

DOI: 10.61189/653155lhssdo

· 产学研融合 ·

## 深度学习辅助精细评估肺结节影像

白春学<sup>1\*</sup>, 朱煜<sup>2</sup>

1. 复旦大学附属中山医院呼吸与危重医学科, 上海呼吸物联网医学工程技术研究中心, 上海市呼吸病研究所, 上海 200032

2. 华东理工大学信息科学与工程学院, 上海 200237

**[摘要]** 随着低剂量螺旋CT(low-dose computed tomography, LDCT)筛查的推广及胸部CT在体检、慢病管理和多学科诊疗中的广泛应用,肺结节检出率持续升高。肺结节管理的重点亦由“发现病灶”逐步转向“精细评估、风险分层和动态随访”。传统影像判读主要依赖结节直径、密度、边缘及增长趋势等指标,虽在临床实践中具有重要价值,但在小结节、血管旁结节、贴胸膜结节、亚实性结节及多发结节等复杂场景中,仍存在评估者差异大、重复性不足和纵向比较困难等问题。深度学习能够直接从二维、三维乃至多时点CT影像中自动提取多尺度表征特征,已被广泛应用于肺结节自动检出、精准分割、表型刻画、良恶性风险评估、动态随访及进展预测等多个环节。本文围绕肺结节影像精细评估的临床基础与管理框架,系统综述深度学习在肺结节检出与分割、影像表型精细刻画、良恶性分层决策、时间序列随访分析及临床转化方面的研究进展,并结合 Fleischner 学会指南、ACR Lung-RADS、BTS 指南及近年亚实性结节相关共识,讨论深度学习在真实世界应用中的主要瓶颈,包括外部验证不足、标注真值异质性、可解释性有限、概率校准不足及流程整合困难等。

**[关键词]** 肺结节;深度学习;低剂量螺旋CT;磨玻璃结节;影像精细评估;风险分层

**[中图分类号]** R 563 **[文献标志码]** A

### Deep learning-assisted fine assessment of pulmonary nodule images

Bai Chunxue<sup>1\*</sup>, Zhu Yu<sup>2</sup>

1. Department of Respiratory and Critical Care Medicine, Zhongshan Hospital, Fudan University; Shanghai Center for Respiratory Internet of Things Medical Engineering Technology; Shanghai Respiratory Research Institute; Shanghai 200032, China

2. School of Information Science and Engineering, East China University of Science and Technology, Shanghai 200237, China

**[Abstract]** With the increasing adoption of low-dose computed tomography (LDCT) screening and the widespread use of chest CT in health examinations, chronic disease management, and multidisciplinary care, the detection rate of pulmonary nodules has risen substantially. Accordingly, the focus of pulmonary nodule management has shifted from simple lesion detection to refined evaluation, risk stratification, and longitudinal follow-up. Traditional radiologic assessment mainly relies on nodule diameter, density, margin characteristics, and interval growth. Although these approaches remain clinically valuable, they are limited by interobserver variability, suboptimal reproducibility, and difficulty in longitudinal comparison, especially in small nodules, juxta-vascular nodules, pleural-based nodules, subsolid nodules, and multiple nodules. Deep learning can automatically extract multi-scale imaging representations from two-dimensional, three-dimensional, and longitudinal CT data. In recent years, it has been widely applied to pulmonary nodule detection, precise segmentation, phenotypic characterization, malignancy risk prediction, dynamic follow-up, and progression forecasting. This review summarizes the clinical basis of refined pulmonary nodule imaging assessment and the current management framework, and systematically discusses recent advances in deep learning for nodule detection and segmentation, radiologic phenotype characterization, malignancy risk stratification, longitudinal follow-up, and clinical translation. In addition, based on the Fleischner Society guidelines, ACR Lung-RADS, BTS guideline, and recent consensus statements on subsolid nodules, this review analyzes the major barriers to real-world implementation, including insufficient external validation, heterogeneity of reference standards, limited interpretability, poor probability calibration, and incomplete workflow integration.

**[Key Words]** pulmonary nodule; deep learning; low-dose computed tomography; ground-glass nodule; refined imaging

**[收稿日期]** 2026-02-26

**[接受日期]** 2026-03-07

**[基金项目]** 四大慢病重大专项(2024ZD0529300). Supported by Noncommunicable Chronic Diseases-National Science and Technology Major Project (2024ZD0529300).

**[作者简介]** 白春学, 博士, 主任医师、教授。

\*通信作者 (Corresponding author). E-mail: bai.chunxue@zs-hospital.sh.cn

assessment; risk stratification

肺癌长期位居全球肿瘤死亡原因前列,而肺结节则是肺癌早期发现、风险评估和干预决策中最常见、最关键的影像学切入点。随着低剂量螺旋CT (low-dose computed tomography, LDCT)筛查推广和胸部CT在体检、慢病管理及多学科诊疗中的广泛应用,临床面对的不再只是“能否发现肺结节”,而是“如何对肺结节进行精细评估、准确分层和长期管理”。美国国家肺癌筛查试验(National Lung Screening Trial, NLST)证实,LDCT筛查可使肺癌死亡率下降20%;荷兰-比利时NELSON研究进一步表明,基于体积CT的筛查可显著降低高危人群肺癌死亡率<sup>[1-2]</sup>。美国预防服务工作组2021建议将筛查起始年龄降至50岁、吸烟暴露阈值降至20包年;美国癌症学会随后也更新了相关建议<sup>[3-4]</sup>。肺结节诊治中国专家共识<sup>[5]</sup>鉴于早期肺癌患者平均年龄在50岁左右,建议将肺癌筛查年龄降至40岁,且具有下述任一危险因素者(1)吸烟指数 $\geq 400$ 支/年(或20包/年);(2)环境或高危职业暴露史(如石棉、铍、铀、氡等接触者);(3)合并慢性阻塞性肺病、弥漫性肺纤维化或曾患肺结核者;(4)曾患恶性肿瘤或有肺癌家族史者,尤其一级亲属家族史。推荐每年采用胸部LDCT对肺癌高危人群进行筛查。这些循证证据提示,未来进入临床管理路径的肺结节数量将持续增加,肺结节影像评估的标准化、自动化和精细化需求也将随之增强。

从临床实践看,肺结节的来源多种多样。一部分来自高危人群LDCT筛查,一部分来自体检或住院期间的偶然发现,另有相当数量出现于慢性阻塞性肺病、肺纤维化、感染后改变或肿瘤随访患者的常规影像检查中。不同来源结节的人群风险背景、结节类型构成、随访依从性和处置目标上存在明显差异。例如,筛查场景更强调结构化分层与总体流程效率,而偶发结节管理更重视个体化判断和既往病史整合。因此,肺结节影像学评估不仅是影像学技术问题,也是场景化管理问题。

肺结节管理的重点,已由“发现病灶”逐步转向“影像学性质判断、恶性风险估计、复查节奏制定以及进一步介入决策”。Fleischner学会2017指南为偶发肺结节管理建立了系统框架,ACR Lung-RADS v2022则在肺癌筛查场景中强化了结构化分类、增长定义和处置建议<sup>[6-7]</sup>。英国胸科协会(British Thoracic Society, BTS)指南及近年亚实性结节相关共识进一步强调,肺结节评估正在由“直径驱动”走

向“体积、实性成分、增长模式和时间维度共同驱动”的综合模式<sup>[8-10]</sup>。然而,传统影像判读在小结节、血管旁结节、贴胸膜结节、亚实性结节及多发结节等复杂场景中,仍面临评估者差异大、重复性不足及纵向比较困难等瓶颈。

传统阅片模式的局限并非仅体现在“看得见”与“看不见”,更体现在“看得准不准、量得稳不稳、前后比得清不清”。对于部分实性结节而言,微小实性成分的识别可能直接影响其管理等级;对于多发结节而言,如何识别主导病灶、如何评价不同病灶之间的风险差异,也是日常工作中的常见难题。再如,对于生长缓慢的磨玻璃结节,即使经验丰富的医师也可能在前后扫描层厚、重建核、窗位设定略有不同的情况下,难以准确判断其是否真实进展。这些问题使得肺结节管理越来越需要依赖定量化、客观化和可重复的技术支持。

深度学习作为人工智能的重要分支,能够直接从二维、三维及多时点CT数据中自动提取多尺度表征特征,近年来已逐步从肺结节自动检出扩展至精准分割、影像表型刻画、良恶性概率估计、进展预测和管理流程支持等多个环节<sup>[11-16]</sup>。与传统CAD不同,深度学习不再主要依赖人工构建的特征,而是能够通过大规模数据训练自主学习结节的空间模式、纹理异质性和时间变化规律,从而在复杂场景中展现出更高的适应能力。尤其是三维卷积网络、注意力机制、变换器结构以及时间序列建模方法的发展,使肺结节影像分析从单张图像识别走向跨层、跨期、跨模态综合理解成为可能。

因此,系统梳理深度学习辅助精细评估肺结节影像的研究进展及临床转化方向,具有重要理论意义和现实价值。本文将围绕肺结节影像精细评估的临床基础与管理框架、深度学习在检出与精准分割中的应用、影像表型刻画、良恶性风险评估、动态随访与进展预测,以及临床转化的关键挑战进行综述,以期为相关研究与实践提供参考。

## 1 肺结节影像精细评估基础及现有管理框架

1.1 肺结节影像精细评估的临床内涵 肺结节影像精细评估并不等同于单纯测量结节直径,而是包括结节数量、位置、大小、体积、体积倍增时间、密度类型、实性成分比例、边缘特征、内部结构、与周围血管及支气管关系,以及前后影像变化趋势等多个维度<sup>[6-10]</sup>。尤其对于亚实性结节和磨玻璃结节,影像

学评估不仅影响恶性风险判断,还关系到随访周期、活检策略及手术方式选择。

从放射学角度看,肺结节精细评估至少包含3个层面。(1)病灶本体特征,即结节大小、体积、密度、边缘、分叶、毛刺、空泡、钙化和脂肪等;(2)病灶与周围结构的关系,例如胸膜牵拉、血管束、支气管截断或扩张等;(3)病灶在时间维度上的行为,包括增长速度、实性成分变化和影像稳定性。临床上,真正决定管理路径的往往不是单一指标,而是这些信息的综合结果。例如,一个直径不大的部分实性结节,如果在短期内实性成分逐渐增加,其风险可能明显高于一个体积更大但多年稳定的纯磨玻璃结节。

此外,肺结节精细评估还具有明显的学科交叉特征。影像科关注的是识别与量化,呼吸科关注的是风险与随访,胸外科关注的是可切除性与手术时机,肿瘤科则更看重其与系统治疗策略的衔接。因此,现代意义上的“精细评估”不应只停留在影像描述,而应能直接服务于后续决策。换言之,精细评估的终点不是报告写得更长,而是让下一步处理更合理、更可解释、更一致。

**1.2 现有指南与风险分层体系** Fleischner学会2017指南<sup>[6]</sup>主要适用于偶发肺结节管理,强调依据结节性质、大小、单发或多发及患者风险因素制定随访方案。Lung-RADS v2022<sup>[7]</sup>则服务于筛查场景,突出结构化分类、增长阈值及风险升级路径。BTS指南<sup>[8]</sup>更重视体积和体积倍增时间在风险管理中的作用。此外,针对亚实性结节的共识<sup>[9-10]</sup>进一步提出应重视其长期随访及侵袭性演变风险。这些指南为肺结节管理提供了基础框架,但对于复杂影像表型、连续影像变化和跨中心一致性仍存在局限。

Fleischner学会指南的价值在于构建了偶发结节管理的普适性框架,其核心思想是将结节分类与患者基础风险结合起来。然而,该指南主要基于静态影像指标和经验性阈值,对于影像异质性强、形态复杂或时间行为不典型的结节,其指导仍需依赖医师个体判断。Lung-RADS则更强调筛查体系中的结构化输出,有助于提升不同中心间的沟通一致性和质量控制水平,但其标准化分类也可能对个别复杂病例存在一定简化倾向。

BTS指南更早地强调了体积测量与体积倍增时间的重要性,这一点与深度学习和自动分割的发展方向高度契合。实际上,体积及倍增时间比单纯直径更能反映结节真实生长情况,尤其对于不规则结

节和部分实性结节更具优势。近年围绕亚实性结节的共识进一步表明,对这类病灶不能仅靠一次测量做决策,而应重视其长期行为与侵袭性演变。因此,可以说现有指南已经提出了临床问题,但如何以更高效、更客观、更一致的方式实现这些管理要求,正是AI,尤其是深度学习,需要回应的核心命题。

**1.3 传统风险预测模型的贡献与不足** Mayo、Brock、Herder等经典模型<sup>[17-22]</sup>通过整合年龄、吸烟史、结节大小、部位、边缘形态和PET信息,实现了肺结节恶性概率的初步量化。这些模型在临床实践中具有重要价值,但其输入变量相对有限,对结节内部纹理异质性、局部微环境及纵向变化利用不足,也较难适应多发结节和亚实性结节等复杂场景。因此,基于原始CT数据直接建模的深度学习方法,逐渐成为新的研究热点。

经典模型最大的贡献,是把“经验性印象”转化为“概率化思维”。这种思维转变极其重要,因为它使肺结节管理开始具有风险分层意识,而不仅是二元化判断。然而,这些模型的变量多来自人工提取或临床登记表,难以反映影像中更细致、更连续的信息。例如,结节边缘不规则程度、内部纹理渐变、局部血管受牵拉的模式,并不能被简单地压缩为一个“有/无毛刺”或“上叶/非上叶”的变量。

此外,传统模型大多是在特定人群或特定场景中建立的,其可推广性也有限。筛查人群、偶发结节人群和既往肿瘤史患者的人群构成差异明显,单一模型往往无法在所有场景中都保持稳定表现。特别是在亚实性结节和多发结节中,经典模型的适配性往往更弱。因此,深度学习并不只是为了追求更高的AUC,而是为了弥补传统模型在复杂表征、时间维度和场景迁移上的不足。

## 2 深度学习在肺结节检出与精准分割中的应用进展

**2.1 从传统CAD到深度学习检测框架** 肺结节自动检出是精细评估的起点。传统CAD多依赖阈值分割、形状规则和手工特征提取,在血管旁结节、贴胸膜结节及复杂解剖背景中误报较多。深度学习尤其是卷积神经网络的发展,使肺结节检测从“规则驱动”转向“数据驱动”,显著提升了候选结节识别和假阳性排除能力<sup>[11-12,23-27]</sup>。

传统CAD的问题之一在于,肺内正常解剖结构本身就高度复杂。血管分叉、支气管断面、局灶性瘢痕及胸膜邻近密度改变,都可能在CT上呈现类似结节的外观。基于规则的方法需要预先设定许多

判别条件,但这些条件很难穷尽真实世界中的所有变异。深度学习则不同,它通过学习大量正负样本,逐渐形成对“结节模式”的统计性理解,能够更好地区分真正病灶与复杂背景。

值得注意的是,肺结节检测并不只是一个单纯的技术起点,而是整个后续流程的“入口质量控制点”。如果初始检出漏掉关键病灶,则再精细的后续分类和随访模型也无从发挥作用;如果误报过多,则会增加医师负担并降低系统可用性。因此,近年来很多研究不再单纯追求最高灵敏度,而是更强调在合理误报水平下的综合性能,这种理念也更贴近真实临床场景。

**2.2 公共数据库与技术演进** LIDC-IDRI 和 LUNA16 是肺结节 AI 研究最重要的公共基础。LIDC-IDRI 提供多专家标注的 1018 例胸部 CT 数据,为检测、分割和分类研究奠定了标准化资源基础;LUNA16 则建立了统一的算法评价框架<sup>[23-25]</sup>。在此基础上,肺结节检测模型经历了从 2D CNN、2.5D 多视图融合到 3D CNN 以及检测-分割联合模型的演进,多尺度特征融合、注意力机制和变换器思路也被逐步引入<sup>[11-12,23,26-27]</sup>。

2D CNN 的优势在于实现简单、训练成本相对较低,但它对层间三维信息利用不足,容易忽视结节在连续层面上的空间一致性。2.5D 方法通过同时输入轴位、冠状位和矢状位视图,在一定程度上弥补了二维信息不足。随着算力提升,3D CNN 逐步成为肺结节检测和分类的重要方向,因为肺结节本质上是三维结构,其形态、边缘及与周围组织关系都需要在体数据层面进行理解。

近年来,多尺度特征提取成为重要趋势。因为肺结节大小跨度很大,既有几毫米的小结节,也有边界复杂的较大病灶,单一尺度的卷积核往往难以兼顾。注意力机制和特征金字塔网络等方法,使模型能够同时关注局部细节和全局背景。更进一步,变换器类方法开始尝试增强长距离依赖建模能力,这对识别结节与周围血管、支气管关系具有潜在价值。

**2.3 精准分割在精细评估中的基础地位** 分割的准确性直接影响体积测量、倍增时间计算、实性成分量化、纹理提取和纵向配准结果。近年来,U-Net、3D U-Net、nnU-Net 及注意力分割网络已成为肺结节分割主流技术路线<sup>[11-12,28-30]</sup>。对于实性结节,这些模型在边界勾画和体积测量方面已取得较好表现;对于亚实性结节和磨玻璃结节,由于其边界模糊、密度渐变,精细分割仍是当前难点,也是后续侵

袭性预测和个体化管理的关键基础。

在临床中,许多重要指标都建立在准确分割之上。体积是评估生长速度的重要参数,体积倍增时间则与恶性风险密切相关;实性成分比例是部分实性结节分层的关键指标;纹理分析和影像组学特征提取同样依赖稳定、可重复的病灶勾画结果。因此,分割质量的提高不仅是图像处理精度问题,更直接关系到临床风险估计的可信度。

对于亚实性结节,精细分割尤其重要。原因在于这类病灶的边界与正常肺实质往往呈渐变过渡,且内部可能同时包含磨玻璃成分和实性成分。如果分割过粗,可能低估实性成分比例;如果分割不稳定,则前后随访比较的结果也会失真。深度学习分割模型在这一方向上的持续进步,意味着未来肺结节管理可以从“经验性估计”逐步过渡到“像素级定量”。

**2.4 困难场景下的检出与分割挑战** 深度学习模型在小结节、血管旁结节、贴胸膜结节、裂旁结节及多发结节场景中,仍面临鲁棒性考验<sup>[11-12,23-27]</sup>。尤其在 GGO 和部分实性结节中,边界模糊与病理异质性增加了检出和分型难度<sup>[28,30-32]</sup>。因此,未来研究应更加重视困难亚群的亚组分析和真实世界泛化验证,而非仅报告整体性能指标。

小结节的核心问题是信号弱、体素少,轻微噪声即可影响模型判断。血管旁结节则容易与血管截面或分叉混淆,这要求模型同时具备病灶识别与解剖理解能力。贴胸膜结节和裂旁结节则涉及边界缺失或与胸膜平行的复杂几何关系,也对分割算法提出更高要求。多发结节场景下,则不仅是“检出多个病灶”的问题,还涉及如何排序风险、如何识别主导病灶、如何避免多病灶之间的特征混淆。

这些困难场景之所以重要,是因为真实临床最常见的问题恰恰往往出现在这些边缘地带。一个模型若只在标准化数据集中的“典型病例”上表现良好,但面对复杂病例性能急剧下降,其临床价值有限。因此,今后研究应更加注重多中心、多协议、多场景验证,以及困难亚型的单独性能呈现,这 will 比单纯提高平均指标更有意义。

### 3 深度学习驱动的肺结节影像表型精细刻画

**3.1 从尺寸测量到多维影像表型** 传统肺结节评估多聚焦于长径、平均径或体积,而深度学习能够进一步学习结节形状复杂度、边缘不规则性、分叶、毛刺、胸膜牵拉、血管集束及周围肺实质特征等多维影像表型<sup>[13,33-36]</sup>。这种能力推动肺结节影像学从

“描述征象”向“刻画表型”转变,更接近结节真实的生物学行为。

表型刻画的意义在于,它能把以往由医师主观描述的征象转化为更连续、更定量的特征空间。传统报告中“边缘欠光整”“可见轻度分叶”“伴轻度胸膜牵拉”等表述,虽然临床上有价值,但不同医师之间理解并不完全一致。深度学习能够从原始数据中提取连续特征,有助于降低这种表述差异带来的信息损失。

更重要的是,某些影像风险信号本身可能并不容易用传统术语准确概括。模型所学习到的高维特征,未必能一一对应到现有影像征象,但它们可能真实地反映结节与恶性生物学行为之间的复杂关联。因此,深度学习带来的不仅是自动化,更是影像理解维度的扩展。

**3.2 深度学习与影像组学的互补关系** 影像组学通过人工定义的形状、灰度和纹理参数量化结节影像特征,而深度学习可自动学习更复杂、非线性的高维表征。两者并非相互替代,而是具有明显互补性<sup>[31,33-36]</sup>。近年来,影像组学与深度学习融合模型在肺结节侵袭性预测、高危病理分层和良恶性鉴别中取得了较为稳定的表现。

影像组学的优势在于可解释性较强、特征定义清晰,便于统计分析和模型外推;其不足则在于高度依赖分割质量和预处理标准,并且难以穷尽图像中的全部复杂信息。深度学习的优势在于自动学习能力强,能够发现人工未显式定义的模式;不足则是可解释性相对有限,且对数据量和域一致性要求较高。两者结合,实际上是在性能与可解释性之间寻求平衡。

从临床接受度角度看,这种融合路径也更有现实意义。纯黑箱模型虽然可能性能更高,但若缺乏解释和可视化支撑,往往难以获得临床充分信任。若能够将深度特征与影像组学特征、传统影像征象共同纳入模型,则有望形成更易被放射科、呼吸科和胸外科共同理解的决策框架。

**3.3 亚实性结节与磨玻璃结节的精细定量** 亚实性结节尤其是纯磨玻璃结节和部分实性结节,是肺腺癌谱系影像学的重要对象。深度学习可通过三维表征与细粒度分割,更准确识别微小实性成分、密度梯度和内部异质性,从而提高对 AAH/AIS、MIA 及 IA 的区分能力<sup>[9-10,28,30-32]</sup>。这对于决定继续观察、有限切除或手术升级具有重要意义。

磨玻璃结节管理之所以困难,是因为其影像与病理之间存在连续谱,而非简单二分。很多纯磨玻

璃结节长期稳定,适合观察;但一旦出现实性成分增加,其侵袭性风险明显上升。部分实性结节更是如此,微小实性核心的识别可能直接改变治疗策略。因此,深度学习在这一场景中的最大价值,不只是“分类”,而是帮助进行更细粒度的侵袭性评估。

此外,磨玻璃结节通常进展缓慢,传统人工判断容易受主观因素影响。借助深度学习与精细分割技术,可以更加稳定地量化其密度分布和结构演变,为长期监测提供客观依据。这对于避免过度治疗和避免延误治疗同样重要。

**3.4 结节微环境与局部解剖语境的信息价值** 越来越多研究提示,结节周围血管、支气管、胸膜及局部肺实质纹理异常同样承载重要风险信息。深度学习特别适合联合学习“结节+周围微环境”的多尺度特征<sup>[13,33-36]</sup>。未来肺结节精细评估将不再仅限于病灶本体分析,而有望逐步向包含局部解剖语境的综合建模转变。

临床上,放射科医师在判断结节时,本来就不会只盯着病灶本身,而会同时观察周围血管是否聚拢、胸膜是否受牵拉、邻近支气管是否变形。深度学习若仅输入孤立裁剪的小块图像,可能反而丢失了这些重要上下文信息。多尺度建模和更大视野输入,能够让模型更接近真实阅片逻辑。

这种“语境化评估”还有助于解释一些传统上难以量化的征象。例如,局部肺实质细微结构紊乱、血管走行异常等,在人眼看来只能模糊感知,而模型可能把它们转化为稳定的判别特征。这意味着未来肺结节影像评估将逐步从“病灶中心主义”走向“病灶—环境整体主义”。

## 4 深度学习辅助肺结节良恶性风险评估与分层

**4.1 从经验判断到概率化输出** 临床真正需要的不是单纯的“良/恶性”标签,而是能够支持随访、PET-CT、活检和手术决策的恶性概率估计。深度学习在肺结节风险评估中的优势,在于可基于原始CT影像直接输出概率分层结果,并与既往影像、临床变量进一步整合<sup>[13-16]</sup>。

概率化输出的意义在于,它更符合临床真实决策过程。绝大多数肺结节并不会在首次评估时被简单地定义为“肯定良性”或“肯定恶性”,而是落在不同程度的不确定区间中。对临床而言,最重要的是确定风险区间并匹配相应管理策略。深度学习若能提供可靠的概率估计,而非单纯标签,将更有助于与既有指南阈值和实际工作流衔接。

同时,概率输出还有助于沟通。对于患者而言,明确说明“当前风险较低,建议定期复查”往往比笼统地说“问题不大”更容易理解;对于MDT团队而言,风险量化结果也更利于形成共识。因此,深度学习在风险分层中的价值,不仅体现在算法层面,也体现在医患沟通和团队协作层面。

4.2 代表性研究及其启示 Ardila等<sup>[15]</sup>基于当前及既往LDCT体数据建立三维深度学习模型,实现了较高水平的肺癌风险预测。Venkadesh等则分别展示了单时点和引入既往CT的深度学习恶性风险评估框架,提示时间维度可显著提高三年恶性风险预测能力<sup>[13-14]</sup>。这些工作共同表明,肺结节风险不是静态概念,而是与时间轴密切相关的动态风险对象。

这些代表性研究的启示在于:(1)三维体数据本身包含比二维切片更多的信息,尤其在结节立体形态和与周围结构关系方面更具优势。(2)既往CT不应被视为附加资料,而应成为模型的重要输入,因为结节变化本身往往比单次静态外观更具有临床意义。(3)深度学习模型并非只能用于“发现病灶”,同样可以用于更高层级的风险推断和临床决策支持。

但是,应客观看待这些研究。许多高影响力成果仍主要集中于筛查人群或较为理想的数据条件下,而真实世界中的偶发结节、多发结节和复杂背景病例更加多样。因此,代表性研究更多是证明“可行性与潜力”,而非意味着问题已经完全解决。

4.3 多模态融合的趋势 近年来,研究者开始将深度学习影像特征与临床信息、影像组学、PET/CT及cfDNA等分子标志物整合,构建多模态风险模型<sup>[34,37,39-40]</sup>。这类模型更接近真实临床决策逻辑,有望提高阳性预测值并减少不必要的侵入性检查。

单纯依赖CT影像,虽然可以在很大程度上反映结节特征,但临床决策显然不会忽略患者年龄、吸烟史、恶性肿瘤既往史、家族史以及代谢和分子层面的信息。多模态融合的价值就在于将这些信息纳入同一风险框架,使模型更贴近真实世界。尤其对于边界性结节,多模态模型可能比单一影像模型更具区分力。

此外,多模态融合也有助于提升模型的临床可接受性。医生更容易接受一个“综合信息后的风险评估系统”,而不是完全忽略临床背景的纯图像黑箱模型。因此,未来的高价值模型很可能不是“影像替代一切”,而是“影像驱动下的多源信息整合”。

4.4 与指南分层路径的衔接 深度学习风险模型若

要真正进入临床,理想输出形式不应只是抽象概率值,而应转化为与Fleischner、Lung-RADS和BTS管理路径相匹配的“概率+建议”双层结果<sup>[6-8,13-16]</sup>。例如,对低危结节建议延长复查间隔,对中危结节建议短期复查或PET-CT,对高危结节建议MDT评估和组织学取材。

这实际上提出了一个重要转化原则:AI输出必须“可落地”。很多研究止步于模型性能比较,却没有回答临床最关心的问题——这个结果如何改变下一步处理。真正理想的系统应能够将风险评估自动映射到建议路径中,例如生成“建议6个月复查”“建议进一步PET-CT评估”“建议肺结节专门诊会诊”等结构化意见。只有当AI输出能够嵌入真实决策流程时,其价值才能被充分释放。

## 5 深度学习在肺结节动态随访中的价值

5.1 肺结节管理的时间轴特征 肺结节管理并非一次性判断,而是一个长期动态过程。对于许多边界性结节,真正决定临床策略变化的往往是后续随访中体积增长、密度变化和实性成分增加<sup>[6-10]</sup>。因此,任何高价值的肺结节AI系统都应具备一定的纵向分析能力。

从临床管理角度说,初次检查只是风险评估的起点,而不是终点。许多结节首次影像表现并不典型,尤其是亚实性结节,真正能提示侵袭性升级的往往是后续变化。因此,肺结节AI若仍停留在对单次CT的静态分类上,便只能解决部分问题。只有把时间维度纳入分析,模型才更接近临床真实需求。

此外,纵向管理还涉及随访依从性和流程连续性。若系统能自动调取既往影像并进行比较,不仅有助于提高分析精度,也能减少人为遗漏,提高专病管理效率。

5.2 双时点及多时点模型的优势 通过自动配准当前与既往CT,并联合输入多时点影像,深度学习模型能够更加敏感地识别细微增长和结构变化。Venkadesh等<sup>[14]</sup>的研究显示,既往CT信息可显著提高恶性风险评估;近年关于AI预测肺结节增长的综述也表明,时间序列学习正成为该领域的重要发展方向<sup>[41]</sup>。

双时点和多时点模型的本质,是让模型从“看图”转向“看变化”。这种变化可以是体积上的,也可以是密度、形态或实性成分上的。对部分生长缓慢但持续进展的病灶而言,这种能力非常关键。未来,如果模型能够进一步预测“未来一段时间内是否可能增长”“实性成分是否可能增加”,将更有助

于制定个性化随访方案。

**5.3 亚实性结节长期监测中的特殊价值** 亚实性结节常呈慢性进展过程,既存在过度治疗风险,也存在延误治疗风险。深度学习结合精细分割与时间序列分析后,有望辅助识别真正需要升级处理的病灶,并帮助制定更个体化的复查间隔<sup>[9-10,28,30-32,41]</sup>。

这是肺结节AI最具潜在临床价值的场景之一。因为亚实性结节的管理本身充满灰区:过早干预可能造成不必要手术和肺功能损失,过晚处理则可能错过最佳治疗窗口。深度学习若能帮助更早、更准确地识别风险升级征象,便可能显著改善这类患者的管理质量。

**5.4 从单点算法到管理平台** 未来,肺结节AI更可能以平台化形式存在,即完成自动检出、自动分割、自动调取既往影像、自动比较变化、自动生成风险提示和随访建议<sup>[7,14,16,38,41]</sup>。这一模式更适用于筛查中心、肺结节专病门诊、体检中心及MDT管理流程。

平台化的意义在于把分散的算法模块整合为完整流程。对临床而言,最需要的不是十个彼此孤立的高性能模型,而是一个能够稳定服务日常流程的系统。未来若能把结构化报告、随访提醒、异常升级提示和多学科转诊接口整合进去,肺结节管理将真正迈向闭环化。

## 6 临床转化挑战、争议与未来发展方向

**6.1 外部验证不足与数据偏倚** 虽然肺结节深度学习研究快速增长,但系统综述反复指出,当前模型的主要局限仍是外部验证不足、样本选择偏倚和方法学异质性<sup>[11-12,16,38]</sup>。许多模型在公开数据库或单中心回顾性数据中性能较好,但缺乏跨设备、跨地区和跨人群的真实世界验证。

数据偏倚不仅影响模型泛化,还可能导致某些人群或某些结节类型的性能被系统性高估或低估。例如,训练集若主要来自筛查人群,则模型应用于偶发结节时可能表现下降;若训练集以实性结节为主,则对亚实性结节的识别能力可能不足。因此,外部验证不只是“再测一次性能”,而是检验模型是否具备临床推广基础的关键步骤。

**6.2 真值标注与标签异质性问题** 肺结节良恶性标签并不总是基于统一病理金标准。很多研究依赖随访、专家共识或有限病理验证,尤其在磨玻璃结节和炎症性病变中,标签不确定性更高<sup>[11-12,28,30,38]</sup>。这直接影响模型训练质量和结果可迁移性。

在肺结节领域,所谓“真值”本身常常是一个概

率性概念。不是所有结节都能获得病理证实,而长期稳定也并不等于绝对良性。标签的不确定性会给模型学习带来噪声,使其可能学习到不一致甚至相互矛盾的模式。因此,未来需要更高质量的标注体系,包括多专家共识、病理对照、长期随访确认及标签置信度设计等。

**6.3 可解释性、校准性与责任归属** 即使模型具有较高AUC,若缺乏可解释性和良好校准,临床使用仍会受到限制。近年来研究强调应加强决策曲线分析、概率校准、人机协同评估和真实 workflow 中的净获益验证<sup>[16,38]</sup>。现阶段,深度学习更适合作为辅助决策工具,而非完全替代医生。

可解释性并不一定意味着必须完全揭示模型内部机制,但至少应能以热图、区域贡献或特征可视化方式说明模型主要依据什么作出判断。校准性同样重要,因为一个概率输出若系统性偏高或偏低,将直接影响随访和干预决策。责任归属问题则关系到医疗安全与监管合规,尤其在AI建议与医师判断不一致时更需明确边界。

**6.4 基础模型、多模态融合与流程重构** 未来肺结节深度学习的发展可能主要集中于:(1)肺部CT基础模型与预训练框架,提高跨任务迁移能力;(2)图像、临床、分子和文本等多模态融合<sup>[34,37,39-40]</sup>;(3)时间序列学习与平台化部署<sup>[14,41]</sup>。其最终目标不是单纯提高某一性能指标,而是与指南、结构化报告、专病门诊和MDT流程深度耦合,成为肺结节管理流程的一部分。

基础模型的发展有望缓解小样本和标注昂贵的问题,使模型能够在多个任务之间共享更通用的表征能力。多模态融合将使风险评估更贴近真实临床。流程重构则是AI真正落地的关键,即不再把AI作为附加工具,而是作为管理流程的一部分嵌入日常工作。可以预见,未来高水平肺结节AI不只是“一个模型”,而更可能是“一个持续学习、持续迭代、持续服务临床流程的系统”。

## 7 小结

深度学习正在推动肺结节影像评估从“发现病灶”走向“精细表型刻画、恶性概率分层、动态随访预测和全流程管理”。深度学习已在结节检出、精准分割、GGO评估、风险预测及时间序列分析等方面显示出较高潜力。但现阶段其最佳定位仍是增强型临床决策支持工具,而非脱离指南和医师判断的独立替代方案。未来,随着多中心真实世界验证、可解释性提升、多模态整合、时间序列学习及平

台化部署的发展,深度学习有望成为肺结节精准管理体系中的关键技术支柱之一。

更进一步看,深度学习对肺结节领域的意义,不仅在于提高某一任务的技术指标,更在于推动整个管理理念的变化。它正在使肺结节影像学由经验性、离散化、静态化分析,逐步走向量化、连续化和全流程化管理。未来若能实现与临床指南、结构化报告、专病门诊和MDT决策的深度耦合,其价值将不再局限于“辅助读片”,而可能成为肺结节精准防治体系的重要基础设施。

**伦理声明** 无

**利益冲突** 所有作者声明不存在利益冲突。

**作者贡献** 白春学:选题、撰写、修改;朱煜:撰写、修改。

#### 参考文献

- [1] National Lung Screening Trial Research Team, Aberle D R, Adams A M, et al. Reduced lung-cancer mortality with low-dose computed tomographic screening [J]. *N Engl J Med*, 2011, 365(5): 395-409.
- [2] de Koning H J, van der Aalst C M, de Jong P A, et al. Reduced lung-cancer mortality with volume CT screening in a randomized trial[J]. *N Engl J Med*, 2020, 382(6): 503-513.
- [3] US Preventive Services Task Force, Krist A H, Davidson K W, et al. Screening for lung cancer: US Preventive Services Task Force recommendation statement[J]. *JAMA*, 2021, 325(10): 962-970.
- [4] Wolf A M D, Oeffinger K C, Shih T Y, et al. Screening for lung cancer: 2023 guideline update from the American Cancer Society[J]. *CA Cancer J Clin*, 2024, 74(1): 50-81.
- [5] 中华医学会呼吸病学分会,中国肺癌防治联盟专家组.肺结节诊治中国专家共识(2024年版)[J].*中华结核和呼吸杂志*, 2024, 47(8): 716-729.
- [6] MacMahon H, Naidich D P, Goo J M, et al. Guidelines for management of incidental pulmonary nodules detected on CT images: from the Fleischner Society 2017 [J]. *Radiology*, 2017, 284(1): 228-243.
- [7] American College of Radiology. Lung-RADS® v2022[EB/OL]. [2025-01-02]. <https://www.professionalradiology.com/media/user/resources/Lung-RADS%202022.pdf>.
- [8] Callister M E, Baldwin D R, Akram A R, et al. British Thoracic Society guidelines for the investigation and management of pulmonary nodules [J]. *Thorax*, 2015, 70 (Suppl 2): ii1-ii54.
- [9] Chen H, Garrana S, Moreira AL, et al. Imaging and management of subsolid lung nodules [J]. *Radiol Clin North Am*, 2025, 63(4): 517-535.
- [10] Raad R A, Garrana S, Moreira A L, et al. Imaging and management of subsolid lung nodules [J]. *Radiol Clin North Am*, 2025, 63(4): 517-535.
- [11] Gao C, Wu L Y, Wu W, et al. Deep learning in pulmonary nodule detection and segmentation: a systematic review[J]. *Eur Radiol*, 2025, 35(1): 255-266.
- [12] Cheo H M, Ong CYG, Ting Y. A systematic review of AI performance in lung cancer detection on ct thorax [J]. *Healthcare (Basel)*, 2025, 13(13): 1510.
- [13] Venkadesh K V, Setio A A A, Schreuder A, et al. Deep learning for malignancy risk estimation of pulmonary nodules detected at low-dose screening CT[J]. *Radiology*, 2021, 300 (2): 438-447.
- [14] Venkadesh K V, Aleef T A, Scholten E T, et al. Prior CT improves deep learning for malignancy risk estimation of screening-detected pulmonary nodules [J]. *Radiology*, 2023, 308(2): e223308.
- [15] Ardila D, Kiraly A P, Bharadwaj S, et al. End-to-end lung cancer screening with three-dimensional deep learning on low-dose chest computed tomography[J]. *Nat Med*, 2019, 25(6): 954-961.
- [16] Herber S K, Müller L, Pinto Dos Santos D, et al. Diagnostic performance of artificial intelligence models for pulmonary nodule classification: a multi-model evaluation [J]. *Eur Radiol*, 2026, 36(1): 537-547.
- [17] Swensen S J, Silverstein M D, Ilstrup D M, et al. The probability of malignancy in solitary pulmonary nodules. Application to small radiologically indeterminate nodules [J]. *Arch Intern Med*, 1997, 157(8): 849-855.
- [18] Al-Ameri A, Malhotra P, Thygesen H, et al. Risk of malignancy in pulmonary nodules: a validation study of four prediction models[J]. *Lung Cancer*, 2015, 89(1): 27-30.
- [19] Choi H K, Ghobrial M, Mazzone P J. Models to estimate the probability of malignancy in patients with pulmonary nodules [J]. *Ann Am Thorac Soc*, 2018, 15(10): 1117-1126.
- [20] Susam S, Çinkoğlu A, Ceylan K C, et al. Comparison of Brock University, Mayo Clinic and Herder models for pretest probability of cancer in solid pulmonary nodules [J]. *Clinical Respiratory J*, 2022, 16(11): 740-749.
- [21] Senent-Valero M, Librero J, Pastor-Valero M. Solitary pulmonary nodule malignancy predictive models applicable to routine clinical practice: a systematic review [J]. *Syst Rev*, 2021, 10: 308.
- [22] Balekian A A, Silvestri G A, Simkovich S M, et al. Accuracy of clinicians and models for estimating the probability that a pulmonary nodule is malignant[J]. *Ann Am Thorac Soc*, 2013, 10(6): 629-635.
- [23] Setio A A A, Traverso A, de Bel T, et al. Validation, comparison, and combination of algorithms for automatic detection of pulmonary nodules in computed tomography images: the LUNA16 challenge [J]. *Med Image Anal*, 2017, 42: 1-13.
- [24] Armato S G 3rd, McLennan G, Bidaut L, et al. The Lung Image Database Consortium (LIDC) and Image Database Resource Initiative (IDRI): a completed reference database of lung nodules on CT scans [J]. *Med Phys*, 2011, 38(2):

- 915-931.
- [25] Fedorov A, Hancock M, Clunie D, et al. DICOM re-encoding of volumetrically annotated Lung Imaging Database Consortium (LIDC) nodules[J]. *Med Phys*, 2020, 47(11): 5953-5965.
- [26] Zhang H, Peng Y J, Guo Y F. Pulmonary nodules detection based on multi-scale attention networks[J]. *Sci Rep*, 2022, 12(1): 1466.
- [27] Wang Q, Shen F Y, Shen L Y, et al. Lung nodule detection in CT images using a raw patch-based convolutional neural network [J]. *J Digit Imaging*, 2019, 32(6): 971-979.
- [28] Shah H P, Naqvi A S, Rajput P, et al. Artificial intelligence-based deep learning algorithms for ground-glass opacity nodule detection: a review[J]. *Narra J*, 2025, 5(1): e1361.
- [29] Mehrnia S S, Safahi Z, Mousavi A, et al. Landscape of 2D deep learning segmentation networks applied to ct scan from lung cancer patients: a systematic review [J]. *J Imaging Inform Med*. 2025. 38(6): 3711-3740.
- [30] Wu W, Gao C, Wu L Y, et al. Diagnostic accuracy of deep learning for the invasiveness assessment of ground-glass nodules with fine segmentation: a systematic review and meta-analysis [J]. *Quant Imaging Med Surg*, 2025, 15(4): 2722-2738.
- [31] Sun Q, Yu L, Song Z Q, et al. Deep learning and radiomics fusion for predicting the invasiveness of lung adenocarcinoma within ground glass nodules[J]. *Sci Rep*, 2025, 15: 29285.
- [32] Shao M H, Wang J, Zhu L, et al. Deep learning for Three-Class Classification of ground-glass nodules on non-enhanced chest CT: a multicenter comparative study of CNN architectures [J]. *Eur J Radiol Open*, 2025, 15: 100690.
- [33] Sun Y, Ge X Y, Niu R, et al. PET/CT radiomics and deep learning in the diagnosis of benign and malignant pulmonary nodules: progress and challenges[J]. *Front Oncol*, 2024, 14: 1491762.
- [34] Du H, Shen J, Chen F, et al. Integrating CT-based radiomics and deep learning for invasive prediction of ground-glass nodules in lung adenocarcinoma: a multicohort study [J]. *Insights Imaging*, 2025, 16(1): 271.
- [35] Warkentin M T, Al-Sawaihey H, Lam S, et al. Radiomics analysis to predict pulmonary nodule malignancy using machine learning approaches[J]. *Thorax*, 2024, 79(4): 307-315.
- [36] Ye G, Wu G, Li K, et al. Development and validation of a deep learning radiomics model to predict high-risk pathologic pulmonary nodules using preoperative computed tomography [J]. *Acad Radiol*, 2024, 31(4): 1686-1697.
- [37] Wang H, Huang H, Li F, et al. Risk-stratified classification of pulmonary nodule malignancy via a machine learning model integrating imaging and cell-free DNA: a model development and validation study (DECIPHER-NODL) [J]. *Lancet Reg Health West Pac*, 2025, 64: 101730.
- [38] Geppert J, Asgharzadeh A, Brown A, et al. Software using artificial intelligence for nodule and cancer detection in CT lung cancer screening: systematic review of test accuracy studies [J]. *Thorax*, 2024, 79(11): 1040-1049.
- [39] Xiao D, Forero Y, Kammer M N, et al. Radiomic ‘Stress Test’: exploration of a deep learning radiomic model in a high-risk prospective lung nodule cohort [J]. *BMJ Open Respir Res*, 2025, 12(1): e002687.
- [40] Huang P, Lin C, Li Y, et al. Prediction of lung cancer risk at follow-up screening with low-dose CT: a training and validation study of a deep learning method [J]. *Lancet Digital Health*, 2019, 1: e353-e362.
- [41] Huang W, Ye M, Wang B, et al. Narrative review: the research advances of artificial intelligence in the prediction of pulmonary nodule growth [J]. *J Thorac Dis*, 2026, 18(2): 165.

#### 引用本文

白春学,朱煜.深度学习辅助精细评估肺结节影像[J].元宇宙医学,2026,3(1):72-80.

Bai C X,Zhu Y. Deep learning-assisted fine assessment of pulmonary nodule images[J]. *Metaverse Med*,2026,3(1):72-80.