

# 肺癌预测模型及其进展

张熙明, 朱 辉\*

新疆医科大学第一附属医院胸外科, 新疆 乌鲁木齐

收稿日期: 2024年2月7日; 录用日期: 2024年2月29日; 发布日期: 2024年3月8日

## 摘 要

肺癌是我国及世界范围内发病率和死亡率最高的恶性肿瘤之一。随着低剂量计算机断层扫描的普及, 我国肺癌的检出率逐年升高, 全球癌症生存趋势监测报告数据显示, 2000~2014年间肺癌的年龄标化5年生存率波动在10%~20%之间, 在中国, 肺癌的年龄标化5年生存率在2003~2015年间略有上升的趋势, 但仍低于20.0%; 2012~2015年中国人群肺癌的5年生存率仅为19.7%。影响肺癌5年生存率的关键因素在于临床诊断时的肿瘤分期。临床研究数据显示, 早期肺癌的5年生存率可达61.2%, 其中小于1 cm的I期肺癌的5年生存率达92%; 而晚期肺癌的5年生存率仅7.0%。因此, 早期诊断对于改善肺癌患者的预后至关重要。然而, 中国肺癌病例的诊断以晚期居多, 早期诊断率较低。预计至2025年, 我国每年新发的肺癌患者将达到100万, 其中约有75%的肺癌患者就诊时已属晚期, 形成了发病率高、病死率高、5年存活率低的“二高一低”特点。肺癌发病时间较短且致病原因复杂, 可在短时间内病灶转移, 早期阶段以肺结节的形式存在, 常因为其隐匿性而错过了最佳的治疗时间, 而肺癌相关的预测模型的研究也在近些年较为热门, 因为其有助于帮助患者早期诊断肺部结节的良恶倾向, 本文主要讨论肺癌相关预测模型的应用和其和人工智能相结合的发展。

## 关键词

肺肿瘤, 预测模型, 人工智能

# Lung Cancer Prediction Model and Its Progression

Ximing Zhang, Hui Zhu\*

Thoracic Surgery Department of First Affiliated Hospital of Xinjiang Medical University, Urumqi Xinjiang

Received: Feb. 7<sup>th</sup>, 2024; accepted: Feb. 29<sup>th</sup>, 2024; published: Mar. 8<sup>th</sup>, 2024

\*通讯作者。

## Abstract

Lung cancer is one of the most common malignant tumors with the highest morbidity and mortality in China and worldwide. With the popularity of low-dose computed tomography, the detection rate of lung cancer in China has increased year by year. According to the Global Cancer Survival Trend Monitoring Report, the age-standardized 5-year survival rate of lung cancer fluctuated between 10% and 20% from 2000 to 2014. In China, the age-standardized 5-year survival rate of lung cancer showed a slight upward trend from 2003 to 2015, but it was still lower than 20.0%; the 5-year survival rate of lung cancer in the Chinese population was only 19.7% from 2012 to 2015. The key factor affecting the 5-year survival rate of lung cancer is the tumor stage at the time of clinical diagnosis. Clinical research data show that the 5-year survival rate of early lung cancer can reach 61.2%, among which the 5-year survival rate of stage I lung cancer with a size less than 1 cm is 92%, while the 5-year survival rate of stage III lung cancer is only 7.0%. Therefore, early diagnosis is crucial for improving the prognosis of lung cancer patients. However, the diagnosis of lung cancer cases in China is mainly in late stage, and the early diagnosis rate is low. It is estimated that by 2025, the number of new lung cancer patients in China will reach 1 million annually, and about 75% of them are in late stage at the time of diagnosis, forming the characteristics of high morbidity, high mortality, and low 5-year survival rate. Lung cancer has a short onset time and complex pathogenic factors, and the lesions can metastasize in a short time. In the early stage, they exist in the form of pulmonary nodules, which often miss the best treatment time because of their insidiousness. In recent years, the research on lung cancer-related prediction models has become popular, because they help patients diagnose the benign and malignant tendency of pulmonary nodules in an early stage. This paper mainly discusses the application and development of lung cancer-related prediction models and their combination with artificial intelligence.

## Keywords

Lung Neoplasms, Prediction Model, Artificial Intelligence

Copyright © 2024 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## 1. 引言

肺癌是世界上最常见,死亡率最高的癌症,中国2005~2014年间的肺癌诊断数据显示,43.6%的肺癌患者在诊断时已处于IIIB~IV期,II~IIIA期占37.4%,I期肺癌仅占19.0% [1]。2020年中国癌症新发病例数第一的癌症也是肺癌,新增了82万的病例。之后是结直肠癌56万,胃癌48万,乳腺癌42万,肝癌41万,食管癌32万,甲状腺癌22万,胰腺癌12万,前列腺癌12万,宫颈癌11万[2] [3]。肺癌的广泛发病率及晚期肺癌预后差对患者及社会造成了沉重的相关负担。而且它仍然呈现上升的趋势。2008年全国30个城市的住院花费合计,肺癌造成了226.95亿的合计住院费用,并且造成了858.85亿的总经济负担(就诊率高,住院天数长) [4]。而根据英国的一项研究提示:早期的肺癌筛查能够在总的可用的医疗成本相对低时,降低肺癌死亡率并导致更多的肺癌诊断。而对于结节良恶性的判断往往需要经验丰富的专科医生,而因为医疗资源的分布不均匀,部分地区难以获得优质的医疗资源[5]。而作为一个人口老龄化所伴随的影响,等到2024年时,全球的癌症负担将增加百分之五十。但癌症的生存率仍然受到诊断

阶段的严重影响, 大多数肺癌(57%)是在癌症转移到肺外时诊断出来的[2]。社会经济地位和教育程度较低会导致肺癌风险和更严重的后果, 尤其是在男性中。越贫穷的因为吸烟及生活地区等的影响[6], 越容易得病, 而预后就越差。

## 2. 肺癌的诊断及特点

使用低剂量计算机断层扫描(LDCT), 通过观察肺癌的影像学特征, 筛查肺癌是一种有效的筛查方式, 可以降低肺癌的死亡率[5] [6]世界卫生组织(World Health Organization, WHO)将肺腺癌按其发展演变过程分为非典型腺瘤样增生(Atypical Adenomatous Hyperplasia, AAH)、原位腺癌(Adenocarcinoma In Situ, AIS)、微浸润腺癌(Minimally Invasive Adenocarcinoma, MIA)、浸润性腺癌(Invasive Adenocarcinoma, IAC)。随着高分辨率 CT 的广泛应用, 越来越多磨玻璃结节(Ground-Glass Nodule, GGN)型肺腺癌被发现。GGN 定义为在高分辨 CT 上局灶性肺组织呈云雾状密度增高影, 直径  $\leq 3$  cm, 但不掩盖其内正常的结构, 如支气管影、血管影; 通常根据其内是否含有实性成分可再分为纯磨玻璃结节(pure Ground-Glass Nodule, pGGN)和混合磨玻璃结节(mixed Ground-Glass Nodule, mGGN)。LDCT, 即低密度薄层 CT 扫描, 能以更薄的扫描层厚来提供更高的临床诊断价值, 但考虑到高假阳性、过度诊断、经济负担和反复放疗的潜在危害, LDCT 筛查应针对高危人群, 而不是全部应用此手段。2021 年, 美国预防服务工作组(United States Preventive Services Task Force, USPSTF)对有 20 包年吸烟史、目前吸烟或戒烟时间位于 15 年范围内的 50~80 岁成人进行 LDCT 肺癌筛查[6]。尽管简单实用, 但这种对吸烟者进行肺癌筛查的风险分层是过于理想化的, 因为极低风险的个体不太可能从高风险筛查中获益, 并排除了更有可能从肺癌筛查中获益的个体, 比如较为年轻的潜在的肺癌的患者以及女性的肺癌患者, 手术可能对他们来说能有更少的打击和更快的恢复速度, 但他们往往易被忽略, 因为依据以前的研究他们的发病几率更低, 所以更容易从筛查中忽略[7]; 因为来自许多亚洲国家的数据已经确定不吸烟者确实构成了不容忽视的非小细胞肺癌(NSCLC)患者的一部分。如部分针对亚洲人群(主要源于台湾, 香港, 甘肃)的研究发现, 在从不吸烟的人群中, 女性的肺癌发病率更高。在这一组患者中, NSCLC 的发病机制中存在非烟草相关危险因素, 可能的危险因素还包括暴露于烹饪油烟, 激素, 月经周期和病毒感染[8]。Lynch 等的报道, NSCLC 患者亚组(主要是腺癌)在表皮生长因子受体基因中携带特异性激活突变, 该突变与酪氨酸激酶抑制剂吉非替尼的临床反应性相关, 表皮生长因子受体酪氨酸激酶结构域的突变在女性、从不吸烟的人群、亚洲人和腺癌中更常见。所以综上所述, 这些研究表明, 吸烟者和从不吸烟者在流行病学特征和组织学亚型方面存在差异, 这些不吸烟的人罹患肺癌的风险可能与东亚血统、女性和有肺癌腺癌病史有关[9]。另外一项针对英国 2170 名肺癌患者的研究发现, 不吸烟者的肺癌患者的比例占英国 2014 年肺癌发病率为 28%, 高于 2008 年的 13%。在美国的队列研究中发现了相同的结果。根据《美国医学会肿瘤学杂志》(JAMA Oncology)的一项研究, 美国高达 12% 的肺癌患者从未吸烟, 不吸烟者在肺癌患者中所占比例越来越大。这些研究表明, 在一些国家, 非吸烟者罹患肺癌的风险正在增加[9]。另外的规律则是, 发现在不吸烟的癌症患者中, 腺癌患者的比例较高, 与外国的文献研究一致。大量有关国内的肺癌患者研究表明, 在不吸烟的癌症患者中腺癌的比例约为 60% [10]。

## 3. 经典肺癌预测模型及其后续发展的种类

因此, 准确识别待筛查的高危人群对于最大限度地提高肺癌筛查的疗效至关重要。而临床上为了达到早期发现早期诊断的目的, 根据肺癌的相关影像学及其他特征, 提出了肺癌预测模型的理论。即通过收集相关的肺部占位的影像特征, 包括毛刺, 胸膜牵拉, 直径等的特点, 结合相关其他的检查结果及患者的个人史, 并根据患者的最终病理结果, 来整理出一个计算患者肺部结节的恶性概率的(公式)模型。而

肺恶性肿瘤的预测模型也在近些年来得到了蓬勃的发展, 最早期的模型当属 Mayo 肺癌风险评估模型, 即 1997 年由美国梅奥诊所建立的 Mayo 模型, 该模型是基于 6 个临床和影像特征(年龄、吸烟史、既往恶性肿瘤史、结节直径、毛刺和结节是否位于上叶)而建立的。在此基础上, 国外学者又建立了 Brock 模型、Herder 模型、VA 模型等的多着重于影像特征等的模型[11][12][13], 而随着相关研究的进行, 不同的研究结果提示: 不同地域和人种之间的差异, 会导致相关数据推导出的模型在危险因素的侧重上有所不同的倾斜。这提示更高效率的预测模型还需要结合当地的患者数据, 才能得到更高的有效性[14]。为了解决现有指南的这一局限性, 近年来开发了几种肺癌风险预测模型, 主要基于已确定的危险因素, 如吸烟、职业暴露、肺癌家族史和呼吸系统疾病等。此外, 随着相关研究的进展, 也有学者发现除了基本的影像因素外, 一些血液指标, 或者是基因检测的结果, 也对于肺癌的诊断及预测具有重要的意义, 尤其是多种复合因素的影响。通过该方法得出的预测模型, 往往具备更高的诊断准确性[15]。肠道菌群的相关研究一直是近些年的热点, 而有研究指出, 肠道菌群对于肺癌也具有一定的预测作用, 表明肺癌可能与病原体的增加和某些益生菌微生物的减少有关。比如双歧杆菌属和粪杆菌属健康人群中更丰富, 而肺癌患者则表现出升高的芽孢杆菌。双歧杆菌所属的菌落具有多种益生菌功能, 并达到抗癌作用, 减少 TNF- $\alpha$  和脂多糖(LPS)引起的炎症。TNF- $\alpha$  可通过诱导上皮-间充质转化促进肺癌转移。粪杆菌的主要成员普氏栖粪杆菌(*Faecalibacterium prausnitzii*)通过抑制 NF- $\kappa$ B 活化和 IL-8 分泌与抗炎功能相关。肺部痰液中的芽孢杆菌可能与肺癌风险增加有关, 婴儿链球菌在肺癌患者中的分布显著减少, 它是一种定植于呼吸道的微生物, 通过解毒多环芳烃(PAH)调节肺部先天免疫, 与慢性肺病有因果关系。而通过对于上述菌群的操作分类单元(Operational Taxonomic Unit, OTU)的监测, 从而根据菌群组成的微生物谱系, 构建一个被称为: 患者鉴别指数(DPI)的指数, 来辅助预测早期肺癌的可能结果[16]。而根据巴基斯坦的学者的研究, 肺癌是因为 DNA 的异常而导致了体细胞被影响产生突变, 所以突变的 DNA 逆转录产生的 RNA 会产生特点的蛋白质, 他们根据测量特定蛋白质的氨基酸序列来预测肺癌[17], 宁波的学者的预测模型结合了高血压, 纤维蛋白原和 BUN, 先前的研究表明, 高血压在癌症患者中很常见, 包括非小细胞肺癌的患者。对这种现象的一种解释是, 高血压可能会增加血浆血管内皮生长因子水平。而纤维蛋白原  $\beta$  链和纤维蛋白原  $\gamma$  链联合用于预测肺结节良恶性具有良好的敏感性和特异性发现纤维蛋白原与肺癌显著相关。BUN 确定为肺癌的危险因素, 研究表明, 尿素氮与血清白蛋白的比率可用于预测严重肺癌患者的预后, 故将其纳入肺癌评估模型[18]。但是与之伴随而来的就是随着相关危险因素的增多, 数据分析的难度也就因此大大的提升了。在过去的几十年里, 癌症研究技术的也不断发展促成了许多大型合作癌症项目, 这些项目产生了大量的临床数据。这些数据库有助于研究人员从诊断、治疗到临床结果的反应等方面研究肺癌的综合模式。特别是目前关于组学分析的研究, 如基因组学、转录组学、蛋白质组学和代谢组学, 扩展了我们的研究工具和能力。癌症研究正在向多种数据类型和巨型数据的整合转变。使用多样化和高维数据类型完成临床分析也需要大量的时间和专业知, 并且分析呈指数级增长的癌症相关数据库对研究人员构成了重大挑战[19]。而随着研究的逐渐深入, 模型技术及人工智能(AI)的发展, 使得于诊断相伴生的各类数据也由原来庞杂的“累赘”变成了可以进行统计分析的各类影响因子。除了对于肺结节性质本身的预测模型, 目前也出现了关于肺恶性肿瘤相关预后的模型, 其中也包括结合了人工智能技术的预测模型。

#### 4. 肺癌预测模型结合人工智能的发展

而人工智能和机器学习的出现, 使得庞杂的数据分析所造成的困难迎刃而解, 而根据数据模型的概念来看, 越多的基础数据, 最后推导出的模型的准确性和有效性相对来说就会越高, 较早的肺癌预测模型仅仅具备数以百人计的案例数量, 而大数据使得以千人, 万人为单位的模型建立变得不再困难。人工智能的应用是一种更进一步的解决方案, 因为它有着许多的应用及优点[20]。人工智能的一些首批应用是



该领域致力于在 X 射线和计算机断层扫描(CT)检查中自动检测肺结节, 即计算机辅助诊断(CAD)系统, 以帮助医生解释医学影像数据, 这已被证明是对于医生有明确参考价值的意见, 其性能现在达到或超过放射科医生的水平[19]。而随着技术水平与模型种类的增强, 人工智能可以帮助识别人类无法使用或者使用起来很困难的特征并执行重复性任务。人工智能(AI)的一个优势是, 它能够对肿瘤影像学特征进行定量评估, 而不是定性评估, 这一过程也称为“影像组学”。研究表明, 与常规临床测量评估相比, 结合了 AI 的模型结果具有更强的预测能力, 能够通过非手术的方式去描述肿瘤表型[21]。传统的诊断方式通常是临床医生依据 CT 影像特征的临床参数用于确定治疗类型和预测良恶性结果, 这些参数包括 CT 图像中的结节大小、结节类型、结节位置、结节计数、结节边界和肺气肿信息, 临床变量包括患者的年龄、性别、肺癌家族史、吸烟暴露等[6]。然而, 这些特征大多是随着临床医生的个人水平和状态主观界定的, 通常无法实现对恶性结节外观的完整和定量描述。并且仅局限于单个时间点, 未考虑肿瘤的表型变化或演变过程, 影像组学通常旨在从诊断图像中提取定量且理想情况下可重现的信息, 包括人眼难以识别或量化的复杂模式, 影像组学可用于捕捉组织和病变特性, 例如形状和异质性, 并且在连续成像中, 它们随时间的变化。研究表明, 影像组学特征与细胞水平的异质性指数密切相关[22]。活检仅捕获肿瘤一小部分的异质性, 并且通常仅在单个解剖部位捕获异质性, 而影像组学则捕获整个肿瘤体积的异质性, 以非侵入性方式跟踪病变的这种演变, 并可随时间纵向跟踪同一区域, 从而提供在单个时间点通过静态图像获得的肿瘤特征之外的其他肿瘤特征。CT 复查随访作为肺癌诊断治疗中的重要一部分, 通过与人工只能手段结合, 可通过新旧对比提供有关患者的更多信息。深度学习(DL)是人工智能的一个子领域, 可以算作更加自主化及深入化的人工智能模型的分支。与传统的机器学习方法相反, 传统的机器学习方法需要从图像中手工提取特征和基于人类的数据结构, 而基于深度学习的算法则以原始数据为依据, 并开发模式识别所需的自身表示。另外由于图片存档和通信系统的发展, 大型图像数据集的应用变得更可行。大型图像数据集对于训练深度学习模型至关重要[15]。通过将 AI 方法与影像组学、病理学、基因组学、转录组学、蛋白质组学和临床数据相结合, 使得能在免疫治疗和靶向治疗中显示出预测价值, 显著提高了肺癌患者的精准治疗。使用深度学习方法进行肿瘤评估, 无需手动划定轮廓或靶区即可提取表型变化, 减少那些容易因为不同的医生而导致的观察差异, 从而提供临床参考。而人工智能也能通过自我训练和自我学习, 使得预测的效果变得更好。这一切都能使一些容易被医生主观因素影响的相关指标被注意到, 包括结节的实性成分, 密度等的难以被肉眼直观分辨的影像, 以及一些类似于病变的容积, 相关血液指标的长期变化, 也能人工智能及相关的分析软件被有效的注意并分析。另外人工智能的发展可以通过提供准确有效的决策支持, 在临床上适用于加强患者护理。跟踪肿瘤进展以预测化疗和放疗后的生存率和反应对于治疗评估和适应性治疗计划以改善患者预后至关重要[20] [23]。传统的机器学习技术涉及推导工程特征以定量描述图像, 并成功检测生物标志物以进行反应评估和临床结果预测。深度学习的最新进展已证明在无需人类特征定义的图像分析中的成功应用。卷积神经网络(CNN)的使用允许自动提取成像特征并识别复杂数据中的非线性关系。在数百万张摄影图像上训练的 CNN 网络可以通过迁移学习应用于医学图像。这在肿瘤检测和术前分期方面的癌症研究已得到证实[19]。

## 5. 肺癌预测模型的应用及展望

而目前出现的诸多预测模型中, 不乏适应地区特点及具有较高的诊断价值, 预测模型可依托网络等技术较为便捷的使得更多人群受益。而肺癌的诊断往往需要明确的病理结果支持, 而根据美国胸科医生协会的推荐, 预测模型可用作是否对患者施行有创性检查的一个参考。所以肺癌的预测模型因此在近年间逐渐变成了一个热点话题。另外除了上述这些通过临床数据进行预测的相关模型外, 瑞典的研究提示, 他们可以通过临床的调查问卷, 来预测具备可疑肺癌风险的患者, 并且通过一些基础的问题, 比如

患者是否具备声嘶, 是否近期感染, 以及咯血和吸烟等。这些基础的问题可以在一些基层医疗机构或者是全科医学科起到意义[10]。上述这些预测模型涵盖了整个诊疗过程的方方面面, 这也意味着预测模型能对一些医疗数据进行更好的整合和利用, 并结合人工智能起到指导诊疗的作用。虽然它无法给予一个完整准确的诊断, 但是可以在早预防和早诊断上起到积极的作用。而也正如前文所述, 越早的进行医疗干预, 肺癌所造成的伤害和经济花费越小。

早期干预, 如家庭支持、血液样本生化分析和 CT 扫描, 可以使低风险患者受益, 而定期随访可以确保监测肺结节, 及时发现和治疗病情变化。准确的肺结节评估可以帮助外科医生确定肺癌风险, 并确保对高危患者进行及时治疗, 同时减少对低危患者的侵入性治疗。准确预测特定患者患肺癌的风险是困难的, 它需要经验丰富的临床医生的意见, 而这对于很多地区的患者来说是难以轻易得到的, 涉及基于流行病学和临床病理学数据的人工智能模型或许能起到更好的作用[24]。

## 6. 小结

故预测模型能有效地提示患者的肺部占位的良性及恶性的倾向, 从而指导患者是否需要进一步的医疗干预, 这对于一些优质医疗资源匮乏的地区的患者, 具有明显的意义。他们可以通过网络将自己的相关数据上传至预测模型, 并根据此结果来指导自己的诊疗。这将避免了漫长的求诊过程和路途花费, 使得医疗资源能够更有效地得到利用。并且通过与人工智能技术的结合, 预测模型将从更加细微之处出发, 从患者的病史及各项检查检验指标入手, 对临床工作进行指导并帮助临床医生去减少因为个人倾向所产生的诊断上的偏倚及误差。

## 参考文献

- [1] 罗汶鑫, 杨澜, 王成弟, 等. 肺癌筛查与早期诊断的研究现状与挑战[J]. 中国科学: 生命科学, 2022, 52(11): 1603-1611.
- [2] Bade, B.C. and Dela Cruz, C.S. (2020) Lung Cancer 2020: Epidemiology, Etiology, and Prevention. *Clinics in Chest Medicine*, **41**, 1-24. <https://doi.org/10.1016/j.ccm.2019.10.001>
- [3] 国家癌症中心发布最新一期全国癌症统计数据[J]. 上海护理, 2022, 22(4): 72.
- [4] 兰蓝, 李晓松, 周婷, 等. 中国疾病经济负担测量——以 2008 年数据为例[J]. 中国卫生事业管理, 2017, 34(2): 123-127.
- [5] 樊淼, 郭乾乾, 李之淳, 等. 2009-2020 年中国西部地区政府医疗卫生支出效率分析[J]. 现代预防医学, 2023, 50(10): 1831-1836. <https://doi.org/10.20043/j.cnki.mpm.202211224>
- [6] Gould, M.K., Fletcher, J., Iannettoni, M.D., Lynch, W.R., Midthun, D.E., Naidich, D.P., Ost, D.E. and American College of Chest Physicians (2007) Evaluation of Patients with Pulmonary Nodules: When Is It Lung Cancer? ACCP Evidence-Based Clinical Practice Guidelines (2nd Edition). *Chest*, **132**, 108S-130S. <https://doi.org/10.1378/chest.07-1353>
- [7] Liao, Z., Zheng, R. and Shao, G. (2022) A Lung Cancer Risk Prediction Model for Nonsmokers: A Retrospective Analysis of Lung Nodule Cohorts in China. *Journal of Clinical Laboratory Analysis*, **36**, E24748. <https://doi.org/10.1002/jcla.24748>
- [8] Brenner, A.V., Wang, Z., Kleinerman, R.A., Lei, S., Metayer, C., Wang, W. and Lubin, J.H. (2003) Menstrual and Reproductive Factors and Risk of Lung Cancer among Chinese Women, Eastern Gansu Province, 1994-1998. *Journal of Epidemiology*, **13**, 22-28. <https://doi.org/10.2188/jea.13.22>
- [9] Toh, C.K., Gao, F., Lim, W.T., Leong, S.S., Fong, K.W., Yap, S.P., Hsu, A.A., Eng, P., Koong, H.N., Thirugnanam, A. and Tan, E.H. (2006) Never-Smokers with Lung Cancer: Epidemiologic Evidence of a Distinct Disease Entity. *Journal of Clinical Oncology*, **24**, 2245-2251. <https://doi.org/10.1200/JCO.2005.04.8033>
- [10] Nemlander, E., Rosenblad, A., Abedi, E., Ekman, S., Hasselström, J., Eriksson, L.E. and Carlsson, A.C. (2022) Lung Cancer Prediction Using Machine Learning on Data from a Symptom E-Questionnaire for Never Smokers, Formers Smokers and Current Smokers. *PLOS ONE*, **17**, E0276703. <https://doi.org/10.1371/journal.pone.0276703>
- [11] Swensen, S.J., Silverstein, M.D., Ilstrup, D.M., Schleck, C.D. and Edell, E.S. (1997) The Probability of Malignancy in Solitary Pulmonary Nodules. Application to Small Radiologically Indeterminate Nodules. *Archives of Internal Medicine*, **157**, 849-855. <https://doi.org/10.1001/archinte.1997.00440290031002>

- [12] Gould, M.K., Ananth, L., Barnett, P.G. and Veterans Affairs SNAP Cooperative Study Group (2007) A Clinical Model to Estimate the Pretest Probability of Lung Cancer in Patients with Solitary Pulmonary Nodules. *Chest*, **131**, 383-388. <https://doi.org/10.1378/chest.06-1261>
- [13] Swensen, S.J., Silverstein, M.D., Edell, E.S., Trastek, V.F., Aughenbaugh, G.L., Ilstrup, D.M. and Schleck, C.D. (1999) Solitary Pulmonary Nodules: Clinical Prediction Model versus Physicians. *Mayo Clinic Proceedings*, **74**, 319-329. <https://doi.org/10.4065/74.4.319>
- [14] Wu, Z., Huang, T., Zhang, S., Cheng, D., Li, W. and Chen, B. (2021) A Prediction Model to Evaluate the Pretest Risk of Malignancy in Solitary Pulmonary Nodules: Evidence from a Large Chinese Southwestern Population. *Journal of Cancer Research and Clinical Oncology*, **147**, 275-285. <https://doi.org/10.1007/s00432-020-03408-2>
- [15] Yang, D., Zhang, X., Powell, C.A., Ni, J., Wang, B., Zhang, J., Zhang, Y., Wang, L., Xu, Z., Zhang, L., Wu, G., Song, Y., Tian, W., Hu, J.A., Zhang, Y., Hu, J., Hong, Q., Song, Y., Zhou, J. and Bai, C. (2018) Probability of Cancer in High-Risk Patients Predicted by the Protein-Based Lung Cancer Biomarker Panel in China: LCBP Study. *Cancer*, **124**, 262-270. <https://doi.org/10.1002/cncr.31020>
- [16] Zheng, Y., Fang, Z., Xue, Y., Zhang, J., Zhu, J., Gao, R., Yao, S., Ye, Y., Wang, S., Lin, C., Chen, S., Huang, H., Hu, L., Jiang, G.N., Qin, H., Zhang, P., Chen, J. and Ji, H. (2020) Specific Gut Microbiome Signature Predicts the Early-Stage Lung Cancer. *Gut Microbes*, **11**, 1030-1042. <https://doi.org/10.1080/19490976.2020.1737487>
- [17] Sattar, M., Majid, A., Kausar, N., Bilal, M. and Kashif, M. (2022) Lung Cancer Prediction Using Multi-Gene Genetic Programming by Selecting Automatic Features from Amino Acid Sequences. *Computational Biology and Chemistry*, **98**, Article ID: 107638. <https://doi.org/10.1080/19490976.2020.1737487>
- [18] Liao, Z., Zheng, R. and Shao, G. (2022) A Lung Cancer Risk Prediction Model for Nonsmokers: A Retrospective Analysis of Lung Nodule Cohorts in China. *Journal of Clinical Laboratory Analysis*, **36**, E24748. <https://doi.org/10.1002/jcla.24748>
- [19] Li, Y., Wu, X., Yang, P., Jiang, G. and Luo, Y. (2022) Machine Learning for Lung Cancer Diagnosis, Treatment, and Prognosis. *Genomics Proteomics Bioinformatics*, **20**, 850-866. <https://doi.org/10.1016/j.gpb.2022.11.003>
- [20] Yin, X., Liao, H., Yun, H., Lin, N., Li, S., Xiang, Y. and Ma, X. (2022) Artificial Intelligence-Based Prediction of Clinical Outcome in Immunotherapy and Targeted Therapy of Lung Cancer. *Seminars in Cancer Biology*, **86**, 146-159. <https://doi.org/10.1016/j.semcancer.2022.08.002>
- [21] Mayerhoefer, M.E., Materka, A., Langa, G., Häggström, I., Szczypiński, P., Gibbs, P. and Cook, G. (2020) Introduction to Radiomics. *Journal of Nuclear Medicine*, **61**, 488-495. <https://doi.org/10.2967/jnumed.118.222893>
- [22] Choi, E.R., Lee, H.Y., Jeong, J.Y., Choi, Y.L., Kim, J., Bae, J., Lee, K.S. and Shim, Y.M. (2016) Quantitative Image Variables Reflect the Intratumoral Pathologic Heterogeneity of Lung Adenocarcinoma. *Oncotarget*, **7**, 67302-67313. <https://doi.org/10.18632/oncotarget.11693>
- [23] Zhang, R., Wei, Y., Wang, D., Chen, B., Sun, H., Lei, Y., Zhou, Q., Luo, Z., Jiang, L., Qiu, R., Shi, F. and Li, W. (2023) Deep Learning for Malignancy Risk Estimation of Incidental Sub-Centimeter Pulmonary Nodules on CT Images. *European Radiology*. <https://doi.org/10.1007/s00330-023-10518-1>
- [24] Xu, Y., Hosny, A., Zeleznik, R., Parmar, C., Coroller, T., Franco, I., Mak, R.H. and Aerts, H.J.W.L. (2019) Deep Learning Predicts Lung Cancer Treatment Response from Serial Medical Imaging. *Clinical Cancer Research*, **25**, 3266-3275. <https://doi.org/10.1158/1078-0432.CCR-18-2495>