及可重复的优势，已成为诊断HCC的有力手段。CEUS结合肝脏影像报告与数据系统（liver imaging reporting and data system, LI–RADS）可以对HCC进行无创诊断，通过特异的影像学征象将肝结节分为LR–1类（绝对良性）至LR–5类（绝对HCC）和LR–M类（明确或可疑恶性但非HCC特异性）[5–6]。然而，研究[7]发现，CEUS LI–RADS无法同时实现对HCC的高敏感度和特异度，尤其是LR–M类中有48%的病变为HCC，须依赖穿刺活检，增加了并发症的风险。有研究[8]通过优化分类标准和联合肿瘤标志物如甲胎蛋白（alpha–fetoprotein, AFP），可从LR–M类中检出部分HCC，但流程的复杂性可能会增加临床工作的负担，存在耗时较长、依赖医师经验等局限性[9–10]。

在当前大数据驱动的人工智能时代，深度学习（deep learning, DL）技术的快速迭代与工程化应用为肝癌的早期检测和无创诊断提供了有创新性的技术路径。基于深度神经网络的智能算法能够从大量医学影像数据中自动提取从微观像素级纹理到宏观结构的多尺度特征向量，并通过端到端的反向传播算法实现模型参数的梯度下降优化[11]。卷积神经网络（convolutional neural network, CNN）作为具有局部连接和权值共享特性的经典深度学习架构，主要由卷积层、池化层和全连接层组成。卷积层通过卷积核提取数据的局部特征，是CNN的核心；池化层对特征图下采样，降低维度和计算量；全连接层整合特征并输出预测结果。为增强学习能力和泛化性，CNN还融合了激活函数、损失函数、反向传播、正则化和迁移学习等技术，在医学图像的目标检测、语义分割和分类等任务中已展现出优异的特征提取能力和泛化性能[12–14]。而目前医学影像研究大多采用静态图像分析的处理流程，其系统架构难以对造影剂在肝结节中灌注过程的时变血流动力学参数进行动态建模，从而限制了模型性能。为此，本研究通过分析CEUS视频构建了一个时空混合DL模型，旨在从LI–RADS LR–M类肝结节中准确诊断出HCC，以期为临床提供更安全、可靠的肝癌无创诊断方案。

1 资料与方法

1.1 一般资料

本研究对中山大学附属第一医院和广西医科大学第一附属医院2019年1月至2023年2月期间收治的LR–M类肝癌高危患者的CEUS视频进行回顾性

分析。纳入标准：①患者伴有HCC高危因素，如肝硬化、慢性乙型病毒性肝炎或既往HCC病史；②经LI–RADS算法判定为LR–M类；③经组织学检查明确病理类型，具备CT/MRI典型的增强模式并伴有明确的肝外恶性肿瘤病史。排除标准：①年龄小于18周岁；②超声造影图像质量不合格。本研究经医院伦理委员会审核批准（伦理编号：2025–E0560）。

1.2 CEUS扫描

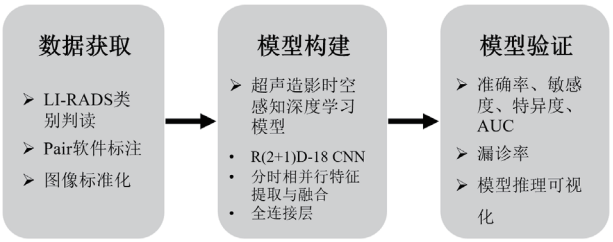
本研究采用了LOGIQ E9（GE Healthcare, Chicago, IL, USA）、EPIQ7（Philips Healthcare, Amsterdam, Netherlands）、Resona7（Mindray Medical, Shenzhen, China）、Aplio 500（Canon Medical Systems, Otawara, Japan）等彩色多普勒超声诊断仪，使用宽频凸阵探头，频率为1 ~ 6 MHz，机械指数(mechanical index, MI)为0.07 ~ 0.10。①常规灰阶超声检查：扫查整个肝脏后重点观察肝内目标病变，选取该病变最大切面，对病变大小进行测量。②CEUS检查：以常规超声图像选取的切面作为最佳切面，经肘前静脉团注适量声诺维超声造影剂（SonoVue, Bracco Suisse SA, Geneva, Switzerland），后以5 mL生理盐水冲管。在造影剂注入的同时启动计时器，并持续采集90 s动态视频，此后每分钟记录10 s，一直间断观察至6 min或微泡在循环中消除。

1.3 CEUS视频预处理

（1）LI–RADS类别判读：CEUS视频删除患者有关信息，并以avi格式储存，由两名具有10年以上肝脏CEUS工作经验的超声医师按照美国放射学会指南《Contrast–Enhanced Ultrasound Liver Imaging Reporting and Data System v 2017（CEUS LI–RADS v2017）》对肝结节进行LI–RADS分类，如果两名医师无法达成一致，则由一名有超过15年肝脏CEUS工作经验的专家作出最终裁决。

（2）结节标注：由超声医师使用Pair软件（https://aipair.com.cn/en/）标注首帧中结节的感兴趣区域（region of interest, ROI），基于跟踪算法实现自动标注，标注过程中由医师对ROI进行调整与修改，保证标注的准确性。

（3）图像标准化：统一调整为112 px × 112 px分辨率，训练阶段采用随机翻转、动态范围扰动和增加噪声进行数据扩增，避免过拟合，验证和测试阶段使用原始数据。研究设计见图1。



注：LI–RADS为肝脏影像报告与数据系统；CNN为卷积神经网络；AUC为曲线下面积。

图1 深度学习诊断系统开发设计流程

1.4 DL模型构建

本研究的目标是构建深度学习模型以便从CEUS视频中自动提取肝结节灌注过程中的时空动态特征，因此需要三维卷积神经网络（3D CNN）进行特征提取，然而传统的3D CNN直接耦合时空特征，即通过3D卷积核同时沿空间（H × W）和时间（T）维度滑动，学习空间和时间特征的联合表示，计算复杂度高，且在医学图像这种小样本场景下，可能会因数据量不足导致过拟合[15]。本研究使用R(2+1)D–18 CNN模型架构，采用空间—时间解耦策略，将传统三维卷积分解为空间2D卷积和时间1D卷积的级联[16]。具体而言，中间特征通道数通过以下公式动态调整以平衡参数量：

$$M_i = \left| \frac{t d^2 N_{i-1} N_i}{d^2 N_{i-1} + t N_i} \right| \tag{1}$$

其中，M_i为参数量，t为时间卷积核大小，d为空间卷积核尺寸，N_{i-1}和N_i分别为输入和输出通道数。

空间2D卷积提取肝结节在特定时相的形态学特征如结节增强程度、增强形态等，而时间1D卷积则捕捉病灶在动态增强过程中的血流动力学变化如增强时序、廓清速度等，实现更高效、更可解释的特征建模；空间—时间解耦策略也使得模型添加了更多ReLU激活层，更容易训练优化。模型还整合了残差神经网络（residual neural network, ResNet）的残差学习机制，其核心思想是“捷径连接（shortcut connection）”，将输入直接传递到输出端，与经过卷积层处理后的特征相加，可以表示为：$F(x)+x$，其中，x是输入，$F(x)$是经过卷积层等变换后的输出，这种结构使得梯度可以直接通过捷径连接传播，有效缓解了梯度消失问题；同时提出“恒等映射（identity mapping）”的概念，即当某些层对输入特征的变换并未带来性能提升时，网络能够通过捷径

连接直接传递输入，实现对冗余变换的跳过，增强了网络的特征表达能力[17]。

临床上，CEUS视频分为动脉期、门脉期和延迟期，动脉期为注射造影剂后10～20 s至30～45 s，主要观察病灶增强模式，门脉期和延迟期分别为30～45 s至120 s和120 s以后至微泡消失，这2个期相主要观察病灶的廓清情况。为准确建模各时相的差异化影像特征，使之符合临床诊断逻辑，模型采用分时相并行特征提取架构，研究根据临床应用指南将肝脏CEUS视频按照固定时间点划分为上述的3个时期，对各期相视频进行时域独立特征提取，并构建跨时相融合模块将3个期相的特征进行拼接，最后由全连接层与Dropout正则化执行最终决策。

1.5 DL模型训练与验证

本研究采用帧采样策略优化模型输入，以平衡血流灌注信息完整性与高帧频数据冗余性之间的矛盾。基于CEUS视频的3期动态特征，设计了3种等间隔关键帧提取方案：①15帧视频（3期各5帧）；②30帧视频（3期各10帧）；③45帧视频（3期各15帧）。通过系统比较3种采样策略的模型表现，在保证时空特征完整提取的同时优化计算效率，最终实现模型性能的最佳平衡。

模型使用Pytorch1.10构建，在配置16 GB内存的Nvidia RTX 4060Ti GPU上进行训练。采取动态早停机制（early stopping）调整训练周期防止模型过拟合，每批次处理8个病例，模型的输入为（batchsize, channel, frame, height, width）。学习率初始设定为0.001，采用余弦退火的方法更新学习率，使用Adam优化器（权重衰减系数为0.01）和BCEWithLogitsLoss二元交叉熵损失函数进行网络权重的调整和算法优化。

1.6 DL模型可视化

本研究采用梯度加权分类激活映射技术（gradient–weighted class activation mapping, Grad–CAM）对在模型决策中具有贡献性的关键区域进行可视化呈现，以展示其推理过程[18]。Grad–CAM技术提取CNN最后一层特征图的梯度权重，将得到的权重线性组合输入ReLU激活函数，从而生成热力图，表示为：

$$L^{c}_{\text{Grad-CAM}} = \text{ReLU}\left(\sum_{k=1}^{K} \alpha_k^c A^k\right) \quad (2)$$

其中，$L^{c}_{\text{Grad-CAM}}$为类别c的热力图，α_k^c为特征图A^k对类别c的权重，A^k为卷积层输出的第k个特征图。

1.7 统计学方法

使用SPSS 27.0和Medcalc 23.0.6软件进行数据统计分析。采用均值±标准差（$\bar{x} \pm s$）形式来描述连续变量，而分类变量则以n（%）进行表示。使用准确率、敏感度、特异度和受试者工作特征（receiver operating characteristic, ROC）曲线来评价模型在诊断HCC方面的表现，以χ^2检验比较组间差异，DeLong检验比较ROC曲线的性能。以P<0.05表示差异具有统计学意义。

2 结果

2.1 患者基线资料

本研究共纳入222例患者，来自中山大学附属第一医院的176例患者按7∶3比例划分为训练集（n=124）和验证集（n=52），训练集平均年龄为（56.7±10.7）岁，其中HCC 49例（39.5%），验证集平均年龄为（53.9±12.5）岁，其中HCC 24例（46.2%）；测试集（n=46）来自广西医科大学第一附属医院，平均年龄为（56.3±11.1）岁，其中HCC 19例（41.3%）。见表1。

表1 患者基线资料

项目	训练集（n=124）	验证集（n=52）	测试集（n=46）
年龄（岁，$\bar{x} \pm s$）	56.7±10.7	53.9±12.5	56.3±11.1
性别［n（%）］			
男	92（74.2）	34（65.4）	32（69.6）
女	32（25.8）	18（34.6）	14（30.4）
病灶大小（cm，$\bar{x} \pm s$）	4.72±2.17	5.77±3.28	4.16±2.03
病理或综合诊断［n（%）］			
HCC	49（39.5）	24（46.2）	19（41.3）
非HCC	75（60.5）	28（53.8）	27（58.7）
AFP［ng/mL, n（%）］			
≤200	102（82.3）	39（75.0）	40（87.0）
>200	22（17.7）	13（25.0）	6（13.0）

注：HCC为肝细胞癌；AFP为甲胎蛋白。

2.2 不同CNN对比实验

为验证所使用的R(2+1)D–18模型作为主要框架的有效性，本研究选取多种深度CNN进行对比实验，包括传统二维卷积2D ResNet–18模型，三维卷积I3D、3D ResNet–18模型和混合卷积MC3–18模型。我们统计了上述模型的参数量大小以及在15帧视频

作为输入时的分类准确率和推理耗时，具体如表 2 所示。可以看出，本研究使用的 R(2+1)D–18 模型参数量虽然较大，但在分类准确率上有显著优势，达到了 84.62%，相较于 2D ResNet–18 提升了 19.24 个百分点，并且在推理速度上并不逊于其他模型。

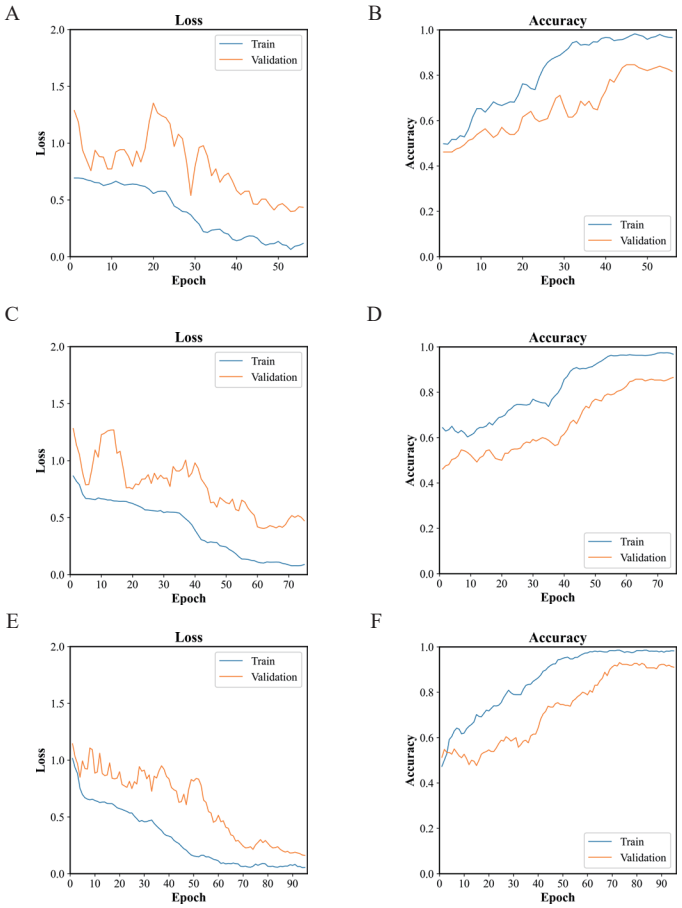
表 2 多种卷积神经网络框架参数和性能比较

模型	模型参数量（MB）	分类准确率（%）	推理耗时（ms）
2D ResNet–18	34.3	65.38	11.65
MC3–18	35.5	73.07	10.52
I3D	38.4	73.07	32.83
3D ResNet–18	99.5	80.77	11.14
R(2+1)D–18	93.9	84.62	11.48

注：模型参数量表示模型所有可训练参数的存储大小；推理耗时为模型推理时处理 1 例样本所消耗的时间。

2.3 DL 模型对 LR–M 类中 HCC 的诊断效能

使用不同帧采样策略作为输入时，模型训练过程中的损失曲线和准确率曲线如图 2 所示，总体来看，随着帧数的增加，损失下降的速度加快，准确率也有所提高，这表明帧数增加后模型的学习能力增强。模型诊断 HCC 的性能如表 3 和图 3 所示。45 帧视频作为输入时模型的表现最优，在验证集和测试集中，敏感度为 91.67% 和 84.21%，特异度为 96.43% 和 81.48%，曲线下面积（area under curve, AUC）为 0.940 和 0.828；在模型计算效率上，15 帧视频作为输入时，模型训练 1 轮耗时 22.99 s，推理平均每例耗时 11.48 ms；30 帧输入时分别为 43.91 s 和 21.76 ms，45 帧输入时为 55.75 s 和 24.96 ms，见表 4。结合模型的诊断效能和训练和推理耗时进行分析，随着输入帧数的增加，模型计算耗时有所增加，但差距仅为秒或毫秒级，在 GPU 加速环境下可忽略不计，但模型的性能得到了大幅度的提升。



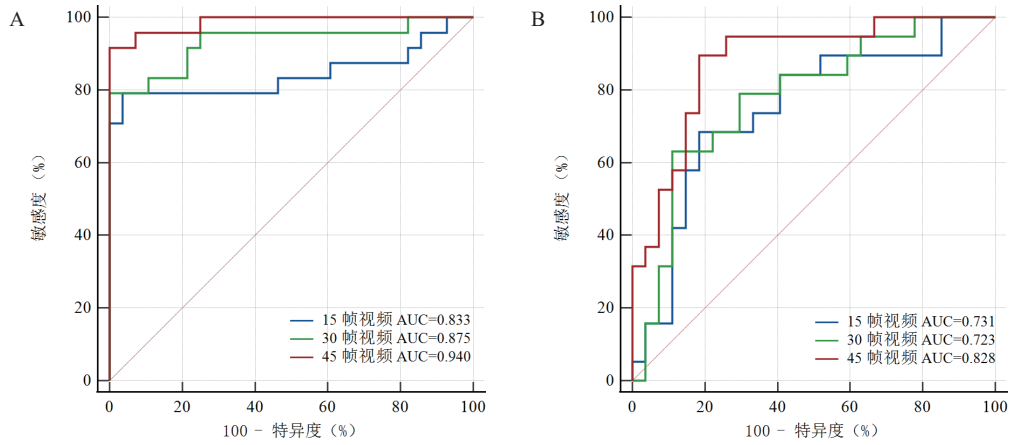
注：Loss 为损失值；Accuracy 为准确率；Epoch 为训练轮次；Train 为模型训练过程；Validation 为模型验证过程。

图 2 不同帧采样策略下模型训练和验证的损失曲线与准确率曲线 A 为 15 帧视频作为输入时训练和验证的损失曲线；B 为 15 帧视频作为输入时训练和验证的准确率曲线；C 为 30 帧视频作为输入时训练和验证的损失曲线；D 为 30 帧视频作为输入时训练和验证的准确率曲线；E 为 45 帧视频作为输入时训练和验证的损失曲线；F 为 45 帧视频作为输入时训练和验证的准确率曲线

表3 不同帧采样策略下模型对HCC的诊断效能比较

项目	准确率（%）	敏感度（%）	特异度（%）	阳性预测值（%）	阴性预测值（%）
验证集					
15帧视频	84.62	66.67	100.00	100.00	77.78
30帧视频	88.46	75.00	100.00	100.00	82.35
45帧视频	94.23	91.67	96.43	96.65	93.10
测试集					
15帧视频	73.91	68.42	77.78	68.42	77.78
30帧视频	73.91	63.16	81.48	70.59	75.86
45帧视频	82.61	84.21	81.48	76.19	88.00

注：HCC为肝细胞癌。



注：HCC为肝细胞癌；ROC曲线为受试者工作特征曲线；AUC为曲线下面积。

图3 验证集和测试集中不同帧采样策略下模型诊断HCC的ROC曲线 A为验证集；B为测试集

表4 不同帧采样策略下模型运行性能比较

	运行所需计算量（GFLOPs）	训练耗时（s/epoch）	推理耗时（ms）
15帧视频	4.37	22.99	11.48
30帧视频	8.05	43.91	21.76
45帧视频	11.76	55.75	24.96

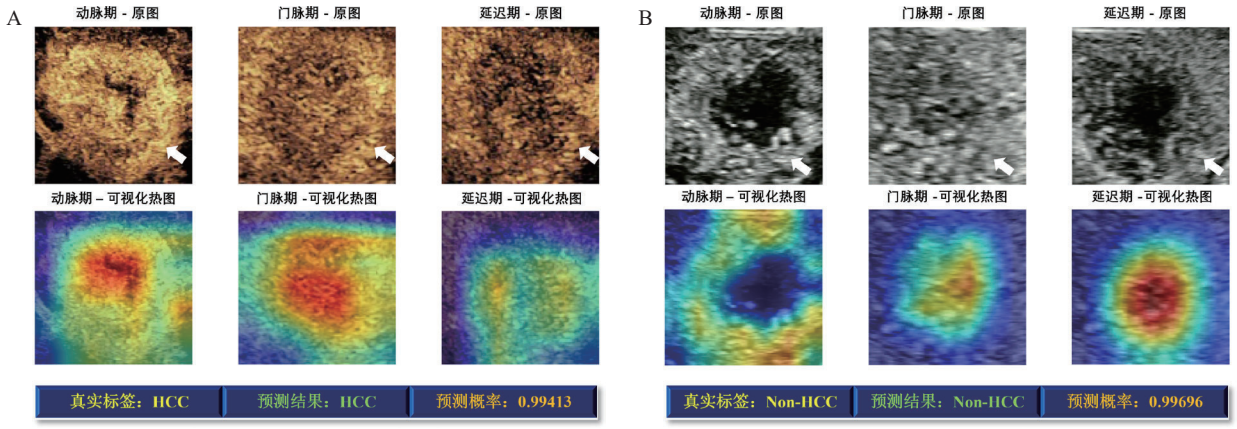
注：运行所需计算量用来衡量模型完成1次计算所需的浮点运算总量，单位为十亿次浮点运算/秒（Giga floating-point operations per second，GFLOPs）；训练耗时为模型训练1轮所消耗的时间；推理耗时为模型推理时处理1例样本所消耗的时间。

2.4 DL模型决策可视化

本研究采用Grad-CAM可视化算法，分析超声造影视频中各区域对模型诊断HCC的贡献度，见图4。结果显示，无论是HCC还是非HCC病灶，生成的热图均能凸显病灶区域，并且对LR-M类的关键影像特征也表现出特异性响应。在动脉期均匀强化病灶中，热力图集中于结节核心区域（图4A），而在环状强化病灶中，热力图显著激活于强化环区域（图4B），与影像特征空间相对应；在门脉期和延迟期，热力图持续聚焦病灶区域，通过与周围肝实质的对比强化差异实现廓清特征识别，这与超声医师的诊断思路高度一致，有助于提高AI辅助诊断的临床可解释性。

2.5 DL模型与肿瘤标志物AFP及两者联合时诊断效能与漏诊率的比较

本研究对比了AFP（>200 ng/mL）指标与DL模型对LR-M类中HCC的诊断效能，见表5和图5。结果显示，DL模型的诊断效能优于AFP，特别是敏感度（验证集91.67% *vs.* 50.00%，χ^2=6.750，P=0.006；测试集84.21% *vs.* 31.58%，χ^2=5.063，P=0.021）和AUC（验证集0.940 *vs.* 0.732，z=3.211，P=0.001；测试集0.828 *vs.* 0.658，z=1.769，P=0.076），HCC的检出率较高。AFP的特异度则较高（验证集96.43%，测试集100.00%），但其较低的敏感度会导致许多病例漏诊。研究还比较了联合AFP和DL模型的辅助诊断方法，将二者并联整合，只要满足任一条件即AFP值>200 ng/mL或



注：DL为深度学习；HCC为肝细胞癌；在可视化结果中，关键特征区域以暖色调（红色）突出显示，次要特征区域以冷色调（蓝色）标识。

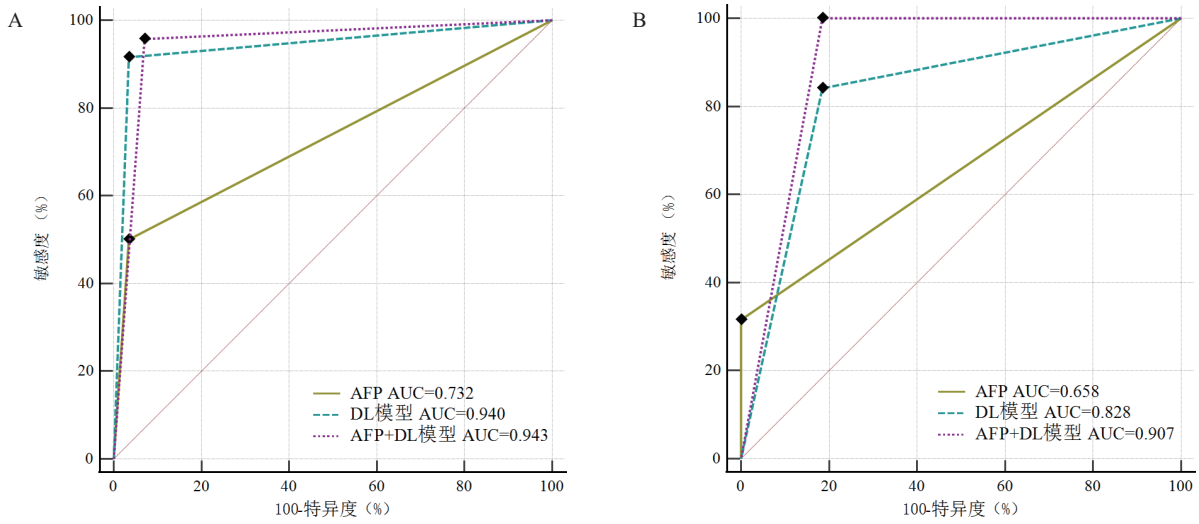
图 4 DL 模型推理决策的可视化结果 A 为 1 例肝细胞癌超声造影的原始图像和基于 DL 模型推理的可视化结果；B 为 1 例转移性肝癌超声造影的原始图像和基于 DL 模型推理的可视化结果

者 DL模型输出 HCC概率 >0.5即诊断 HCC，结果显示联合方法可以优化整体诊断效能，在验证集和测试集中 AUC分别为 0.943和 0.907，平衡了敏感度（95.83%和 100.00%）以及特异度（92.86%和 81.48%）。

表 5 肿瘤标志物 AFP 与 DL 模型对 HCC 的诊断效能比较

	准确率(%)	敏感度(%)	特异度(%)	阳性预测值(%)	阴性预测值(%)
验证集					
AFP	75.00	50.00	96.43	92.31	69.23
DL模型	94.23	91.67[a]	96.43	96.65	93.10
AFP+DL模型	94.23	95.83	92.86	92.00	96.30
测试集					
AFP	71.74	31.58	100.00	100.00	67.50
DL模型	82.61	84.21[b]	81.48	76.19	88.00
AFP+DL模型	89.13	100.00	81.48	79.17	100.00

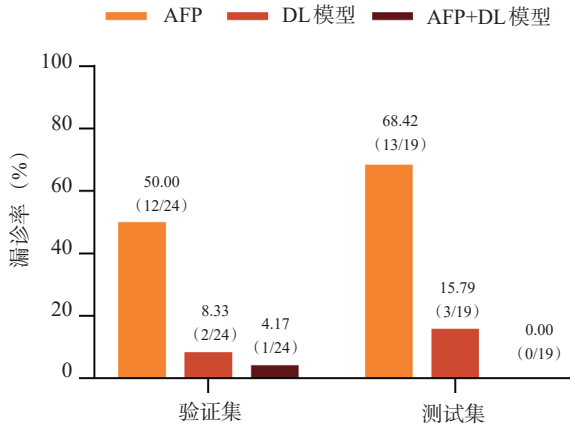
注：AFP为甲胎蛋白；DL为深度学习；HCC为肝细胞癌；[a]与验证集 AFP比较，P=0.006；[b]与测试集 AFP比较，P=0.021。



注：AFP为甲胎蛋白；DL为深度学习；HCC为肝细胞癌；ROC曲线为受试者工作特征曲线；AUC为曲线下面积。

图 5 肿瘤标志物 AFP 与 DL 模型以及两者联合应用诊断 HCC 的 ROC 曲线 A 为验证集；B 为测试集

在临床实践中，经LI-RADS算法判定为LR-M类的肝结节通常需要活检以明确病变性质，本研究通过整合AFP与DL模型，显著提高了HCC诊断准确性，有效减少了不必要的穿刺活检。如图6所示，单独使用AFP辅助诊断时，验证集和测试集的漏诊率为50.00%和68.42%，这部分HCC需要进一步穿刺活检；单独使用DL模型辅助时，漏诊率降至8.33%和15.79%；联合两者进行辅助后漏诊率进一步降低，分别为4.17%和0.00%。可以看出，AFP与DL模型联合进行HCC的辅助诊断，可显著降低漏诊率，意味着更多HCC病例能通过无创方式被准确识别，避免因不确定病变而接受不必要的穿刺活检。



注：AFP为甲胎蛋白；DL为深度学习；HCC为肝细胞癌。

图6 肿瘤标志物AFP与DL模型以及两者联合应用诊断HCC的漏诊率比较

3 讨论

本研究为从LR-M类肝结节中精准诊断HCC，成功构建了一个基于动态增强超声造影视频时空特征感知的R(2+1)D CNN深度学习模型。结果表明，DL模型对HCC的诊断效能显著优于AFP指标，将两者联合可以提升整体诊断效能，实现精准的肿瘤无创诊断。

近年来，许多研究针对CEUS LR-M类中HCC和非HCC恶性肿瘤的影像学特征重叠问题进行探索，通过分析影像的附加征象或结合临床信息优化诊断效能，但是诊断流程的复杂化会加重临床工作的负担，同时有赖于超声医师的诊断经验和能力[10, 19]。本研究创新性地引入人工智能技术，采用时空解耦的深度神经网络架构，提出了一种面向CEUS视频分析的时空特征感知融合诊断系统。基于预训练的R(2+1)D-18 CNN框架，将传统的3D卷积解耦成空间维度的2D卷积（提取形态学特征）和时间维度的1D卷积（捕捉血流动力学特征）的级联结构，并用ResNet的残差学习机制来构建深层网络，以提升自适应时空感知能力[20]。在特征提取策略上，本研究充分考虑了CEUS检查的临床特点，采用分时相并行特征提取架构：先独立分析动脉期、门脉期和延迟期的特征，再进行跨期相特征融合，这种处理方式模仿了超声医师的诊断逻辑，使模型的决策更贴合临床流程。为优化模型输入，系统比较了3种帧采样策略，发现45帧（3期各15帧）输入能在获得最佳诊断性能的同时保证计算效率，虽然增加帧数会轻微延长训练和推理时间，但可以显著提升诊断效能。

本研究还通过Grad-CAM技术对模型决策过程中关注的区域进行可视化，结果显示，模型能够智能地赋予结节区域更高的权重，该可视化系统通过多尺度特征卷积和空间注意力量化技术，实现了AI决策过程与临床诊断思维的对齐，可视化热图有效提升了深度学习模型的临床可解释性，为医疗AI系统的可靠性验证提供了量化评估工具[21]。

文献[22-23]指出，结合AFP（>200 ng/mL）重新分类LR-M类肝结节，可以在一定程度上提高HCC诊断敏感度，但效果不明显；Li等[9]研究发现，AFP在HCC诊断中敏感度较低，仅为41.5%，单独依靠AFP检测易漏诊大量HCC病例。本研究也证实了这一点，通过将DL模型与AFP对HCC的诊断表现进行比较，我们发现，AFP虽然表现出较高的特异度，但仅能检出少量HCC。DL模型在敏感度、AUC等指标上均显著优于AFP，表明DL模型在整体诊断效能上的优越性，当联合应用DL模型与AFP指标时，诊断效能得到进一步优化，漏诊率降低。这种联合策略可兼顾DL模型的高敏感度和AFP的高特异度，有望实现HCC的精准无创诊断，减少不必要的穿刺活检。

本研究有以下局限性：①数据仅来自于两个医疗中心，样本量有限且可能存在选择偏倚，未来需多中心验证。②模型尚未开发针对临床适用的用户

界面，未来考虑进行友好的用户界面开发并试用。③模型决策可解释性仍然有限，无法深入揭示模型关注的具体影像学特征，需结合更精细的解释技术提升可信度和适用性。

4 结论

综上所述，本研究基于超声造影视频构建了一个时空特征感知的R(2+1)D CNN深度学习模型，实现对LR-M类肝结节中HCC的精准识别，并通过联合AFP应用进一步提高诊断水平，为无创诊断HCC提供了高效、精准且具备良好可扩展性的技术手段，具有重要的临床应用潜力。

[参 考 文 献]

[1] Bray F, Laversanne M, Sung H, et al. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries [J]. CA Cancer J Clin, 2024, 74(3): 229–263.

[2] Xie D, Shi J, Zhou J, et al. Clinical practice guidelines and real–life practice in hepatocellular carcinoma: a Chinese perspective [J]. Clinical and Molecular Hepatology, 2023, 29(2): 206–216.

[3] Vitale A, Trevisani F, Farinati F, et al. Treatment of hepatocellular carcinoma in the precision medicine era: from treatment stage migration to therapeutic hierarchy [J]. Hepatology, 2020, 72(6): 2206–2218.

[4] 中华人民共和国国家卫生健康委员会医政司. 原发性肝癌诊疗指南（2024 年版）[J]. 中国实用外科杂志, 2024, 44(4): 361–386.

[5] Chernyak V, Fowler K J, Kamaya A, et al. Liver imaging reporting and data system (LI–RADS) version 2018: imaging of hepatocellular carcinoma in at–risk patients [J]. Radiology, 2018, 289(3): 816–830.

[6] Huang J Y, Li J W, Lu Q, et al. Diagnostic accuracy of CEUS LI–RADS for the characterization of liver nodules 20 mm or smaller in patients at risk for hepatocellular carcinoma [J]. Radiology, 2020, 294(2): 329–339.

[7] Terzi E, Iavarone M, Pompili M, et al. Contrast ultrasound LI–RADS LR–5 identifies hepatocellular carcinoma in cirrhosis in a multicenter restropective study of 1,006 nodules [J]. J Hepatol, 2018, 68(3): 485–492.

[8] Chan S L, Mo F, Johnson P J, et al. Performance of serum alpha–fetoprotein levels in the diagnosis of hepatocellular carcinoma in patients with a hepatic mass [J]. HPB (Oxford), 2014, 16(4): 366–372.

[9] Li C Q, Huang H, Ruan S M, et al. An assessment of liver lesions using a combination of CEUS LI–RADS and AFP [J]. Abdominal Radiology (New York), 2022, 47(4): 1311–1320.

[10] Huang W, Wen R, Wu Y, et al. Can modifications of LR–M criteria improve the diagnostic performance of contrast–enhanced ultrasound LI–RADS for small hepatic lesions up to 3 cm? [J]. J Ultrasound Med, 2023, 42(10): 2403–2413.

[11] Georgevici A I, Terblanche M. Neural networks and deep learning: a brief introduction [J]. Intensive Care Medicine, 2019, 45(5): 712–714.

[12] Hosny A, Parmar C, Quackenbush J, et al. Artificial intelligence in radiology [J]. Nature Reviews Cancer, 2018, 18(8): 500–510.

[13] Hu H T, Wang W, Chen L D, et al. Artificial intelligence assists identifying malignant versus benign liver lesions using contrast–enhanced ultrasound [J]. Journal of Gastroenterology and Hepatology, 2021, 36(10): 2875–2883.

[14] Liu F, Liu D, Wang K, et al. Deep learning radiomics based on contrast–enhanced ultrasound might optimize curative treatments for very–early or early–stage hepatocellular carcinoma patients [J]. Liver Cancer, 2020, 9(4): 397–413.

[15] Ji S, Yang M, Yu K. 3D convolutional neural networks for human action recognition [J]. IEEE Trans Pattern Anal Mach Intell, 2013, 35(1): 221–231.

[16] Tran D, Wang H, Torresani L, et al. A closer look at spatiotemporal convolutions for action recognition [C] //2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 18–23, 2018, Salt Lake City, UT, USA. IEEE, 2018: 6450–6459.

[17] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition [C] //2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 27–30, 2016, Las Vegas, NV, USA. IEEE, 2016: 770–778

[18] Selvaraju R R, Cogswell M, Das A, et al. Grad–CAM: visual explanations from deep networks via gradient–based localization [C] //2017 IEEE International Conference on Computer Vision (ICCV), 2017, Venice, Italy. IEEE, 2017: 618–626.

[19] Ding J, Qin Z, Zhou Y, et al. Impact of revision of the LR–M criteria on the diagnostic performance of contrast–enhanced ultrasound LI–RADS [J]. Ultrasound Med Biol, 2021, 47(12): 3403–3410.

[20] Shao Y, Dang Y, Cheng Y, et al. Predicting the efficacy of neoadjuvant chemotherapy for pancreatic cancer using deep learning of contrast–enhanced ultrasound videos [J]. Diagnostics (Basel), 2023, 13(13): 2183.

[21] Reyes M, Meier R, Pereira S, et al. On the interpretability of artificial intelligence in radiology: challenges and opportunities [J]. Radiol Artif Intell, 2020, 2(3): e190043.

[22] Gong W, Wu J, Wei H, et al. Combining serum AFP and CEUS LI–RADS for better diagnostic performance in Chinese high–risk patients [J]. La Radiologia medica, 2023, 128(4): 393–401.

[23] Wen R, Peng Y, Liang Y, et al. CEUS LI–RADS in combination with the serum biomarker–based ASAP model improves the diagnostic performance of HCC in high–risk patients [J]. Ultrasound Med Biol, 2024, 50(11): 1739–1744.