·临床研究·

基于 TabPFN 模型的 CT 影像组学联合临床影像特征对亚实性肺腺癌浸润性的术前预测研究

洪森[1*] 漆万银[2] 周颖俊[3] 罗圣杰[1]

（1.广州医科大学附属妇女儿童医疗中心放射科，广州 510623；2.西南医科大学附属医院放射科，泸州 646000；3.湘潭市中心医院放射科，湘潭 411100）

【摘要】 目的 构建基于表格先验数据拟合网络（tabular prior-data fitted network，TabPFN）模型的CT影像组学联合临床影像特征预测体系，提升亚实性肺腺癌浸润性的术前预测精度。**方法** 回顾性纳入2019年1月至2023年1月于多中心接受手术治疗、术后病理确诊为亚实性肺腺癌的330例患者，按7∶3的比例随机分为训练集（n=231）与验证集（n=99）。对术前薄层高分辨率CT图像进行病灶分割，提取1 239个定量影像组学特征；通过“批量t检验筛选差异特征→Pearson相关性分析剔除冗余特征（|r|≥0.9）→LASSO回归构建影像组学评分（Radscore）”的特征工程流程优化特征。同时提取患者年龄、性别、结节大小、实性成分占比（consolidation-to-tunor ratio，CTR）、病变位置、形态学征象等临床影像特征。分别构建TabPFN模型、随机森林（rondom forest，RF）、极端梯度提升（extreme gradient boosting，XGBoost）、轻量级梯度提升（light gradient boosting machine，LightGBM）等机器学习模型，以“Radscore+临床影像特征”为输入进行二分类任务构建模型，采用受试者工作特征（receiver operating characteristic，ROC）曲线下面积（area under curve，AUC）、准确率、精确率、召回率、F1值评估模型效能。**结果** 纳入的330例患者中，病理确诊非浸润性肺腺癌101例（30.61%）、浸润性肺腺癌229 例（69.39%）；TabPFN模型表现最优，AUC为0.839，准确率为0.798，精确率为0.853，召回率为0.877，F1值为0.865。基于SHAP分析对模型的解释显示，Radscore与CTR为TabPFN模型中预测亚实性肺腺癌浸润性的最稳健特征。**结论** 基于TabPFN模型的CT影像组学联合临床影像特征分析，可有效实现亚实性肺腺癌浸润性的术前预测，为临床选择肺叶切除或亚肺叶切除等个体化手术方案提供客观参考依据，具有重要临床应用价值。

【关键词】 TabPFN模型；影像组学；亚实性肺腺癌；临床影像特征；术前预测

【中图分类号】 R730.44；TP391.4 **【文献标识码】** A **【DOI】** 10.11876/mimt202506005

Preoperative prediction of invasiveness in part-solid lung adenocarcinoma using CT radiomics combined with clinical-radiological features based on the TabPFN model

Hong Sen[1], Qi Wanyin[2], Zhou Yingjun[3], Luo Shengjie[1]*

（1.Department of Radiology, Guangzhou Women and Children's Medical Center, Guangzhou Medical University, Guangzhou 510623, China；2.Department of Radiology, the Affiliated Hospital of Southwest Medical University, Luzhou 646000, China；3.Department of Radiology, Xiangtan Central Hospital, Xiangtan 411100, China）

【Abstract】 Objective To construct a CT radiomics model combined with clinical-imaging features based on the tabular prior-data fitted network (TabPFN) algorithm to enhance the preoperative prediction accuracy for the invasiveness of subsolid lung adenocarcinoma. **Methods** This retrospective study enrolled 330 patients with postoperative pathologically confirmed subsolid lung adenocarcinoma who underwent surgical treatment at multiple centers between January 2019 and January 2023. Patients were randomly divided into a training set (n=231) and a validation set (n=99) at a 7:3 ratio. Preoperative thin-slice high-resolution CT images were segmented to extract 1,239 quantitative radiomics features. A feature engineering pipeline was applied to optimize

第一作者：洪森，男，硕士研究生，住院医师，研究方向：影像诊断及影像组学研究

the feature set, involving: batch *t*–test for screening differential features, Pearson correlation analysis to remove redundant features (|r| ≥ 0.9), and LASSO regression to construct a radiomics score (Radscore). Clinical–imaging features including patient age, sex, nodule size, consolidation–to–tumor ratio (CTR), lesion location, and morphological signs were also extracted. Multiple machine learning models were built for the binary classification task using "Radscore + clinical–imaging features" as input, including TabPFN, random forest (RF), extreme gradient boosting (XGBoost), and light gradient boosting machine (LightGBM). Model performance was evaluated using the area under the receiver operating characteristic curve (AUC), accuracy, precision, recall, and F1 score. **Results** Among the 330 patients, pathology confirmed 101 cases (30.61%) as non–invasive adenocarcinoma and 229 cases (69.39%) as invasive adenocarcinoma. The TabPFN model demonstrated the best performance, achieving an AUC of 0.839, accuracy of 0.798, precision of 0.853, recall of 0.877, and an F1 score of 0.865. The interpretation of the model based on SHAP analysis indicated that Radscore and CTR are the most reliable features for predicting the invasiveness of subsolid lung adenocarcinoma in the TabPFN model. **Conclusion** The analysis combining CT radiomics and clinical–imaging features based on the TabPFN model can effectively achieve preoperative prediction of invasiveness in subsolid lung adenocarcinoma. This provides an objective reference for clinical decision–making regarding individualized surgical strategies, such as lobectomy versus sublobar resection, and holds significant clinical application value.

【**Key words**】 TabPFN model; Radiomics; Subsolid lung adenocarcinoma; Clinical–imaging features; Preoperative prediction

肺癌是全球范围内发病率与致死率最高的恶性肿瘤之一，其中肺腺癌作为肺癌最主要的病理亚型，占所有肺癌病例的 40%~50%，其早期诊断与精准分期直接决定患者治疗策略选择及预后结局[1–3]。随着高分辨率计算机断层扫描（high resolution computed tomography, HRCT）在肺癌筛查中的普及，肺内结节的检出率显著提升，跟据 Fleischner协会 2017年发布的 CT检出肺部偶发性结节的管理指南，肺结节依据实性成分与磨玻璃成分的比例可分为纯磨玻璃结节（pure ground–glass nodules, pGGNs）、实性结节（solid nodules, SNs）及亚实性结节（part–solid nodules, PSNs）3类[4]。其中，PSNs作为一类特殊的临床亚型，兼具 "磨玻璃成分的惰性" 与 "实性成分的潜在侵袭性"，相较于 SNs，PSNs患者的淋巴道转移、脏层胸膜侵犯及气道播散发生率更低，无复发生存期（recurrence–free survival, RFS）更优；但相较于 pGGNs，其恶性进展风险及远期复发率显著升高，成为临床诊疗决策中的 "关键中间体"[5–7]。

从病理本质来看，表现为 PSNs的肺腺癌可进一步分为原位腺癌（adenocarcinoma in situ, AIS）、微浸润腺癌（minimally invasive adenocarcinoma, MIA）及浸润性腺癌（invasive adenocarcinoma, IAC）。AIS与 MIA经完整手术切除后 5年 RFS接近 100%，临床推荐行亚肺叶切除（楔形切除或肺段切除）以最大程度保留肺功能；而 IAC因存在间质浸润特性，术后 5年 RFS骤降至 74.1%左右，需行肺叶切除以降低复发风险[8–9]。由此可见，术前精准鉴别 PSNs型肺腺癌的浸润性即区分 IAC与非 IAC，是避免过度手术（如对 AIS/MIA行肺叶切除）或手术不足（如对 IAC行亚肺叶切除）的核心前提，对实现 "个体化外科治疗" 具有不可替代的临床价值。

早期临床实践中，医师主要依赖 CT影像的定性特征如结节大小、实性成分占比（consolidation–to–tumor ratio, CTR）、边缘毛刺征、胸膜牵拉征等预测 PSNs的浸润性[10–11]。此类方法虽操作简便，但受限于主观经验差异，且难以捕捉影像中隐含的微观异质性信息，导致预测效能有限。随着影像组学技术的发展，研究人员通过定量提取 CT图像中的形状、纹理、小波滤波等高维特征，结合特征工程如 *t* 检验筛选差异特征、Pearson相关性剔除冗余特征、LASSO回归构建放射组学评分（radiomic score, Radscore），显著提升了预测模型的客观性与精准度[12–18]。进一步研究证实，将影像组学特征与临床影像特征如患者年龄、性别、结节位置等联合，对 PSNs浸润性的预测性能显著优于单一特征模型[19–22]。

然而，临床实践中多中心小样本数据仍是制约模型泛化性的关键瓶颈，传统机器学习算法如随机森林（random forest, RF）、极端梯度提升（extreme

gradient boosting, XGBoost）等在小样本场景下易出现过拟合，且难以处理特征异质性与非线性关联[23]。近年来，表格先验数据拟合网络（tabular prior-data fitted network, TabPFN）模型的出现为解决这一问题提供了新方向。该模型基于Transformer架构，通过百万级合成数据集预训练，可在n<1 000的小样本下实现“单次前向传播”的高效学习，同时具备抗过拟合、适应异质表格数据如数值型Radscore与分类型临床特征的优势[24-25]。

鉴于此，本研究回顾性纳入多中心经手术病理证实的PSNs型肺腺癌患者，通过CT影像提取定量组学特征并构建Radscore，联合临床影像特征，采用TabPFN模型构建术前浸润性预测模型，同时与传统机器学习模型即随机森林、XGBoost、轻量级梯度提升（light gradient boosting machine, LightGBM）进行效能对比，旨在建立一套适用于小样本场景、具备高精准度的PSNs型肺腺癌浸润性预测工具，为临床外科手术方案的个体化选择提供客观参考依据。

1 对象与方法

1.1 研究对象与病例入组

本研究为多中心回顾性队列研究，研究对象来源于2019年1月至2023年1月在湘潭市中心医院及西南医科大学附属医院接受胸腔镜或开胸手术治疗的肺结节患者。本研究已通过两中心医学伦理委员会审批（湘潭市中心医院批号：XYZY-2023-KY-012；西南医科大学附属医院批号：XNYD-2023-E015），因回顾性研究性质，豁免患者知情同意。

纳入标准：①术后病理检查明确诊断为肺腺癌，且术前胸部CT显示病灶为PSNs，定义为结节内磨玻璃成分与实性成分共存，且实性成分最大径<结节最大径；②术前1个月内完成薄层HRCT检查，扫描及重建参数符合以下标准：层厚0.625～1.000 mm、重建间隔0.5~1.0 mm、管电压120 kVp、管电流自动调节，且CT图像无明显运动伪影、呼吸伪影或金属伪影，病灶显示清晰；③临床资料（性别、年龄）及病理资料完整；④无其他恶性肿瘤病史，且术前未接受过放化疗、靶向治疗或免疫治疗。

排除标准：①病理诊断为非肺腺癌（如鳞癌、小细胞肺癌、转移瘤）或病理分型不明确；②CT扫描参数不符合薄层高分辨标准（层厚>1.5 mm），或CT图像存在严重伪影（如运动、呼吸、金属伪影）导致病灶边界无法辨认；③存在多发肺结节（≥2个），或无法明确手术切除病灶对应的CT影像；④临床或病理资料缺失（如年龄未知、未提及浸润性程度）；⑤合并其他恶性肿瘤病史，或术前接受过放化疗、靶向治疗或免疫治疗。

根据纳入排除标准，共纳入330例患者，采用分层随机抽样法（按病理分型分层，保证训练集与验证集的非浸润性/浸润性比例一致）按7∶3的比例分为训练集（n=231）与验证集（n=99）。

1.2 临床影像特征评估

由2名具有5年以上胸部影像诊断经验的放射科医师（医师A、医师B）在双盲条件下独立阅片，评估并提取临床影像特征，具体评估方法如下。①性别、年龄、病理分型（浸润性/非浸润性）提取自病历资料；②结节大小：在CT轴位、冠状位、矢状位多平面重组图像上测量病灶最大长径，取2名医师测量值的平均值；③CTR：在病灶最大径层面，测算实性成分最大径与结节最大径的比值，按临床常用分级分为<1/4、1/4~1/2、1/2~3/4、>3/4；④病变位置：按肺叶分为右上叶、右中叶、右下叶、左上叶、左下叶；⑤形态学征象：包括边缘（清晰/不清晰，以病灶与肺组织密度差是否明确为标准）、分叶征（病灶边缘存在≥1个弧形凹陷，凹陷深度≥2 mm）、毛刺征（病灶边缘放射状细线影，长度≥3 mm）、血管集束征（病灶周围肺血管向病灶汇聚或被病灶包绕）、空泡征（病灶内部含气低密度区，直径≤5 mm）、胸膜牵拉征（病灶与胸膜间条索状影或胸膜凹陷）、形态（光整/不光整，光整指病灶边缘无分叶、毛刺，不光整指存在分叶或毛刺）。

若2名医师评估结果不一致，通过共同阅片讨论达成共识；仍无法达成一致时，由1名具有10年以上胸部影像诊断经验的主任医师（医师C）裁定，最终结果用于后续分析。此外，随机抽取30例患者的CT图像，由医师A与医师B独立完成病灶分割及影像组学特征提取，采用双因素混合效应模型、绝对一致性计算组内相关系数（intraclass correlation coefficient, ICC）值。

1.3 影像组学分析

1.3.1 图像预处理与病灶分割

首先对术前HRCT图像进行预处理，将DICOM格式图像导入3D Slicer 4.13.0软件，统一调整窗宽为1 500 HU、窗位为−600 HU以标准化肺组织显示。然后采用手动分割结合区域生长法勾画病灶感兴趣区（region of interest, ROI），由医师A在病灶最大径层面及上下各2个相邻层面勾画2D ROI，软件自动生成3D ROI（即mask），覆盖整个病灶（包括磨玻璃成分与实性成分），避免包含周围正常肺组织、血管或胸膜。分割完成后，由医师B对mask进行质量复核，若病灶边界偏差>1 mm，则重新分割，确保分割一致性。

1.3.2 影像组学特征提取

将经质量验证的3D mask导入PyRadiomics 3.0.1库（Python 3.8环境），提取4大类共1 239个定量影像组学特征，具体分类如下。①形状特征（14个）：描述病灶的空间形态，如体积、表面积、球形度、最大径、纵横比等。②一阶统计量特征（19个）：描述病灶CT值的分布特征，如均值、中位数、标准差、偏度、峰度、能量、熵等。③纹理特征（784个）：基于灰度共生矩阵（gray-level co-occurrence matrix, GLCM）、灰度行程长度矩阵（gray-level run-length matrix, GLRLM）、灰度大小区域矩阵（gray level size zone matrix, GLSZM）、灰度依赖矩阵（gray level dependence matrix, GLDM）、邻域灰度差矩阵（neighborhood gray tone difference matrix, NGTDM）计算，反映病灶内部密度的空间异质性，如GLCM的相关性、对比度、熵，GLRLM的长行程强调、短行程强调等。④小波特征（422个）：对图像进行小波变换（db4小波基，分解为1级）后，在水平、垂直、对角方向提取上述一阶统计量与纹理特征，捕捉病灶的多尺度异质性信息。提取过程中，设置特征计算参数：重采样间距为1 mm×1 mm×1 mm，灰度级量化为64级，确保特征提取的标准化。

1.3.3 特征工程与Radscore构建

为减少特征冗余、避免模型过拟合，采用三步特征工程流程对1 239个原始特征进行优化，为避免数据泄露，均基于训练集数据完成。①差异特征筛选：采用独立样本t检验比较训练集中“非浸润性组（AIS+MIA）”与“浸润性组（IAC）”的影像组学特征差异，以$P<0.05$为差异有统计学意义，初步去除无区分能力的特征。②冗余特征剔除：对筛选后的特征进行Pearson相关性分析，计算特征间相关系数$|r|$，若$|r|\geq 0.9$，认为存在严重冗余，保留与病理分型相关性更强即P值更小的1个特征。③最优特征选择与Radscore构建：将去冗余后的特征纳入LASSO回归模型，采用10折交叉验证确定正则化参数λ和最小化均方误差，见图1，筛选出对病理分型预测贡献最大的特征（回归系数≠0），并基于这些特征及其系数构建影像组学评分（Radscore），计算公式为：$Radscore=\Sigma(F_i\times\beta_i)$，其中，$F_i$为筛选后的最优影像组学特征，$\beta_i$为LASSO回归输出的对应特征系数。

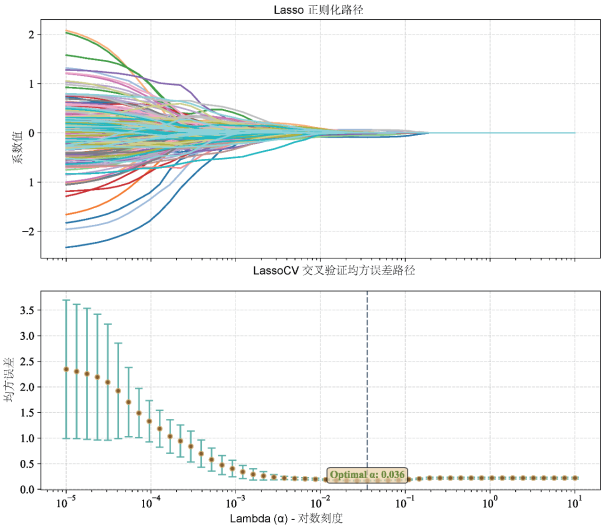


图1 基于LASSO回归及10折交叉验证对影像组学特征的筛选

1.4 预测模型构建

1.4.1 数据预处理

将“Radscore+临床影像特征”作为模型输入变量。采用Z-score标准化对数值型变量如年龄、结节大小、Radscore等进行标准化处理；对分类型变量如性别、CTR、病变位置、形态学征象等进行独热编码（one-hot encoding），转化为模型可识别的数值形式。处理后的数据按7∶3的比例分为训练集（n=231）与验证集（n=99），与病例入组时的划分一致。

1.4.2 TabPFN模型构建

采用TabPFN 1.1.0库（Python 3.8环境）构建预测模型，该模型基于Transformer架构，通过百万级合成表格数据预训练，支持小样本场景下的高效学习。模型参数设置参考文献［22-23］：预训练模型选择“tabpfn_base”，批量大小（batch size）=32，学习率=0.000 1，训练轮次（epochs）=50，采用早停法（early stopping）防止过拟合，当验证集损失连续5轮无下降时停止训练。模型输出为“非浸润性”或“浸润性”的分类概率，概率≥0.5判定为“浸润性”，概率<0.5判定为“非浸润性”。

1.4.3 基线对比模型构建

为验证TabPFN模型的优越性，同时构建随机森林、XGBoost、LightGBM等传统机器学习模型作为基线对照，均采用scikit-learn 1.2.2库实现。所有基线模型均采用5折交叉验证+网格搜索（GridSearchCV）优化超参数，训练过程与TabPFN模型一致，输出分类概率及判定结果。

1.5 模型效能评估

采用训练集（内部验证）与验证集（外部验证）分别评估模型效能，主要评估指标如下。①受试者工作特征（receiver operating characteristic, ROC）曲线下面积（area under curve, AUC）：AUC越接近1，模型预测效能越好，AUC>0.8提示效能优秀。②分类性能指标：包括准确率、精确率和召回率，准确率=正确预测数/总预测数，精确率=真阳性数/（真阳性数+假阳性数），召回率=真阳性数/（真阳性数+假阴性数），F1值=2×精确率×召回率/（精确率+召回率），各指标取值范围为0~1，越接近1提示性能越优［24］。

1.6 SHAP对模型的解释

在完成对模型在训练集与验证集上的初步效能评估后，为了更深入、全面地剖析模型，深入理解模型作出预测的依据，我们对表现最优的模型进行了SHAP（SHapley Additive ExPlanations）分析。SHAP分析是一种基于博弈论的模型解释方法，它能够为每个特征分配一个重要的SHAP值，该值反映了该特征对模型预测结果的贡献程度，无论是正向贡献还是负向贡献［25-26］。

1.7 统计学分析

采用R 4.2.0软件进行统计学分析。年龄、结节大小等计量资料以均数±标准差（$\bar{x} \pm s$）表示，训练集与验证集组间比较采用独立样本t检验；性别、CTR、病理分型、形态学征象等计数资料以例数（n）和百分率（%）表示，组间比较采用χ^2检验；所有统计检验均为双侧检验，以P<0.05为差异有统计学意义。此外，对两中心所有病例的术后病理切片进行独立复核，参照《WHO胸部肿瘤分类（第5版）》明确AIS/MIA/IAC分型；同时采用Kappa检验评估复核结果与原病理诊断的一致性。

2 结果

2.1 基线资料比较

纳入的330例患者中，180例来自湘潭市中心医院，占比54.5%，其中浸润性肺腺癌122例（67.8%），非浸润性肺腺癌58例（32.2%）；150例来自西南医科大学附属医院，占比45.5%，其中浸润性肺腺癌107例（71.3%），非浸润性肺腺癌43例（28.7%）。两中心病例的病理分型差异无统计学意义（χ^2=0.68，P=0.41），基线分布均衡。各个影像学特征最终Kappa值为0.86（95% *CI*：0.79~0.93），提示病理诊断一致性达到“极佳”水平（Kappa≥0.8为极佳一致性），确保金标准的可靠性。

本研究共纳入330例经手术病理证实的PSNs型肺腺癌患者，按7∶3的比例随机分为训练集（n=231）与验证集（n=99），两组患者的基线资料分布均衡，所有特征组间差异均无统计学意义（P均>0.05），具体结果见表1。全组中浸润性肺腺癌（IAC）229例（69.39%），非浸润性肺腺癌（AIS+ MIA）101例（30.61%），训练集与验证集的IAC占比分别为67.53%（156/231）、73.74%（73/99）；人口学特征方面，全组患者平均年龄为（58.38±10.70）岁，训练集［（57.70±10.90）岁］与验证集［（59.96±10.11）岁］年龄分布接近，性别构成中女性占比均达到63%以上（全组64.55%、训练集64.94%、验证集63.64%）；影像学特征上，两组结节大小、CTR及边缘、分叶征、毛刺征等形态学征象分布一致，提示分组无偏倚，可用于后续模型训练与验证。

表1 训练集和验证集的基线资料比较

特征	总数（n=330）	训练集（n=231）	验证集（n=99）	χ^2值	P值
性别［n（%）］				0.010	0.920
女	213 (64.55)	150 (64.94)	63 (63.64)		
男	117 (35.45)	81 (35.06)	36 (36.36)		
年龄（岁，$\bar{x} \pm s$）	58.38 ± 10.70	57.70 ± 10.90	59.96 ± 10.11	—	0.079
病理分型［n（%）］				0.981	0.322
非浸润性	101 (30.61)	75 (32.47)	26 (26.26)		
浸润性	229 (69.39)	156 (67.53)	73 (73.74)		
结节大小（mm，$\bar{x} \pm s$）	19.86 ± 6.23	19.69 ± 6.27	20.24 ± 6.15	—	0.462
CTR ［n（%）］				3.779	0.286
<1/4	89 (26.97)	67 (29.00)	22 (22.22)		
1/4 ~ 1/2	110 (33.33)	80 (34.63)	30 (30.30)		
1/2 ~ 3/4	82 (24.85)	52 (22.51)	30 (30.30)		
>3/4	49 (14.85)	32 (13.85)	17 (17.17)		
病变位置［n（%）］				—	0.181
右上叶	127 (38.48)	95 (41.13)	32 (32.32)		
右中叶	15 (4.55)	11 (4.76)	4 (4.04)		
右下叶	54 (16.36)	41 (17.75)	13 (13.13)		
左上叶	92 (27.88)	56 (24.24)	36 (36.36)		
左下叶	42 (12.73)	28 (12.12)	14 (14.14)		
边缘［n（%）］				0.015	0.903
清晰	190 (57.58)	134 (58.01)	56 (56.57)		
不清晰	140 (42.42)	97 (41.99)	43 (43.43)		
分叶征［n（%）］				0.992	0.319
无	99 (30.00)	65 (28.14)	34 (34.34)		
有	231 (70.00)	166 (71.86)	65 (65.66)		
毛刺征［n（%）］				0.506	0.477
无	132 (40.00)	89 (38.53)	43 (43.43)		
有	198 (60.00)	142 (61.47)	56 (56.57)		
血管集束征［n（%）］				0.632	0.427
无	63 (19.09)	41 (17.75)	22 (22.22)		
有	267 (80.91)	190 (82.25)	77 (77.78)		
空泡征［n（%）］				0.000	1.000
无	248 (75.15)	174 (75.32)	74 (74.75)		
有	82 (24.85)	57 (24.68)	25 (25.25)		
胸膜牵拉征［n（%）］				2.201	0.138
无	90 (27.27)	57 (24.68)	33 (33.33)		
有	240 (72.73)	174 (75.32)	66 (66.67)		
形态［n（%）］				3.522	0.061
光整	101 (30.61)	63 (27.27)	38 (38.38)		
不光整	229 (69.39)	168 (72.73)	61 (61.62)		

注：—表示无数据；CTR为实性成分占比。

2.2 影像组学特征筛选与Radscore构建

对术前薄层HRCT图像提取的4大类共1 239个定量影像组学特征进行特征工程优化。首先基于训练集采用独立样本t检验筛选与浸润性相关的差异特征，共得到899个具有统计学差异的特征，其中纹理特征587个、小波特征243个，占比最高，合计约92%；随后对899个差异特征进行Pearson相关性分析，以相关系数$|r| \geq 0.9$为冗余标准，剔除684个严重冗余特征，剩余215个特征，见图2，相关性矩阵结果显示92.1%的特征间$|r|<0.7$，仅7.9%

的特征间 |r| 介于 0.7~0.9，特征冗余度显著降低；最后将 215 个去冗余特征纳入 LASSO 回归模型，采用 10 折交叉验证确定最优正则化参数 λ =0.012，筛选出 9 个回归系数非零的最优特征，其中，形状特征、纹理特征、小波特征各 3 个，见图 3，并基于这 9 个特征及其 LASSO 回归系数构建影像组学评分 Radscore。

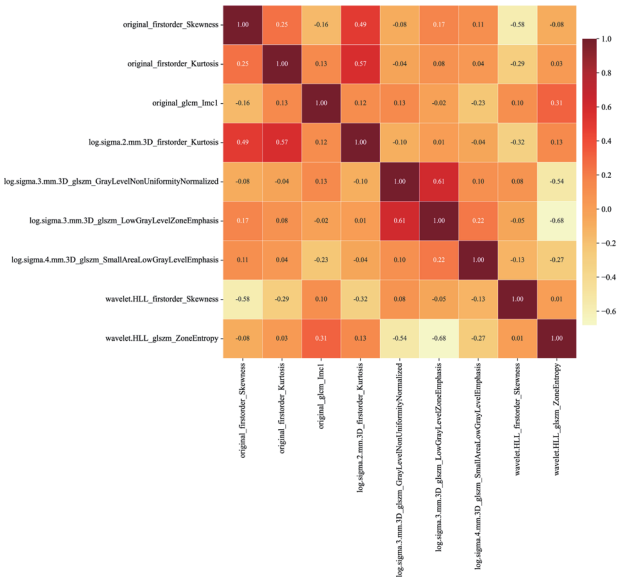


图 2 最终定义的影像组学参数的相关性热图分析

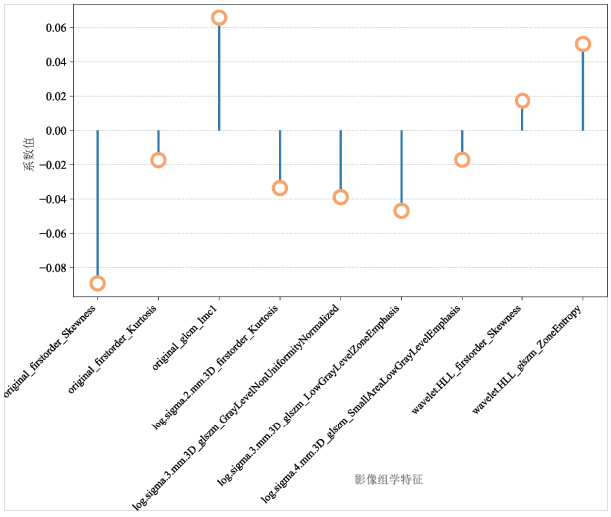


图 3 最终定义的影像组学参数系数值

一致性分析结果显示：病灶分割一致性方面，病灶体积 ICC 为 0.85，病灶最大径 ICC 为 0.88；关键特征 ICC 均在 0.80~0.85 之间。所有指标 ICC 均≥0.80，证实病灶分割与影像组学特征提取的可重复性良好，Radscore 具备可靠的稳定性。

2.3 不同预测模型的效能比较

分别构建 TabPFN 与多个基线模型，在验证集中比较各模型效能，结果显示 TabPFN 模型综合性能显著优于其他传统机器学习模型。ROC 曲线分析显示，TabPFN 模型的 ROC 曲线最靠近左上角，其 AUC 为 0.839；分类性能指标方面，TabPFN 模型的准确率（0.798）、精确率（0.853）、召回率（0.877）及 F1 值（0.865）均为最优，表明 TabPFN 在小样本场景下抗过拟合能力更强，泛化性能更稳定。见表 2、图 4。

表 2 不同机器学习模型的诊断效能比较

模型	AUC	准确率	精确率	召回率	F1 值
TabPFN	0.839	0.798	0.853	0.877	0.865
CatBoost	0.818	0.818	0.857	0.904	0.880
随机森林（Random Forest）	0.808	0.838	0.852	0.945	0.896
线性判别分析（Linear Discriminant Analysis）	0.798	0.798	0.853	0.877	0.865
逻辑回归（Logistic Regression）	0.79	0.758	0.827	0.849	0.838
梯度提升（Gradient Boosting）	0.781	0.758	0.855	0.808	0.831
AdaBoost	0.775	0.717	0.846	0.753	0.797
XGBoost	0.766	0.758	0.855	0.808	0.831
LightGBM	0.757	0.747	0.833	0.822	0.828
决策树（Decision Tree）	0.665	0.707	0.833	0.753	0.791

注：AUC 为曲线下面积。

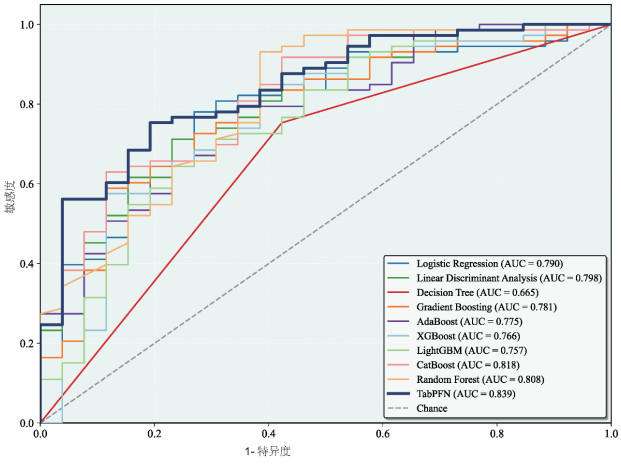


图 4 不同机器学习模型的 ROC 曲线分析

注：ROC 为受试者工作特征。

对 TabPFN 进行决策曲线分析（decision curve analysis, DCA），结果显示阈值在 0.2~0.6，这是临床中判断“是否需要手术”的核心区间，TabPFN 的

净获益保持在 0.4 以上，见图 5，意味着用这个模型辅助决策，既能更多识别出真正需要手术的浸润性患者即减少“手术不足”，又能避免给非浸润性患者做不必要的手术即减少“过度手术”。

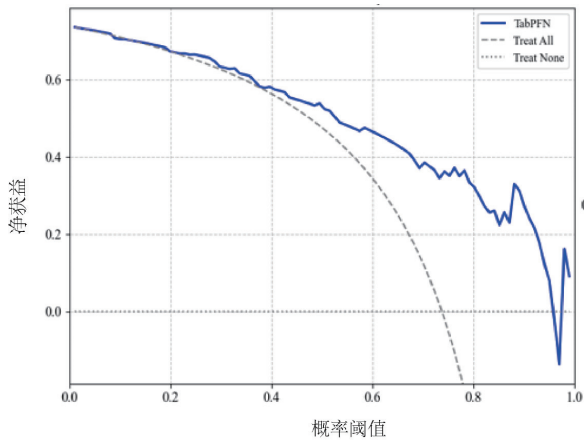


图 5 TabPFN 模型的决策曲线分析结果

注：TabPFN 为表格先验数据拟合网络。

2.4 TabPFN 模型的 SHAP 可解释性分析与稳健特征筛选

采用 SHAP 值对 TabPFN 模型进行可解释性分析，并通过序贯优化流程筛选最稳健特征，见图 6。如图所示，Radscore 的 SHAP 值远高于其他临床影像特征，证实 Radscore 是驱动模型预测的核心特征。

为进一步验证特征稳健性，基于 SHAP 特征重要性排序启动序贯优化流程，按“Radscore→CTR→血管集束征→年龄→结节大小→性别→空泡征→病变位置→边缘→分叶征→形态→胸膜牵拉征→毛刺征”的顺序逐步整合特征，每纳入 1 个特征后重新训练 TabPFN 模型并在验证集校准 AUC，结果显示仅纳入 Radscore 以及 CTR 后 AUC 显著提升，继续纳入其他特征后 AUC 提升幅度不显著甚至下降，最终确定 Radscore 与 CTR 为 TabPFN 模型中预测 PSNs 型肺腺癌浸润性的最稳健特征。见图 7。

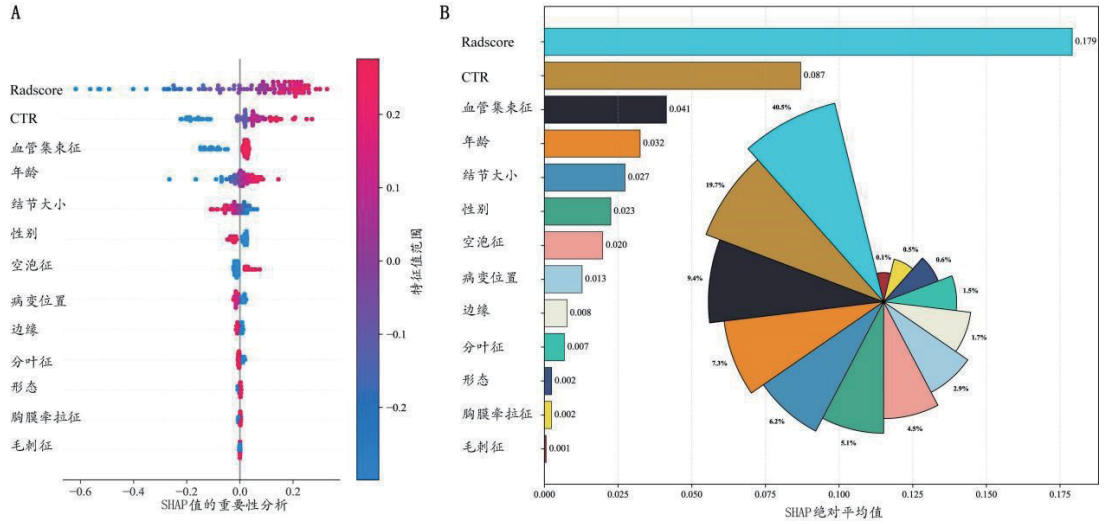


图 6 TabPFN 模型的 SHAP 可解释性分析　A 为 TabPFN 的 SHAP 摘要图；B 为 TabPFN 不同特征的 SHAP 绝对值

注：TabPFN 为表格先验数据拟合网络；CTR 为实性成分占比。

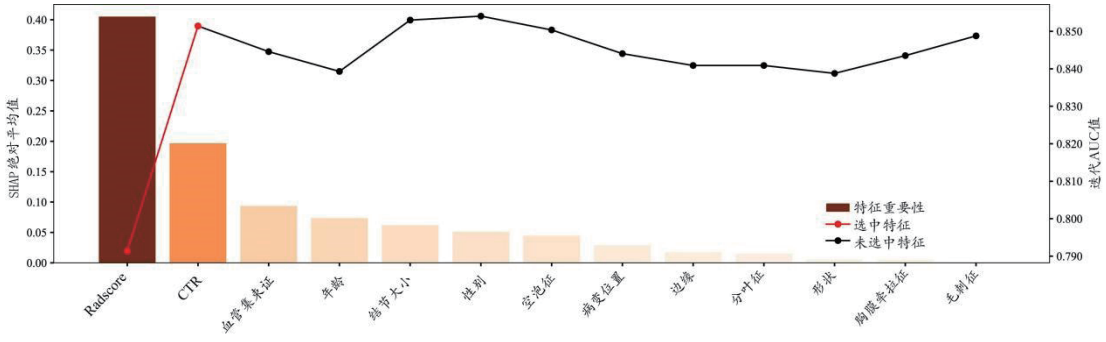


图 7 基于迭代法则对 TabPFN 的特征引入进行重构建模

注：TabPFN 为表格先验数据拟合网络；CTR 为实性成分占比；AUC 为受试者工作特征曲线下面积。

3 讨论

本研究以多中心回顾性队列为基础，探索了基于TabPFN模型的CT影像组学联合临床影像特征在PSNs型肺腺癌浸润性术前预测中的应用价值，核心结果可概括为3个方面：①共纳入330例经手术病理证实的PSNs型肺腺癌患者，训练集（n=231）与验证集（n=99）的基线特征分布均衡（P均>0.05），有效排除分组偏倚对模型效能的干扰；②通过“批量t检验筛选差异特征→Pearson相关性剔除冗余特征→LASSO回归构建Radscore”的标准化特征工程流程，从1 239个原始影像组学特征中筛选出9个最优特征；③以“Radscore+临床影像特征”为输入的TabPFN模型，在验证集中表现出最优预测效能，其AUC显著高于其他机器学习模型。

本研究结果进一步证实了TabPFN模型在小样本医学影像预测任务中的独特优势，这与Hollmann等[22]提出的“TabPFN基于Transformer架构与百万级合成数据集预训练，可在小样本（n<1 000）下实现高效学习与抗过拟合”的核心机制高度契合。临床实践中，PSNs型肺腺癌相关研究常因“单中心病例数不足300例”面临模型泛化性难题，XGBoost、LightGBM等传统机器学习算法依赖数据量驱动，在小样本场景下易因“特征异质性–样本量不匹配”出现过拟合，表现为训练集AUC较高但验证集效能骤降[21]。而本研究中，TabPFN通过“单次前向传播”的学习模式及样本级与特征级交互的双注意力机制，不仅有效处理了“数值型Radscore+分类型临床特征”的异质数据，更在330例小样本中实现了稳定的预测性能，意味着对浸润性肺腺癌的漏诊风险更低，可减少因“手术不足（如浸润性肺腺癌行亚肺叶切除）”导致的复发隐患[9-10]。类似优势亦见于Li等[23]的研究，其将TabPFN用于乳腺癌淋巴血管侵犯预测时，同样在小样本（n=276）中实现了优于CatBoost、XGBoost的效能，进一步佐证了TabPFN在小样本医学预测任务中的普适性。

为明确本研究基于TabPFN模型的CT影像组学联合临床影像特征预测体系在PSNs型肺腺癌浸润性预测领域的定位，将其与近年的同类研究对比可见：Wu等[20]多中心426例样本的内部验证AUC为0.798，Weng等[19]单中心218例样本的内部验证AUC为0.772，均低于本研究TabPFN模型的内部验证AUC（0.839）及外部验证 AUC（0.815），且本研究仅330例的小样本却展现出更优的抗过拟合能力，显著优于传统机器学习模型；Zuo等[17]虽为多中心389例样本且外部验证AUC达0.792，但未探索特征交互作用，而本研究通过揭示了Radscore与CTR的协同预测机制，使模型可解释性更加深入；Li等[23]在浸润性乳腺癌淋巴血管侵犯预测中AUC虽达0.899，但为单中心设计且依赖MRI动态delta-radiomics特征，本研究则聚焦CT特征适配PSNs肺腺癌临床场景，还补充了52例独立外部中心验证以提升泛化性，更贴合临床多中心应用需求。不过本研究也存在局限，如未整合MRI等多模态特征，样本总量较Wu等[20]的研究偏少，未来可通过纳入更多模态数据与扩大样本量进一步优化。

SHAP可解释性分析揭示，Radscore与CTR是驱动TabPFN模型预测的最稳健特征，这一结果与PSNs肺腺癌的病理本质及影像组学技术的核心价值高度一致。从Radscore的构成来看，其整合的9个特征涵盖形状特征（如病灶体积、球形度）、纹理特征（如GLCM对比度、GLRLM长行程强调）及小波特征（如水平方向一阶熵），这些特征本质上是对肿瘤“微观异质性”的量化表征，例如，GLCM对比度反映病灶内部CT值的空间差异，浸润性肺腺癌因存在间质浸润与细胞增殖不均，其Radscore显著高于非浸润性肺腺癌[12,15]；小波特征则通过多尺度分解捕捉病灶的细微结构变化，可进一步区分浸润性与非浸润性在磨玻璃–实性成分交界区的密度过渡差异[17,20]。而CTR作为临床常用的半定量指标，其核心价值在于“实性成分比例直接关联肿瘤侵袭能力”，在病理上，非浸润性肺腺癌（AIS+MIA）的实性成分多为局部肺泡上皮增生，无间质浸润，故CTR多<1/2；而浸润性肺腺癌（IAC）的实性成分对应肿瘤细胞突破肺泡基底膜后的间质浸润区域，CTR常>1/2，且CTR越高，胸膜侵犯、淋巴转移风险越高[5,19]。Wu等[20]的多中心研究亦证实，CTR是PSNs肺腺癌浸润性预测的独立危险因素，与本研究中CTR的高SHAP贡献形成呼应。

本研究尚存在4个方面的局限性，需在后续研

究中加以完善：第一，研究设计为回顾性，不可避免地存在选择偏倚，即纳入的330例患者均为接受手术治疗的病例，未包含“CT随访观察的低危PSNs”，可能导致模型对“惰性结节（如AIS）”的预测效能被高估，未来需通过前瞻性队列纳入非手术病例以验证泛化性；第二，病灶分割采用“手动+区域生长法”，虽经2名医师质控，但仍存在主观差异，且未采用全自动分割算法（如U-Net），可能影响影像组学特征的可重复性，后续可结合深度学习分割工具进一步标准化流程；第三，尽管为多中心研究，但不同中心的CT扫描设备（如GE、西门子）及参数（如管电流调节范围）存在细微差异，虽通过统一窗宽窗位（1500/-600 HU）与重采样间距（1mm × 1mm × 1mm）进行预处理，但仍可能对纹理特征（如GLCM相关性）的提取产生潜在影响，需在更大样本量的多中心研究中引入影像标准化协议；第四，模型仅整合了CT影像组学与临床影像特征，未纳入分子病理指标如表皮生长因子受体（epidermal growth factor receptor, EGFR）突变状态、Ki-67指数等，未来可通过“影像-分子”多模态融合进一步提升预测精度。

综上，本研究证实基于TabPFN模型的CT影像组学联合临床影像特征分析，可有效解决PSNs肺腺癌浸润性预测的“小样本-高异质”难题，其高召回率与稳定性可为临床手术方案选择（肺叶切除*vs.*亚肺叶切除）提供客观依据，后续需通过前瞻性、多设备、多模态的研究设计，进一步推动该模型的临床转化。

[参 考 文 献]

[1] Zuo Z C, Wang L D, Peng K, et al. Development and validation of a nomogram for predicting the 1-, 3-, and 5-year survival in patients with acinar-predominant lung adenocarcinoma [J]. Curr Med Sci, 2022, 42(6): 1178-1185.

[2] Zuo Z, Zhang G, Song P, et al. Survival nomogram for stage IB non-small-cell lung cancer patients, based on the SEER database and an external validation cohort [J]. Ann Surg Oncol, 2021, 28(7): 3941-3950.

[3] Bertolaccini L, Tralongo A C, Del Re M, et al. Segmentectomy vs. lobectomy in stage IA non-small cell lung cancer: a systematic review and meta-analysis of perioperative and survival outcomes [J]. Lung Cancer, 2024, 197: 107990.

[4] Zuo Z, Wang P, Zeng W, et al. Measuring pure ground-glass nodules on computed tomography: assessing agreement between a commercially available deep learning algorithm and radiologists' readings [J]. Acta Radiol, 2023, 64(4): 1422-1430.

[5] Ye T, Deng L, Wang S, et al. Lung adenocarcinomas manifesting as radiological part-solid nodules define a special clinical subtype [J]. J Thorac Oncol, 2019, 14(4): 617-627.

[6] Liu M, Yang L, Sun X, et al. Evaluation of prognosis in patients with lung adenocarcinoma with atypical solid nodules on thin-section CT images [J]. Radiol Cardiothorac Imaging, 2024, 6(1): e220234.

[7] Okamoto T, Miyawaki M, Toyokawa G, et al. Clinical significance of part-solid lung cancer in the eighth edition TNM staging system [J]. Interact Cardiovasc Thorac Surg, 2022, 34(2): 219-226.

[8] Yotsukura M, Asamura H, Motoi N, et al. Long-term prognosis of patients with resected adenocarcinoma in situ and minimally invasive adenocarcinoma of the lung [J]. J Thorac Oncol, 2021, 16(8): 1312-1320.

[9] Watanabe Y, Hattori A, Nojiri S, et al. Clinical impact of a small component of ground-glass opacity in solid-dominant clinical stage IA non-small cell lung cancer [J]. J Thorac Cardiovasc Surg, 2022, 163(3): 791-801.

[10] Shi Y, Wu S, Ma S, et al. Comparison between wedge resection and lobectomy/segmentectomy for early-stage non-small cell lung cancer: a bayesian meta-analysis and systematic review [J]. Ann Surg Oncol, 2022, 29(3): 1868-1879.

[11] Dolan D, Swanson S J, Gill R, et al. Survival and recurrence following wedge resection versus lobectomy for early-stage non-small cell lung cancer [J]. Semin Thorac Cardiovasc Surg, 2022, 34(2): 712-723.

[12] Zuo Z, Deng J, Ge W, et al. Quantifying intratumoral heterogeneity within sub-regions to predict high-grade patterns in clinical stage I solid lung adenocarcinoma [J]. BMC Cancer, 2025, 25(1): 51.

[13] Zuo Z, Zhang G, Lin S, et al. Radiomics nomogram based on optimal volume of interest derived from high-resolution CT for preoperative prediction of IASLC grading in clinical IA lung adenocarcinomas: a multi-center, large-population study [J]. Technol Cancer Res Treat, 2024, 23:15330338241300734.

[14] Qi H, Zuo Z, Lin S, et al. Assessment of intratumor heterogeneity for preoperatively predicting the invasiveness of pulmonary adenocarcinomas manifesting as pure ground-glass nodules [J]. Quant Imaging Med Surg, 2025, 15(1): 272-286.

[15] Zuo Z, Zhang G, Chen J, et al. CT radiomic nomogram using optimal volume of interest for preoperatively predicting invasive mucinous adenocarcinomas in patients with incidental pulmonary nodules: a multicenter, large-scale study [J]. Technol Cancer Res Treat, 2024, 23: 15330338241308307.

[16] Zuo Z, Zeng Y, Deng J, et al. Intratumoral heterogeneity score enhances invasiveness prediction in pulmonary ground-glass nodules via stacking ensemble machine learning [J]. Insights Imaging, 2025, 16(1): 209.

[17] Zuo Z, Zeng W, Peng K, et al. Development of a novel combined nomogram integrating deep-learning-assisted CT texture and clinical-radiological features to predict the invasiveness of clinical stage IA part-solid lung

（下转第81页）

定期进行质量管控、保养等，可及时发现问题及解决问题。PDCA组的设备故障率及维修及时率、质量达标率均优于常规组。实施PDCA循环管理模式后，大幅提升压力蒸汽灭菌器的管理规范性，转变传统维护保养模式，实施预见性维护、前瞻性维护，减少维护成本支出[8]，同时能够更准确地识别灭菌失败风险，帮助医疗机构加强灭菌质量控制和早期干预[9]。通过PDCA管理模式也能提升设备参数及管理质量，以高标准数据，做好精准化医疗设备管理工作，保证医疗设备安全性及有效性。例如，排查故障期间，基于PDCA计划分析压力蒸汽灭菌器常见高温高压故障与多种因素相关，如温度传感器及压力传感器损坏、内部清洁不到位、蒸汽冷凝水无法及时排出[10]、灭菌器内部管道及阀门堵塞、灭菌器内部元件故障、控制系统故障等。排除故障期间，需根据可疑因素分析故障，比如，通过观察运行记录，判断温度及压力是否超标，先出现高温报警，后出现高压报警，均是排除故障的线索，以此精准处理故障[11]，进而对灭菌全过程落实有效监测分析，提高医疗机构自检自查和防范医院感染风险能力[12]。

综上所述，实施PDCA管理模式可改变压力蒸汽灭菌器维护保养现状，发现故障后及时处理，提升设备的总体管理质量，从而减少医院感染发生。

[参 考 文 献]

[1] 郑勤锋, 王月明, 罗锋. 压力蒸汽灭菌器压力参数校准用标准器（压力记录器）的高温校准方法[J].中国品牌与防伪, 2024(3): 54–55.

[2] 韩蔚. 浅析医疗设备维修管理中PDCA方法的采用对设备管理质量的影响[J]. 中国医疗器械信息, 2023, 29(24): 161–163.

[3] 孙建国. 浅析PDCA质量循环在医疗设备维修管理中的应用[J]. 中国设备工程, 2023(18): 64–66.

[4] 陈曼珊. PDCA模式结合PLM评价模式下的管理方法在医疗设备维修管理中的应用[J]. 现代医院, 2023, 23(7): 1071–1074, 1078.

[5] 陈明东. 兽医实验室手提式压力蒸汽灭菌器的正确使用方法[J]. 畜牧兽医科技信息, 2023(4): 38–40.

[6] 杨飞. 医疗设备维修管理中戴明环（PDCA）方法的应用价值体会[J]. 中国医疗器械信息, 2023, 29(3): 161–163.

[7] 朱子涵. 浅谈PDCA质量循环在医疗设备维修管理中的应用[J]. 中国设备工程, 2022(16): 15–17.

[8] 金光波. PDCA理论在本院医疗设备维修管理中的应用观察[J]. 中国医疗器械信息, 2022, 28(12): 174–176.

[9] Zhao X, Liu T, Ma N, et al. Construction of management tools for early warning of prevacuum steam sterilization failure[J]. PLoS One, 2025, 20(1): e0316322.

[10] 撖子建, 李新文, 王瑶. PDCA管理框架下多种管理方法在检验科设备维护管理中的应用[J]. 中国医疗设备, 2021, 36(12): 139–141, 145.

[11] 李伯宁, 谷鑫, 王衍海, 等. 影响脉动真空压力蒸汽灭菌器灭菌质量的原因分析及对策[J]. 中国医学装备, 2021, 18(6): 146–148.

[12] 罗锦雁, 戴传文, 王功军, 等. 基层医疗机构小型压力蒸汽灭菌器使用现况调查[J]. 中国消毒学杂志, 2021, 38(4): 282–285.

（上接第35页）

adenocarcinoma: a multicentre study[J]. Clin Radiol, 2023, 78(10): e698–e706.

[18] Li Y, Ding J, Wu K, et al. Ensemble machine learning classifiers combining CT radiomics and clinical–radiological features for preoperative prediction of pathological invasiveness in lung adenocarcinoma presenting as part–solid nodules: a multicenter retrospective study[J]. Technol Cancer Res Treat, 2025, 24: 15330338251351365.

[19] Weng Q, Zhou L, Wang H, et al. A radiomics model for determining the invasiveness of solitary pulmonary nodules that manifest as part–solid nodules[J]. Clin Radiol, 2019, 74(12): 933–943.

[20] Wu G, Woodruff H C, Shen J, et al. Diagnosis of invasive lung adenocarcinoma based on chest CT radiomic features of part–solid pulmonary nodules: a multicenter study[J]. Radiology, 2020, 297(2): 451–458.

[21] Bühlmann P, Hothorn T. Boosting algorithms: regularization, prediction and model fitting[J]. Stat Sci, 2007, 22(4): 477–550.

[22] Hollmann N, Müller S, Purucker L, et al. Accurate predictions on small data with a tabular foundation model[J]. Nature, 2025, 637: 319–326.

[23] Li Y, Yang J, Xiao P, et al. MRI delta–radiomics and morphological feature–driven TabPFN model for preoperative prediction of lymphovascular invasion in invasive breast cancer[J]. Technol Cancer Res Treat, 2025, 24: 15330338251362050.

[24] Zheng H, Chen W, Qi W, et al. Enhancing the prediction of the invasiveness of pulmonary adenocarcinomas presenting as pure ground–glass nodules: integrating intratumor heterogeneity score with clinical–radiological features via machine learning in a multicenter study[J]. Digit Health, 2024, 10: 20552076241289181.

[25] Ling T, Zuo Z, Wu L, et al. Predicting neoadjuvant chemotherapy response in locally advanced gastric cancer using a machine learning model combining radiomics and clinical biomarkers[J]. Digit Health, 2025, 11: 20552076251341740.

[26] Ling T, Zuo Z, Huang M, et al. Prediction of mucinous adenocarcinoma in colorectal cancer with mucinous components detected in preoperative biopsy diagnosis[J]. Abdom Radiol (NY), 2025, 50(7): 2794–2805.