

王甘霖, 靳明明, 黄钢. 深度学习与机器学习预测 NSCLC 的 N 分期[J]. 智能计算机与应用, 2024, 14(4): 184-189. DOI: 10.20169/j.issn.2095-2163.240430

深度学习与机器学习预测 NSCLC 的 N 分期

王甘霖¹, 靳明明², 黄钢^{1,2}

(1 上海理工大学 健康科学与工程学院, 上海 200093; 2 上海健康医学院 分子影像学重点实验室, 上海 201318)

摘要: 本文利用深度学习和机器学习技术, 提出一种基于 PET 和 CT 图像的非小细胞肺癌 (Non-Small Cell Lung Cancer, NSCLC) 肿瘤 N 分期预测方法, 以提高 N 分期的准确率和稳定性。使用深度学习残差神经网络 3D Resnet50 提取 NSCLC 的 CT 与 PET 影像特征, 并将特征降维到 32 维; 将多模态特征进行拼接, 输入机器学习模型进行 NSCLC 的 N 分期预测。最终, 本文所提出的基于深度学习和机器学习预测 NSCLC 肿瘤分期预测方法的预测准确率在训练集达到了 1.0, 测试集上达到了 0.852, 平均受试者工作特征曲线下面积 (Receiver Operating Characteristic, ROC) 超过了 0.91。实验表明, 基于深度学习和机器学习的 NSCLC 肿瘤预测方法可以提高非小细胞肺癌 N 分期的准确率和稳定性, 具有良好的应用前景和实际应用价值。

关键词: 3D Resnet50; 多模态; NSCLC; N 分期

中图分类号: R318

文献标志码: A

文章编号: 2095-2163(2024)04-0184-06

Deep learning and machine learning for predicting N stage of NSCLC

WANG Ganlin¹, JIN Mingming², HUANG Gang^{1,2}

(1 School of Health Science and Engineering, University of Shanghai for Science and Technology, Shanghai 200093, China; 2 Shanghai University of Medicine and Health Sciences, Shanghai Key Laboratory of Molecular Imaging, Shanghai 201318, China)

Abstract: In this paper, a PET and CT image-based tumor N-staging prediction method for Non-Small Cell Lung Cancer (NSCLC) is proposed using deep learning and machine learning techniques to improve the accuracy and stability of N-staging. Deep learning residual neural network 3D Resnet50 is used to extract the CT and PET image features of NSCLC, and the features are downscaled to 32 dimensions; the multimodal features are spliced and input into a machine learning model for N-staging prediction of NSCLC. Ultimately, the prediction accuracy of the deep learning and machine learning based NSCLC tumor staging prediction method proposed in this paper reached 1.0 on the training set and 0.852 on the test set, and the average subject area under the curve (Receiver Operating Characteristic, ROC) exceeded 0.91. The experiments showed that the NSCLC tumor prediction method based on deep learning and machine learning can improve the accuracy and stability of N-staging of non-small cell lung cancer, which has good application prospects and practical application value.

Key words: 3D Resnet50; multimodal; NSCLC; N staging

0 引言

根据世界卫生组织的数据, 肺癌是全球最为常见且致命的恶性肿瘤之一, 其中 NSCLC 占据肺癌病例比例最高, 为 80%~85%^[1]。美国癌症联合委员会第八版的分期系统采用肿瘤、淋巴结和转移的分期标准对患者进行分类。根据这一分类标准, 在 NSCLC 患者中大约有 30%~40% 的患者会经历淋巴结转移, 这些转移引起的并发症可能对生活产生难

以预测的影响, 甚至可能威胁生命^[2]。王会等^[3]在 2020 年的一项研究表明, NSCLC 作为全球最常见的肺癌类型之一, 每年导致约 160 万人死亡。由于 NSCLC 表现出高度的异质性, 不同患者之间存在显著的病理特征和分子表型差异, 因此制定个体化评估策略对于提高疗效至关重要^[4]。

目前, 对于 NSCLC 淋巴结转移的评估主要依赖影像学和病理学方法。对于病理学方法通过分析肿瘤组织或体液中的分子标志物, 如基因、蛋白质和非

作者简介: 王甘霖(1994-), 男, 硕士研究生, 主要研究方向: 人工智能在医学影像中的应用。

通讯作者: 靳明明(1987-), 男, 博士, 讲师, 主要研究方向: 分子生物学。Email: asdjningming@126.com; 黄钢(1961-), 男, 博士, 教授, 主要研究方向: 核医学。Email: huanggang@sumhs.edu.cn

收稿日期: 2023-07-02

哈尔滨工业大学主办 ◆ 专题设计与应用

编码 RNA 等,来评估肿瘤的生物学特性和预后情况,这种检测方法能够提供临床检测所无法提供的分子水平信息^[5-8];病理学方法具有侵入性风险,其对人体伤害较大^[9]。影像学方法通过结合代谢信息和解剖结构信息,为 NSCLC 的分期和治疗提供了更好的评估^[10-11]。然而,这种方法需要依赖医生进行主动识别,因此医生在进行分期工作时面临着巨大的工作量,导致其分辨可能不准确^[12]。此外,医生在解读影像时可能会受到个人经验、知识水平和解读技巧的影响,这些因素都可能导致对影像的误解。例如,对于患者的影像,医生需要综合考虑代谢特征和解剖结构信息,其主观性可能会影响到对肿瘤分期和治疗计划的决定^[13]。因此,除提高医生对影像技术的理解和解读能力外,开发更为客观的影像解读工具,对于减少人为因素导致的误差具有重要意义。

近年来,深度学习和机器学习在医学图像处理领域得到了广泛应用。2017年,Wang H等^[14]提出了一种基于深度学习的计算机辅助诊断方法,采用反向传播人工神经网络(BP-ANN)基于168名患者的PET/CT图像提取深度学习特征,通过准确性和ROC曲线下面积等标准进行了评估,准确率和ROC下面积分别为0.86和0.91,与医生相比,深度学习方法表现出更高的敏感性,但特异性较低;2018年,Kim团队^[15]研究了NSCLC治疗后淋巴结残留的PET摄取转移情况,利用逻辑回归模型评估了CT放射组学方法在预测这些转移方面的潜在价值,该预测模型的ROC下面积值达到0.77,显示出该方法在预测转移风险方面具有一定的有效性;2018年,Gu-Wei Ji团队^[16]运用支持向量机(SVM)和最小绝对收缩与选择算子算法,对177名患者的CT图像进行影像组学特征分析,研究结果显示,基于放射组学特征的列线图在预测淋巴结转移上展现出良好的性能,主要队列和验证队列的ROC下面积值均达到0.81和0.80,显示出较高的准确性和可靠性;2020年,Cong等^[17]结合临床参数和影像组学特征,开发和验证了针对NSCLC患者淋巴结转移的预测模型,利用最小绝对收缩和选择算子算法进行特征选择,模型预测结果的ROC下面积值高达0.86,显示出该模型在预测淋巴结转移风险方面的优越性能;2023年,Zhang等^[18]使用多视角影像组学和深度学习网络ResNet18建立了预测术前临床I-II期NSCLC患者N2淋巴结转移的模型,在预测NSCLC分期方面的AUC为0.86。这些方法基于大量的肺部影像数

据,通过分析图像特征和建立模型来实现对NSCLC的分期预测。本文则是结合了PET和CT图像的多模态特征,利用深度学习和机器学习对NSCLC分期做出了较准确的预测。

1 方法

1.1 深度学习模型架构及特征提取

本文使用3D Resnet50模型分别对样本PET和CT的影像特征进行采集,深度学习模型3D Resnet50架构如图1所示。首先,图像(Images)输入到3D Resnet50的三维卷积层(Conv3D),经过三维卷积处理后,进行三维批量归一化以减少内部协变量偏移,数据送入线性整流函数(ReLu)进行特征处理,线性整流函数的输出 $f(x)$,见公式(1):

$$f(x) = \max(0, x) \quad (1)$$

其中, x 是输入。

由于线性整流函数本身的特点,经过ReLu处理的特征具有非线性特点,同时抑制了不必要的特征;数据送入最大池化层(MaxPool)进行特征压缩并保留关键特征,重要特征被逐步送入残差层(Layer),残差层Layer1的架构图如图2所示,由残差块和下采样构成,残差层使输入特征直接跨越多个层级传递到输出,从而保留原始特征的信息。残差的核心思想如公式(2)所示:

$$Output = F(x) + x \quad (2)$$

其中, $F(x)$ 可以是多个卷积层、批量归一化层等构成的复杂结构,而 x 则是输入直接经过跳跃连接传递的值。

每个残差层都由残差块(ResBlock)和下采样(Downsample)构成,不同的是,Layer3在下采样后使用4次残差块以保留更多的原始特征。经过残差层的数据输入到平均池化层,更进一步地降低特征维度,保留了更多的全局信息。由于本文使用的是预训练模型,因此特征从平均池化层获取,相较于全连接层(FC),平均池化层提取的特征更加具有语义信息,汇聚了整个特征图的信息,相比之下,全连接层提取的特征更加依赖于具体任务的细节,不具备通用性。

对于平均池化层的输出维度,每一个样本的维度为2048,维度太大,会极大增加机器学习模型的计算量,因此利用主成分分析(Principal Component Analysis, PCA)对PET、CT的每个样本数据分别降维到32维,以提高机器学习模型的计算效率^[19]。

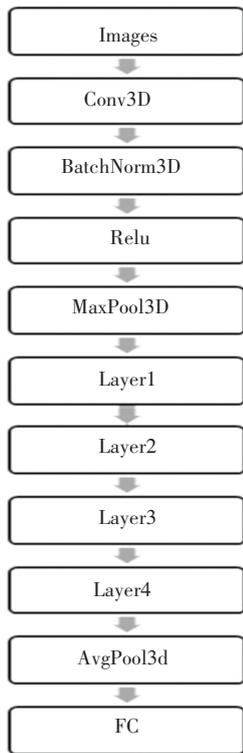


图1 3D Resnet50 架构图

Fig. 1 3D Resnet50 architecture diagram

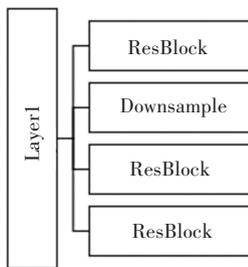


图2 Layer1 的架构图

Fig. 2 Layer1 architecture diagram

由于深度学习模型在特征处理方面采用了端到端的方式,因此模型提取的特征通常较难解释。为了更好地理解模型的工作方式,本文采用了梯度加权类激活映射(Gradient-weighted Class Activation Mapping, GradCAM)来展示特征提取的过程,也就是生成的热图,如图3所示。相关特征:3D Resnet50模型在训练过程中学习到了与肿瘤相关的特征,包括形状、纹理、边缘等与肿瘤区域有关的视觉特征,因此,GradCAM会将红色部分聚集在肿瘤区域。局部响应:GradCAM是基于局部梯度计算的,而模型的决策可能受到多个因素的影响。因此,在某些情况下,即使肿瘤区域的大部分红色区域聚集在一起,模型在预测时同时关注了其他与肿瘤无关的图像特征,仍然可能存在一小部分红色区域不在肿瘤区域

内。噪声或误差:在生成 GradCAM 图像的过程中,可能存在一定程度的噪声或误差,由于模型的复杂性或数据的特点导致一小部分红色区域出现在非肿瘤区域中。

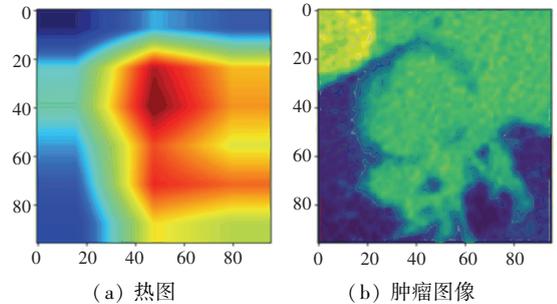


图3 CT 图像热图

Fig. 3 CT image thermal map

1.2 数据集与设备

数据集是来自斯坦福大学医学院与帕洛阿尔托退伍军人事务医疗系统 211 名病人的数据^[20]。该数据集统计了临床信息包括年龄、性别、体重、抽烟史、癌症位置、肿瘤大小(Tumor size)、N 分期情况(Lymph Node involvement)、远处转移情况(Metastasis)、组织病理分级、淋巴血管侵犯、辅助治疗等等。影像数据包括肺部 CT 和 PET 数据,每个病人约 90~300 张图片。在数据集中,病人的 N 分期信息见表 1。

表1 病人 N 分期信息

Table 1 Patient's N-stage information

N 分期	病人数量
N0	129
N1	18
N2	15
N3	0
NX	0
缺失	49

设备是 GTX 3080Ti 16 G 显存的 GPU、windows11 64 位系统、32 G 内存、24 核 Inter(R) Core(TM) i9-12900HX 芯片。软件是 python3.9, cuda11.6, pytorch1.13.1。

1.3 数据预处理

首先筛选掉没有 N 分期信息 49 个样本,去掉影像质量模糊的 28 个样本,最终有 134 个样本参与运算,其中 N0 样本 105 个, N1 样本 17 个, N2 样本 12 个。对样本数据进行批量重采样和数据归一化,原始影像的格式由 Dicom 格式转为 Nifti 格式,数据形式为 $D \times 512 \times 512$ 单通道的三维矩阵,每一个三维矩阵代表一位样本的影像信息,其中 D 代表每个样本的二维影像数量。

1.4 特征筛选与模型训练

对深度学习提取的多模态特征进行正则化以提高模型的泛化能力,接着对特征进行相关系数可视化,如图 4 所示。特征筛选首先采用皮尔森检验,为了筛选掉高度重复性的特征,本文把相关系数大于 0.9 的特征中随机的一个特征筛选掉。为防止模型过拟合,采用最小绝对收缩和选择算子(LASSO)算法来构建惩罚系数,进一步筛选出权重系数绝对值大于 10^{-6} 特征,LASSO 路径如图 5 所示,横坐标为惩罚系数,纵坐标为权重系数,筛选出来的特征如图 6 所示。

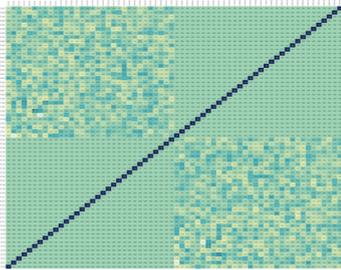


图 4 PET 与 CT 的联合特征可视化

Fig. 4 Joint feature visualization of PET and CT

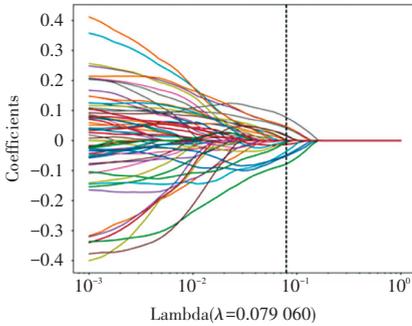


图 5 Lasso 路径图

Fig. 5 Lasso path map

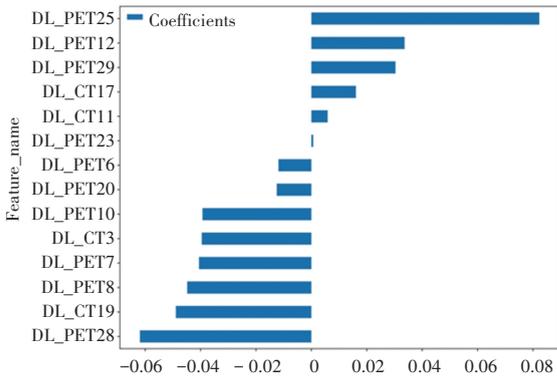


图 6 Lasso 算法筛选出的特征

Fig. 6 Features filtered by Lasso algorithm

本文将数据集按照 4 : 1 的比例划分为训练集和测试集。采用的机器学习模型有 SVM、KNN、RandomForest、ExtraTrees、XGBoost、LightGBM、NaiveBayes、AdaBoost、GradientBoosting、MLP。这些模

型由 sklearn 的官方 API 提供,将训练集数据输入到模型中进行训练,训练好的模型用测试集测试,根据测试集的准确率和受试者工作曲线选取最佳模型,获得最佳方法。

2 评价指标

准确率、受试者工作曲线评估分类模型的性能。

1) 准确率 (Accuracy)

准确率是分类问题中最常用的评估指标,表示模型预测正确的样本占总样本的比例,如公式(3)所示:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (3)$$

其中, TP 为真正例,表示模型将正类别样本正确地预测为正类别的数量; TN 为真负例,表示模型将负类别样本正确地预测为负类别的数量; FP 为假正例,表示模型将负类别样本错误地预测为正类别的数量; FN 为假负例,表示模型将正类别样本错误地预测为负类别的数量。

2) 受试者工作曲线 (Receiver Operating Characteristic, ROC)

ROC 曲线是以假正例 (False Positive Rate, FPR) 为横轴,真正率 (True Positive Rate, TPR) 为纵轴,绘制的一条曲线。通过改变分类模型的阈值,可以得到不同点对应的 TPR 和 FPR 。通常情况下,ROC 曲线越靠近左上角,模型的性能越好。 FPR 和 TPR 如公式(4) ~ 公式(5) 所示:

$$TPR = \frac{TP}{TP + FN} \quad (4)$$

$$FPR = \frac{FP}{TN + FP} \quad (5)$$

3) 微平均 (Micro-Average ROC)

在多类别分类问题中,如果要计算 ROC 曲线的微平均,需要将所有类别的真正例数 TP 、假正例数 FP 、真负例数 TN 和假负例数 FN 进行汇总,然后计算微平均的 TPR 和 FPR ,如公式(6) ~ 公式(7) 所示:

$$TPR_{Micro} = \frac{\sum_{i=1}^c TP_i}{\sum_{i=1}^c (TP_i + FN_i)} \quad (6)$$

$$FPR_{Micro} = \frac{\sum_{i=1}^c FP_i}{\sum_{i=1}^c (FP_i + TN_i)} \quad (7)$$

其中, C 是类别总数, TP_i, FP_i, TN_i, FN_i 分别表示第 i 个类别的真阳性数、假阳性数、正例数和负例数。同样, 曲线越靠近左上角, 模型的性能越好。

4) 宏平均 (Macro-Average ROC)

宏平均 ROC 通过计算每个类别的真正例率和假正例率的平均值得到, 这种方法对每个类别的性能都给予了同等重视, 不考虑类别的样本不平衡情况。因此, Macro-Average ROC 用于评估分类器在各个类别上的平均性能, 计算如公式 (8) ~ (9) 所示:

$$TPR_{Macro} = \frac{1}{C} \sum_{i=1}^C TPR_i \quad (8)$$

$$FPR_{Macro} = \frac{1}{C} \sum_{i=1}^C FPR_i \quad (9)$$

其中, TPR_i 表示第 i 个类别的真阳性率; FPR_i 表示第 i 个类别的假阳性率; C 代表样本总的个数。曲线越靠近左上角, 模型的性能越好。

3 实验结果

每个机器学习模型的准确度, 如图 7 所示。可见 RandomForest、ExtraTree、XGBoost 结果较好, 在训练集的准确率达到 0.991、1.0、1.0 的准确度, 测试集的准确率分别为 0.778、0.852、0.778。其他的模型除了 LightGBM, 测试集和训练集的准确度普遍都在 0.75 以上。

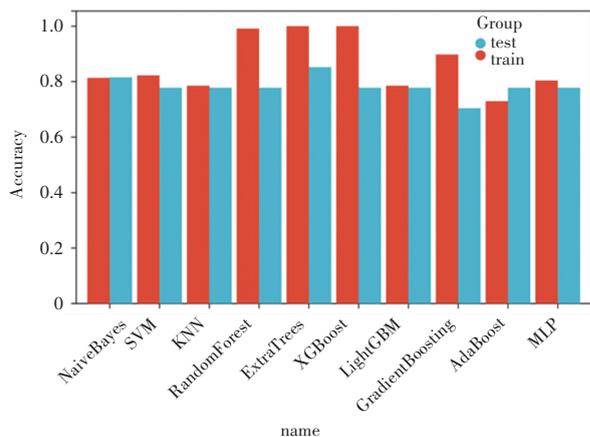


图 7 各个机器学习模型的准确度

Fig. 7 Accuracy of each machine learning model

RandomForest、ExtraTrees、XGBoost 模型的 ROC 曲线如图 8~图 10 所示。其中, class 0、class 1、class 2 代表 N0、N1、N2 3 个类别。测试集上的结果表明, RandomForest 的微平均和宏平均的 ROC 达到了 0.8 以上, 各个类别的 ROC 达到了 0.72 以上; ExtraTrees 的微平均和宏观 ROC 均达到了 0.87 以上, 各个类别的 ROC 达到了 0.87 以上; XGBoost 的微平均和宏

平均 ROC 达到了 0.76 以上, 各个类别的 ROC 达到了 0.64 以上。ExtraTrees 的整体平均 ROC (单类别 ROC 平均数) 最高, 为 0.93。

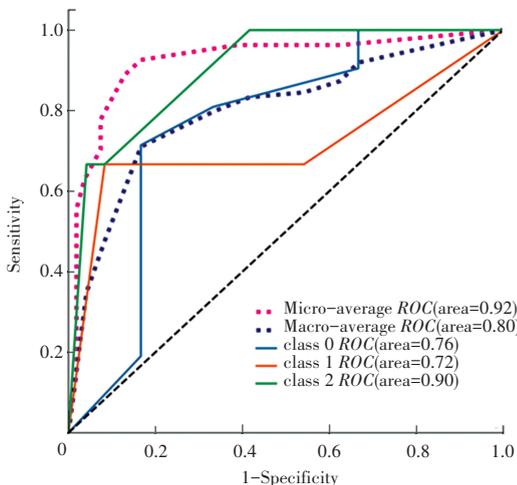


图 8 随机森林 ROC 图示

Fig. 8 ROC diagram of RandomForest

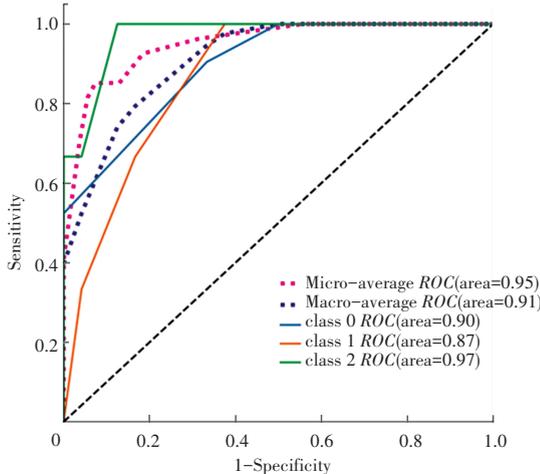


图 9 ExtraTrees 的 ROC 图示

Fig. 9 ROC diagram of ExtraTrees

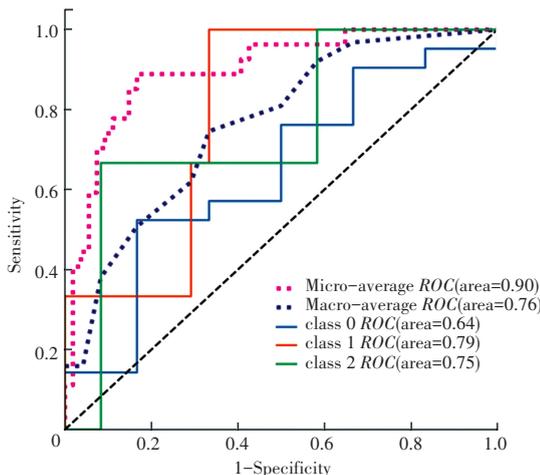


图 10 XGBoost 的 ROC 图示

Fig. 10 ROC diagram of XGBoost

4 结束语

本文利用深度学习和机器学习提出了一种基于 PET 和 CT 图像的非小细胞肺癌肿瘤 N 分期预测方法。在使用 3D Resnet50 提取特征所构建的数据集中,3 个最佳机器学习模型 RandomForest, ExtraTrees 和 XGBoost 均显示出了很高的预测准确性,其中 ExtraTrees 模型在 NSCLC 的 N 分期预测方面表现最好。在测试集中,ExtraTrees 模型的准确率比其它两个模型高出 0.074,ExtraTrees 模型的宏平均 ROC 超过了 0.91,表明其在 NSCLC 的 N 分期预测方面的可靠性和准确性。本文 NSCLC 的 N 分期预测提供了新的思路和方法,为临床医生提供了更加准确可靠的分期预测工具和指导意见,为临床实践中 NSCLC 的治疗和预后评估提供有益帮助。

但本文提出的方法也存在一些局限性。首先,本文未收集到 N2 分期以上的病人数据,因此本文只对 NSCLC 的 N2 及以下分期的预测具有借鉴意义。此外,病人的样本数据分布不平衡,对准确性和可靠性产生一定的影响。在今后的研究中,需要扩大 NSCLC 样本数据的范围,涵盖更多的 N 分期和性别之类的临床数据,以进一步提高分类预测的准确性和可靠性。

参考文献

[1] PIRLOG R, CHIROI P, RUSU I, et al. Cellular and molecular profiling of tumor microenvironment and early-stage lung cancer [J]. International Journal of Molecular Sciences, 2022, 23(10): 5346.

[2] 邓仁芳,曾月,潘越,等. 驱动基因阳性非小细胞肺癌免疫治疗进展[J]. 中国肺癌杂志,2022,25(3):201-206.

[3] 王会,周雯敏,郭乔如,等. 慢性阻塞性肺疾病转化为肺癌的研究进展[J]. 药学学报,2020,55(7):1410-1418.

[4] 何建行. 基于分子标志的非小细胞肺癌术后复发预测专家共识 [J]. 中国肺癌杂志,2022,25(10):701-714.

[5] CHEN Qing, SUN Zhuo, LI Jinfang, et al. Set domain-containing protein 5 enhances the cell stemness of non-small cell lung cancer via the PI3K/Akt/mTOR pathway [J]. Journal of Environmental Pathology, Toxicology and Oncology, 2021, 40(2):55-63.

[6] GUO F, GUO Lili, LI Yongwen, et al. MALAT1 is an oncogenic long non-coding RNA associated with tumor invasion in non-small cell lung cancer regulated by DNA methylation [J]. International Journal of Clinical and Experimental Pathology, 2015, 8(12):15903-15910.

[7] XU W, HANG M, YUAN C Y, et al. MicroRNA-139-5p

inhibits cell proliferation and invasion by targeting insulin-like growth factor 1 receptor in human non-small cell lung cancer [J]. International Journal of Clinical and Experimental Pathology, 2015, 8(4):3864-3870.

[8] LIU Juncai, LIU Xiaochun. UBE2T silencing inhibited non-small cell lung cancer cell proliferation and invasion by suppressing the wnt/ β -catenin signaling pathway [J]. International Journal of Clinical and Experimental Pathology, 2017, 10(9):9482-9488.

[9] LI Yanhua, WEI Suju, WANG Junyan, et al. Analysis of the factors associated with abnormal coagulation and prognosis in patients with non-small cell lung cancer [J]. Chinese Journal of Lung Cancer, 2014, 17:789-796.

[10] OWENS C, HINDOCHA S, LEE R, et al. The lung cancers: staging and response, CT, 18F-FDG PET/CT, MRI, DWI: review and new perspectives [J]. The British Journal of Radiology, 2023, 96(1148): 1148.DOI:10.1259/bjr.20220339

[11] BURY T, PAULUS P, DOWLATI A, et al. Staging of the mediastinum: value of positron emission tomography imaging in non-small cell lung cancer[J].The European Respiratory Journal, 1996, 9(12):2560-2564.

[12] Betancourt-Cuellar S L, Carter B W, Palacio D, et al. Pitfalls and limitations in non-small cell lung cancer staging[J].Seminars in Roentgenology, 2015, 50(3):175-182.

[13]STEINERT H. PET and PET-CT of lung cancer[J].Methods in Molecular Biology, 2011, 727:33-51.

[14] WANG H, ZHOU Z, LI Y, et al. Comparison of machine learning methods for classifying mediastinal lymph node metastasis of non-small cell lung cancer from 18 F-FDG PET/CT images [J]. Ejnmmi Research, 2017, 7: 1-11.

[15]KIM C H, LEE H Y, KIM H K, et al. P2.01-56 metastases in residual pet uptake of lymph nodes after treatment: added value of ct radiomic approach for prediction [J]. Journal of Thoracic Oncology, 2018, 13(10): S686.

[16]JI Guwei, ZHANG Yudong, ZHANG Hui, et al. Biliary tract cancer at CT: A radiomics-based model to predict lymph node metastasis and survival outcomes [J]. Radiology, 2019, 290(1): 90-98.

[17]CONG M, FENG H, REN J L, et al. Development of a predictive radiomics model for lymph node metastases in pre-surgical CT-based stage IA non-small cell lung cancer [J]. Lung Cancer, 2020, 139: 73-79.

[18]ZHANG Haifei, LIAO Meiyang, GUO Qun, et al. Predicting N2 lymph node metastasis in presurgical stage I-II non-small cell lung cancer using multiview radiomics and deep learning method [J]. Medical Physics, 2023, 50(4): 2049-2060.

[19]UKAI S, DAISAKI H, ISHIYAMA M, et al. Reproducibility of the principal component analysis (PCA)-based data-driven respiratory gating on texture features in non-small cell lung cancer patients with (18) F-FDG PET/CT [J]. Journal Applied Clinical Medical Physics, 2023, 24(5): e13967.

[20]BAKR S, GEVAERT O, ECHEGARAY S, et al. A radiogenomic dataset of non-small cell lung cancer [J]. Scientific Data, 2018, 5(1): 180202.