

# 人工智能医疗诊断中的数据整合挑战与对策

刘禹佳 廉欢 白锦林 聂炆 欧政宏

南宁理工学院，广西壮族自治区，541006；

**摘要：**随着人工智能（AI）技术的飞速发展，在医疗诊断领域的使用逐步被视为提高医疗服务的品质和效益的关键途径。但是人工智能在医疗领域中的高效运用却面临着很多挑战，数据整合就是一个至关重要的难点。通过对人工智能医疗诊断数据融合现状进行分析，揭示在数据质量，隐私保护和医疗机构之间数据共享方面存在的问题并给出对策与建议。

**关键词：**人工智能；医疗诊断；数据整合；数据质量

**DOI：**10.69979/3041-0673.26.02.017

## 引言

随着人工智能技术快速发展，将人工智能技术应用于医疗诊断领域正在成为医疗智能化变革的主要推动力。但由于医疗数据具有分散性，复杂性和隐私性等特点，数据整合是人工智能进一步应用的关键性瓶颈。优质、可共享和安全的数据基础是精准诊断和智能决策的先决条件。文章围绕人工智能医疗诊断，系统地论述了在数据整合方面所遇到的挑战及应对措施，以期对医疗智能化提供可行路径和理论支撑。

## 1 人工智能医疗诊断的特点

### 1.1 医学数据的多样性与复杂性

医学领域的数据类型多样，涵盖影像，电子病历和基因组数据几个层次。这些数据既来自不同医疗设备与系统，又具有极其复杂的形式，例如影像数据大多是二维或者三维的，基因组数据是通过海量序列信息表现出来。医疗影像是数据的重要来源，一般需结合医学专业知识对其进行细致解读，电子病历中包括病人历史病史，诊疗记录以及处方，这类数据关联性极强，需要人工智能技术对多元数据处理<sup>[1]</sup>。

### 1.2 数据来源的分散性与异构性

在医疗行业，数据广泛分布在各种医疗系统与设备之中，其中就包括了医院电子病历系统，影像数据系统以及检验数据平台。这些数据零散地分布在不同医院，诊所等医疗机构，各平台数据格式及存储方式有明显区别，造成其互操作性不强。数据异构性特别是医疗影像与实验室数据融合时往往会遇到标准化难题。即便同一医院内，不同科室使用的设备与系统也可能存在兼容性问题，进一步增加了数据整合的难度。将人工智能技术运用于诊断过程，有赖于从各种数据源中获得综合而统

一的信息，需要人工智能系统既能接收到各种源的信息。

### 1.3 数据处理的精度要求与实时性

在医疗诊断过程中，数据处理对于精度，实时性等方面都有着极高的要求。人工智能医疗诊断系统下，准确的算法与优质的数据处理能力为高效诊断提供了依据。以医学影像为例，对其进行处理既要具备高分辨率图像数据又要有高效算法的支持才能在短期内完成疾病检测和诊断建议等。而且对实时性要求很高的临床诊断特别是急诊、手术等高风险场景下，人工智能系统需要能迅速做出反应，及时给出诊断结果<sup>[2]</sup>。所以在对数据进行处理时，既需要确保其精准度又需要符合临床工作时效性需求。

## 2 人工智能医疗诊断的现状分析

### 2.1 国内外人工智能医疗诊断发展现状

人工智能应用于医疗诊断领域，历经由基础研究逐渐走向临床实践。国内外许多研究机构与企业在医学影像识别，电子病历分析领域都取得了重要的突破。外国的技术进步相对较早，特别是在人工智能辅助诊断系统的实际运用方面，美国和欧洲的医疗机构已经在一定程度上实现了临床应用。比如 IBM 的 Watson 和 Google Health 已经投入了大量的资源来开发医疗人工智能产品，并在癌症诊断，影像分析和基因组学研究等方面显示出了初步成效。国内虽然起步略晚，但是近几年在政策支持与市场需求驱动下人工智能医疗诊断技术得到了快速的发展。国内几家公司已经引入医疗影像分析系统并结合深度学习算法给出辅助诊断的建议。在促进国内人工智能在医疗应用中迅速发展的同时，政策层面上的扶持是关键要素，在国家颁布的《“健康中国 2030”规划纲要》以及有关政策文件中，为人工智能技术应用

于医疗领域打下基础。我国人工智能医疗诊断总体发展水平仍然落后于国外先进国家,在技术成熟度,临床验证和数据共享上还面临着挑战。

## 2.2 医疗数据应用的成熟度与挑战

医疗数据应用成熟度对人工智能医疗诊断系统应用效果具有直接影响。尽管医疗行业积累了大量的医疗数据,但由于数据格式的多样性、来源的分散性以及标准化不足,导致现有数据难以充分发挥其潜力。电子病历与影像数据为最常用的数据来源之一,但是这些数据通常具有结构化程度较低,数据不连贯,标注不完整等特点,这使人工智能在进行数据处理与分析时,面临着更多的挑战。对于 AI 系统而言,数据质量与精度是决定系统性能好坏的根本,缺少标准化、标注不正确或者不完整的数据都会造成训练模型出现偏差从而影响诊断结果准确性<sup>[3]</sup>。

## 2.3 现有人工智能辅助诊断产品的市场状况

人工智能辅助诊断产品市场发展势头迅猛,特别是在医疗影像分析和疾病筛查方面逐步显示出潜能。国内外很多公司都推出了以人工智能为基础的影像识别,语音识别等辅助诊断工具和自然语言处理技术,这类产品可以在某种程度上缓解医生工作压力和提升诊断效率。但目前已有的产品多处于试点或者早期推广的阶段,还没有推广至大规模的临床应用。目前市面上虽有多款人工智能的医疗辅助诊断产品,但是这类产品大多是针对单一领域的,比如放射科的影像分析和皮肤癌筛查,没有一个综合的,跨学科整合平台很难适应医院多科室,多专业综合需求。一些大型医疗设备生产商,例如 GE 和飞利浦,已经成功地将人工智能技术与他们的传统医疗设备融合,从而提供了更加全面的诊断支持系统,但是这类产品推广速度慢,而且价格昂贵,多针对大医院、研究机构等,很难推广到基层医疗机构使用。

## 3 人工智能医疗诊断中数据整合存在的问题

### 3.1 数据质量与准确性

人工智能医疗诊断系统有效性对数据质量与准确性具有较高依赖性,然而目前医疗数据中存在着广泛的质量问题。一是问题的提出医疗数据在采集时通常没有严格标准化,数据录入与采集方法也不一致,造成数据格式不同、不一致。比如电子病历对患者信息描述方式会因为不同医院或者医生的喜好而不一样,从而影响了数据的有效集成。医学影像数据往往会出现图像模糊和噪声干扰等品质问题,从而影响了算法对图像进行识别时的精度。二是数据标注是否准确也是个大问题。

医疗数据标注一般依靠人工医生做出专业判断,错误或者不全面的标注会使训练数据产生偏差,影响模型学习效果并最终造成人工智能系统诊断结果错误。在对某些新兴疾病或者稀有病症进行鉴定时,标注数据不足或者不正确也会影响到系统诊断能力<sup>[4]</sup>。因此如何保证数据采集,处理及标注的质量及精度已成为制约人工智能广泛应用于医疗领域的主要障碍之一。

### 3.2 数据隐私与安全保护

医疗数据一般都会涉及个人的敏感健康信息,所以人工智能医疗诊断过程中如何保证数据隐私和安全就成为严峻的课题。尽管关于数据隐私保护的规定逐渐完善,如 GDPR,《中华人民共和国个人信息保护法》法案已经对医疗数据保护有明确的要求,但是在实际运行过程中仍然面临着许多挑战。第一,医疗数据具有多样性、复杂性等特点,增加了隐私保护的难度。不同种类的医疗资料,如影像资料、基因组信息和电子病历等,都有其特定的隐私保护需求,在整合和分享这些数据时,如何对不同数据类型建立合理保护措施以避免数据泄露或者误用是当前急需解决的问题。二是跨医院、跨地区数据共享过程中,经常要有多方参与,怎样在确保各参与方遵守规定的情况下,实现数据流通、避免非法访问或者盗用数据,已经成为当前数据整合中面临的主要问题。

### 3.3 医疗机构间的数据共享与互操作性

数据共享和互操作性对人工智能医疗诊断数据整合具有重要意义,然而在实践中医务机构之间数据共享还面临着许多技术和制度上的障碍。不同医院及诊所一般采用不同管理系统,数据标准及存储方式等,造成各机构间数据不能实现无缝衔接。尽管存在如 HL7、DICOM 这样的国际标准,但很多医疗机构还没有完全接受这些建议,这导致了跨部门的数据交流和分享变得更为复杂。医疗数据互操作性还表现在技术层面上,目前人工智能系统通常很难有效应对不同医疗机构采用的数据格式及编码系统,造成各系统间数据兼容性较差,从而影响人工智能模型学习效果和诊断能力。为使数据有效融合,迫切需要突破技术瓶颈、促进医院系统互联互通、推进标准化建设等措施来保障各类医疗机构间医疗数据流通畅通准确。

## 4 人工智能医疗诊断中的数据整合对策建议

### 4.1 提高数据采集与标注的标准化

规范的数据采集和标注是保证人工智能医疗诊断系统有效运行的先决条件。医疗数据具有多样性、复杂

性等特点,需要在收集时严格执行统一标准,降低数据不一致性、误差等。在影像数据处理中,需要对影像格式,分辨率以及采集方法等进行统一,使不同装置以及平台之间的数据可以有效地兼容。标注工作也要尽量标准化,特别是医学影像标注工作,一定要依靠专家级医生精准标注才能确保数据优质高可靠性。标准化标注在提升数据利用价值的同时也避免了人工智能模型在训练过程中出现偏差的问题,保证了训练后模型泛化能力更强,诊断精准度更高<sup>[5]</sup>。强化数据采集与标注标准化工作,既可以提高人工智能系统性能,又可以为今后跨机构数据共享与模型协作等提供基础支撑,保障各种诊疗场景下应用成效。

## 4.2 加强医疗机构间的数据共享与协作

数据共享和协作,是人工智能医疗诊断系统是否能获得广泛应用的关键要素之一。现有医疗生态下,各医疗机构间信息孤岛严重,制约数据资源流动性与共享效率。要解决这一难题,需要促进各个医院与医疗机构之间技术层面的合作,使用统一标准与接口协议来突破数据壁垒。特别是医学影像,病理数据方面,各医院所采用的设备及系统是有区别的,所以有必要建立统一数据交换格式及通信协议以保证数据能无缝连接。加强跨机构合作既有利于增强人工智能诊断模型训练效果,又可使人工智能系统能够在各地域,各医院之间进行全面验证,增强适用性与普遍性。通过政府政策引导,行业标准制定和技术创新等方式加强医疗机构之间数据共享和合作,将极大地提升人工智能在医疗领域中的现实应用价值。

## 4.3 强化数据安全保护与隐私政策的实施

数据安全保护与隐私政策执行是人工智能在医疗诊断应用过程中至关重要的一环,它直接影响着患者个人隐私与系统可靠性。医疗数据一般都涉及高度敏感信息,所以在对其进行收集,储存,处理以及传输时都要有严密的安全措施以避免其被泄露,篡改或者误用。医疗行业要加强同网络安全技术公司和数据加密专家等方面的协作,构建更加完备的数据安全体系。加密技术和区块链技术作为现代信息安全手段,能够有效地确保数据传输过程中数据的安全,避免恶意攻击或者是数据窃取。尽管各国和地区的隐私保护法律,如《通用数据保护条例》(GDPR)和《中华人民共和国个人信息保护法》为数据的安全性提供了坚实的法律支撑,但这些法律的实施效果和在特定行业的适用性还需要进一步

强化。

## 5 结论

人工智能应用于医疗诊断具有广阔前景,然而数据整合问题仍是其推广和应用所面临的核心难题。医疗数据是否优质准确直接关系到诊断系统的性能,特别是数据采集及标注是否规范更是保证系统准确可靠的根本所在。另外,数据隐私保护及安全问题还涉及患者隐私及系统稳定性等问题,迫切需要强化相关法律法规执行及技术防护措施。医疗机构之间数据共享及互操作性问题尽管受技术及制度等诸多约束,但是跨机构数据整合及协作会极大增强人工智能系统诊断能力及广泛应用性。从整体上看,这些问题的解决将会促使人工智能技术有效普及医疗领域,从而带动全球医疗水平不断提升。

## 参考文献

- [1] 李韬,张琚,李亚哲,等.人工智能在医疗健康领域的创新应用,风险挑战与治理对策[J].医学信息学杂志,2025,46(1):2-8,16.
- [2] 崔蕾,郑新昕.人工智能时代法律适用问题研究:挑战与对策[J].公安研究,2024(12):38-47.
- [3] 侯小花,安洁,方莎娜.人工智能赋能教育心理学教学的机制,挑战与对策研究①[J].2025.
- [4] 陈虹旭,赖睿,徐建国,等.人工智能赋能神经外科教学:机遇,挑战与对策[J].华西医学,2024,39(9):1470-1473.
- [5] 张恩宁,王辰,郑志昕,等.人工智能技术在免疫学教学中的挑战及对策研究[J].Chinese Journal of Immunology,2025,41(6).

作者简介:刘禹佳(2004.11-)女,汉族,本科,广西柳州人,研究方向:人工智能;

廉欢(2006.01-)男,满族,本科,河北省秦皇岛市人,研究方向:人工智能;

白锦林(2005.10.28-)男,汉族,本科,河南省南阳市新野县人,研究方向:人工智能;

聂炆(2006.01.01-)男,汉族,本科,陕西省商洛市商南县人,研究方向:人工智能;

欧政宏(2006.03-)男,汉族,本科,广西桂平市人,研究方向:人工智能。

2025 广西壮族自治区大学生创新创业训练计划项目:  
智医-智能医疗辅助诊断平台。项目编号:S202513645  
089。