

基于多模态数据的儿童川崎病专病平台建设与应用

杨恺熠^{1,2,3} 赵永根^{1,2,3} 耿志敏⁴ 高立超⁴ 胡 坚⁴ 齐国强^{1,2,3}

¹ 浙江大学医学院附属儿童医院数据信息部 杭州 310052

² 浙江-芬兰儿童健康人工智能联合实验室 杭州 310052

³ 国家儿童健康与疾病临床医学研究中心(儿童健康医工信交叉研究中心) 杭州 310052

⁴ 浙江大学医学院附属儿童医院心血管科 杭州 310052)

[摘要] **目的/意义** 基于多模态数据构建川崎病专病平台，支持临床诊疗优化与科研探索。**方法/过程** 以浙江大学医学院附属儿童医院 2019—2024 年川崎病患儿诊疗数据为基础，采用抽取、转换和加载技术汇聚多源异构数据，利用自然语言处理、计算机视觉和数据治理技术进行数据清洗、标准化、结构化，建立专病数据库。在此基础上，构建集成患儿管理、智能搜索、科研项目与随访管理等功能模块的专病平台。

结果/结论 建成含 10 689 例确诊患儿、覆盖 18 个模块共 300 个数据项的川崎病专病库，可为临床诊疗和病程管理提供有力支持，为开展大样本临床研究和科研创新奠定基础。

[关键词] 儿童川崎病；专病平台；多模态数据；数据质控

[中图分类号] R -058 **[文献标识码]** A **[DOI]** 10.3969/j.issn.1673-6036.2025.12.013

Construction and Application of a Pediatric Kawasaki Disease Specialized Platform Based on Multimodal Data

YANG Kaiyi^{1,2,3}, ZHAO Yonggen^{1,2,3}, GENG Zhimin⁴, GAO Lichao⁴, HU Jian⁴, QI Guoqiang^{1,2,3}

¹ Department of Data and Information, Children's Hospital, Zhejiang University School of Medicine, Hangzhou 310052, China; ² Sino-Finland Joint AI Laboratory for Child Health of Zhejiang Province, Hangzhou 310052, China; ³ Pediatric Medicine Engineering and Information Research Center, National Clinical Research Center for Child Health, Hangzhou 310052, China; ⁴ Department of Pediatric Cardiology, Children's Hospital, Zhejiang University School of Medicine, Hangzhou 310052, China

[Abstract] **Purpose/Significance** To construct a specialized Kawasaki disease (KD) platform based on multimodal data, so as to support optimized clinical diagnosis and treatment as well as scientific research exploration. **Method/Process** Based on the diagnostic and treatment data of KD children in the Children's Hospital, Zhejiang University School of Medicine from 2019 to 2024, multi-source heterogeneous data are aggregated by using extraction, transformation and loading (ETL) technology. Data are cleaned, standardized and structured using natural language processing (NLP), computer vision and data governance techniques, a specialized disease database is established. On this basis, a specialized disease platform integrating functional modules such as patient management, intelligent search,

[修回日期] 2025-11-17

[作者简介] 杨恺熠，助理工程师，发表论文 1 篇；通信作者：齐国强，工程师。

[基金项目] 国家自然科学基金面上项目（项目编号：62576314）；浙江省“尖兵”“领雁”研发攻关计划项目（项目编号：2024C03180）。

research project and follow-up management is constructed. **Result/Conclusion** A specialized KD database containing 10 689 confirmed cases and 300 data items across 18 modules is established. The platform can provide strong support for clinical diagnosis and treatment as well as disease course management, laying a foundation for conducting large-sample clinical study and scientific research innovation.

[Keywords] pediatric Kawasaki disease; specialized disease platform; multimodal data; data quality control

1 引言

川崎病，亦称皮肤黏膜淋巴结综合征，是以全身性血管炎为特征的急性发热性疾病，主要影响5岁以下儿童^[1-2]。目前，医学界对川崎病的认知仍存在显著局限，其病因与长期预后尚未明确，且临床诊断与治疗方案也缺乏统一规范^[3]。近年来，全球川崎病发病率持续上升，中国也呈相似上升态势^[4]。尽管中国儿童川崎病治疗指南和共识不断更新，临床研究取得积极成果，但临床管理仍面临早期诊断困难、个性化治疗不足、信息获取滞后、多学科协作效率低等挑战^[5]。深入挖掘医疗数据潜力，建立基于多模态数据的川崎病专病平台，有助于提升儿童川崎病临床服务质量，并推动医疗领域创新。

目前，国内已建设儿童注意缺陷多动障碍^[6]、儿童哮喘^[7]等儿科疾病专病库。这些系统在一定程度上支撑了疾病监测与基础研究，但多数局限于结构化数据收集，缺乏对多模态医疗文本深度解析与智能关联能力，也未形成从临床检索到科研分析的一体化闭环。在知识图谱与人工智能（artificial intelligence, AI）辅助诊疗方面，虽有一些儿科常见疾病知识库问世，但专门面向川崎病、兼具临床与科研功能的系统平台仍较匮乏。

浙江大学医学院附属儿童医院（以下简称浙大儿院）近6年收治川崎病患儿10 689人，在儿童川崎病诊治方面学科优势明显。得益于近年医院信息化建设，诊治患儿同时积累了大量诊疗信息，涵盖患儿人口学信息、门诊病历、住院病历、检验检查报告、护理记录、随访记录，为专病库构建和研究提供了良好数据支持。因此，本研究基于大数据技术，汇聚、结构化与标准化处理海量川崎病诊疗数据，构建集专病数据库、智能查询及AI分析于一体的川崎病专病平台，旨在辅助临床诊疗、规范病

程管理、助力科研探索。

2 建设方案

2.1 系统架构

儿童川崎病专病平台以浙大儿院川崎病患儿临床诊断数据为基础，综合考虑临床诊疗与科研需求，基于Linux操作系统，运用Kubernetes和Docker虚拟化技术构建大规模Hadoop集群，自主研发运维管理平台，自动监控平台各组件问题，为医疗数据处理深度调优、提高效率。其总体架构，见图1。数据汇聚层利用数据仓库技术，从分散业务系统抽取原始诊疗数据。数据治理层利用自然语言处理、计算机视觉等技术，对跨媒体、多模态数据进行标准化清洗和结构化治理。数据管理层，经质控合格的数据正式入库，构建以患儿主索引（enterprise master patient index, EMPI）为核心的川崎病专病数据库。数据应用层涵盖平台管理、患者管理、项目管理等功能模块，支持临床医生与科研人员进行高效数据检索、分析及导出等操作。

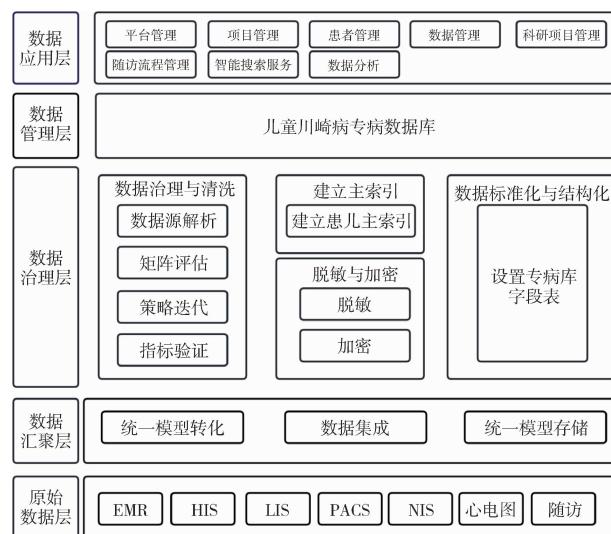


图1 川崎病专病平台系统架构

2.2 数据采集与整合

纳入 2019 年 1 月 1 日—2024 年 12 月 31 日经临床确诊（ICD - 10 代码：M30.3）的川崎病患儿，建立研究队列，追溯整合其全周期诊疗数据，涵盖初诊门诊、住院治疗、随访复诊及并发症管理等临床轨迹。

数据标准化方面，严格遵循 ICD - 10 和 ICD - 9 - CM - 3 标准对疾病诊断与手术操作编码，检验检查项目基于 LOINC 术语体系，依托 SNOMED CT 构建临床症状标准化表述，药品数据基于国家药品编码与 ATC 分类体系进行标准化整合。上述融合了疾病、手术、检验、症状与药品等多维度标准的术语治理体系，保障了多源临床数据语义一致性与结构规范性，为专病数据资源深度挖掘与多中心研究奠定基础。

数据处理与集成方面，基于抽取、转换和加载（extraction - transformation - loading, ETL）技术，从电子病历（electronic medical record, EMR）系统、医院信息系统（hospital information system,

HIS）、检验信息系统（laboratory information system, LIS）、影像存储与传输系统（picture archiving and communication system, PACS）及护理信息系统（nursing information system, NIS）等临床业务系统抽取诊疗数据，以保障数据信息完整性。

2.3 数据治理

2.3.1 数据清洗 采用 5 阶段标准化清洗框架^[8]。第 1 阶段判别数据源，解析多源数据结构与接口规范。第 2 阶段构建可用性评估矩阵，该矩阵包括两个核心维度：一是纵向质量扫描，以量化关键字段的质量缺陷；二是横向关联分析，以追溯单次诊疗全流程的数据逻辑链。第 3、4 阶段依据评估结果动态制定并迭代优化清洗策略。第 5 阶段通过预设的完整性指标验证第 2 阶段结果。清洗后，就诊数据为 196 681 人次（门诊 184 291 人次，住院 12 390 人次），确诊患儿共 10 689 例。数据清洗流程，见图 2。该结构化流程实现了从数据质控到价值挖掘的方法论闭环。

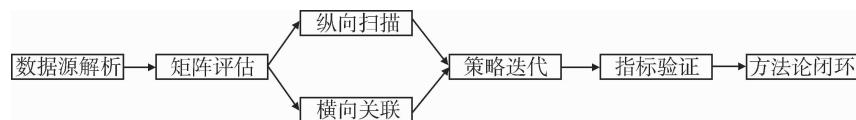


图 2 数据清洗流程

2.3.2 建立患儿主索引 采用患儿主索引技术，准确识别患儿身份，整合同一患儿不同时间的就诊记录（如多次门诊或住院），关联其同次就诊分散在不同业务系统（如门诊、住院、检验、护理等）中的数据，并按类别整理，实现跨系统信息检索和共享^[9]。例如，汇集同一住院患儿住院前的门诊症状描述与治疗情况，为住院治疗方案的制定提供关键依据，同时帮助医生全面掌握病情发展历程。

2.3.3 数据脱敏与加密 为保护患儿隐私，严格遵守中国《个人信息保护法》和美国《健康保险携带和责任法案》等法律法规，对患儿信息中 18 类敏感内容（如姓名、身份证号、电话、家庭住址等），采用专门脱敏和加密技术（如 AES 等国际通用算法或我国商用密码算法）处理^[10-11]。脱敏是

将敏感信息转化为无法直接识别的形式（如用 *** 代替部分数字），加密是为数据加上“密码锁”。为数据库中每个敏感数据字段设置具体脱敏规则，仅授权工作人员设置或修改。

2.3.4 数据标准化与结构化 从各业务系统收集大量原始数据，针对川崎病特点详细整理，重点处理心脏超声检查结果、冠状动脉病变情况等关键信息。邀请心内科临床专家参与，结合医院实际需要和现有数据情况，参考行业通用术语标准及最新《中国儿童川崎病诊疗循证指南（2023 年）》^[12]，统一命名和规范整理川崎病关键信息。最终确定川崎病专病数据库详细结构，含 18 个主要类别、300 个具体数据项，见表 1。在操作层面，由心内科医生逐一明确每个数据项的含义、格式、取值范围与内容限制等规则。

表 1 川崎病专病库结构

序号	模块	字段数量(个)
1	患者人口学信息	36
2	就诊信息(就诊基本信息、个人信息)	16
3	诊断	11
4	病程记录	9
5	生命体征	11
6	门急诊及入院记录(门急诊及入院记录、主诉、现病史、个人史、既往史、家族史、体格检查、专科查体)	69
7	检验报告	13
8	心电图检查	4
9	普放检查	10
10	超声检查	30
11	住院用药	24
12	非药品医嘱	10
13	门诊处方	21
14	诊疗概要	7
15	出院记录	5
16	生存结局	6
17	微生物报告	8
18	药敏试验	10

2.4 数据质控

采用3级质控保障原始数据质量。第1级为自动化校验，以患儿主索引为唯一标识，扫描川崎病关键字段(如冠状动脉病变指标)完整性，实时标记缺失、异常数据，由规则引擎修正或人工补全。第2级利用智能算法校正逻辑错误，验证诊断、治疗及随访数据跨系统一致性(如住院与门诊用药冲突)。第3级为人工复核，随机抽取10%病例溯源至原始病历核验，确保数据可靠可用^[13]。

2.5 多模态数据整合

在数据管理层，设计系统化多模态数据整合技术路径，汇聚分散的文本、影像与结构化数据。文本信息结构化提取方面，先训练川崎病专用命名实体识别模型，精准抽取出院小结关键体征、治疗反应和病变特征，自动转化为标准化字段填入电子病例报告表(case report form, CRF)。影像学特征量化方面，采用基于注意力机制的视觉-文本联合模型，通过对学习实现心脏超声影像与报告的精准映射，自动测量冠状动脉内径、计算Z值(标准差值)，识别冠状动脉瘤等关键特

征^[14]。跨模态数据融合方面，依托EMPI和就诊时间轴，对齐提取的影像学特征、文本信息和实验室检验结果，构建融合影像、文本与时序3维度的科研数据集。

3 系统功能

3.1 科研流程管理

平台支持线上科研项目全流程管理，助力临床研究高效实施，见图3。核心功能包括：一是项目构建，自定义研究目的、设计方案与受试者纳排标准；二是数据治理，智能计算关键指标，实现全周期随访管理；三是质控分析，多维度数据核查与统计分析；四是成果输出，可视化呈现统计结果。此外，构建多模态科研支持体系，兼容回顾性与前瞻性研究设计，提供对电子病历文本、医学影像、实验室检验时序数据等多模态数据导入导出、批量编辑、预审核及深度分析功能。将临床诊疗产生的多源数据(如电子表单、影像报告等)同步录入专病数据库，实现数据资产化整合与复用。

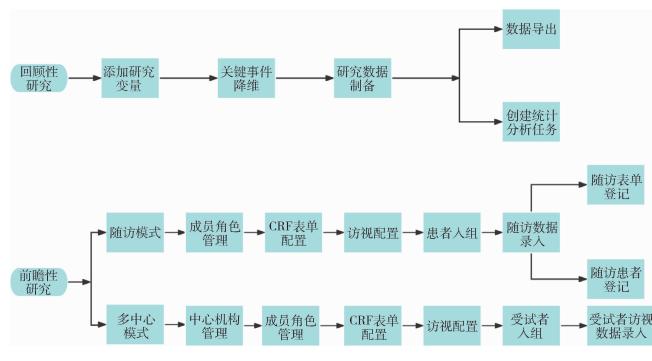


图3 科研项目管理流程

3.2 随访流程管理

通过随访流程管理实现科研数据动态闭环管理，构建川崎病研究全周期随访管理体系。该功能支持按项目创建任务，配置专属人员、时间窗口及患儿队列规模，基于可视化面板实时追踪随访进度(如成功率、失访率、冠状动脉病变转归率、IVIG维持治疗脱落率)。可针对单项目独立设计随访流程，关联定制化表单，预设川崎病分期特异随访策略(如急性

期症状追踪、恢复期冠脉超声复查)。CRF 定制模块提供可配置的电子病例报告表，兼容前瞻性研究设计。表单字段聚焦川崎病核心指标，如冠脉 Z 值动态变化、IVIG 不良反应事件，确保多中心数据标准化采集。随访计划与执行功能依据川崎病临床分期(急性期/亚急性期/恢复期)^[15]，批量生成患儿随访日历并支持灵活配置。可设定随访频次(周/月/季度)、启动条件(入组即时启动/出院后时间点触发/手动指定)、终止规则(计划完成/患儿死亡/出院满5年)；生成每日随访清单并推送提醒，高效录入专项数据(如心脏超声复查结果)；实时汇总分析随访数据，支持动态评估病情进展、优化治疗方案。系统自动生成随访报告，便于多学科团队协作，提升川崎病管理水平。还能内置指南推荐方案(如《中国儿童川崎病诊疗循证指南(2023)》冠脉随访计划)，加速研究启动。随访数据实时同步至专病库，为治疗疗效评估(IVIG 耐药性分析)及长期预后研究(冠状动脉瘤转归轨迹)提供结构化支撑。

3.3 患者管理

提供个案级全息诊疗档案探查与专科定制化数据管理功能。数据分析检测异常时，用户可直接访问集成化医疗档案，查看电子病历、检验报告、费用清单及随访记录等多维度数据；点击 AI 提取的疾病特征(如冠状动脉病变分期)，自动溯源至 HIS/PACS 原始数据，动态可视化川崎病患儿初诊至随访全周期诊疗轨迹。针对专科科研需求，支持灵活配置川崎病专属指标(如冠脉病变分级、IVIG 耐药预测值)并自定义分组规则。基于患儿 360 视图整合多源数据(含电子病历、冠脉超声、用药记录及随访结果)，实现智能特征与原始记录关联定位(如点击冠脉瘤风险评分定位至对应检验单)。发现统计异常(如耐药患儿聚集)时，即时穿透至病案详情核查病程日志，快速完成根因分析。该设计打通了从宏观分析到微观验证的全链路，承载与驱动川崎病研究“趋势发现-个案验证-方案优化”全流程。

3.4 智能搜索服务

基于人工智能的智能搜索服务，为川崎病研究

提供全面高效智能化支持。该服务依托高性能分布式处理架构，具备 TB 级医疗数据实时(秒级)检索能力，支持全文检索、数值范围筛选与层级关联查询等多种检索方式。内置医疗语义理解引擎，能自动识别并关联川崎病多种表达方式，结合智能纠错机制，显著提升目标病例召回精度。在病例筛选场景，科研人员可通过定向字段搜索功能，快速定位主诉、现病史等关键信息。以疾病树形式清晰展示川崎病不同分期及相关并发症分布情况。可构建包含发热天数、草莓舌表现、冠状动脉扩张等特征的多层次条件组合，结合事件序列分析，追踪 IVIG 治疗后冠脉变化等时间线信息，实现漏斗式病例筛选，并实时查看样本量，辅助科研判断。该服务进一步实现了科研流程的闭环管理。筛选结果可生成患儿列表并一键溯源至原始病历，结构化数据也能便捷导出至专业统计工具。系统内置描述性统计功能，支持用户自定义纳排标准，并允许将常用筛选条件保存至科研项目中，从而贯通从病例筛选、数据导出到统计分析的全流程，整体提升川崎病研究的效率与规范性。

4 应用效果

4.1 患病人群基本特征

截至 2024 年 12 月，川崎病专病库累计纳入就诊患儿 196 681 人次(门诊 184 291 人次，住院 12 390 人次)，确诊患儿 10 689 例。性别分布：男性患儿 6 463 例(60.46%)，女性患儿 4 226 例(39.54%)。就诊科室集中于内科(72.1%)、心血管科(18.3%)及发热门诊(9.6%)，高频症状前 3 位为发热(89.2%)、咳嗽(76.8%)及皮疹(63.4%)。地域分布：96.0% 患儿来自浙江省，来自杭州市的患儿占比 54.72%。

4.2 专病平台对科研的支持

临床研究人员可通过该平台开展回顾性与前瞻性队列研究，直接提取患儿人口学特征、临床基线资料及多维度诊疗数据(含实验室检验、影像学报告、用药记录等)，驱动川崎病药物疗效预测模型

构建与疾病发生机制解析。平台深度集成支持向量机、XGBoost、随机森林等主流机器学习算法，支持从特征工程、模型训练到验证评估的全流程闭环，支撑高效科研；并推动临床实现从数据挖掘到机制研究的转化医学探索。

4.3 数据管理与应用效能升级

基于大数据技术构建川崎病专病库，实现了多源诊疗数据智能汇聚与结构化整合。平台启用前，临床统计需信息中心跨系统检索数据，手动整理至Excel表再统计分析，常出现重复建库申报，效率较低。平台上线后，临床研究者可在统一的数据池中自主配置查询条件，通过差异化纳排标准快速构建研究队列，规避多库查询风险。该模式既保障了生产库数据安全性，又显著提升了数据利用效率与科研敏捷性。

5 结语

智能多模态专病数据库建设是医院构筑高质量科研级数据资产的关键路径^[16]。本研究成功构建集成多模态数据的川崎病专病数据库与智能科研平台，通过医学自然语言处理引擎，实现从临床数据到科研证据高效转化，为开展大样本临床队列研究、优化诊疗路径提供了数据基础。目前平台已支持回顾性与前瞻性专项课题研究，有效促进研究成果向临床实践转化。然而，本研究仍存在局限：数据标准化主要基于院内体系，跨中心数据协同能力有待提升；预测模型构建与多中心验证有待完善。后续研究将推进区域多中心数据标准统一，扩展动态随访数据维度，深化人工智能在预后预测与个性化诊疗方案推荐中的应用，提升平台科研价值与临床支撑能力，为儿童川崎病诊疗优化和学科发展提供持续动力。

作者贡献：杨恺熠负责研究设计、数据收集与整理、论文撰写与修订；赵永根负责数据整理与分析；耿志敏、高立超、胡坚负责临床指导、数据质控；齐国强负责论文审核。

利益声明：所有作者均声明不存在利益冲突。

参考文献

- 姚固先. 儿童川崎病临床疗效浅析 [J]. 贵州医药, 2024, 48 (9): 1407–1409.
- KAWASAKI T. Acute febrile mucocutaneous syndrome with lymphoid involvement with specific desquamation of the fingers and toes in children [J]. Arerugi, 1967, 16 (3): 178–222.
- 张清友, 杜军保. 川崎病诊断中的若干问题 [J]. 中华实用儿科临床杂志, 2020, 35 (13): 961–964.
- JIAO F, PAN Y, DU Z, et al. Guideline for the diagnosis and treatment of incomplete Kawasaki disease in children in China [J]. BMC pediatrics, 2024, 24 (1): 477.
- 潘炎, 焦富勇. 中国儿童川崎病管理的DeepSeek 视角 [J]. 中国当代儿科杂志, 2025, 27 (5): 524–528.
- 陆昕玥, 黄坚, 陈凌栋, 等. 儿童注意缺陷多动障碍专病库构建与应用 [J]. 医学信息学杂志, 2024, 45 (9): 75–81.
- 刘芳, 张玉蓉, 李小玲, 等. 儿童哮喘专病库的构建与应用研究 [J]. 护理学杂志, 2022, 37 (19): 25–28.
- 王耀国, 李鹏, 刘迷迷, 等. 临床专病数据库建设现状与思考 [J]. 医学信息学杂志, 2024, 45 (3): 65–69.
- 罗浩, 李华康, 王飞. 基于多模态数据模型的心血管疾病专病库构建与应用 [J]. 中国医疗设备, 2025, 40 (6): 61–68.
- 高玉宝. 《中华人民共和国个人信息保护法》背景下的大数据发展 [J]. 图书馆理论与实践, 2022 (4): 4–11.
- 连斐. 云计算时代下数字医院的信息安全 [J]. 网络安全技术与应用, 2024 (5): 99–101.
- 黄玉娟, 李夙凌, 田园, 等. 中国儿童川崎病诊疗循证指南(2023年) [J]. 中国当代儿科杂志, 2023, 25 (12): 1198–1210.
- 武亚琴, 卢晟晔, 赵慧颖, 等. 医院重症医学专病库的建设与应用 [J]. 医院管理论坛, 2023, 40 (3): 93–96.
- 齐恩林, 周悦, 赵静, 等. 川崎病患儿的临床特征、实验室指标及冠状动脉内径Z值变化分析 [J]. 医学研究杂志, 2025, 54 (5): 118–123.
- 张媛媛, 赵艳君, 洪霞, 等. 川崎病儿童健康管理专家共识 [J]. 上海医学, 2024, 47 (7): 389–397.
- 李哲明, 朱珠, 黄坚, 等. 面向科研专病库的可视化前置数据采集系统建设探索与实践 [J]. 中国卫生信息管理杂志, 2024, 21 (1): 69–74.