

# 机器学习在呼吸医学中的研究动态与趋势

赵羿帆<sup>1</sup> 陶敏佳<sup>2</sup> 张静薇<sup>2</sup> 赵慧敏<sup>3</sup> 张昕仪<sup>3</sup> 王丽<sup>4\*</sup> 张春英<sup>3\*</sup>

1 华北理工大学口腔医学院 2 华北理工大学临床医学院

3 华北理工大学理学院 4 哈尔滨医科大学附属第一医院心内七科

DOI:10.12238/bmtr.v6i5.10077

**[摘要]** 目的: 机器学习(ML)在医学研究,尤其是呼吸医学研究中的重要性,已成为学者们关注的焦点。为了总结当前的研究热点和发展趋势,本文对2013年至2023年的研究论文进行可视化分析,从而为进一步的科学研究以及产业建设提供参考。方法: 本文通过检索web of science核心数据集中有关呼吸医学中应用机器学习的论文,发表时间限制为2013年至2023年。使用citespace软件对纳入的论文样本进行了作者、机构、国家/地区、关键词方面的分析,运用文献计量学方法阐述本领域中的知识框架。结果: 经过去重筛选,我们的研究共纳入了来自113个国家/地区的18808位作者的2593篇文章。分析显示,从2018年开始,由机器学习驱动的呼吸医学研究数量显著增加。美国、中国、印度、英国、意大利为发文量排名前五的国家,高频关键词有: 深度学习、分类、诊断、肺癌、CT。关键词突现分析显示目前研究热点为: 支持向量机、X射线成像、颗粒物、医学图像处理、预测模型。结论: 通过预测模型的构建与突破实现呼吸系统疾病的诊断、程度分级、预后预测、药物筛选是目前的研究热点。

**[关键词]** 呼吸医学; 机器学习; 人工智能; 文献计量学; CiteSpace

中图分类号: R332.3 文献标识码: A

## Research Trends and Trends of Machine Learning in Respiratory Medicine: A Quantitative Analysis Based on Citespace

Yifan Zhao<sup>1</sup> Minjia Tao<sup>2</sup> Jingwei Zhang<sup>2</sup> Huimin Zhao<sup>3</sup> Xinyi Zhang<sup>3</sup> Li Wang<sup>4\*</sup> Chunying Zhang<sup>3\*</sup>

1 School of Stomatology, North China University of Science and Technology

2 School of Clinical Medicine, North China University of Science and Technology

3 School of Science, North China University of Science and Technology

4 Department 7 of Cardiology First Affiliated Hospital of Harbin Medical University

**[Abstract]** Objective: The importance of machine learning (ML) in medical research, especially in respiratory medicine research, has become a focus of attention for scholars. In order to summarize the current research hotspots and development trends, this article conducts a visual analysis of research papers from 2013 to 2023, providing reference for further scientific research and industrial construction. Method: This article retrieved papers on the application of machine learning in respiratory medicine from the Web of Science core dataset, with publication time limited to 2013 to 2023. We used CiteSpace software to analyze the included paper samples in terms of authors, institutions, countries/regions, and keywords, and used bibliometric methods to explain the knowledge framework in this field. Result: After de duplication screening, our study included 2593 articles from 18808 authors from 113 countries/regions. Analysis shows that since 2018, the number of respiratory medicine research driven by machine learning has significantly increased. The United States, China, India, the United Kingdom, and Italy are the top five countries in terms of publication volume, with high-frequency keywords including deep learning, classification, diagnosis, lung cancer, and CT. Keyword emergence analysis shows that the current research hotspots are support vector machine, X-ray imaging, particulate matter, medical image processing, and prediction models. Conclusion: The construction and breakthrough of predictive models for the diagnosis, severity grading, prognosis prediction, and drug screening of respiratory diseases are currently research hotspots.

**[Key words]** respiratory medicine; Machine learning; artificial intelligence; Bibliometrics; CiteSpace

## 引言

呼吸医学,也被称为肺部医学或呼吸系统疾病学,是一门研究呼吸系统(包括鼻腔、咽、喉、气管、支气管以及肺)的结构、功能、疾病及其治疗方法的学科。呼吸系统疾病是一类病变主要发生在气管、支气管、肺部以及胸腔的常见疾病,包括肺炎、肺栓塞、气胸、支气管哮喘、慢性阻塞性肺疾病(COPD)、肺水肿、胸腔积液、间质性肺疾病、尘肺病、肺结节病等。呼吸系统疾病常用检查方法包括痰液分析、血液分析、胸部影像学检查(包括X线、CT、MR)等。早期发现对于施加干预、改善治疗方案和延长生存期至关重要。

机器学习,作为人工智能领域的一个核心分支,主要关注如何让计算机系统通过经验改善性能。机器学习的基本思想是,通过训练算法分析大量数据,使得算法能够学习数据模式和特征,进而对新数据进行预测和分类。这一特点正是该技术能广泛应用于临床呼吸医学的重要原因。人工智能通过深度学习技术、自然语言识别技术等方法,可通过分析医学影像、临床病历等信息,应用于呼吸系统疾病的临床诊断<sup>[1]</sup>。

近年来,呼吸医学与机器学习领域内的学者展开了广泛的合作,运用机器学习技术对呼吸系统疾病的预测、筛查、诊断、鉴别、评估等方面展开了深入的研究。本文使用citespace软件对Web of science核心数据集中近十年来呼吸医学中机器学习技术的发展与应用进行可视化分析,运用合作网络从作者、机构、国家角度进行分析,运用聚类对研究学者、关键词等要素进行归纳梳理。通过文献计量学方法对目前的研究现状进行可视化分析,深入了解目前呼吸医学中机器学习的研究热点与发展趋势,以期推动机器学习技术在呼吸医学中的进一步突破,早日实现产业化。

## 1 数据来源与研究方法

本文以webofscience核心合集数据库为文献数据来源,整理机器学习在呼吸系统疾病领域的研究,检索式为((TS=(respiratory)orTS=(lung)orTS=(pulmonary))ANDTS=(disease)AND(TS=(machinelearning)ORTS=(artificialintelligence)))AND(LA=("ENGLISH")ANDPY=(2013-2023)),共检索出文献4011篇。在数据库中把"ARTICLE"作为限制条件,排除掉了745篇其他文档类型的研究。然后笔者团队对剩下的3266篇文献进行人工筛选,对标题和摘要进行详细阅读与评估,排除了46篇以动物为实验样本的研究以及570篇与本文研究内容不相关的文献。以纯文本文件方式导出这2650篇研究的全记录和引用文献,将此导入citespace软件进行去重,去掉了57篇重复文献,最终纳入了2593篇文献作为本文分析的数据来源。

本文使用陈超美教授带领团队研发的citespace6.2.R4软件对论文作者、国家/地区、机构、关键词等内容进行文献计量分析<sup>[2]</sup>。

## 2 结果与分析

2.1年度发文量统计。某一研究领域的文献数量变化可以反映该领域的发展情况,对纳入的论文样本进行文献数量统计,如图1。2013至2023年的发文量可分为三个阶段。第一阶段(2013

年~2017年),年发文量均在30篇以内,处于平稳、缓慢增长的阶段。总文献数量占纳入的总文献数量的3.9%,处于较低水平,反映了在本阶段,呼吸医学中机器学习的研究尚处于起步阶段。第二阶段(2018年~2021年),本阶段论文的发表数量较前一阶段增长趋势迅猛,呈现明显的倍增态势。2018年、2019年、2020年、2021年的年发文量分别达到了61篇、118篇、325篇、580篇,总文献数量也因此得到了快速的增加,占纳入总数的45.6%。第三阶段(2022年~2023年),年发文量增速有所放缓,分别是677篇与731篇,呈现欣欣向荣的状态。前期建立起的研究基础为现阶段研究者提供了良好的参考,总文献数量得到了迅速扩充。这样的特点与趋势显示,在呼吸医学中运用机器学习的方法来解决目前问题已得到了呼吸系统疾病相关从业者以及计算机科学学者们的广泛关注,并已开展了可观规模的合作。

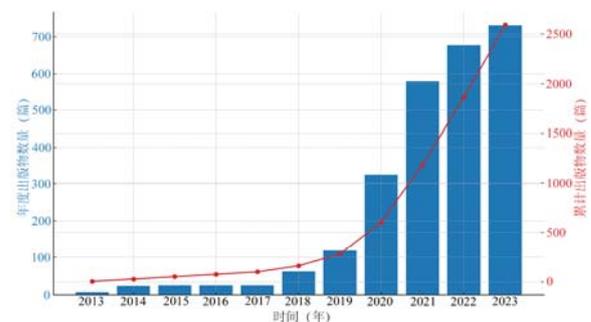


图1 年度及累计发表量

2.2发文作者分析。发文量在10篇及以上的学者有Huang Tao、Silverman Edwin K、Chowdhury Muhammad E H、Estepar Raul San Jose、Khandakar Amith、Heidari Ali Asghar、Chen Huiling这两位学者的平均被引次数达到了67.86与54.33,分列第一与第二,这显示了这两位学者非凡的影响力。Silverman Edwin K是发文量前十作者中,总被引次数最高的作者,达到了28566次,同时其篇平均被引次数(39.02)也处于较高水平,在前十作者中位居第三。

2.3发文国家及机构分析。对纳入论文所涉及的113个国家/地区进行分析,发文量居于前五的国家或地区依次为美国(738篇)、中华人民共和国(567篇)、印度(305篇)、英国(200篇)、意大利(138篇),中心性居于前五的国家或地区为英国(0.28)、印度(0.19)、意大利(0.14)、澳大利亚(0.13)、瑞士(0.12)。美国在发文量居于本领域榜首且中心性较高(0.1),反映了美国学者在呼吸疾病中的机器学习这一领域已进行了大量的工作,形成了良好的研究基础。英国学者的中心性大幅领先其他国家,表示英国学者的论文成果被国际广泛认可并引用,其他国家/地区的学者与之开展了积极的合作。

发文量前五的机构依次为哈佛大学(112篇)、加州大学系统(83篇)、哈佛医学院(76篇)、布莱根妇女医院(60篇)、中国科学院(55篇)。中心性前五的机构依次是哈佛大学(0.12)、巴黎公共援助医院(0.09)、加州大学旧金山分校(0.08)、波士顿大学(0.08)、布莱根妇女医院(0.07)。尤其突出的是,哈佛大学在

论文的数量和中心性上均取得了领先地位,在一众机构中表现最为亮眼。

2.4 关键词共现分析。关键词是论文研究者研究时所围绕的、能高度概括文章研究核心的词,可以通过关键词方便、快捷地了解论文的研究内容<sup>[3]</sup>。我们通过对纳入的2593篇文献共计547种关键词进行统计,制作了表1。

表1 高频关键词及中心度(前20)

关键词	频次(次)	中心性
deep learning	320	0.02
classification	284	0.12
diagnosis	235	0.11
disease	191	0.05
lung cancer	177	0.06
computed tomography	139	0.03
COPD	134	0.04
prediction	112	0.03
risk	106	0.03
mortality	103	0.03
model	99	0.02
expression	81	0.05
identification	80	0.09
system	80	0.03
ct	80	0.04
pneumonia	78	0.01
cancer	77	0.03
feature extraction	72	0.01
convolutional neural network	72	0
survival	71	0.02

Top 25 Keywords with the Strongest Citation Bursts

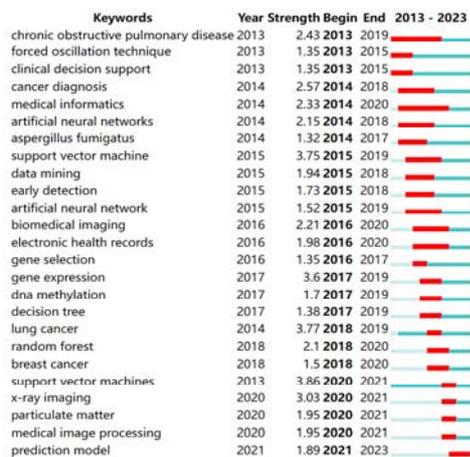


图2 关键词突现图

在这十年间,如图2所示,突现的关键词经历了明显的更替,25个关键词涵盖疾病(慢性阻塞性肺病、癌症、烟曲霉感染、肺癌)、分析指标(强迫震荡法进行肺功能测定、医学信息学、电子健康记录、基因的选择与表达、DNA甲基化检测、X射线成像、空气中的颗粒物)、机器学习的具体技术(人工神经网络、支持向量机、数据挖掘、决策树、随机森林、)、临床应用(临床决策支

持、癌症诊断、医学图像处理、预测模型)等方面。其中,“预测模型”这一关键词的热度仍在持续,值得学者们广泛关注。

### 3 讨论

呼吸医学领域机器学习的论文数量自2018年以来迎来了爆发式增长,2018年-2021年的发文量均倍增且目前仍呈现良好的增长趋势。这一现象揭示了国际上的研究者对于机器学习在呼吸医学中的应用研究保持着蓬勃的热情,笔者预计这一趋势仍将延续下去。

开发生物标志物可以提高疾病的早期诊断能力、精确评估疾病严重程度以及监测治疗效果。将呼吸系统疾病与其他疾病鉴别出来有助于提供准确的诊断、制定有效的治疗计划,并避免不必要的治疗和其带来的副作用。对呼吸系统疾病进行严重程度分级能够指导医疗资源的合理分配和治疗方案的选择。通过评估疾病的严重程度,医生能够更准确地预测治疗结果,制定个性化治疗计划,并为患者和家属提供更明确的预后信息。将机器学习技术应用于药物靶点的筛选和药物开发过程中,可以极大提高识别有效药物靶点的速度和准确性。机器学习模型能够处理和分析大量生物信息学数据,识别出潜在的药物候选分子,加快从实验室研究到临床应用的转化速度。

综上所述,机器学习技术在呼吸医学领域已经展现出了巨大的潜力和实际应用价值。通过利用大量的医疗数据,如患者的临床信息、影像数据、生理信号等,机器学习模型能够辅助医生进行更准确的诊断、疾病分类以及治疗效果预测。随着技术的进步和数据可访问性的提高,预计机器学习将在呼吸医学领域扮演越来越重要的角色,为患者提供更加精准和高效的医疗服务。

本文使用citespace软件对Web of science数据库中近十年关于机器学习在呼吸医学领域中的应用研究进行了统计与可视化分析,为领域内学者们的后续研究提供参考。

#### 【基金项目】

河北省创新创业训练计划(R2023149)。

#### 【参考文献】

[1]Exarchos K P,Beltsiou M,Votti C-A,etal.Artificial intelligence techniques in asthma: a systematic review and critical appraisal of the existing literature[J].European Respiratory Journal,2020,56(3).

[2]Chen C,Song M.Visualizing a field of research: A methodology of systematic scientometric reviews[J]. PLOS ONE, 2019, 14(10):e0223994.

[3]Mukherjee D,Lim W M,Kumar S,etal.Guidelines for advancing theory and practice through bibliometric research[J]. Journal of Business Research,2022,148:101-115.

#### 作者简介:

赵羿帆(2002--),女,汉族,河北邯郸人,本科在读,研究方向:医学领域内的文献计量学分析。