

# 影像组学在预测肺腺癌 EGFR、TP53 突变中的应用

王卯<sup>1</sup> 王锐<sup>1\*</sup> 王锦玉<sup>2</sup>

1. 承德医学院附属医院 河北省承德市 067000

2. 河北大学附属医院消化科 河北省保定市 071000

**摘要：**肺癌是全球癌症相关死亡率最高的原因，其中肺腺癌是最常见的病理类型，占有肺癌的40%<sup>[1]</sup>。许多研究表明，实施肺癌筛查计划后，肺癌尤其是肺腺癌的检出率逐年增加，肺腺癌已成为肺癌最常见的亚型，基因突变率最高。肺癌驱动基因的发现，尤其是表皮生长因子受体（EGFR）基因，增加了个体化靶向治疗的采用。肺腺癌患者中，EGFR突变的患者对EGFR酪氨酸激酶抑制剂（TKI）反应良好，而没有这些突变的患者在疾病任何阶段均不适合接受EGFR-TKI治疗<sup>[2]</sup>。然而，部分患者表现出对EGFR-TKI的原发性耐药，部分患者即便初始有反应，也存在耐药性不可避免地发展。TP53的治疗前改变是与一线无进展生存期和总生存期降低相关的最常见的并发基因组事件。因此，在对肺腺癌患者进行靶向药物治疗之前，检测相关基因的表达状态极为重要。通过活检或手术切除获得的肿瘤组织的分子检测是鉴定基因突变的金标准，但是组织样本的采集伴随着侵入性，不可避免会存在一定的采样误差。影像组学可以从图像中提取定量特征，全面分析生物学信息，评估肿瘤异质性。在反映肿瘤基因表型等方面表现出了相应潜力，在预测肺腺癌基因突变方面表现出良好的应用前景，为临床提供了一种无创、简便的检测手段。

**关键词：**影像组学；EGFR；TP53；共突变

## 1. CT 影像组学概述、程序和分类

影像组学由LAMBIN等在2012年首次提出<sup>[3]</sup>，主要涉及4个步骤。①图像采集、②图像分割、③特征提取、④数据分析。传统的影像组学（TR）主要包括图像采集、重建、分割、特征提取、过滤、模型建立和性能评估等程序。在模型建立过程中使用多个分类器。根据不同的分类器，可将影像组学分为基于统计的传统的影像组学（TSR）、基于机器学习的影像组学（MLR，特别是浅层学习）和深度学习（DL），以反映组学技术的迭代。TSR可以通过提取TR特征（例如，基于形状的特征、直方图特征和纹理特征）并应用最小绝对收缩和选择运算符（LASSO）等算法集成的数学统计方法来识别具有非零系数的最有价值的特征。然后为每个病灶计算和建立影像组学评分（Rad评分），并采用所选特征的线性组合，相应地按其系数加权。然后采用经典的逻辑回归（LR）方法作为分类器来构建放射组学模型。这种方法是最早、最简单、最易解释的组学技术。MLR是最成熟和主流的方法，它在特征提取和最佳影像组学特征筛选后使用机器学习算法建立分类或预测模型。常见的机器学习分类器包括随机森林（RF）、支持向量机（SVM）、决策树（DT）、贝叶

斯网络（BN）和k最近邻（KNN）算法。在MLR之后，有一种独特的方法称为基于深度学习的放射组学（DLR），它可以使用人工神经网络（ANN）提取DL特征，例如卷积神经网络（CNN），然后使用机器学习算法构建模型。DL与TSR和MLR代表的TR不同，它依赖于多层非线性神经网络，绕过了繁琐的特征提取和选择程序，自动从图像中学习特征，从而根据“端到端”工作流程建立预测模型，无需任何人工干预<sup>[4]</sup>。

## 2. 影像组学与EGFR

利用常规CT成像特征和基于CT图像的影像组学预测肺腺癌中EGFR突变及其突变亚型是一个热门话题。常规CT成像特征维数低，预测效果有限。影像组学为预测肺腺癌中的EGFR基因突变提供了一种新的定量分析方法，具有阳性预测价值，并有可能成为鉴定EGFR突变、其亚型和耐药基因突变的替代生物标志物。然而，并非所有基于CT图像的影像组学预测模型都显示出特别好的性能。在Liu等人的研究中，使用临床变量结合常规CT成像特征（AUC=0.778）优于单独使用临床变量（AUC=0.690）<sup>[5]</sup>。Mei等人通过放射组学和临床特征组合来预测肺腺癌的EGFR外显子19和

21 突变, 获得了中等诊断性能(外显子 19 突变、外显子 21 突变和全部外显子突变的 AUC 分别为 0.655、0.675 和 0.664)。人们认为, EGFR 突变亚型的预测性能不佳, 主要是由于 TSR 技术的固有限制, 导致训练模型过度拟合。这种情况部分源于影像组学特征的高维性, 单一的 LASSO 回归可能会导致关键信息的丢失。影像组学作为一个新兴领域迅速发展, 但 TSR 继续被使用主要是因为它具有良好的可解释性和易于在临床实践中应用等优点。TSR 可以有效地将数据转化为分数, 通常以列线图的形式出现, 从而可以直观地计算并更容易理解和分析数据之间的关联和差异。然而, 它的局限性在于, 传统 logistic 回归模型的预测效果较差, 相较之下, 这也是 MLR 的使用越来越普遍的原因。

大多数机器学习算法旨在开发最佳模型来解决一个问题。机器学习分类器的选择需要综合考虑数据特征、模型性能、计算资源和使用情况等多个因素, 并且需要进行实验来确定最佳分类器。RF 算法是最经典的机器学习算法之一, 是一种基于决策树的集成学习算法, 它整合了几个弱分类器获得的结果, 以提高模型性能。Gao 等人发现, 随机森林(RF)在 CT、PET 和 PET/CT 影像组学模型中表现优于逻辑回归(LR)和支持向量机(SVM), 其测试集的 AUC 分别为 0.726、0.678 和 0.704。此外, Yang et al 揭示了 SVM 模型优于 RF 模型和 DT 模型, 分别达到 0.805 和 0.859 的 AUC 值来预测测试集中的 19del 和 21L858R 突变。基于亚型预测结果, MLR 的总体结果优于 TSR。

随着 AI 技术的不断发展, 以 CNN 为代表的 DL 已经显示出令人满意的性能。尽管基于机器学习的影像组学方法具有优势, 但手绘特征需要对病变组织进行耗时的边界分割, 因此一些研究提出了使用深度学习方法来学习某些基因相关特征, 从而避免了特征工程的四个复杂过程。DLR 作为 TR 和 DL 之间的过渡, 使用多层神经网络对高级特征和图像数据进行提取和分类, 从而提高数据分析的准确性。该领域关于 DLR 的研究很少, 将来可能需要更多的研究来不断探索和验证该提案的可行性和性能。

### 3. 影像组学与 EGFR 与 TP53 共突变

具有表皮生长因子受体(EGFR)突变的晚期 NSCLC 患者应接受 EGFR-TKI 作为其主要治疗形式。因此, 早期识别 EGFR 突变对于确定合适的靶向治疗方案至关重要。然而, 并非所有患者都会从这种方法中受益。TP53 是一种肿瘤抑

制基因, 该基因突变会导致体内基因监测丧失, 从而导致肿瘤增殖。在非小细胞肺癌(NSCLC)患者中, TP53 突变常与 EGFR 突变并存, 这会降低患者对 EGFR-TKI 的反应, 并导致不良预后。

以前的基因状态预测模型主要关注 EGFR、KRAS 和 ALK 等基因, 而对 TP53 的研究相对较少。此外, 关于 TP53 基因突变的预测往往是单独进行的, 缺乏对基因共突变状态影响的综合考虑。Zhang 等人开发并验证了一种深度学习放射组学模型“Deep-RadScore”, 用于预测预后因素、基因突变和免疫分子表达水平<sup>[6]</sup>。先前的一项研究应用了三分类策略, 利用影像组学来探索 EGFR 和 TP53 的基因共突变状态, 包含临床特征、影像诊断和影像组学特征的综合模型达到了 AUC=0.791<sup>[7]</sup>。在另一些学者的研究中<sup>[8]</sup>, 采用了逐步二分类策略来构建模型, 并利用影像组学和深度学习来预测 EGFR 突变队列中是否合并 TP53 突变。目前, 使用 DL 模型进行 EGFR 共突变状态预测的研究同样很少, 在未来需要进行更深入的研究工作。

### 4. 小结

许多研究表明共突变在 EGFR 突变型肺腺癌中对 EGFR-TKI 的反应及耐药起着重要作用, 其中, EGFR-TP53 共突变作为晚期肺腺癌中最常见的共突变类型, 迫切需要了解 EGFR-TP53 共突变与 EGFR-TKI 疗效及预后之间的关系, 指导临床医生为这部分患者选择最佳的治疗策略。影像组学是一种成熟的研究方法, 形成了完整的理论体系和研究过程, 但基于人工智能的医学影像组学仍处于起步阶段, 尚未得到广泛应用。随着机器学习方法的进步和 DL 技术的不断发展, 通过放射组学鉴定肺腺癌 EGFR-TP53 共突变状态将继续成为临床医学深入研究的领域, 为患者提供更精确的影像学诊断和个体化治疗指导。

### 参考文献:

- [1] Siegel RL, Miller KD, Wagle NS, Jemal A. Cancer statistics, 2023. CA: a cancer journal for clinicians. 2023;73(1):17-48.
- [2] Ettinger DS, Wood DE, Aggarwal C, Aisner DL, Akerley W, Bauman JR, et al. NCCN Guidelines Insights: Non-Small Cell Lung Cancer, Version 1.2020. Journal of the National Comprehensive Cancer Network : JNCCN. 2019;17(12):1464-72.
- [3] Lambin P, Rios-Velazquez E, Leijenaar R, Carvalho

S, van Stiphout RG, Granton P, et al. Radiomics: extracting more information from medical images using advanced feature analysis. *European journal of cancer (Oxford, England : 1990)*. 2012;48(4):441–6.

[4] LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature*. 2015;521(7553):436–44.

[5] Liu Y, Kim J, Qu F, Liu S, Wang H, Balagurunathan Y, et al. CT Features Associated with Epidermal Growth Factor Receptor Mutation Status in Patients with Lung Adenocarcinoma. *Radiology*. 2016;280(1):271–80.

[6] Zhang X, Zhang G, Qiu X, Yin J, Tan W, Yin X, et al. Exploring non-invasive precision treatment in non-small cell lung cancer patients through deep learning radiomics across

imaging features and molecular phenotypes. *Biomarker research*. 2024;12(1):12.

[7] Zhu Y, Guo YB, Xu D, Zhang J, Liu ZG, Wu X, et al. A computed tomography (CT)-derived radiomics approach for predicting primary co-mutations involving TP53 and epidermal growth factor receptor (EGFR) in patients with advanced lung adenocarcinomas (LUAD). *Annals of translational medicine*. 2021;9(7):545.

[8] Wang XY, Wu SH, Ren J, Zeng Y, Guo LL. Predicting Gene Comutation of EGFR and TP53 by Radiomics and Deep Learning in Patients With Lung Adenocarcinomas. *Journal of thoracic imaging*. 2024.