

# Research on the Application of Big Data-Based Electronic Information Technology in Medical Diagnosis

Jiaping Li

The Second People's Hospital of Lianyungang City, Lianyungang, Jiangsu, 222002, China

## Abstract

With the rapid advancement of information technology, the application of big data in the field of medical diagnosis has become increasingly widespread and profound. This paper aims to explore the core technologies of big data-based electronic information technology in medical diagnosis, including the collection and integration of medical big data, storage and management, analysis and mining, as well as data security and privacy protection. Additionally, the article provides a detailed discussion on the practical implementation and effectiveness of these technologies in specific application scenarios such as medical imaging-assisted diagnosis, disease risk prediction and early screening, clinical decision support systems, personalized medical diagnosis and treatment optimization, and the diagnosis of rare and complex diseases. Finally, the paper outlines future development trends, aiming to provide insights for advancing the innovation and progress of medical diagnostic technologies.

## Keywords

big data; Electronic information technology; Medical diagnosis; application

## 基于大数据的电子信息技术在医疗诊断中的应用研究

李家靖

连云港市第二人民医院, 中国·江苏 连云港 222002

## 摘要

随着信息技术的飞速发展, 大数据在医疗诊断领域的应用日益广泛且深入。本文旨在探讨基于大数据的电子信息技术在医疗诊断中的核心技术, 包括医疗大数据采集与整合、存储与管理、分析与挖掘以及数据安全与隐私保护等方面。同时, 文章详细阐述了这些技术在医学影像辅助诊断、疾病风险预测与早期筛查、临床决策支持系统、个性化医疗诊断与治疗方案优化以及罕见病与疑难杂症诊断等具体应用场景中的实践与成效。最后, 文章展望了未来发展趋势, 以期为推动医疗诊断技术的创新与发展提供参考。

## 关键词

大数据; 电子信息技术; 医疗诊断; 应用

## 1 引言

随着大数据时代的来临, 海量的医疗数据不断涌现, 这些数据蕴含着丰富的信息, 为医疗诊断带来了新的机遇和挑战。基于大数据的电子信息技术凭借其强大的数据处理和分析能力, 逐渐在医疗诊断中崭露头角, 成为推动医疗诊断技术进步的重要力量。深入研究基于大数据的电子信息技术在医疗诊断中的应用, 对于提高医疗诊断的准确性、效率以及个性化程度, 改善患者的就医体验和健康状况, 具有重要的现实意义和广阔的发展前景。

## 2 基于大数据的电子信息技术在医疗诊断中的核心技术

### 2.1 医疗大数据采集与整合技术

医疗数据来源广泛且类型多样, 包括电子健康档案 (EHR)、医学影像 (如 PACS 系统中的 DR、CT、MRI、

超声)、实验室检验数据 (LIS)、病理报告、临床用药记录、可穿戴设备产生的生理数据 (如心率、血压、血糖)、基因组学数据、医疗物联网设备实时监测数据以及患者行为和社交媒体数据等。电子信息技术通过标准化的数据接口、互操作性协议 (如 HL7/FHIR)、物联网感知技术、云计算平台等, 实现对这些多模态、异构数据的高效采集。进一步, 通过数据清洗 (去除噪声、填补缺失值)、数据转换 (统一格式和编码标准)、数据融合 (关联不同来源数据) 等技术, 将分散的医疗数据整合到统一的数据仓库或数据湖中, 为后续的分析应用奠定基础。例如, 通过整合患者的历次就诊记录、影像资料和基因数据, 可以构建全面的患者健康画像。

### 2.2 大数据存储与管理技术

医疗数据具有海量性、高增长率和长期保存的需求。传统的关系型数据库在处理大规模非结构化和半结构化医疗数据时面临挑战。基于大数据的电子信息技术采用分布

式文件系统（如 Hadoop Distributed File System, HDFS）、NoSQL 数据库（如 MongoDB 用于存储非结构化病历，Cassandra 用于高可用场景）、NewSQL 数据库以及云存储服务，提供高容量、高可靠性、高扩展性和低成本的数据存储解决方案。同时，结合数据生命周期管理策略，对不同重要性和访问频率的数据进行分级存储和管理，确保数据的安全性、完整性和可访问性。例如，利用 Hadoop 生态系统存储 PB 级别的医学影像数据和基因测序数据，并通过 YARN 进行资源调度管理。

## 2.3 大数据分析 with 挖掘技术

大数据分析 with 挖掘技术主要包括：

(1) 统计分析与机器学习：运用描述性统计、推断性统计方法对数据进行初步探索。更重要的是，利用机器学习算法，如监督学习（支持向量机 SVM、决策树、随机森林、逻辑回归、深度神经网络）用于疾病预测、风险评估和影像识别；无监督学习（聚类分析如 K-Means、关联规则挖掘如 Apriori 算法）用于发现疾病亚型、未知的疾病关联因素；半监督学习和强化学习也在特定医疗场景中得到应用。例如，使用随机森林算法分析患者的临床指标和生活习惯数据，预测其患糖尿病的风险。

(2) 深度学习技术：尤其在医学影像诊断方面取得了突破性进展。卷积神经网络（CNN）能够自动提取医学影像中的关键特征，实现对肿瘤、骨折、眼底病变等的精准检测、分割和分类。循环神经网络（RNN）和 Transformer 模型则适用于处理时序医疗数据（如心电图信号、ICU 监测数据）和自然语言处理任务（如病历文本分析、临床笔记挖掘）。例如，基于 CNN 的模型在肺结节 CT 影像检测中的准确率已接近甚至超越资深放射科医师。

(3) 自然语言处理（NLP）技术：医疗领域存在大量非结构化文本数据，如病历、病程记录、文献资料等。NLP 技术通过分词、命名实体识别（识别疾病、症状、药物、检查等实体）、关系抽取（提取实体间关系，如“疾病—症状”关系）、情感分析、文本分类和信息抽取等，将非结构化文本转化为结构化数据，从中挖掘有价值的诊断线索和知识。例如，从海量病历中自动抽取特定疾病的临床表征和治疗方案，辅助医生决策。

## 2.4 数据安全和隐私保护技术

医疗数据包含大量个人敏感信息，其安全与隐私保护至关重要。基于大数据的电子信息技术采用数据加密（传输加密如 SSL/TLS、存储加密如 AES）、访问控制（基于角色的访问控制 RBAC、基于属性的访问控制 ABAC）、数据脱敏（去标识化、匿名化处理，如替换或删除患者身份信息）、隐私计算（如联邦学习允许数据在本地处理，仅共享模型参数；安全多方计算；差分隐私在数据发布时添加噪声以保护个体隐私）等技术，在充分利用数据价值的同时，严格遵守相关法律法规（如 HIPAA、GDPR 及我国的《个人信息保护法》

《数据安全法》），防止数据泄露和滥用。例如，多家医院可以通过联邦学习联合训练疾病预测模型，而无需共享原始患者数据。

## 3 基于大数据的电子信息技术在医疗诊断中的具体应用场景

### 3.1 医学影像辅助诊断

医学影像是医疗诊断的重要依据，基于大数据和深度学习的电子信息技术在该领域应用尤为成熟。通过构建包含大量标注影像数据的训练集，训练深度学习模型（如 CNN），能够实现对 X 光片、CT、MRI、超声、病理切片等影像的自动分析。具体应用包括：肺结节、肝癌、乳腺癌等肿瘤的早期筛查与良恶性鉴别；眼底照片中糖尿病视网膜病变、青光眼等的分级诊断；脑影像中脑卒中病灶的快速定位与体积测算；骨龄评估、骨折检测等。这些辅助诊断系统能够帮助医生提高诊断速度和准确性，尤其在基层医疗机构和医疗资源匮乏地区，可弥补专业影像医师不足的短板，实现诊断能力的下沉。例如，AI 辅助诊断系统能在 CT 影像中快速标记出微小肺结节，提醒医生重点关注。

### 3.2 疾病风险预测与早期筛查

利用大数据分析技术，对海量人群的健康数据、生活行为数据、环境数据、遗传数据等进行建模分析，可以识别疾病的危险因素和预警指标，构建疾病风险预测模型。这些模型能够对个体未来患特定疾病（如心血管疾病、糖尿病、阿尔茨海默病、某些癌症）的风险进行评估。例如，通过分析居民的体检数据（血压、血脂、血糖）、饮食习惯、运动情况、吸烟史以及家族病史等，预测其未来 5 年内患高血压的概率。结合风险预测结果，可以对高风险人群进行早期干预和精准筛查，实现疾病的早发现、早诊断、早治疗，从而降低疾病发生率和死亡率，节约医疗成本。

### 3.3 临床决策支持系统（CDSS）

基于大数据的 CDSS 能够整合患者的实时临床数据（如生命体征、检验结果、用药情况）、历史病历数据、医学文献、临床指南等信息，通过自然语言处理、知识图谱构建和推理引擎等技术，为医生提供实时的、个性化的诊疗建议。例如，当医生开具处方时，系统可以自动检查药物相互作用、剂量合理性，并基于患者的过敏史进行预警；在诊断过程中，系统可以根据患者的症状和检查结果，提供可能的诊断列表和鉴别诊断依据，并推荐进一步的检查项目。这有助于减少医疗差错，规范医疗行为，提高诊疗的科学性和一致性，尤其对年轻医生和基层医生具有重要的指导作用。

### 3.4 个性化医疗诊断与治疗方案优化

每个人的基因背景、生活环境和健康状况存在差异，基于大数据的电子信息技术能够推动医疗诊断向个性化方向发展。通过整合患者的基因组数据、蛋白质组数据、代谢组数据与临床表型数据，利用生物信息学和机器学习方法，深入理解疾病的分子机制和个体差异。在诊断方面，可以实

现疾病的精准分型,例如对癌症进行分子分型,明确驱动基因,从而进行更精确的诊断。在治疗方面,可以基于患者的基因特征预测药物疗效和不良反应,辅助医生选择最适合患者的治疗药物和剂量,实现“量体裁衣”式的精准治疗。例如,在肿瘤治疗中,通过基因测序数据和大数据分析,为患者匹配最有效的靶向药物。同时,根据患者的代谢特征和药物反应数据,动态调整治疗方案,提高治疗的有效性和安全性。对于慢性疾病患者,系统能够持续跟踪其健康数据,包括血糖、血压、血脂等指标,结合大数据分析,为患者提供个性化的健康管理建议,如饮食调整、运动计划等,有效控制疾病进展,提升患者生活质量。此外,个性化医疗还体现在预防医学领域,通过分析个体的遗传信息、生活习惯、环境因素等,预测个体患病风险,提前采取干预措施,实现疾病的早期预防 and 健康管理。

### 3.5 罕见病与疑难杂症诊断

罕见病由于病例稀少、临床表现复杂多样,诊断难度极大,容易出现误诊和漏诊。基于大数据的电子信息技术通过收集和整合全球范围内的罕见病病例数据、基因数据和文献资料,构建罕见病知识库和诊断模型。当遇到疑似罕见病患者时,系统可以根据患者的症状、体征、实验室检查结果等,与知识库中的病例进行比对和匹配,快速缩小诊断范围,为医生提供潜在的诊断方向和依据。例如,利用自然语言处理技术分析患者的复杂病历文本,并与国际罕见病数据库连接,辅助医生识别出罕见的遗传性疾病。例如,针对一些症状不典型、病因复杂的疑难杂症,系统能够综合多源数据,运用机器学习算法挖掘潜在关联,为医生提供多种可能的诊断假设。同时,通过大数据分析还可以发现不同罕见病之间的相似症状模式,帮助医生在面对未知病症时,从相似病症的诊断思路入手,逐步排查,提高诊断的准确性和效率。此外,基于大数据的电子信息技术还能实时跟踪罕见病和疑难杂症的最新研究成果与临床案例,及时更新知识库,为医生提供最前沿的诊断参考。

## 4 未来发展趋势

### 4.1 边缘计算与实时诊断的结合

随着可穿戴设备、便携式诊断设备和物联网技术的普

及,医疗数据采集将更加实时和泛在。结合边缘计算技术,在数据产生的本地进行实时分析和快速响应,能够实现对慢性患者的持续监测、急性病症的早期预警和紧急情况下的快速诊断支持,提升医疗的及时性和连续性。

### 4.2 联邦学习等隐私计算技术的广泛应用

为解决数据孤岛和隐私保护的矛盾,联邦学习、安全多方计算、可信执行环境等隐私计算技术将在医疗诊断领域得到更广泛的应用。未来将形成跨机构、跨区域甚至跨国界的联邦学习模型训练网络,在不共享原始数据的前提下,共同提升诊断模型的性能,推动 AI 诊断技术的普惠发展。

### 4.3 智能化、个性化、全程化的健康管理及诊断

基于大数据和人工智能的医疗诊断将从疾病诊断向更广义的健康管理延伸,实现从预防、筛查、诊断、治疗到康复的全程化、个性化健康管理闭环。通过持续监测和分析个体健康数据,提供动态的健康评估、风险预警和个性化的干预建议,真正实现“以疾病为中心”向“以健康为中心”的转变。

## 5 结语

综上所述,基于大数据的电子信息技术在医疗诊断领域已经展现出了巨大的潜力和价值。它不仅提高了诊断的准确性和效率,还为个性化医疗和精准治疗提供了有力支持。随着技术的不断进步和数据的持续积累,大数据将在未来医疗诊断中发挥更加核心的作用,推动医疗行业向更高水平发展。

### 参考文献

- [1] 吴威骏.基于机器学习的交互式智能医疗诊断辅助系统[D].华东师范大学,2024.
- [2] 潘明.大数据在医疗行业的应用[J].中国新通信,2021,23(04):90-91.
- [3] 高蕊.随机森林算法改进及其在医疗诊断系统中的应用[D].南京邮电大学,2020.
- [4] 韩温,冈倪,鲁慧,刘庆燕,王计帅.基于健康医疗大数据的智能辅助诊断系统研究[J].中国卫生产业,2020,17(10):159-161.
- [5] 吴江朋.基于血液检测数据的计算机辅助医疗诊断系统构建[D].兰州大学,2020.