

文献计量学视角下的卵巢癌预测模型研究：基于公共数据库的分析

应璐琦 赵鲁文*

承德医学院附属医院

DOI:10.32629/ffcr.v4i2.19985

[摘要] 卵巢癌早期诊断难、疗效也不理想,是导致女性死亡的主要恶性肿瘤之一。基于公共数据库建立临床预测模型,能把多组学数据和临床信息整合起来,给卵巢癌精准诊疗开了一条新路。本研究用PubMed数据库2006到2026年收录的431篇文献,应用文献计量学方法梳理了这个领域的现状和发展脉络。结果显示:这个领域产出增长快,中美是核心贡献国,已建起多中心、跨学科的国际合作网;热点在预后评估、标志物筛选和基因调控这一块,机器学习应用正成为新增长点;整体研究正从基础机制探索往临床转化和智能决策支持转化。以后得强化多组学整合、优化算法、深化合作,推动成果落地。本研究给学者提供了全景图,对资源配置和协同创新也有参考价值。

[关键词] 卵巢癌; 临床预测模型; 公共数据库; 文献计量学; 机器学习

中图分类号: R711.75 **文献标识码:** A

Ovarian Cancer Prediction Model Research from the Perspective of Bibliometrics: An Analysis Based on Public Databases

Luqi Ying Luwen Zhao*

Affiliated Hospital of Chengde Medical University

[Abstract] Ovarian cancer is hard to diagnose early and responds poorly to treatment, making it a leading cause of cancer death in women. Clinical prediction models using public databases integrate multi-omics and clinical data, offering a new path for precision care. This study analyzed 431 PubMed publications (2006–2026) via bibliometrics to map the field. Results show rapid output growth, with China and the USA as core contributors forming an international network. Hotspots include prognosis, biomarkers, and gene regulation, with machine learning as a new growth area. Research is shifting from basic mechanisms to clinical translation and intelligent support. Future work should strengthen multi-omics integration, optimize algorithms, and deepen cooperation for clinical application. This study offers scholars a broad overview and insights for resource allocation and collaboration.

[Key words] ovarian cancer; clinical prediction model; public database; bibliometrics; machine learning

前言

卵巢癌作为女性生殖系统三大常见恶性肿瘤之一,因其隐匿性强、早期症状缺乏特异性,发病率与病死率均位于前列。临床预测模型整合多维信息,为早诊、治疗及预后提供科学依据,是个体化医疗的关键支撑。随着癌症基因组图谱(TCGA)、基因表达综合数据库(GEO)和监测、流行病学与最终结果数据库(SEER)等公共数据库的完善,使大规模多组学数据获取成为可能;大数据技术与人工智能算法不断升级,也改善了临床预测模型性能。

该领域研究快速增长,内容涵盖基因表达、临床特征和影像组学等多个维度。例如,基于BRCA1/2突变的风险评估模型^[1]、中性粒细胞与淋巴细胞比值(NLR)等炎症标志物的预后价值^[2]、腹膜癌指数(PCI)对减瘤术效果的预测^[3],以及RIF1蛋白表达与铂类化疗敏感性有关^[4]、循环肿瘤细胞对生存期的影响^[5]等,均为模型构建提供了新视角。机器学习算法为药物反应预测提供了新的方法学支撑^[6]。然而,现有研究分散,整体格局的系统梳理缺乏,国际合作、研究力量分布及热点演变等量化分析不足,限制了对领域全局的把握。

文献计量学是一种通过分析大量文献来评估研究发展趋势和揭示重点研究方向的统计方法,可量化分析科研产出与合作网络,客观揭示领域趋势^[7]。可视化分析有助于揭示研究主题的演进脉络与前沿热点。因此,围绕基于公共数据库的卵巢癌预测模型展开全面分析,可描绘研究现状,促进跨学科合作与临床转化。

1 资料与方法

本研究选择PubMed数据库作为数据源,检索式综合了“卵巢癌”“预测模型”“公共数据库”等核心概念,时间从2006年1月1日至2026年1月16日。排除了会议摘要、评论、社论等非研究型文献,仅保留原创研究和系统综述。

采用R语言bibliometrix包(4.3.0版)进行文献计量分析。内容包括:用词频和共现分析识别核心主题;用聚类算法对文献、作者和关键词归类;构建引文和合著网络定位关键文献和核心作者;用时序分析追踪主题演变、捕捉新兴热点。2025年6月获取了2024年度期刊影响因子和JCR分区数据。

2 结果

2.1 文献发表趋势

本研究共纳入431篇文献,原创研究393篇(占91.2%),综述38篇(占8.8%)。二十年间,年发文量从寥寥几篇涨到2025年74篇峰值,增长了50多倍。多项式拟合曲线显示这一增长趋势具有显著统计学意义($R^2=0.77$) (见图1)。

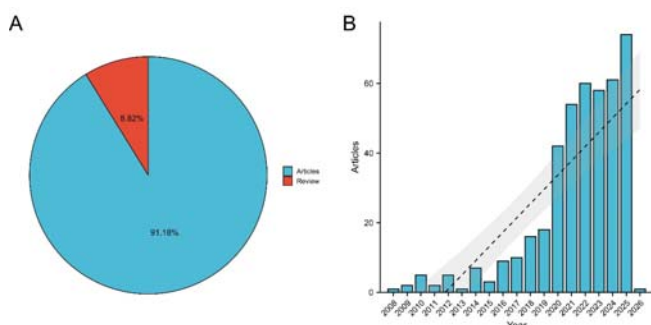


图1 2006—2026年卵巢癌临床预测模型研究年发文量趋势

注:柱状图显示各年发文量,曲线为多项式拟合趋势线。

2.2 国家/地区与研究机构贡献

31个国家/地区参与了研究。中国发文最多156篇占36.2%;美国98篇占22.7%,两国是研究核心。日本多国合作论文比例最高,达46.9%,体现出活跃的国际合作态势;法国、德国、荷兰的多国合作比例也超40%,表明欧洲国家在国际合作中同样积极。年增长率中国以18.5%领先,伊朗15.2%、韩国13.2%紧随其后,显示亚洲国家近年发展迅猛。

国家合作网络有6个主要集群。最大的集群以中、美、日、法、希为核心,其中美国是枢纽,连接着北美、亚洲和欧洲多个国家。第二集群以德国为中心,辐射奥地利、比利时、加拿大等国;第三集群由意大利、伊朗、韩国、荷兰、西班牙构成;此外还有新加坡-澳大利亚-孟加拉国集群、印度-埃及-沙特阿拉伯集群,以及由匈牙利等组成的东欧集群。这种多元化合作格局

表明该领域已形成跨地域、多中心的全球协作网络。

研究机构层面,发文前十中有七所来自中国。香港以23篇领跑,复旦大学(19篇)、中国医科大学盛京医院(17篇)紧随其后。机构合作形成多个研究集群,例如以华中科技大学、中南大学为核心的集群聚焦于肿瘤免疫治疗预测模型,以南京医科大学、苏州大学为主的集群侧重于临床特征模型的构建与验证。这种协作机制为多中心研究及成果转化注入了强劲动力。

2.3 期刊分布

431篇文献分布于27种不同期刊。FRONTIERS IN GENETICS以13篇发文量居首;FRONTIERS IN ONCOLOGY发文11篇,CANCERS发文9篇,分列二、三位。尽管BIOMOLECULES仅发表4篇但其JCR影响因子高达4.8,体现对高质量研究的吸引力。发文前十的期刊涵盖肿瘤学、遗传学、分子生物学、计算生物学等多个学科,肿瘤学期刊处于知识传播核心。

2.4 中国学者贡献与团队网络

高产作者前十位中九位来自中国。LI, Li (广西医科大学肿瘤医院)和ZHANG, Yu (陕西省人民医院)各发文7篇,并列榜首;LI, Na (山东省放射肿瘤学重点实验室)发文6篇紧随其后。作者合作网络揭示出三个主要集群:第一集群以LI, Li, WANG, Hui等为代表,聚焦于基于基因表达谱的预后模型构建;第二集群以ZHANG, Yu, MA, Ding等为核心,侧重于临床特征与影像组学相结合的预测工具开发;第三集群则由LI, Na, CHEN, Jing等组成,主要研究生物标志物与机器学习算法的融合应用。这三个集群既独立又协作,形成了一个以中国学者为核心的较为稳固的学术生态圈。

2.5 研究热点演进

高频关键词分析显示,“预后”(145次)、“中年”(70次)、“肿瘤基因表达调控”(64次)、“肿瘤生物标志物/遗传学”(50次)、“基因表达谱”(45次)等词频最高,反映当前研究主要聚焦于预后判断、生物标志物挖掘和分子机制解析。此外,“列线图”“机器学习”“转录组”“计算生物学”等关键词频繁出现,表明方法学创新正成为新的关注点。关键词聚类形成三个主要板块:人群特征与基因调控、生物标志物与计算方法、基因表达谱与临床预测工具。

引文突现分析结果显示(见图2),“肿瘤生物标志物/遗传学”(突现强度19.112)、“转录组”(9.248)、“机器学习”(5.195)等主题突现强度高,且爆发期集中在2024—2026年,是当前乃至未来的前沿热点。其中“机器学习”的突现标志着方法论变革:从发现单一关联转向整合多组学数据、构建智能化预测模型。这一趋势与整个肿瘤学研究向“多模态数据融合”和“临床问题驱动”转型的大方向完全一致。

3 讨论

发文量指数增长($R^2=0.77$)凸显该领域科研活力。增长主要靠三个因素:多组学整合方法成熟,为复杂模型提供了技术支撑^[7];公共数据库的开放共享使大规模研究成为可能^[7];精准医疗的普及推动研究向临床问题驱动转型^[8]。91.2%的原创

研究占比,揭示了该领域“数据驱动、实证为本”的学科特质,这与肿瘤研究从基因组发现迈向临床应用的整体趋势高度一致^[8]。

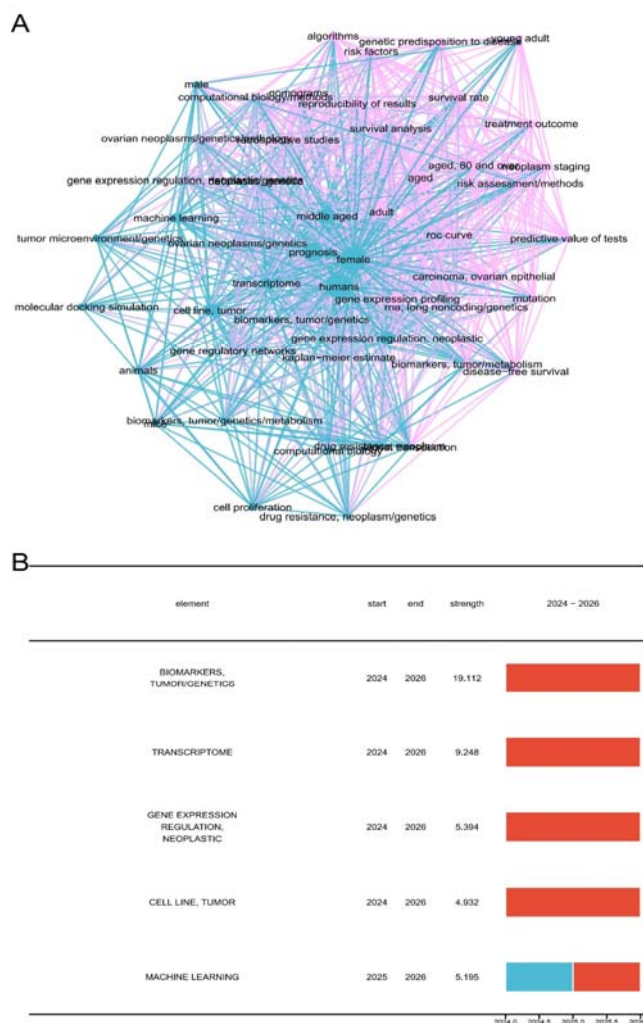


图2 关键词突现分析(2024—2026年)

注：显示突现强度前五的关键词及其爆发期。

从空间分布看,以中美为核心、辐射全球的合作网络已经成型。美国强在数据库建设和方法学创新,中国靠资源投入、患者队列和政策引导。文献计量学研究显示,中国在多个肿瘤领域的国际合作网络中均扮演着日益重要的枢纽角色^[9]。日本、法国、德国等国多国合作论文比例超40%,体现了用国际合作弥补本国资源不足的智慧。六大集群的形成,是全球科研资源优化配置的结果:北美与东亚聚焦临床转化,欧洲深耕基础机制,形成优势互补的协同创新格局^[9]。国内复旦大学、华中科技大学等机构已形成多个紧密合作的学术集群,围绕肿瘤免疫治疗预测模型等特定问题协同攻关,为多中心前瞻性研究提供了平台。

高产作者榜单中九位来自中国,这折射出中国在临床科研上的系统性优势:患者队列大,为模型验证提供了真实世界数据;“临床-基础-转化”平台促进了多学科协作;政策支持为创新提供了动力^[9]。三大研究集群的分工——基因表达谱模型、临

床影像组学、生物标志物与机器学习融合——恰好映射了精准医学从分子分型到影像特征再到多组学整合的核心逻辑。这种分工协作模式既避免了同质化竞争,又形成了完整的技术链条,反映出中国在该领域已形成较为成熟的研发生态^[10]。

关键词分析勾勒出研究重心的清晰演变路径。“机器学习”的突现标志着范式变革——从“发现关联”向“构建决策系统”跃升。传统模型多基于单一类型数据(如临床指标或单个基因),其预测能力受限于信息维度;而机器学习的引入使得整合基因组、转录组、影像组、临床表型的多模态预测成为可能。这一趋势与整个肿瘤学研究向“多模态数据融合”和“临床问题驱动”转型的大方向完全一致^[10],预示着卵巢癌预测模型正从实验室探索走向更贴近临床需求的“智能辅助工具”。

4 结论与局限性

本研究系统描绘了2006—2026年基于公共数据库的卵巢癌临床预测模型研究的全景图谱。该领域发展迅速,呈现多学科融合、国际合作紧密、热点从基础机制向智能化演进的特征。中美是领跑者,中国学者贡献突出。机器学习与多组学结合正把研究推向临床价值导向的新阶段。

本研究存在局限:数据源仅限PubMed,可能遗漏交叉学科文献;文献计量学重数量分析,引文频次难以完全代表研究质量。未来需在多组学整合、真实世界验证、数据标准统一三个方向突破。政策上应倡导数据共享,支持跨学科团队,推动研究成果惠及患者。

【参考文献】

[1] Jervis S, Song H, Lee A, et al. A risk prediction algorithm for ovarian cancer incorporating BRCA1, BRCA2, common alleles and other familial effects. *J Med Genet* [J]. 2015;52(7):465-75.

[2] Zhao Z, Zhao X, Lu J, et al. Prognostic roles of neutrophil to lymphocyte ratio and platelet to lymphocyte ratio in ovarian cancer: a meta-analysis of retrospective studies. *Arch Gynecol Obstet* [J]. 2018;297(4):849-857.

[3] Lluca A, Serra A, Rivadulla I, et al. Prediction of suboptimal cytoreductive surgery in patients with advanced ovarian cancer based on preoperative and intraoperative determination of the peritoneal carcinomatosis index. *World J Surg Oncol* [J]. 2018;16(1):37.

[4] Liu YB, Mei Y, Tian ZW, et al. Downregulation of RIF1 Enhances Sensitivity to Platinum-Based Chemotherapy in Epithelial Ovarian Cancer (EOC) by Regulating Nucleotide Excision Repair (NER) Pathway. *Cell Physiol Biochem* [J]. 2018;46(5):1971-1984.

[5] Zhou Y, Bian B, Yuan X, et al. Prognostic Value of Circulating Tumor Cells in Ovarian Cancer: A Meta-Analysis. *PLoS One* [J]. 2015;10(6):e0130873.

[6] Huang C, Mezencev R, McDonald JF, et al. Open source machine-learning algorithms for the prediction of optimal cancer drug therapies. *PLoS One* [J]. 2017;12(10):e0186906.

[7]Denis M, Tadesse MG. Evaluation of hierarchical models for integrative genomic analyses. *Bioinformatics*[J].2016;32(5):738–46.

[8]Zhang X,Li Y, Akinyemiju T, et al.Pathway-Structured Predictive Model for Cancer Survival Prediction: A Two-Stage Approach.*Genetics*[J].2017;205(1):89–100.

[9] Huang Q, Su W, Li S, et al.Global trends and academic landscapes of ultrasound applications in lung cancer research: a bibliometric analysis (2000–2024).*Cancer Imaging*[J].2025;

25(1):137.

[10]Zhang W,Gu J,Chen Y, et al.Emerging research themes in ferroptosis research for non-small cell lung cancer: a bibliometric and visualized analysis.*Front Immunol*[J].2025;16:1563-108.

作者简介：

应璐琦(1998--),女,汉族,浙江舟山人,研究生在读,研究方向为妇产科学。