# 基于区块链的医学科学数据溯源模型构建研究[1]

乔柏珏，方安，周易，杨晨柳，陈凌云，胡佳慧
（北京协和医学院/中国医学科学院/医学信息研究所/图书馆，北京 100020）

[摘要] **目的/意义** 针对医学科学数据在跨机构协作中面临的数据孤岛、隐私安全与信任缺失问题，构建一种可信、可追溯的数据流通与管理模型。**方法/过程** 提出融合联盟链、PROV 模型与 FHIR 标准的可信溯源框架，通过分层架构设计，形成覆盖数据处理、区块链存证、智能合约执行与跨链互操作的全链路溯源体系，实现医学科学数据的标准化描述细粒度隐私保护与可验证流转。**结果/结论** 基于区块链的溯源支持数据存证、授权审计与完整性验证，为跨机构医学数据治理与可信协作提供可行技术路径。

[关键词] 医学科学数据；区块链；数据溯源；跨机构协作；可信数据空间

## Research on the Construction of a Blockchain-based Provenance Model for Medical Science Data

Baijue Qiao, An Fang, Yi Zhou, Chenliu Yang, Lingyun Chen,Jiahui Hu
(Institute of Medical Information, Chinese Academy of Medical Sciences & Peking Union Medical College, Beijing)

[Abstract] **Purpose/Significance** To address the challenges of data silos, privacy security risks, and lack of trust in cross-institutional collaboration of medical scientific data, this study aims to construct a trustworthy and traceable model for data circulation and management. **Method/Process** A trusted traceability framework integrating consortium blockchain, the PROV model, and the FHIR standard is proposed. Through a layered architecture design, a full-chain traceability system covering data processing, blockchain-based certification, smart contract execution, and cross-chain interoperability is developed to enable standardized representation, fine-grained privacy protection, and verifiable circulation of medical scientific data. **Results/Conclusion:** The proposed method supports data certification, authorization auditing, and integrity verification, providing a feasible technical path for cross-institutional data governance and trustworthy collaboration.

[Keywords] medical scientific data; blockchain; data provenance; cross-institutional collaboration; trusted data space



## 1 引言

数据资源已成为驱动创新的重要新质生产力要素。2025 年 11 月，国家卫生健康委办公厅、国家发展改革委办公厅、工业和信息化部办公厅、国家中医药局综合司、国家疾控局综合司联合印发《关于促进和规范“人工智能+医疗卫生”应用发展的实施意见》[1]，将建立“卫生健康行业高质量数据集和可信数据空间”列为总体目标。2026 年 3 月，《中华人民共和国国民经济和社会发展第十五个五年规划纲要》[2]明确提出“建立健全数据产权、流通利用、收益分配、安全治理等数据要素基础制度”，并部署实施“可信数据空间发展行动计划”与“国家区块链网络建设工程”。在医疗信息化与精准医学快速发展的背景下医学科学数据呈现爆发式增长，涵盖电子病历、影像、基因组、可穿戴设备等多源异构类

1基金项目：中国医学科学院医学与健康科技创新工程项目（项目编号：2021-I2M-1-056）通讯作者：胡佳慧，hu.jiahui@imicams.ac.cn

型，已成为临床研究、诊疗决策与健康管理的核心战略资产。充分释放和挖掘其价值，对我国医疗健康事业发展至关重要。

然而，医学科学数据体量巨大、模态多样、价值密度高，其“数据要素”潜能尚未充分释放[3]。当前医学科学数据的利用面临三大挑战：一是“数据孤岛”现象普遍，数据难以安全流通与整合；二是数据质量参差不齐，存在不一致、不完整与非标准化问题；三是数据敏感性与合规要求形成数据利用壁垒[4-5]。这些难题不仅制约了高质量数据集的构建，同时也影响了人工智能技术在医疗领域的规范化应用。医学科学数据作为国家重要的战略资源和核心生产要素，其治理效能关系“健康中国”战略的实施与数字经济的发展。在此背景下，亟需构建一套可信、可溯的医学科学数据流通与管理模型。

数据溯源通过对数据进行多方向、多维度的追踪管理，有助于提升数据的可靠性和可信性，为解决上述困境提供了有效路径。区块链作为一种去中心化的分布式数据库技术，基于密码学原理和共识机制，以区块链式结构实现数据的存储、验证与传播[6]，具备防篡改、去中心化、可追溯等特性，为构建可信溯源体系提供了坚实支撑。目前，区块链已在制造业、食品溯源等多个领域得到应用验证[7-8]。本研究将其引入医学数据管理，旨在构建一套兼顾数据标准化描述、隐私保护、可信存证与跨机构互操作的医学科学数据溯源方案以促进透明、可信且高效的医学科学数据流通与管理。

# 2 相关研究

区块链技术在解决医疗数据孤岛、保障隐私安全及重建数据信任等方面展现出显著潜力，已成为医疗信息化领域的研究热点。

针对数据管理与共享，目前研究致力于通过区块链的去中心化、不可篡改等特性，提升医疗数据的安全性和共享效率。例如，Azaria 等[9]提出去中心化病历管理系统 MedRec，通过模块化设计增强互操作性；Peterson 等[10]构建基于区块链的健康交换网络，解决机构异构与语义分歧问题；陈嘉莉等[11]总结了符合本土需求的访问控制与完整性验证方案。

为平衡数据利用与隐私安全，研究者探索了多种技术路径。Myrzashova 等[12]将区块链与联邦学习结合，构建去中心化协作框架，以缓解医疗数据集中训练导致的隐私泄露问题 Verma 团队[13]提出基于权限评分的动态访问控制架构，兼顾隐私保护与访问效率；Zhao 等[14]通过属性加密实现细粒度权限管理。

在医疗设备、药品供应链及影像数据等领域，区块链的溯源价值已得到初步验证。Wu[15]等利用区块链技术实现了医疗器械全生命周期溯源，Abidin[16]团队构建了药品供应链透明化网络，Gao[17]等开发了医疗影像共享与溯源系统。

综上，当前研究已在特定领域取得一定进展，为构建跨机构协同的数据信任体系奠定了重要基础。但现有研究大多以单一视角展开，例如仅针对医学影像数据的溯源[17]，缺乏对医学科学数据标准化、溯源数据储存以及跨链互操作多个层面的系统研究。特别是面对日益增长的多模态医学科学数据共享需求，实现其全流程、可互操作、可验证的溯源仍面临挑战。因此，本文引入联盟链及相关隐私保护技术，从体系层面构建覆盖医学科研全流程的科学数据可信溯源模型。

# 3 基于区块链的医学科学数据溯源

## 3.1 相关技术

区块链通过加密链式结构、分布式节点共识与智能合约，实现数据的可信存储与验证[18-20]。根据运行模式的不同，可分为私有链、公有链和联盟链。其中，联盟链采用准入机制，适合跨机构协作场景。与公有链相比，其节点规模可控、效率更高；相较于私有链，联盟链能更好地支持多类型机构间安全、可控数据共享。

PROV 数据溯源模型是由万维网联盟（W3C）制定的开放标准，旨在为数据的来源与演变历史提供规范化描述框架。该模型围绕实体（Entity）、活动（Activity）与代理（Agent）三个核心概念，对数据生命周期中的各种关系进行形式化刻画，为评估数据的可信性提供结构化基础。该模型具备良好的可扩展性，可根据具体应用场景进行语义细化与扩展，从而适应不同领域的复杂需求[21]。

快速医疗互操作性资源（Fast Healthcare Interoperability Resources，FHIR）是由 HL7 制定的新一代医疗信息交换标准。该标准通过将复杂的临床与管理信息拆解为模块化组件“资源”，并借助标准化数据元素与语义规范，提升跨系统数据共享的互操作性，是智慧医疗信息集成的核心基础。

## 3.2 语义框架

为实现对医学复杂研究活动的结构化描述，提升跨机构数据共享与可信互操作的可行性与效率，本研究以 PROV 为参考，对其节点类型和依赖关系进行语义扩展，设计面向医学科学数据的溯源框架，如图 1 所示。其中，实体涵盖从原始数据到最终知识产出的全类型数据载体，包括原始数据（如原始影像文件）、派生数据（如从电子病历中提取的标准

10.12201/bmr.202605.00089V1

化病理特征表）、聚合数据（如队列研究数据表）、知识成果（如发表的学术论文）及引用数据（如数据库）；代理则囊括科研活动中的各类参与主体与工具，包括研究人员等参与主体、医院等机构、专业分析工具等软件算法，以及测序仪等物理设备；活动贯穿数据完整生命周期，涵盖数据采集、处理、分析及融合等一系列关键步骤；新增研究项目节点将其作为完整科研活动的顶层容器，将所有相关的数据、活动和代理聚合在明确的研究语境之下，用于从宏观层面界定和组织资源；新增数据集节点，面向共享与管理的多实体逻辑集合，将不同来源的分散数据实体封装成统一单元，便于后续授权、引用和全流程审计各节点间的溯源关系由边进行表征，核心关系类型包括管理与归属、成员与聚合、衍变与生成、使用、委托等。

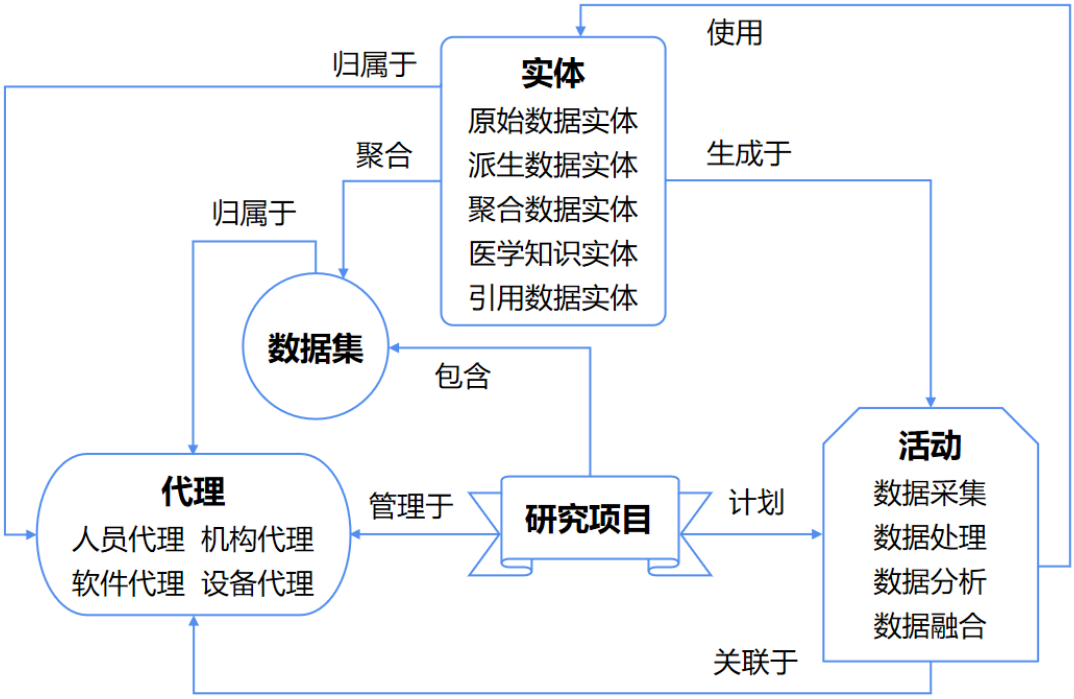


图 1 医学科学数据溯源语义框架

上述语义框架通过引入医学科研环境下的节点类型与依赖关系，以语义扩展实现多模态医学科学数据从采集、治理、分析到知识生成的全链路、细粒度、可验证溯源，为跨机构、全生命周期的可信数据协作提供标准化语义基础。

## 3.3 溯源流程

针对医学科学数据多源异构、隐私敏感及跨机构流转的特点，对其全生命周期进行动态追踪与验证。该流程通过对临床观测、实验研究等产生的原始数据进行标准化转换与溯源信息捕获，形成完整的可信数据链并分布式存储，最终供研究人员、伦理委员会等授权主体进行查询与验证，以评估数据的可靠性、合规性及科研价值。溯源流程主要包括数据注册与存证、数据授权与共享、数据使用与派生、审计与验证等过程。

### 3.3.1 注册与存证

依据 FHIR 标准，将来自各医疗信息系统的原始数据转换为结构化资源，同时通过溯源模型记录其生成上下文信息，构建机器可读的溯源文档。原始数据存储于链下分布式系统中，而其哈希指纹与关键元数据则通过智能合约锚定至联盟链，完成数据的身份注册与可信存证。

### 3.3.2 授权与共享

用户发起数据访问请求时，通过访问控制智能合约自动验证其属性是否满足数据加密时设定的 CP-ABE 访问策略。验证通过后，向用户分发与其属性绑定的解密密钥。整个过程无需数据所有者介入，实现细粒度、动态化的安全共享，并形成完整的权限审计轨迹。

### 3.3.3 使用与派生

获得授权的研究人员使用密钥解密数据后，任何基于原始数据的分析、计算或整合操作均会触发新一轮的溯源捕获。系统自动生成新的溯源记录，并将新数据的存证信息再次

10.12201/bmr.202605.00089V1

上链。这使得数据的整个谱系得以延伸，形成环环相扣、不可篡改的溯源链，完整记录其科研价值衍生过程。

#### 3.3.4 审计与验证

针对数据质量评估或合规性审查需求，授权用户可通过溯源查询服务获取完整的操作历史，最终组装成完整的、可视化的数据审计报告。此外，用户也可随时通过数据验证合约比对文件哈希，即时验证数据的完整性。

### 3.4 系统架构

为实现医学科学数据在跨机构协作中的可信溯源、隐私安全与高效互操作，本研究基于联盟链、PROV 模型与 FHIR 标准，设计医学科学数据可信溯源系统架构，如图 2 所示，包括基础支撑层、数据处理层、区块链层、智能合约层、核心服务层及应用层。

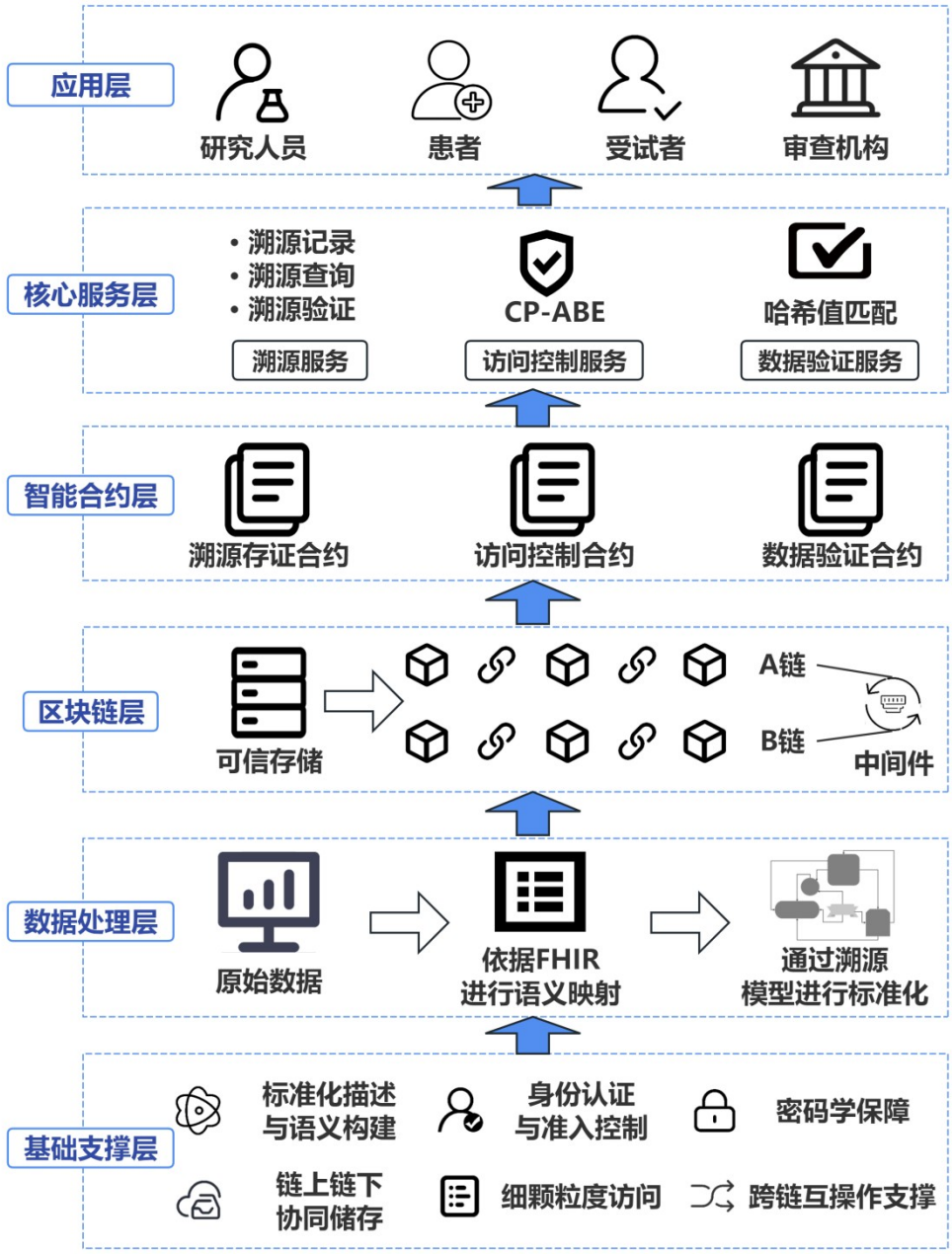


图 2 基于区块链的医学科学数据溯源架构

#### 3.4.1 基础支撑层

基础支撑层主要包括标准化描述与语义建模、身份认证与准入控制、密码学保障、链上链下协同存储、细粒度访问控制以及跨链互操作支撑等，用于支撑多源异构医学数据的

10.12201/bmr.202605.00089V1

统一表示、可信存证、隐私保护与跨机构共享，为上层各模块的协同运行、可信交互与安全治理提供了统一的技术底座。

#### 3.4.2 数据处理层

数据处理层是溯源流程的起点，主要负责数据标准化与溯源信息采集。依据 FHIR 标准，对来自不同医疗信息系统的多源异构原始数据进行映射和转换，并通过溯源模型将其表示为“实体-活动-代理”及其相互关系，从而序列化为机器可读形式，同步自动记录数据生成、转换和使用的上下文信息，为后续区块链存证与全生命周期追溯提供可信基础。

#### 3.4.3 区块链层

针对医学数据来源广泛、类型多样、隐私要求严苛等特点，为兼顾准入控制、系统性能与协作需求，采用联盟链架构以支持多家机构在可控条件下开展数据共享。在存储策略上，采用“链上存哈希，链下存原件”的混合存储模式，即：大体量原始数据文件存储于链下专用系统，仅将其哈希指纹及关键溯源信息摘要等元数据锚定至区块链，从而既利用区块链实现可信存证与公开验证，又避免原始数据直接上链带来的存储膨胀与性能瓶颈，且链下数据一旦被篡改，其哈希值将无法与链上记录匹配，从而实现完整性校验与隐私合规保障。

#### 3.4.4 智能合约层

智能合约层作为系统的自动化核心与互联枢纽，通过模块化智能合约实现数据确权、授权与审计规则的自动执行，并借助跨链中间件支持不同联盟链间的安全交互。其中，跨链中间件承担异构适配、消息中继、证明验证与结果映射等功能，针对不同联盟链在底层平台、接口规范及共识机制上的差异[22]，提供标准化接口；发起跨链请求时，中间件先校验身份与权限，再将请求转换并中继至目标链，目标链仅返回哈希指纹等可验证证据而非原始数据，请求链侧依据信任锚完成校验并映射业务状态，从而实现不依赖同构底层架构的可信跨链协作。

模块化智能合约群主要包括三种智能合约：①溯源存证合约。负责接收来自上层服务的溯源信息及其哈希值记录，并将其永久性地写入区块链。②访问控制合约。与 CP-ABE 配合，验证用户属性并执行动态细粒度授权。③数据验证合约。自动计算用户提交数据文件哈希值并与链上存储的原始哈希值进行比对，验证数据的完整性。此外，当需要进行跨链验证或查询时，跨链中间件则在此基础上进一步实现联盟链间的安全数据与溯源信息交互。

#### 3.4.5 核心服务层

核心服务层将底层区块链操作与密码学计算封装为标准 API，向上提供溯源、访问控制与数据验证三大服务。其中，溯源服务细分为记录、查询与验证，具体为：记录服务接收 PROV 溯源文档并调用合约将哈希上链存证；查询服务根据数据标识检索链上交易并重组为可视化溯源链条；验证服务校验链条中各环节的数字签名。访问控制服务采用 CP-ABE，仅在用户属性满足策略时分发对应密钥，实现链下数据的细粒度解密授权。数据验证服务则自动计算文件哈希并与链上存证比对，返回数据是否被篡改的判定。

#### 3.4.6 应用层

面向不同角色，应用层提供定制化交互功能。科研人员可查询数据、溯源信息并验证完整性；患者/受试者通过授权与知情同意界面，清晰掌握数据的使用主体及目的；审查机构可获取全流程、不可篡改的审计追踪视图，对研究项目的数据使用合规性进行高效监督。

上述各层协同运行，共同保障医学科学数据的全生命周期可信管理，促进跨机构的高信度医学科研协作。

### 3.5 案例分析

以多中心肝细胞癌（Hepatocellular Carcinoma，HCC）影像-基因组关联研究为例，图 3 给出了模型在真实医学场景中的应用流程。

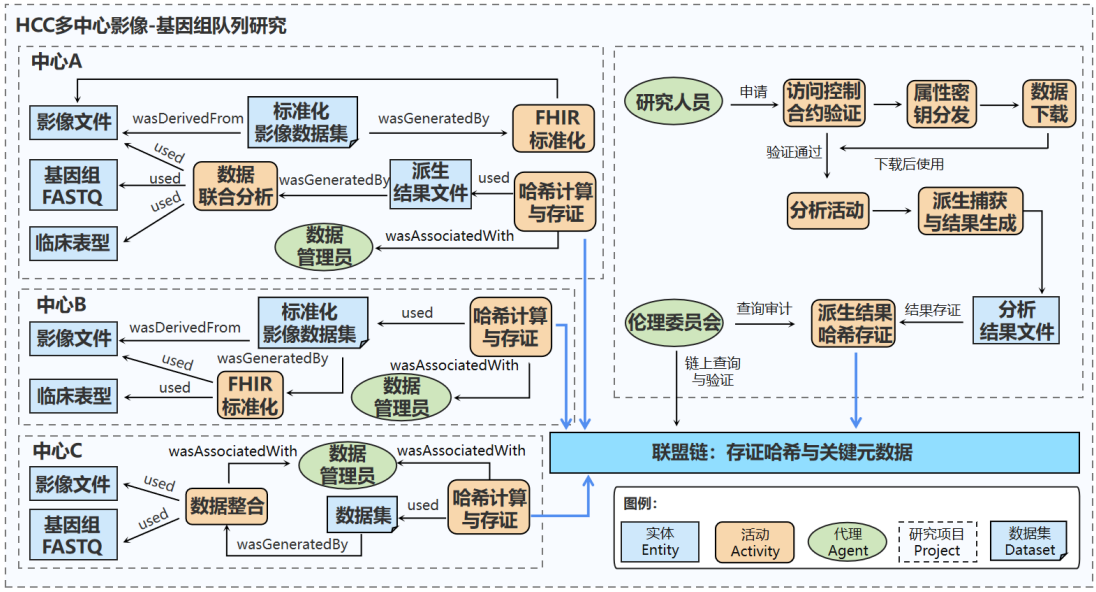


图 3 数据溯源模型应用流程

三家医疗中心分别将 HCC 患者的动态对比增强 CT/MRI 影像（DICOM 格式）、病理切片图像（SVS 格式）、全外显子测序 FASTQ 文件及临床病历（含 AFP 值、肿瘤分期等）等数据上传至本地数据处理层。系统依据 FHIR 标准将影像元数据映射为“ImagingStudy”资源，基因组变异映射为“MolecularSequence”资源，临床信息映射为“Patient”“Observation”资源。同时，PROV 扩展模型自动捕获每个文件的来源（采集设备型号、操作技师等）、生成时间、所属研究项目（“多中心 HCC 影像基因组学队列”），序列化为溯源文档。原始文件存储于各中心链下专用存储，哈希值与关键元数据写入联盟链。研究员 A 申请获取多中心影像与基因组数据时，系统根据其属性（所属机构为“肝癌研究联盟”成员、项目角色为“数据分析师”、已通过伦理审查）进行授权判断，验证通过后分发属性密钥。研究员 A 解密并下载数据后，进行影像组学特征提取，并与基因组变异数据进行关联分析，生成一份包含影像组学标签与基因突变状态的相关性矩阵的结果文件。系统自动捕获该派生数据的生成活动：原始 DICOM 影像及由测序数据处理得到的 VCF 变异文件作为输入实体，影像组学软件和 R 脚本作为代理，新生成的 CSV 文件作为输出实体，并建立“wasDerivedFrom”关系，新文件的哈希值再次上链。伦理委员会成员 B 通过项目 ID 查询链上记录和可视化溯源图谱，核验数据访问、分析派生和结果文件完整性，并逐一验证每个文件的哈希值是否与链上记录一致。若某中心声称其数据未被使用链上记录可提供客观证据。

综上，本研究所构建的溯源模型可有效支持多中心、多模态、隐私敏感的 HCC 研究数据全流程可信溯源，保障医学科研的可重复性与合规审计。

## 4 结语

针对医学科学数据跨机构协作中的核心挑战，本研究构建了以联盟链为信任基石、以 PROV 模型为溯源语义框架、以 FHIR 标准为数据规范化基础的可信溯源体系。通过分层架构设计，形成了覆盖数据标准化预处理、智能合约自动化执行、跨链互操作的全流程方案所提出的模型与技术路径可为医学科学数据的标准化治理、可信存证、权限管理与跨机构协作提供思路参考。

随着区块链技术与人工智能、大数据等新质生产力要素的深度融合，医学科学数据溯源将在以下方面发挥关键作用：一是各研究机构可利用本地私有数据训练模型，仅将加密后的模型参数在链上进行聚合与更新，实现“数据不动模型动”的可信协作，在保障隐私合规的同时，确保 AI 模型训练历程与数据来源可追溯、可验证；二是基于区块链的溯源模型有望成为支撑人工智能驱动的科学发现（AI4S）下一代科研范式的基础设施，通过将科研全流程中的关键步骤，以不可篡改、机器可读的形式记录为“数字研究见证”，助力构建可信、高效的医学科学研究生态。

**作者贡献：**乔柏珏负责模型构建与论文撰写；方安负责系统技术架构设计；周易负责技术调研与分析；杨晨柳、陈凌云负责标准整合与流程设计；胡佳慧负责研究整体指导、论文审阅与修改。
**利益声明：**所有作者均声明不存在利益冲突。

# 参考文献

[1] 国家卫生健康委办公厅, 国家发展改革委办公厅, 工业和信息化部办公厅, 等. 关于促进和规范“人工智能+医疗卫生”应用发展的实施意见[EB/OL]. (2025-11-04)[2025-11-30]. https://www.nhc.gov.cn/guihuaxxs/c100133/202511/d1a42ae835c743b9b3e83ac0253c3e9f.shtml.

[2] 新华社. 中华人民共和国国民经济和社会发展第十五个五年规划纲要[EB/OL]. (2026-03-13)[2026-04-12]. https://www.gov.cn/yaowen/liebiao/202603/content_7062633.htm

[3] 覃惠迪, 张丁钒, 阳春, 等. 医学科学数据管理与共享现状及发展策略探究[J]. 图书馆, 2025(08): 56-64.

[4] 杨珑平, 刘泽坤, 杨俊涛, 等. 医学科学数据共享的理论逻辑与实践探索[J]. 医学信息学杂志, 2025, 46(06): 37-42.

[5] 张怡, 夏寒, 冯骏, 等. 医疗卫生数据开放利用的挑战与对策[J]. 健康发展与政策研究, 2025, 28(04): 393-400.

[6] Arbabi M S, Lal C, Veeraragavan N R, et al. A survey on blockchain for healthcare: challenges, benefits, and future directions[J]. IEEE Communications Surveys & Tutorials, 2023, 25(1): 386-424. doi:10.1109/COMST.2022.3224644.

[7] 邵鑫喆, 库涛, 贾岩峰, 等. 基于区块链的可信制造追溯系统研究[J]. 信息安全研究, 2025, 11(11): 46-56.

[8] 左敏, 刘泓辰, 汪颢懿, 等. 基于零信任机制的粮食溯源区块链访问控制模型[J]. 信息安全研究, 2024, 10(10): 944-951.

[9] Azaria A, Ekblaw A, Vieira T, et al. MedRec: using blockchain for medical data access and permission management[C]//2016 2nd International Conference on Open and Big Data (OBD). New York: IEEE, 2016: 25-30. doi:10.1109/OBD.2016.11.

[10] Peterson K, Deeduvanu R, Kanjamala P, et al. A blockchain-based approach to health information exchange networks[C]//NIST Workshop on Blockchain & Healthcare. Gaithersburg: National Institute of Standards and Technology, 2016.

[11] 陈嘉莉, 马自强, 兰亚杰, 等. 基于区块链技术的医疗信息共享研究综述[J]. 计算机应用研究, 2024, 41(09): 2573-2584. doi:10.19734/j.issn.1001-3695.2023.12.0620.

[12] Myrzashova R, Alsamhi S H, Shvetsov A V, et al. Blockchain meets federated learning in healthcare: a systematic review with challenges and opportunities[J]. IEEE Internet of Things Journal, 2023, 10(16): 14418-14437.

[13] Verma P, Tripathi V, Pant B. MRDACE: an intelligent architecture for secure sharing and traceability of the medical images and patients' records[J]. ACM Transactions on Computing for Healthcare, 2025, 6(3): Article 31.

10.12201/bmr.202605.00089V1

[14] Zhao Z, Yu X, Ma Z. Attribute encryption based blockchain electronic medical record traceability method[C]//Proceedings of the 2024 3rd International Conference on Cryptography, Network Security and Communication Technology (CNSCT). New York: ACM, 2024: 17-24. doi:10.1145/3673277.3673281.

[15] Wu J, Fu Y, Li G. Medical device traceability management scheme based on blockchain[C]//2025 28th International Conference on Computer Supported Cooperative Work in Design (CSCWD). New York: IEEE, 2025: 98-103. doi:10.1109/CSCWD64889.2025.11033298.

[16] Abidin A A P, Alamsyah A, Irawan H. Blockchain for traceability and security in pharmaceutical supply chain[C]//2024 IEEE International Conference on Industry 4.0, Artificial Intelligence, and Communications Technology (IAICT). New York: IEEE, 2024: 235-240. doi:10.1109/IAICT62357.2024.10617533.

[17] Gao Y, Lu G, Wang Z, et al. Research on medical image data sharing traceability system based on blockchain[C]//2023 5th International Conference on Decision Science & Management (ICDSM). New York: IEEE, 2023: 113-116. doi:10.1109/ICDSM59373.2023.00032.

[18] Nakamoto S. Bitcoin: a peer-to-peer electronic cash system[EB/OL]. (2008-10-31) [2025-11-30]. https://bitcoin.org/bitcoin.pdf.

[19] 袁勇, 王飞跃. 区块链技术发展现状与展望[J]. 自动化学报, 2016, 42(04): 481-494. doi:10.16383/j.aas.2016.c160158.

[20] 张宁熙. 区块链技术发展综述及其政务领域应用研究[J]. 信息安全研究, 2020, 6(10): 910-918.

[21] 冯凯璇. 基于区块链的溯源模型及检索方法研究[D]. 杭州: 浙江科技大学, 2024. doi:10.27840/d.cnki.gzjkj.2024.000624.

[22] Hyperledger Cacti Contributors. Using and Developing with Hyperledger Cacti[EB/OL]. (2026-04-13) [2026-04-13]. https://hyperledger-cacti.github.io/cacti/.