

# 基于复杂数据构建肺癌中医证候数据模型评述

李 奕<sup>1</sup> 刘福栋<sup>1</sup> 花宝金<sup>1</sup> 庞 博<sup>2</sup>

(1. 中国中医科学院广安门医院,北京 100053; 2. 中国中医科学院广安门医院国际医疗部,北京 100053)

**【摘要】**肿瘤作为世界各国人口死亡的主要原因和影响预期寿命延长的主要障碍,其整体发病形势依然严峻。肺癌作为全球高发的恶性肿瘤病种,是癌症相关死亡的重要因素。在科技竞争日愈激烈的时代背景下,如何将中医药辨证论治体系、优势与信息时代的技术模式相结合,利用数字化和人工智能技术,全面精准地整合复杂数据中的有效信息,进而形成肺癌防治的数据网络、中医证候数据模型与临床信息处理机制,将是制定肺癌临床诊疗策略极为关键的内容。在结合团队前期研究的基础上,笔者认为有必要在客观面对复杂数据科学标注和分类方法的过程中,探索、总结肺癌中医证候数据模型构建的方法与规律,形成兼顾高效性、实用性、可靠性的肺癌中医证候诊疗辅助决策方案,这将为中医药防治肺癌的守正创新工作提供重要的技术支持与基础保障。

**【关键词】**肺癌;复杂数据;中医证候;数据模型;机器学习;防治结合

**DOI:** 10.16025/j.1674-1307.2022.05.002

原发性支气管肺癌简称肺癌,美国癌症协会(American Cancer Society, ACS)最新发布的数据统计结果显示<sup>[1]</sup>,2022年美国将有约191.8万例癌症新发病例,同时预计有60.9万人将死于癌症,其中肺癌的疾病负担较重,原发性肺癌的5年生存率仅有6%,仍是导致肿瘤死亡的首要原因。与我国最新癌症预测数据相比,2022年中国预计将有479.6万例新发癌症病例和约319.3万例癌症死亡病例,肺癌患者病死率同样居于所有癌症病种之首,肿瘤防控仍面临巨大挑战<sup>[2-3]</sup>。目前肺癌主要治疗方法以手术、化疗、放疗、免疫、靶向治疗等为主,但肺癌起病隐匿,早期难以察觉,患者往往丧失最佳手术治疗时机,与此同时,临床中众多患者也难以忍受放疗与靶向药物带来的不良反应。因此笔者基于复杂数据模型研究,系统挖掘、梳理中医临床诊疗肺癌过程的思维规律,剖析中医古籍、人用经验、标准规范等文本数据中的内在逻辑关系,将分散、无序的中医证候有效地整合为客观、规范、有序的“病-证-症-法-方-药”非线性复杂逻辑,形成准确高效的复杂数据文本学习与数据模型的关键范式,以期提高临床肺癌防治水平、改善患者症状、提高生活质量以

及减少公共卫生负担。

## 1 肺癌中医证候研究现状

肺癌属于中医学之“肺积”“肺岩”“息积”“咳嗽”“咯血”“喘息”“虚劳”等疾病范畴,笔者认为肺癌的发生是患者在体质因素影响的基础上,复加外感、内伤多种病因,致脏腑气血阴阳功能失衡、邪毒积结而产生的一类病症,“正虚”是肺癌发生和进展的关键要素,寒凝、痰浊、瘀毒阻肺为其核心病机,益气扶正、祛邪解毒为肺癌治疗大法,笔者团队多年来致力于肺癌的临床与基础研究,认为预防为主、防治结合是控制肺癌的根本出路<sup>[4-5]</sup>。近年来,有关肺癌的中医证候临床研究、回顾性研究、文献学研究广泛开展,形成了集分期证候、分型证候、证候要素、文献系统评价以及名医经验证候在内的多层次、多维度、多体系中医肺癌证候评价方法,极大地拓展了肺癌研究的视野,同时也指导着临床对肺癌的辨治。

### 1.1 肺癌证候分期、分型研究

王少墨等<sup>[6]</sup>对388例原发性肺癌患者证型变化进行研究,数据显示在不同分期中肺阴虚证始终排于首位,此外脾气虚证I期的发病率明显低

**基金项目:**首都卫生发展科研专项自主创新项目(首发2022-2-4155);中国中医科学院科技创新工程项目(CI2021A01805, CI2021B009)

**作者简介:**李奕,男,24岁,硕士研究生。研究方向:中医药肿瘤防治与名老中医经验传承方法学研究。

**通信作者:**庞博,E-mail:drpangbo@gmail.com

**引用格式:**李奕,刘福栋,花宝金,等.基于复杂数据构建肺癌中医证候数据模型评述[J].北京中医药,2022,41(5):476-479.

于Ⅳ期,肺阴虚证中Ⅳ期的发病率显著低于Ⅰ期,由血瘀证、痰湿证向肺阴虚证的转变是晚期肺癌的基本规律。周建龙<sup>[7]</sup>等采用调查表形式,对383例非小细胞肺癌的西医分型分期与中医证候分布规律进行分析,发现Ⅰ、Ⅱ期肺癌中医证型以气滞血瘀型为主,Ⅲ、Ⅳ期肺癌证型以气阴两虚型为主,证实了中医证型与肺癌TNM分期相关。在肺癌证候分型研究方面,国家中医药管理局相关部门发布的《肿瘤科中医临床路径(试行)》和中华中医药学会标准(ZYYXH/T136-156-2008)《肿瘤中医诊疗指南》等行业标准中均提出中医证候诊断基础上的肺癌分型辨证诊疗方法,确证了肺癌相关分型的研究内容。

### 1.2 肺癌证候要素研究

贺佐梅等<sup>[8]</sup>通过对非小细胞肺癌患者共计7435份辨证病案开展回顾性调查研究,发现非小细胞肺癌患者中医辨证结果以气虚、血瘀、热毒、阴虚、痰湿等为证候的基本要素,并以此5种证候要素构成的5类证候群为其临床症状的基本特征。另一项回顾性研究<sup>[9]</sup>证实肺癌患者在证候要素分布方面以复合证型占主导(占总研究的91.7%),其中尤以气虚与痰湿复合者为多,在基础证候要素方面,以气虚证为主导,如病理分期Ⅰ期的肺癌患者证型多为痰湿证、气虚证,Ⅱ、Ⅲ、Ⅳ期患者以阴虚证、气虚证、痰湿证证候要素多见,此外,各期的核心症状表现也与相对应的证候要素一致。笔者团队<sup>[10]</sup>采用前瞻性研究方法,针对初次治疗的晚期非小细胞肺癌患者病情发展不同节点开展动态证候观察,发现晚期非小细胞肺癌中医证候具有复杂变化的典型特征,随病程进展,复合证候显著增加,证候要素尤以气虚、阴虚、痰湿及血瘀等为著。

### 1.3 肺癌文献学评价研究

笔者团队在分析近年来发表的肺癌常见证候文献学的基础上,认为目前针对肺癌的文献学评价研究较少,研究方向多围绕肺癌常见证型及病位要素等。如郭雯等<sup>[11]</sup>通过借助双步隐树分析LTM-EAST算法、聚类分析等对数据库中肺癌中医证候相关文献进行分析,总结出肺癌痰瘀互结证、气滞血瘀证、痰热壅肺证、痰湿阻肺证等10个常见证型。韩燕等<sup>[12]</sup>借助文献学研究方法,对晚期非小细胞肺癌中医证候分布情况及组合原则进行分析,推算出肺癌气阴两虚、气虚等中医证候类

型22种,肺脾、脾、肺等病位要素14种,气虚、阴虚、痰湿等病性要素15种,证实了晚期肺癌证候的多样性与症状的复杂性。

### 1.4 肺癌证候名医经验研究

在肺癌证候名医经验研究方面,笔者团队全国名中医朴炳奎教授<sup>[13]</sup>结合肺癌及相关伴随症临床诊疗实际,认为肺癌发病以虚损为本,早期多见气阴两虚证,晚期则阴虚内热证和阴阳俱虚证多见,强调肺癌相关证候要从扶正与祛邪相结合、辨证与辨病相结合、局部与整体相结合的角度综合认识,临床诊疗在着眼于气机阻滞、痰毒凝结、血瘀毒聚等病理变化的同时,围绕肺癌癌毒内阻、湿盛痰阻、肺脾气虚、气阴两虚等证候特点分期论治。周仲瑛教授以癌毒理论为辨证核心,认为癌毒阻肺是肺癌全周期的基本证型,在疾病演变、转归的过程中,常可与其他病机、证型并列化生,发展形成复合病机、复合证型,提出了涵盖癌毒阻肺证、痰浊阻肺证、气滞血瘀证、气阴两虚证、脾胃虚弱证的肺癌中医证候认识<sup>[14]</sup>。周岱翰教授则着眼于肺癌复杂病机特点,根据肺癌痰、热、瘀、虚致病特征,从阴虚痰热、肺郁痰瘀等病机入手,提出肺癌的临床诊治要以四型为抓手,补充了肺癌多病机、多证型的具体认识<sup>[15]</sup>。

## 2 肺癌中医证候数据模型存在的主要问题

几十年来肺癌证候的深入研究已经说明,理性思考、面对信息技术发展为中医药带来的深刻变革是亟需关注与重视的内容,将思维创新融入实践创新、技术创新是打破疾病固有认知,重新揭示和预见中医证候数据模型发展变化规律,破解疾病严峻挑战的关键所在。笔者团队在前期传承朴炳奎教授学术思想、临床经验及思维认知规律的基础上,基于机器学习方法与原型范畴理论,开展了一系列有关中医诊治肺癌认知模型构建的基础性工作<sup>[16-17]</sup>。在丰富肺癌治疗理论、探索中医诊治肺癌“经验”向“知识”转化途径、表达先验知识和思维机制上取得了长足的进步。但放眼中医证候数据模型领域,中医证候的辨证标准及数据模型的具体构建目前仍无确切、稳定的系统结论,各标准之间仍然存在着证候分类、方案实施、名称规范方面的明显区别,行业标准也较难统一。其主要原因在于证候量化标准界定范围模糊、大样本多中心随机对照等高级别循证证据的

缺乏、基于文献学研究和证候量表的方法学尚不完善、针对复杂数据信息的处理方法相对空缺等,而中医证候规范化研究中的相关问题亦囿于方法学的局限性。

### 2.1 缺乏大规模复杂数据的科学标注、分类方法

中医药学是一门实践性较强的学科,其思维方式与方法以中国古代哲学思想为基础架构,以医师临床经验为补充,具体来源于医师对诊疗过程的归纳演绎、逻辑推理与悟性判断,是经长时间观察,进行分析与归纳、类比和论证,经过抽象思维的演绎和推导,进而形成的以辨证论治为诊疗特色的复杂理论体系。因其信息来源充足、样本数量庞大、刻画对象特征的指标复杂、带有时空动态特性、包含大量噪声等特点使其具有大规模复杂数据的属性特点,而这种属性特点为中医学文本数据的科学标注、分类方法构成巨大挑战。其中尤以中医古籍文献为代表的文本数据体量极其巨大,跨越历史时期较长,概念、涵义迁移频繁复杂,基于传统思考领悟、文献挖掘为主的研究方法在开展系统整理、认知方面难度极大。因此,如何采用合理有效的方法对大规模复杂数据集进行科学标注、分类,使其转化为能够理解的机器语言与数据模型,进而构建相应的复杂数据文本处理方法,对还原历史语境、传承中医精华、挖掘肺癌特色治疗方法具有重要意义。

### 2.2 中医肺癌证候诊疗及复杂信息临床处理机制聚焦不足

在复杂信息数据中,障碍空间和高维数据下的数据在实际应用中较为常见,是值得关注并引起重视的内容。笔者认为中医肺癌证候是肺癌疾病复杂信息的核心载体,以往聚焦于中医证候平台的研究仍以决策树、支持向量机、贝叶斯网络、多标记学习等传统机器学习为多,目前亟需提高对复杂信息临床处理机制。曹云等<sup>[18]</sup>在深度神经网络对消化系统疾病证候研究方面,实现了在传统机器学习的基础上,对疾病诊断、预测能力的提升,确证了辨证模型的构建具有方法学上的可行性,但仍然受制于模型样本量较小的因素,识别准确率与传统机器学习算法相比仍无显著突破,这为肺癌中医证候数据模型的参考构建带来不小的挑战。因此未来团队将深入挖掘病-症-证-法-方药的复杂耦合关系,争取构建中医证候一体化的

辅助决策平台,解决目前在辨证模型应用方面的局限性。

### 2.3 中医肺癌证候数据库与机器学习算法技术结合面临挑战

计算机和信息技术的发展,带来了海量的复杂数据,在特定的疾病中构建与之相对应的数据库系统就显得尤为重要,数据库不仅仅是一种存储处理对象、介质和处置系统的集合体,同时也在应用系统层面提供数据支持,是数据基础构建中必不可少的核心组成部分<sup>[19]</sup>。中医肺癌证候数据库以证候数据模型为载体,建立高效的中医肺癌证候数据库可为肺癌相关新药研发及肺癌防治临床决策的制定提供依据。在笔者看来,数据信息研究发展至今天,中医证候机器学习模型(如线性感知器、集成学习、支持向量机和随机森林等)能够在以往数据中挖掘具体的“方案”,不断提升数据模型解决以往难以解决的复杂问题的能力。但受制学习模型的简单性,其往往只能从单一的层面描述学习过程,对肺癌证候的精细化和关键信息的处理能力仍有待提升,同时训练数据局限、训练时间过长、泛化能力受阻、适应性低,这极大地限制了其应用范围,因此中医肺癌证候数据库如何与机器学习算法技术紧密结合,提高中医药肺癌研究信息传承的科技化、数字化水平,是在改进机器学习算法、提升计算能力的过程中必须客观面对和深切解答的重要命题。

## 3 应用前景与展望

中医理论持续创新是时代发展的必然要求,中医肺癌理论研究及诊治技术的提升要充分借鉴现代文明的发展成果,在依托大数据、机器学习、数据模型等前沿技术方法的基础上,融入文献学方法、逻辑学方法、深度学习方法、德尔菲法、测试集仿真应用等定性定量相结合的研究方法,探索构建肺癌中医证候知识经验整理策略-知识应用路径-证据体系。笔者团队将进一步借助复杂数据的信息优势,实现从新的数据中推断出潜在主题以及中医肺癌证候的自然语句和词向量表示方法,理性观察、辨析、判断、思考现代数据模型与中医肺癌研究融合的现实意义,使肺癌中医证候的相关研究加权现代科学及信息技术的引导,构建出客观、科学、公认的中医证候诊疗相关复杂数据的科学标注和分类方法,助力肺癌临床诊疗的疗效再提升,中医理论与当代科学的再结合、

再发展。

### 参考文献

- [1] SIEGEL RL, MILLER KD, FUCHS HE, et al. Cancer statistics, 2022[J]. CA Cancer J Clin, 2022, 72(1): 7-33.
- [2] XIA C, DONG X, LI H, et al. Cancer statistics in China and United States, 2022: profiles, trends, and determinants.[J]. Chin Med J (Engl), 2022, 135(5): 584-590.
- [3] 尹周一, 王梦圆, 游伟程, 等. 2022 美国癌症统计报告解读及中美癌症流行情况对比[J]. 肿瘤综合治疗电子杂志, 2022, 8(2): 54-63.
- [4] 庞博, 李奕, 花宝金, 等. 论肺结节中医病机与防治策略[J]. 北京中医药, 2021, 40(12): 1340-1343.
- [5] 刘福栋, 庞博, 花宝金. 基于中医经典与人用经验构建胰腺癌防治方药发掘体系思路探析[J]. 北京中医药, 2021, 40(9): 929-933.
- [6] 王少墨, 董志毅, 屠洪斌, 等. 388 例原发性肺癌中医证候分布状况分析[J]. 上海中医药大学学报, 2013, 27(5): 30-34.
- [7] 周建龙, 梁静, 邓青南. 非小细胞肺癌的中西医分型分期相关性研究[J]. 吉林中医药, 2015, 35(4): 358-360.
- [8] 贺佐梅, 曾普华, 郜文辉, 等. 7435 份非小细胞肺癌病案中医证候分布频数及关联分析[J]. 中医药导报, 2019, 25(16): 40-44.
- [9] 桂越蓉, 王学谦, 李玉潇, 等. 肺癌中医证候要素及症状分布情况研究[J]. 中医肿瘤学杂志, 2021, 3(4): 12-16.
- [10] 李丛煌, 刘瑞, 郑红刚, 等. 晚期非小细胞肺癌中医证候分布及动态演变临床研究[J]. 中医学报, 2015, 30(8): 1085-1088.
- [11] 郭雯, 冯贞贞, 王露, 等. 基于文献的肺癌常见中医证候研究[J]. 中医杂志, 2021, 62(16): 1447-1454.
- [12] 韩燕, 杨国旺, 王笑民. 晚期非小细胞肺癌中医证候分布及组合规律文献研究[J]. 北京中医药, 2014, 33(3): 175-178.
- [13] 郑红刚, 花宝金, 朴炳奎. 朴炳奎辨治肺癌学术思想与经验探析[J]. 中医杂志, 2010, 51(4): 304-306.
- [14] 蔡云, 陈远彬, 叶放, 等. 国医大师周仲瑛从癌毒理论辨治肺癌经验述要[J]. 中华中医药杂志, 2020, 35(6): 2879-2882.
- [15] 林丽珠. 周岱翰教授从痰辨治癌症经验介绍. 新中医, 2006, 46(3): 10-12.
- [16] 庞博, 刘刚, 周雪忠, 等. 基于机器学习的名老中医诊治肺癌认知模型构建方法[J]. 北京中医药, 2015, 34(12): 949-954.
- [17] 庞博, 徐心瑶, 周培培, 等. 基于原型范畴理论名老中医诊治肺癌认知模型优化思路与方法[J]. 北京中医药, 2018, 37(12): 1141-1145.
- [18] 曹云, 卢毅, 陈建新, 等. 基于机器学习的胃食管反流病中医智能辨证模型的应用[J]. 北京中医药大学学报, 2019, 42(10): 869-874.
- [19] 曲维阳, 何丽丽. 一种基于机器学习的数据库集成方法[J]. 中国科技信息, 2022(9): 90-92.

### Review of TCM syndrome data model of lung cancer based on complex data

LI Yi, LIU Fu-dong, HUA Bao-jin, PANG Bo

(收稿日期: 2022-05-10)