

# 融合 SIFRank 和 DTM 视角下近十年我国医学人工智能研究 主题挖掘及其演化分析

吕艳华<sup>1</sup> 王萱<sup>1</sup> 崔云龙<sup>1</sup> 朱鑫鑫<sup>1</sup> 白慧娟<sup>1</sup>

(<sup>1</sup>山西医科大学管理学院, 山西, 030600)

**[摘要]**目的/意义 医学人工智能在促进医学信息化与智能化发展中具有关键作用, 本文梳理我国近十年医学人工智能领域的研究文献, 揭示其主要研究主题与发展趋势, 为后续研究与实践提供数据支撑与方向参考。方法/过程 以中国知网、万方和维普数据库中 2015 年 1 月 1 日至 2025 年 9 月 10 日期间发表的 7059 篇医学人工智能相关文献为研究对象, 提取题目、关键词及摘要作为分析语料, 采用融合 SIFRank 与 DTM 模型的方法进行主题挖掘, 结合主题热度与主题相似度指标, 分析三个时间阶段的研究热点及其演化特征。结果/结论 共识别出 10 个医学人工智能领域的核心研究主题, 并揭示了该领域的演化路径呈现出以肿瘤学为核心应用领域、传统医学数智化形成持续性发展脉络、研究趋于体系化发展三个演化特征。总体而言, 医学人工智能研究当前已进入并将在长期内处于深度融合阶段。

**[关键词]**医学人工智能; SIFRank-DTM 融合模型; 主题挖掘; 热点识别; 主题演化

## Theme Mining and Evolutionary Analysis of Medical Artificial Intelligence Research in China Over the Past Decade from the Perspectives of SIFRank and DTM

Lv Yanhua Wang Xuan Cui Yunlong Zhu Xinxin Bai Huijuan

School of Management, Shanxi Medical University, Shanxi 030600, China

**[Abstract]** **Purpose/Significance** Medical artificial intelligence drives advancements in medical informatization and intelligent development. This paper reviews research literature in the field of medical AI in China over the past decade, revealing its primary research themes and developmental trends to provide data support and directional guidance for subsequent research and practice. **Method/Process** This study analyzed 7,059 medical AI-related publications from January 1, 2015, to September 10, 2025, in the CNKI, Wanfang, and VIP databases. Titles, keywords, and abstracts were extracted as the analysis corpus. The theme mining process began with using SIFRank to extract key phrases from the corpus, and then the Dynamic Topic Model (DTM) was applied to identify and track themes over time. Theme popularity and similarity metrics were integrated to examine research hotspots and their evolutionary characteristics across three distinct time periods. **Result/Conclusion** Ten core research themes in medical artificial intelligence have been identified, revealing three key evolutionary characteristics: oncology as the primary application domain, the continuous digital transformation of traditional medicine, and the trend toward systematic research development. Overall, medical AI research has now entered and will remain in a phase of deep integration for the foreseeable future. Future development in this field must not only pursue the depth and breadth of technological integration into the domestic medical system, but also directly address the challenges and pain points encountered in clinical practice. By overcoming implementation barriers, it will forge a path of continuous innovation that is both pragmatic and feasible.

**[Keywords]** Medical Artificial Intelligence; SIFRank-DTM Hybrid Model; Topic Mining; Hotspot Identification; Topic Evolution

### 1 前言

随着科学技术的发展和社会的进步, 人工智能技术取得突破性进展, 被广泛应用于多

[作者简介] 吕艳华, 副教授, 博士生导师, 发表论文 50 余篇。

[基金项目] 国家社会科学基金一般项目 (项目编号: 20BTQ064)。

个领域。其中，人工智能与医学的结合（AI+医学）改变了传统的医疗模式，带来了革命性的推动<sup>[1]</sup>。近年来，由于人口老龄化、慢性病患率上升、新发突发传染病等因素给我国人民健康带来了巨大挑战，人工智能技术能否解决医学领域面临的难题，已经成为全社会关注的焦点。随着《新一代人工智能发展规划》、《“健康中国 2030”规划纲要》及《“十四五”卫生与健康科技创新专项规划》等多项政策文件陆续发布，新一代人工智能已发展上升为国家战略，明确指出要加快推动人工智能技术在卫生和健康领域的研究与应用<sup>[2-4]</sup>。因此，医学人工智能研究不仅对于医学发展至关重要，更对提升全民健康质量、优化医疗资源效能、推动健康中国建设具有重要意义。然而，医学人工智能的深入发展也带来了海量的文本数据，这些数据中蕴含着丰富的医学知识、临床经验与研究前沿，如何根据这些非结构化的信息揭示该领域的研究热点，精准挖掘有价值的知识脉络，对于推动医学人工智能的知识发现与创新应用至关重要。

当前，文献计量分析（如 CiteSpace、VOSviewer 等）是研究主题热点与演化分析的常用方法，该类方法依托关键词共现、聚类、突现分析、知识图谱等技术，从文献外部特征切入实现研究趋势的可视化分析，操作便捷、结果直观，已在各学科领域得到广泛应用。但文献计量分析存在显著局限性：其一，该类方法的主观性比较强，比如在词频分析中，不同的研究者选择的关键词或主题词很可能不同，划分高频关键词、低频关键词的词频阈值也比较主观，没有客观的衡量标准<sup>[5]</sup>；其二，该类方法依赖显性关键词的统计关联，难以挖掘文本内部隐含的研究主题，可能因关键词的同义不同形、异形同义词现象导致主题识别偏差；其三，文献计量对时间维度的分析多为静态时间切片式，无法实现研究主题随时间的连续动态演化刻画，难以精准捕捉主题间的传承、衍生与融合关系。

随着文本挖掘技术的发展成熟，文本挖掘技术在文献主题分析中得到了较为广泛的应用。主题模型是一种语言模型，是文本挖掘领域中结合机器学习和自然语言处理等相关方法的一种统计模型<sup>[6]</sup>，主题模型能够自动地从大规模文本集中发现和提炼隐含的主题信息<sup>[7]</sup>，逐渐成为主题挖掘领域的主流研究方法。其中，Blei 等<sup>[8]</sup>人提出的动态主题模型

(Dynamic Topic Model, DTM) 引入了时间属性对主题进行建模，能够刻画出主题跟随时间的动态演化过程。此外，有研究表明主题模型中的特征词提取对于主题识别效果具有重要影响<sup>[9]</sup>。但是，目前大多数研究所采用的传统分词工具或者 TF-IDF 算法不能实现对特征词的精确抽取，并且无法保留特征词的完整语义，这些局限可能会削弱后续主题识别的准确性与可解释性。2020 年，Yi Sun 等<sup>[10]</sup>人提出了 SIFRank 算法，这是一种基于预训练语言模型的无监督关键词提取算法，该模型利用预训练的深度双向语言模型

(ELMo, Embeddings from Language Models) 和句子嵌入方法 (SIF, Smooth Inverse Frequency)，巧妙地结合了语义嵌入的丰富性和句法结构的精确性，能够比较准确地提取文本中的关键短语。

为了解决 DTM 主题识别精度不足的问题，本文提出了一种将 SIFRank 算法和 DTM 模型相结合的方法。聚焦我国医学人工智能领域的本土化发展特征，以近十年国内医学人工智能领域相关文献为研究对象，采用 SIFRank-DTM 融合模型进行主题挖掘、热点识别和主题演化分析，以期更加客观、真实地揭示目前国内该领域的研究热点和发展趋势，为未来医学人工智能的研究方向提供参考。

## 2 数据与方法

### 2.1 数据来源与预处理

本文的数据来源于中国知网、万方和维普数据库，以（主题：医学）AND（主题：人工智能）检索式分别在三个数据库中进行检索，限定时间范围为 2015 年 1 月 1 日至 2025 年 9 月 10 日，文献类型为学术期刊，共得到 12970 篇文献，人工筛选并剔除重复、信息不完整、相关性不高的文献，最终保留 7059 篇有效文献。为适应 DTM 的时间序列分析要求，将所有数据划分为三个时间段，分别为 2015-2018 年 526 篇，2019-2022 年 2750 篇，2023-2025 年 3783 篇。从每篇文献中抽取文献标题、关键词和摘要作为研究对象，采用 jieba 分词工具包对其进行预处理，同时加入自定义词典、停用词表和同义词表进行优化，以提高后续分析结果的准确性。

### 2.2 研究方法

#### 2.2.1 SIFRank 算法

本文运用 SIFRank 算法进行特征词提取，SIFRank 由 ELMo 词向量模型和 SIF 句向量模型两部分组成。其中，ELMo 模型是 Matthew E.Peters 等<sup>[11]</sup>提出的一种基于深度学习的词向量表示方法，其通过双向 LSTM 网络对大规模语料进行预训练，能够捕捉词的多义性和语境依赖性；SIF 是一种基于加权平均的无监督句向量生成方法，通过将句向量乘以其逆文档频率 (IDF) 来降低常见词汇的影响。SIFRank 算法综合考虑了平滑因子、词语的 IDF 权重和句子向量的相似度，通过计算 SIFRank 得分来确定关键词，能够比较准确地提取文本中的关键短语。

鉴于本研究使用的数据包题目、关键词和摘要，其中题目和关键词凝练了文献的核

心内容，这一特征使得文本中的核心关键词通常更早出现。SIFRank 作为一种词袋模型，在长文档的关键词提取中有必要考虑位置信息对候选关键词重要性的影响<sup>[10]</sup>。因此本文在处理长文本时，通过位置偏权（SIFRank<sup>+</sup>）对先出现的词赋予更高权重。如图 1 所示，为 SIFRank 的算法流程。

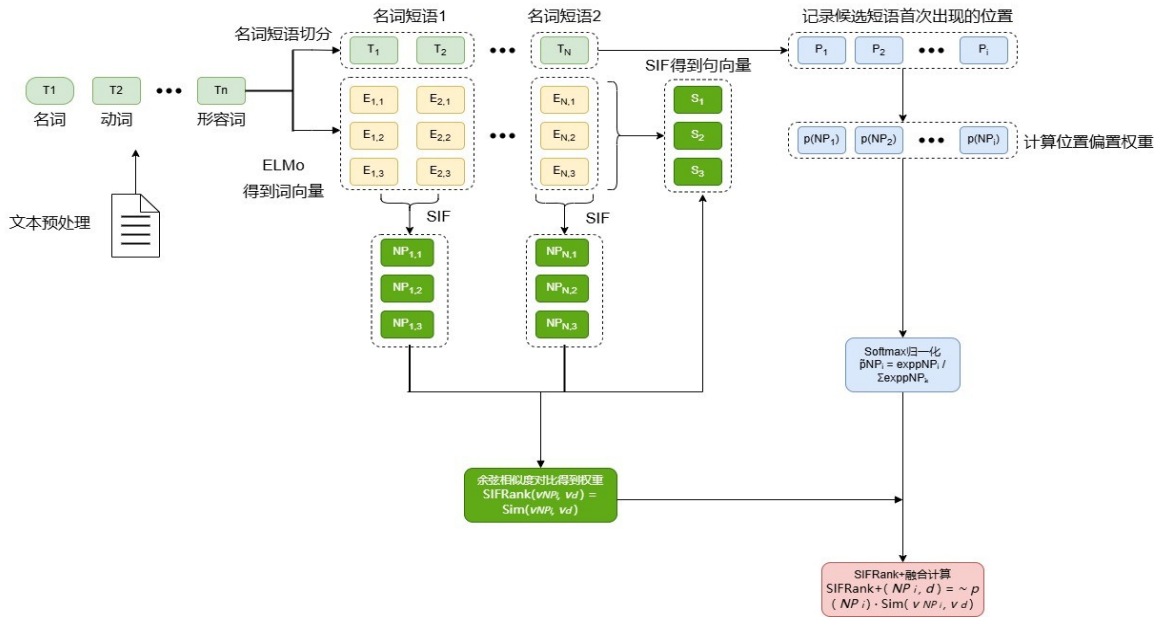


图 1 SIFRank 算法流程图

特征词是否精准、语义是否完整对于主题模型识别效果具有重要影响。表 1 展示了 SIFRank 算法抽取出的部分特征词，表左边为使用句子切分操作后的原句，中间为运用 jieba 分词工具包进行预处理后的结果，右边为 SIFRank 抽取出的特征词。在医学文本里，很多核心概念都是由多个词语组成的短语，比如“左心室舒张功能”“阻塞性睡眠呼吸暂停”等。可以看出，传统的分词工具会把这些完整的短语拆成一个个孤立的词汇，导致语义信息丢失，进而让不同主题变得混淆，SIFRank 算法抽取到的关键词则更为完整，

表 1 特征词展示（部分）

原文（句子切分后结果）	预处理结果	特征词抽取结果
左心室舒张功能评估现状及人工智能应用的研究进展，左心室舒张功能减退是心脏功能障碍的早期表现之一，诸多心脏疾病，包括高血压性心脏病、冠心病和肥厚型心肌病等，均会表现出心肌舒张功能减弱。	左心室 舒张 功能 评估 现状 左心室 舒张 功能 减退 心脏 功能障碍 早期 表现 诸多 心脏 疾病 包括 高血压 心脏病 冠心病 肥厚型 心肌病 均会 表现 心肌 舒张 功能 减弱	左心室舒张功能；心脏功能障碍；心脏疾病；高血压性心脏病；现状
阻塞性睡眠呼吸暂停诊疗的现状与未来，阻塞性睡眠呼吸暂停；睡眠医学；慢性疾患；临床实用性；检查方案；对因治疗；人工智能，阻塞性睡眠呼吸暂停发生率高且仍有上升趋势，其与多种慢性疾病风险增加相关并可造成严重的社会经济及健康负担。	阻塞性 睡眠 呼吸暂停 诊疗 未来 阻塞性 睡眠 呼吸暂停 睡眠 医学 慢性疾患 临床 实用性 检查 方案 治疗 阻塞性 睡眠 呼吸 暂停 发生率 上升 多种 慢性 疾 患 风险 增加 造成 严重 社会 经济 健康 负担	阻塞性睡眠呼吸暂停；慢性疾患；睡眠医学；经济；负担
重症监护病房脓毒症患者临床表型识别与验证，脓毒症；表型识别；机器学习；精准治疗，目的：探讨重症监护病房脓毒症患者临床表型的识别与验证。	重症监护病房 脓毒症 患者 临床 表型 识别 验证 脓毒症 表型 识别 机器学习 精准 治疗 探讨 重症监护病房 脓毒症 患者 临床 表型 识别 验证	重症监护病房；脓毒症患者；表型；脓毒症；机器学习

### 2.2.2 DTM 模型

DTM 模型是一种无监督的动态时序主题模型，它是在潜在狄利克雷模型（Latent Dirichlet Allocation, LDA）的基础上改进而来的一种无监督机器学习主题生成模型<sup>[12]</sup>，它

10.12201/bmr.202604.00081V1

能够分析带有时间属性的文本数据集，识别其主题，并揭示主题的动态演变规律<sup>[13]</sup>。本文在DTM模型初始化阶段加入SIFRank抽取出的关键词表，调整关键词在主题分布中的初始权重，把关键词作为特征增强，为模型提供有价值的先验知识，然后再进行DTM模型训练，这样可以使得主题识别结果具有更强的可解释性。

为确定最佳主题数目，本文综合比较各个时段不同主题数目的一致性和困惑度指标。如图2所示，当主题数目为10时，各时间段一致性得分相对较高，并且困惑度下降趋势趋于平缓，因此最终确定最佳主题数为10。

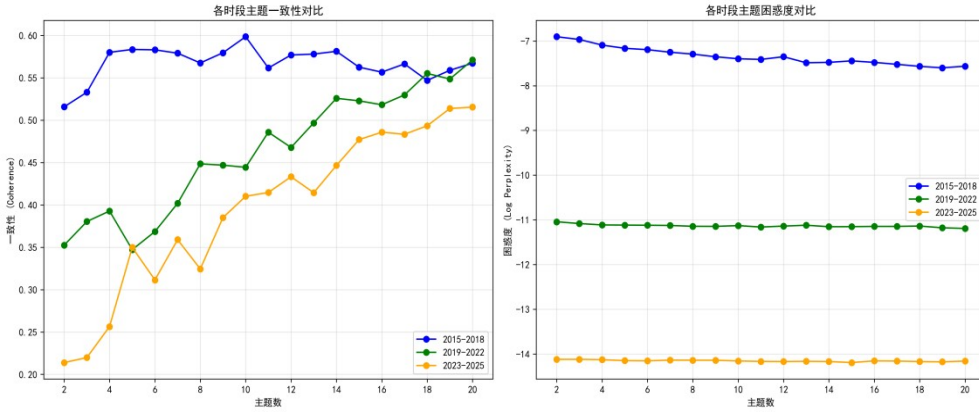


图2 主题一致性和主题困惑度曲线图

### 2.2.3 主题热度计算

主题热度指标主要用于揭示主题的受关注程度<sup>[14]</sup>，通过DTM模型可得“文档-主题”矩阵，进而得出每篇文档可能属于某个主题的概率<sup>[15]</sup>，以下为计算主题热度的公式：

(1)

$$S_{k,t} = \frac{1}{N_t} \sum_{d \in T_t} p(z_k | d)$$

式中， $S_{k,t}$ 表示在当前时间段t下，主题k的强度， $N_t$ 表示该时间段t内的文档总数， $T_t$ 表示时间段t内的文档集合， $p(z_k|d)$ 表示文档d属于主题k的概率。这个公式可以衡量主题k在整体语料中出现的频率或重要性， $S_{k,t}$ 越大则代表主题强度越高。

### 2.2.4 主题相似度计算

主题相似度代表两个文档之间的相似程度，相似度可以通过计算根据两个文档对应的主题概率分布来实现<sup>[16]</sup>。本文运用scikit-learn包中的cosine\_similarity函数来计算不同时间段内主题间的相似度，然后根据相似度来绘制桑基图，以表示医学人工智能的主题演化路径。

### 2.2.5 对比实验设计

为了验证融合模型的优越性，本文设计了消融实验，保持DTM模型的核心参数一致，只改变特征提取方法，设置三组对比实验，具体如下表2所示：

表2 对比试验设计表

实验组	特征提取方法	关键词表	模型
基线	TF-IDF	无	DTM
消融1	TF-IDF	有	DTM
消融2	词频	有	DTM

实验数据从7059篇有效文献中随机选取80%作为训练集，20%作为测试集。选取三个核心量化指标来评价模型性能：主题一致性衡量主题内关键词的语义关联性，数值越高越好；困惑度反映模型对测试集文档的预测能力，对数转换后绝对值越大越好；关键词F1值以人工标注的医学核心术语为标准，衡量特征词提取的准确性，数值越高越好。

## 3 结果

### 3.1 主题识别结果

利用不同阶段的文本数据训练DTM模型，可获得我国各个阶段医学人工智能研究的“主题-词矩阵”及词语权重，根据主题内的高权重词语可归纳该主题的具体类别<sup>[17]</sup>。在本文中，DTM模型共识别出10个研究主题，为揭示各主题的核心内涵，针对同一主题选取

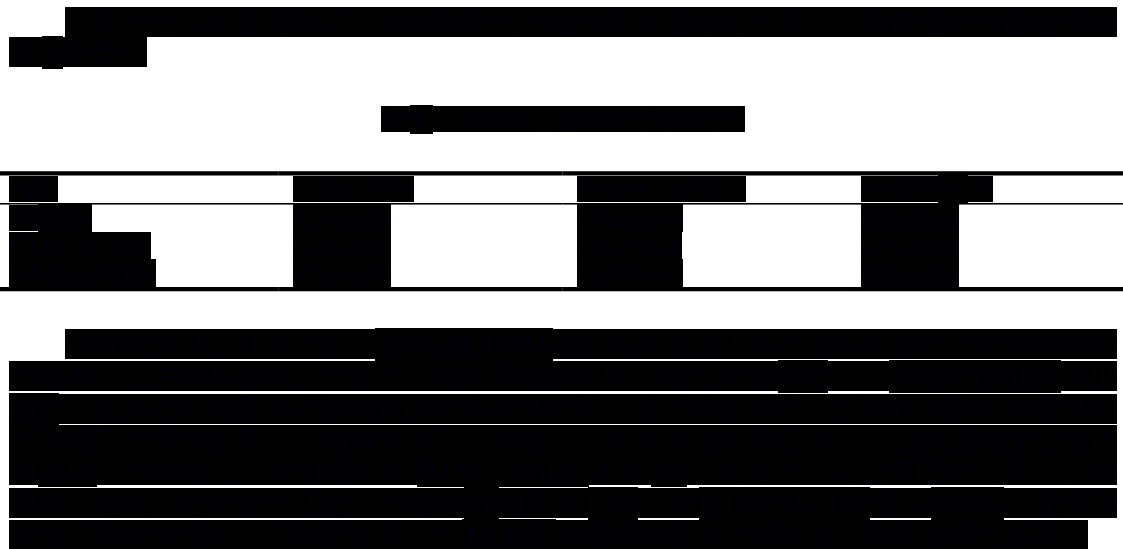
在各阶段均出现或者权重最高的前 10 个特征词，作为该主题的代表性关键词，用以反映医学人工智能领域的主要研究内容。结果如表 3 所示。

表 3 我国医学人工智能研究主题词表

序号	主题	主题关键词
1	人工智能赋能医学教育与实践	医学教育 人才培养 课程 创新 医学人才 学习 实践 临床 智能模拟人 医学生
2	人工智能在医学影像处理与分析中的应用	医学影像 医学图像 超声医学 医学诊断 ct 患者 深度学习 数据信息 机器学习 计算机辅助诊断
3	人工智能驱动的肿瘤诊疗与预后	医学影像 医学诊断 精准医学 治疗 肿瘤 深度学习 机器学习 病理 数据信息 预后
4	医学人工智能的文献计量与可视化分析	热点 文献 citespace 知识图谱 关键词 合作 机构 数据库 医学诊断 心血管疾病
5	人工智能驱动中医诊疗现代化	中医 数据信息 中医药理论 临床 大数据技术 诊疗 针灸 知识图谱 人工神经网络 知识
6	人工智能在药物研发和中医药现代化中的应用	中药 药物 预测 数据信息 生物 设计 基因 细胞 分子 标准
7	医学人工智能的伦理治理研究	医学伦理 社会 虚拟现实技术 chatgpt 人类 医生 道德 隐私问题 风险 机器人
8	人工智能赋能智慧医院建设	数据信息 医院 电子病历 管理 平台 临床 服务 患者 互联网 大数据技术
9	人工智能赋能围术期诊疗与康复协同	康复医疗 机器人 护理 患者 医生 手术 医学知识库 口腔医学领域 骨科 外科
10	多模态人工智能驱动精准诊疗与临床决策革命	大语言模型 临床 机器人 医学图像 深度学习 精准医学 机器学习 医学诊断 患者 评估

从主题分布来看，近十年来人们对于医学人工智能的研究涉及医学教育、医学影像、肿瘤诊疗与预后、文献计量与可视化分析、中医诊疗、药物研发、伦理治理、智慧医院建设等多个领域，这表明随着人工智能技术的迅速发展，人工智能在医学领域得到了广泛的研究与应用，逐步推动我国医疗事业进入智能时代。

### 3.2 融合模型有效性验证结果



### 3.3 热点主题计算结果

为反应不同时间段医学人工智能的研究重心，本文运用公式（1）计算三个时间段内各个主题的主题热度，并绘制主题热力图（图 3）。结合近十年医学人工智能领域文献发表数量增长趋势（图 4）来看，在不同时段医学人工智能研究各有侧重。

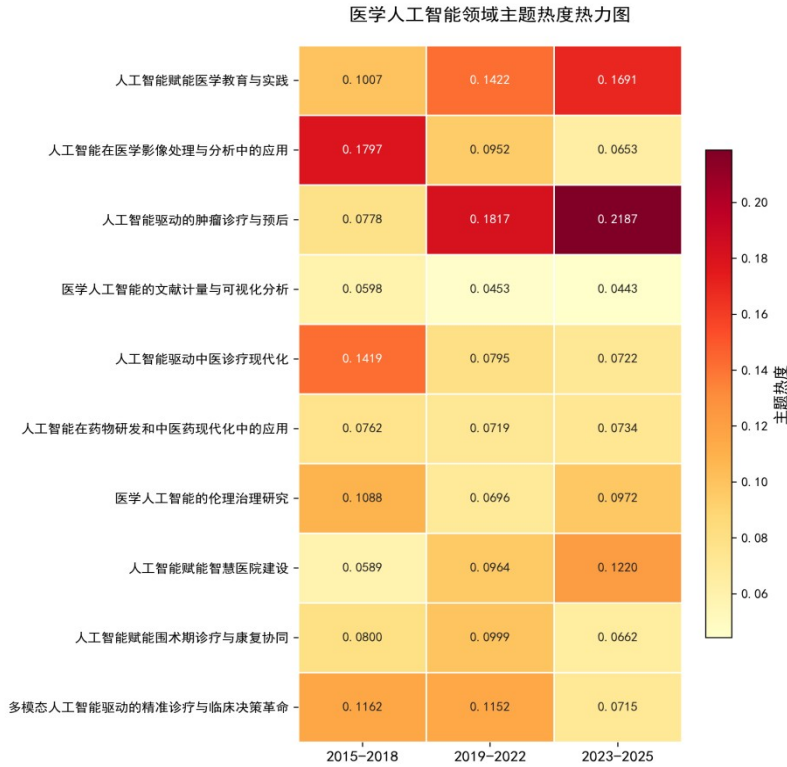


图 3 我国医学人工智能领域主题热度热力图

慢，此阶段发文数量较少且增长速度缓

年我国医学人工智能进入快速发展阶段，发文数量明显增长，这一时期的热点主题是人工智能驱动的肿瘤诊疗与预后、人工智能赋能医学教育与实践、多模态人工智能驱动的精准确诊与临床决策革命，其主题强度分别高达

年我国医学人工智能研究走向深度融合阶段，这一时期的发文数量加速上升并且重点聚焦于人工智能驱动的肿瘤诊疗与预后、人工智能赋能医学教育与实践、人工智能赋能智慧医院建设，其主题强度分别高达

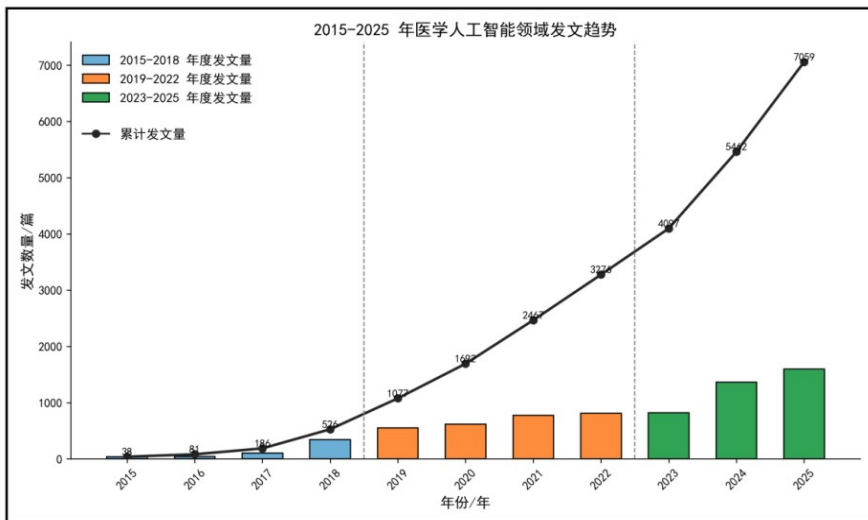


图 4 2015-2025 年医学人工智能领域发文趋势

10.12201/bmr.202604.00081V1

### 3.4 主题演化结果

如图 5 所示，为我国医学人工智能研究在近十年内的主题演化历程。图中，同一时间段的主题位于一列，不同颜色的节点代表不同的主题，节点之间的连线表示不同时间段内主题之间的演化趋势，连线越粗则代表关联越强。

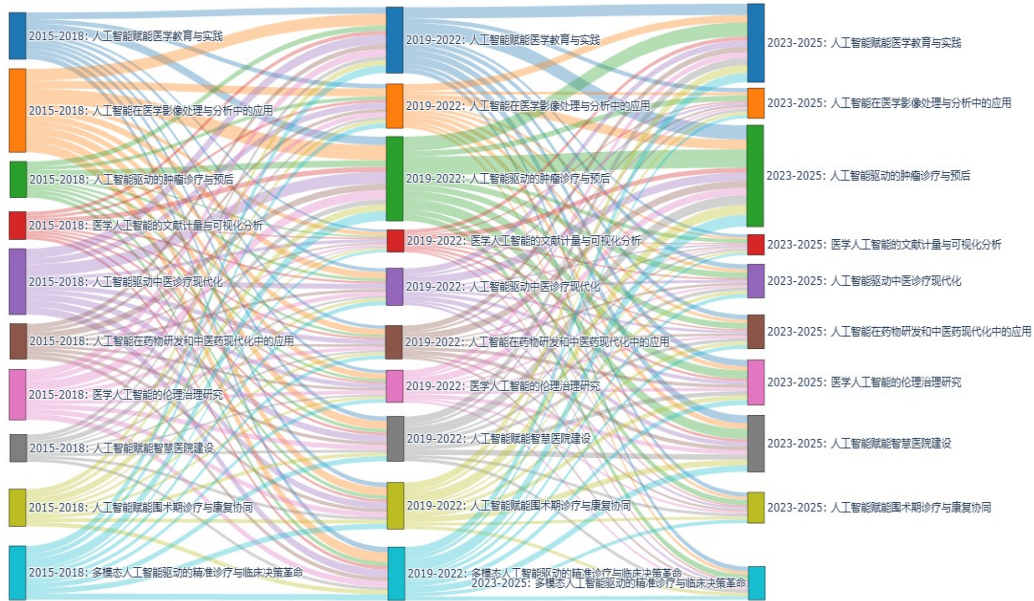


图 5 我国医学人工智能领域主题演化路径

根据图 5 可以看出，我国医学人工智能研究的各主题之间存在显著的演化关系。对于人工智能在肿瘤诊断与治疗中的应用这一主题的研究逐步稳定增多，并且在起步阶段该主题主要由医学影像分析、中医药研究、精准诊疗等技术或方法类主题演化而来，而到了深度融合阶段则向医学教育、医院管理与临床服务模式创新等方向延伸。人工智能驱动中医诊疗现代化以及人工智能在药物研发和中医药现代化中的应用这两个主题则稳定发展、相互融合，并且均匀地渗透在其他主题中。人工智能赋能的医学教育与实践和精准诊疗与临床决策革命、智慧医院建设等主题密切相关。

## 4 分析

### 4.1 热点主题分析

#### 4.1.1 我国医学人工智能研究的起步阶段

在我国医学人工智能的起步阶段，重点聚焦于 [redacted] 影像学检查是临床诊疗的重要组成部分，因此在医学人工智能研究的起步阶段，人们就注重于如何利用人工智能技术更好地识别影像数据中所包含的病理学信息并将其准确用于解决临床问题；中医作为我国的传统医学，有着悠久的历史 and 独特的理论体系，但在临床科研方面面临一些挑战，如数据处理效率低、诊断和治疗的标准化程度有待提高等<sup>[18]</sup>，因此在这一阶段，研究如何利用人工智能技术处理症状描述、舌象脉象记录等具有主观性和模糊性的文本信息，挖掘传统病案在现代医学中的价值等成为热点主题；精准诊疗与临床决策是整个医疗体系中的重要组成部分，在起步阶段国内学者专注于研发机器人技术在医学领域的应用以辅助精准诊疗，因此这一时期国内研究在骨科机器人、神经外科机器人、血管介入机器人等多个专科领域实现了创新。

#### 4.1.2 我国医学人工智能研究的快速发展阶段

[redacted] 人工智能在肿瘤诊断与治疗、医学教育与实践、精准诊疗与临床决策中的应用成为热点主题。 [redacted]，尤其是恶性肿瘤的死亡率非常高，因此运用人工智能技术进行生物信息分析、医学影像分析等为癌症的病理机制研究、早期筛查、诊断和治疗提供支持，是这一阶段的核心主题；随着人工智能技术的高速发展，我国政府也高度重视高等医学院校发展人工智能教育，相继推出系列指导性文件，因此在快速发展阶段，如何推动医学教育的智能化改革、培养具备 AI 素养的复合型医学人才、提高医学生的岗位胜任力等相关研究论文频发；多模态人工智能驱动精准诊疗与临床决策革命在这一时期依然是热点主题，但是在这一阶段人们认识到深度学习的潜力，因此一直致力于将深度学习应用于医学

影像分析、药物研发、基因组学研究、疾病预测等多个领域，有效提高医学诊断的效率和准确率，对医学领域的发展影响深远。

#### 4.1.3 我国医学人工智能研究的深度融合阶段

在我国医学人工智能的深度融合阶段，侧重于对人工智能在肿瘤诊断与治疗、医学教育与实践以及智慧医院建设的应用研究。由于国家对于医学教育智能化改革的重视，以及恶性肿瘤对于人体健康的危害，在这一阶段它们依然是研究重点。此外近年来，随着《关于进一步完善医疗卫生服务体系的意见》《卫生健康行业人工智能应用场景参考指引》等文件的发布，积极运用人工智能技术推动医院管理与临床服务模式创新，建设智慧医院成为研究热点。在这一主题下的关键词主要有大数据技术、平台、医院、数字化、互联网、电子病历等，揭示了人工智能技术为医院的诊疗、管理和服务带来了深刻变革，从实现患者就医流程智能化、疾病筛查和AI辅助诊断，到机器人手术和术后康复检测，再到医院病案数据管理信息化，对于人工智能技术在医院管理中的应用研究几乎覆盖了医院运行的全流程。

### 4.2 主题演化特征分析

#### 4.2.1 我国医学人工智能研究以肿瘤学为核心应用领域

根据主题演化结果可以看出，肿瘤学是我国医学人工智能研究的核心应用领域，并且整体呈现出由微观向宏观、由局部向系统的扩展趋势。在起步阶段，人工智能在医学影像与精准诊疗中的突破为肿瘤的诊断与治疗提供了可靠的技术基础，例如医学影像计算机辅助检测与诊断系统，不仅可以在医学图像上检测异常并且定位呈现出来，还可以帮助医生判断异常的类别以及恶性级别<sup>[19]</sup>，在癌症诊断中与活检相结合，为精确定论提供重要依据<sup>[20]</sup>；再例如基于深度学习算法的纵向影像组学研究，深度学习无需人工定义特征，而是凭借卷积神经网络等网络结构来捕捉图像的空间和纹理信息，可以直接从原始图像数据中自动学习并表征肿瘤结构的复杂特征<sup>[21][22]</sup>，推动了对肿瘤长期疗效评估、预测等任务的深度探索。随着数据资源与技术的逐渐成熟，在深度融合阶段，研究者开始关注肿瘤智能诊疗在医学教育中的核心教学内容以及在医疗流程与临床服务优化中的作用，例如人工智能辅助教学正成为医学教育改革的重要方向，有研究<sup>[23]</sup>表明，住院医师应用虚拟现实技术进行胃肠道内窥镜训练会有效提高他们在临床实践中的精确度；再例如近年来肿瘤患者术后随访管理系统、智能分诊等应用正逐渐兴起，医院运营模式产生了系统性变革，有效提升肿瘤患者的就医体验。在医疗政策方面，

#### 4.2.2 我国传统医学数智化形成持续性发展脉络

中医药作为我国具有独特理论体系和丰富实践经验的传统医学科学，推动其与医学人工智能技术深度融合，实现其数智化转型至关重要。在我国医学人工智能研究起步阶段，人工智能赋能中医诊疗现代化以及人工智能在药物研发和中医药现代化中的应用这两个主题的主要关键词有大数据技术、文献、知识、医学诊断等，说明在这一时期主要聚焦于传统中医药知识的数字化整理与知识建模，开始从“经验中医”迈向“数字中医”，比如有研究<sup>[24]</sup>利用大数据技术构建中医药知识图谱，实现中医药知识资源的有效整合；还有综合运用聚类分析等多种数据挖掘方法，实现名老中医经验传承及对名老中医的经验整理与初步诊疗系统的构建<sup>[25]</sup>等。在医学人工智能快速发展和深度融合阶段，这两个主题开始与医学教育、肿瘤诊断与治疗、精准诊疗与临床决策等主题出现交叉连接，说明医学人工智能在中医药领域中的作用逐步拓宽，并且实现了知识、算法与临床的全面融合，“数字中医”开始向“智慧中医”迈进，比如在中医诊断领域，以舌诊、脉诊、色诊为代表的四诊客观化技术逐渐成熟，形成了舌诊仪、脉诊仪、色诊仪等多种中医诊断仪器<sup>[26]</sup>；还有充分发挥中医治未病思想，将中医药人工智能技术融入健康管理，促进精准化个性医疗的发展等。在医疗政策方面，

#### 4.2.3 我国医学人工智能研究趋于体系化发展

结合主题演化结果以及前面的热点主题分析可以看出，我国医学人工智能研究的核心方向从医学影像、中医药领域、外科手术等单一学科应用转向医学教育与实践、智慧医院建设等多个维度，不同主题之间的交叉密度增加，呈现出多学科融合与应用深化的系统性特征，这意味着医学人工智能研究生态系统正在不断完善，一方面多模态人工智能技术与医学科学的深度融合成为医学创新的基础动力，影像学、肿瘤学、药物研发等领域形成

“算法-数据-临床”闭环；另一方面，医学教育智能化成为医学人工智能研究的重要组成部分，有利于培养复合型医学人才，为医学研究长线发展提供动力；此外，医学人工智能研究延伸至医院管理、医疗服务与公共健康治理，形成全流程智慧医疗体系。在医疗政策方面，

[REDACTED]

## 5 结论

本研究以中国知网、万方与维普数据库中 2015-2025 年医学人工智能相关文献为数据基础，构建融合 SIFRank 与 DTM 模型的主题挖掘框架，通过引入关键词先验增强机制，提高了主题识别结果的语义完整性与可解释性。模型共识别出 10 个具有稳定结构的研究主题，涵盖医学教育、医学影像分析、肿瘤诊疗与预后、中医诊疗、药物研发、智慧医院建设、围术期管理优化、伦理治理等多个方向，较系统地揭示了医学人工智能研究的知识结构。基于此，本研究进一步揭示了在起步阶段、快速发展阶段、深度融合阶段三个不同时期我国医学人工智能研究的热点主题，并归纳出该领域的演化路径呈现出以肿瘤学为核心应用领域、传统医学数智化形成持续性发展脉络、医学人工智能研究趋于体系化发展三个演化特征。

[REDACTED]

作者贡献：吕艳华负责提出研究思路与方法；王萱负责撰写论文，处理数据，修改论文；崔云龙、朱鑫鑫负责处理数据、修改论文；白慧娟负责格式校对，修改论文。  
利益声明：所有作者均声明不存在利益冲突。

## 参考文献

- [1] PeaYao Z , et al.Application of Artificial Intelligence in Medicine: An Overview.[J].Current medical science,2021,41(6):1-11.DOI:10.1007/S11596-021-2474-3.
- [2] “健康中国 2030”规划纲要[EB/OL].(2016-10-25)[2025-10-08].  
[https://www.gov.cn/zhengce/2016-10/25/content\\_5124174.htm](https://www.gov.cn/zhengce/2016-10/25/content_5124174.htm).
- [3] “十四五”卫生与健康科技创新专项规划[EB/OL]. (2022-11-01)  
[2025-10-08]. [https://www.most.gov.cn/xxgk/xinxifenlei/fdzdkgknr/fgzc/gfxwj/gfxwj2022/202301/t20230116\\_184248.html](https://www.most.gov.cn/xxgk/xinxifenlei/fdzdkgknr/fgzc/gfxwj/gfxwj2022/202301/t20230116_184248.html).
- [4] 《新一代人工智能发展规划》政策解读 [EB/OL] . (2018-06-02)  
[2025-10-08].[https://www.sohu.com/a/233833019\\_465947](https://www.sohu.com/a/233833019_465947).
- [5] 马秀敏.中国典型管理期刊文献主题发现与演化分析[D].大连理工大学,2011.
- [6] 王燕鹏.国内基于主题模型的科技文献主题发现及演化研究进展[J].图书情报工作,2016,60(03):130-137.DOI:10.13266/j.issn.0252-3116.2016.03.019.
- [7] Mohr J W,Bogdanov P.Introduction-Topic models:What they Are and Why They

- Matter[J].Poetics,2013,41(6):545-569.
- [8] Blei D M, Lafferty J D. Dynamic topic models[C]//Proceedings of the 23rd international conference on Machine learning. 2006: 113-120.
- [9] Zhu H, Qian L, Qin W, Wei J, Shen C. Evolution analysis of online topics based on 'word-topic' coupling network. *Scientometrics*. 2022; 127(7):3767-3792.
- [10] Yi S, Hangping Q, Yu Z, et al. SIFRank: A New Baseline for Unsupervised Keyphrase Extraction Based on Pre-Trained Language Model[J]. *IEEE Access*, 2020, 8:10896-10906.
- [11] Peters M E, Neumann M, Iyyer M, et al. Deep contextualized word representations[C]// Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. New Orleans: Association for Computational Linguistics, 2018: 2227-2237.
- [12] Blei M D, Ng Y A, Jordan I M. Latent Dirichlet Allocation[J]. *Journal of machine learning research*, 2003, 3(4/5):993-1022.
- [13] Li D, Ding Y, Shuai X, et al. Adding community and dynamic to topic models[J]. *Journal of Informetrics*, 2012, 6(2):237-253. DOI:10.1016/j.joi.2011.11.004.
- [14] 周健,张杰,屈冉,等.基于LDA的国内外区块链主题挖掘与演化分析[J].*情报杂志*,2021,40(09):161-169.
- [15] 邱均平,胡博,徐中阳,等.基于DTM模型的国内外话语权研究主题挖掘及比较分析[J].*情报理论与实践*,2023,46(02):24-34.DOI:10.16353/j.cnki.1000-7490.2023.02.004.
- [16] 王振振,何明,杜永萍.基于LDA主题模型的文本相似度计算[J].*计算机科学*,2013,40(12):229-232.
- [17] 吴瑞鹏,李勇男,刘帅,等.基于DTM的美国人工智能战略热点主题及演化分析[J].*情报杂志*,2023,42(12):134-143.
- [18] 石桥,汪美霞,王艳昕,等.人工智能在中医临床科研中的应用实践与展望[J/OL].*中华中医药学刊*,1-8[2025-10-24].<https://link.cnki.net/urlid/21.1546.r.20251020.0945.004>.
- [19] 郑光远,刘峡壁,韩光辉.医学影像计算机辅助检测与诊断系统综述[J].*软件学报*,2018,29(05):1471-1514.DOI:10.13328/j.cnki.jos.005519.
- [20] endee WR. The impact of future technology on oncologic diagnosis: oncologic imaging and diagnosis. *Int J Radiat Oncol Biol Phys*. 1983 Dec; 9(12):1851-65.
- [21] Ferro A, Bottosso M, Dieci MV, Scagliori E, Miglietta F, Aldegheri V, Bonanno L, Caumo F, Guarneri V, Griguolo G, Pasello G. Clinical applications of radiomics and deep learning in breast and lung cancer: A narrative literature review on current evidence and future perspectives. *Crit Rev Oncol Hematol*. 2024 Nov; 203:104479.
- [22] Shen D, Wu G, Suk H. Deep Learning in Medical Image Analysis[J]. *Annual Review of Biomedical Engineering*, 2017, 19:221-248. DOI:10.1146/annurev-bioeng-071516-044442.
- [23] Grantcharov TP, Carstensen L, Schulze S. Objective assessment of gastrointestinal endoscopy skills using a virtual reality simulator. *JLS*. 2005 Apr-Jun; 9(2):130-3.
- [24] 于彤,刘静,贾李蓉,等.大型中医药知识图谱构建研究[J].*中国数字医学*,2015,10(03):80-82.
- [25] 杨蕴,阮春阳,杨美清,等.人工智能技术促进中医药传承发展[J].*第二军医大学学报*,2018,39(08):873-877.DOI:10.16781/j.0258-879x.2018.08.0873.
- [26] 赵宇平,李楠,闫朋宣,等.中医药人工智能现状研究及发展思考[J].*中国中西医结合杂志*,2020,40(06):746-749.