

基于人工智能的医学检验数据分析方法与应用分析

张莹
西安海棠职业学院，陕西 西安 710038

摘要：随着医疗技术的不断发展以及临床诊疗手段的日益多样化，医学检验数据的规模和复杂程度也在不断攀升。如何从海量且结构多样的医学检验数据中提取有价值的信息，并将其应用于临床诊断与疾病预防，是医学研究和医疗服务亟待解决的重要课题。人工智能技术为医学检验数据分析提供了强大的手段与方法。结合医学检验数据的特点，充分利用人工智能算法进行数据聚类、分类、预测与诊断支持等工作，可有效提高医学检验的准确性与效率，进而为临床诊疗提供科学依据与决策支持。本文围绕医学检验数据分析的需求与挑战，探讨了基于人工智能技术的关键方法，并对实际应用中的困难与前景进行了论述。研究表明，人工智能对医学检验数据的深度挖掘与分析已在多种疾病诊断、药物研发以及公共卫生管理等领域取得了显著的效果。

关键词：医学检验；人工智能；机器学习；深度学习；数据分析

Analysis Method and Application Analysis of Medical Laboratory Data Based on Artificial Intelligence

Zhang,Ying
Xi'an Haitang Vocational College, Xi'an, Shaanxi, 710038, China

Abstract: With the continuous development of medical technology and the increasing diversification of clinical diagnosis and treatment methods, the scale and complexity of medical laboratory data are also rising. How to extract valuable information from massive and diverse medical laboratory data and apply it to clinical diagnosis and disease prevention is an important subject to be solved urgently in medical research and medical service. Artificial intelligence technology provides a powerful means and method for medical laboratory data analysis. Combining with the characteristics of medical laboratory data, making full use of artificial intelligence algorithm for data clustering, classification, prediction and diagnosis support can effectively improve the accuracy and efficiency of medical laboratory, and then provide scientific basis and decision support for clinical diagnosis and treatment. Focusing on the needs and challenges of medical laboratory data analysis, this paper discusses the key methods based on artificial intelligence technology, and discusses the difficulties and prospects in practical application. The research shows that the deep mining and analysis of medical laboratory data by artificial intelligence has achieved remarkable results in many fields such as disease diagnosis, drug research and development, and public health management.

Keywords: Medical examination; Artificial intelligence; Machine learning; Deep learning; Data analysis

DOI: 10.62639/sspis23.20250203

近年来，医学检验在临床诊断与治疗中作用显著，其数据涵盖血常规、生化检验、基因测序等多模态形式。随着医疗信息化发展，高效获取、分析海量医学检验数据成为关键。医学检验数据因多源异构、高度个体化及规模庞大而复杂，传统分析方法难以应对^[1]。此外，噪声和缺失值进一步增加了医学检验数据的处理难度。机器学习与深度学习等人工智能（AI）技术，为医学检验数据提供了高效分析手段，涵盖聚类、分类、预测等任务，在图像识别、自然语言处理及多模态数据融合等领域展现优势。AI 不仅提升诊断准确性，还减少误差，助力个性化医疗与精准治疗。本文将探讨基于 AI 的医学检验数据分析方法，梳理其应用场景及挑战，并展望未来发展趋势。

一、医学检验数据分析的特点与挑战

（一）数据的多元与异构

医学检验往往涉及多种数据类型，如血常规指标、免疫学指标、代谢组学数据，以及临床诊断报告文字记录等。这些数据从结构化到非结构

化，形式多样，且部分类型的数据量极为庞大。基因检测数据可能包含上百万甚至上千万个碱基序列信息，而临床诊断过程中的文字记录则属于典型的非结构化文本数据。异构数据的有效整合和融合分析对算法的可扩展性与通用性提出了更高要求。

（二）数据质量与标准化问题

医学检验数据的采集过程可能受到设备精度、环境干扰、患者配合程度等因素的影响，导致检验数据中存在噪声、缺失值、异常值等。此外，不同医院或科室在采集、记录与存储方式上也可能存在差异，缺乏统一的标准化流程。为确保 AI 模型在医学检验数据上的应用效果，需要在数据预处理阶段做好数据清洗、降噪、缺失值填补以及数据归一化等工作^[2]。然而，这些工作对医疗机构或科研团队在技术与成本方面都有较高门槛，也对 AI 算法在实际落地过程中提出了挑战。

（三）隐私与安全

医学检验数据通常包含患者的个人隐私信

（稿件编号：IS-25-3-1024）

作者简介：张莹（1986-），性别：女，汉，籍贯：陕西省宝鸡市，学历：大学本科，职称：主管检验师。

息,一旦泄露或误用将可能对患者的权益造成严重影响。医疗领域对数据安全与隐私保护有着极为严格的要求,如 HIPAA 法规(在美国)或相关国家的个人信息保护法,都对医疗数据的采集、使用和传输制定了具体规范。在进行人工智能分析时,往往需要将大量患者数据集中存储或传输到云端,这就对数据的脱敏和访问控制提出了极高的安全要求。在算法层面,联邦学习(Federated Learning)等技术的出现,为在保护数据隐私的前提下开展大规模联合建模提供了新的解决思路。

二、基于机器学习的医学检验数据分析方法

在人工智能以及大数据分析的范畴下,机器学习是最早得到广泛应用的关键技术之一,将统计学方法与算法进行有效结合,机器学习可以帮助医疗行业从海量医学检验数据中挖掘出潜在规律并应用于临床诊断和治疗决策中。以下从聚类、关联、分类与预测四个方面进行阐述。

(一) 医学检验数据的聚类分析

聚类分析(Clustering)主要关注的是如何将大量无标签或缺乏先验知识的数据加以归类,寻找数据间的相似性与差异性。在医学检验中,聚类分析可帮助临床医生初步识别患者群体的不同亚型或疾病不同阶段的分布情况。

1. 基于传统聚类算法的改进:常见的 K-means、层次聚类等算法在处理小规模医学数据时有一定效果,但当数据规模迅速扩增或特征维度极高时,传统串行算法常面临计算复杂度过高、内存占用过大的挑战。为此,可借助 MapReduce 或 Spark 等分布式计算框架对 K-means 等算法进行并行化改进,大幅度提高聚类效率。

2. 以密度聚类为代表的高维医学数据聚类:面对基因组学或蛋白组学等高维度检验数据,基于密度的 DBSCAN 或 OPTICS 等算法在发现任意形状的聚类时具有较大优势。但对于极高维数据,需要对算法进行降维或特征提取处理,以保证聚类结果的准确性和可解释性。

(二) 医学检验数据的关联分析

医学检验中,常常需要对检测指标之间的关联性和潜在因果关系进行研究。某些血常规指标与肝功能指标之间可能存在较强的耦合关系,或某些微量元素水平与糖尿病发生风险相关。经典的关联分析算法包括 Apriori 和 FP-Growth 等,它们能够在大规模的检验数据中挖掘频繁模式或强关联规则。

然而,这些传统方法在医学领域应用时也面临以下难点:

数据噪声与缺失值:大规模医学检验数据常常包含缺失值和噪声,需要先进行数据预处理,否则可能增加关联挖掘中的误判率。

多因素耦合作用:疾病产生和发展的过程往往受多种因素交叉影响,简单的关联规则难以准确刻画这种复杂的多因素耦合关系^[3]。

为此,研究者们尝试在关联规则挖掘的基础

上,结合统计学和临床知识,将关联分析与机器学习中的其他方法(如特征选择、聚类)相结合,或引入网络图模型(Graph Model),以构建复杂网络的方式来探讨医学检验指标之间的潜在关系,提高关联分析的准确性与可解释性。

(三) 医学检验数据的分类分析

分类分析是医学检验数据挖掘的核心任务之一,也是临床诊断最常使用的 AI 算法场景。其目标是根据已有的标注或确诊信息,将新的患者样本分类至相应的类别,以此帮助医生快速判别疾病类型或患者群体状态。常用的分类算法包括决策树、随机森林(Random Forest)、支持向量机(SVM)以及基于统计学的朴素贝叶斯分类等。

决策树及随机森林:在可解释性与计算开销之间取得平衡,并在医学检验数据量较为适中的场景下,有较好的泛化能力。

SVM:对小样本或高维数据有一定适应性,但在超大规模数据场景下需要进行适当的特征降维或核函数调整,并利用并行计算框架来减少训练时间。

朴素贝叶斯:适合特征相互独立假设场景,但在医学检验中常常存在特征之间的交互影响,因此需要结合专业知识进行特征工程。

对于分类模型而言,模型的准确度、灵敏度(Sensitivity)和特异度(Specificity)等指标都是衡量其临床应用价值的关键。由于医学检验往往影响临床决策,分类过程中需要注重可解释性,确保医生可以理解模型的推断逻辑,从而提升对模型结果的信任度。

(四) 医学检验数据的预测分析与辅助诊断

在预测分析方面,机器学习不仅可以对单一检验指标进行趋势预测,也可综合多个患者特征指标,对疾病的发生风险、进展阶段或预后结果进行评估。如对于糖尿病患者,机器学习可根据患者的血糖、胰岛素水平、体重指数(BMI)以及生活习惯等特征进行长周期的血糖走势预测,为制定个性化的降糖治疗方案提供参考。在传染病和重大公共卫生事件监测中,对检验数据和流行病学数据进行预测分析,还可以为突发疫情的预警和干预提供决策依据。

三、基于深度学习的医学检验数据分析方法

深度学习(Deep Learning)在近年来的医学领域中发展极为迅速,尤其在医学影像识别、病理切片辅助诊断和基因组分析等方面有着显著突破。深度学习以其多层神经网络结构对数据进行特征提取与模式识别,大大提升了对复杂、非结构化数据的处理能力。

(一) 深度神经网络在医学检验中的应用

1. 卷积神经网络(CNN)在医学图像分析中的应用:在医学检验和医学影像中,大量的图像数据如病理切片、CT、MRI、X光片、超声等数据是临床诊断的重要依据。CNN 擅长提取图像的局部特征,能够帮助医生迅速发现病灶位置或早期病变信号。CNN 在训练大量病理切片图像后,可以自动分辨肿瘤细胞与正常细胞,对乳

腺癌、肺癌等多种癌症的早期筛查提供辅助判断。

2. 循环神经网络 (RNN) 及其变体 (LSTM、GRU) 在时间序列预测中的应用: 医学检验数据如动态血压、心电图 (ECG) 或连续血糖监测 (CGM) 等均具有明显的时间序列特征。RNN、LSTM (长短期记忆网络) 和 GRU (门控循环单元) 利用记忆和序列信息的更新, 对患者生理指标的变化趋势进行预测, 如监测重症监护病房 (ICU) 患者的病情变化, 预测是否会出现心律失常或其他重大并发症。

(二) 医学文本与多模态深度学习

除影像和生理信号外, 临床病历、检验报告等文本数据也是重要的信息载体。深度学习中的自然语言处理 (NLP) 技术, 如 BERT、GPT 等模型, 可以对海量的临床文本和医学文献进行语义理解与知识抽取, 从而实现自动化的疾病判别、医学文献检索与智能问诊系统。

多模态深度学习则是将文本、影像、基因组数据等不同模态的数据进行融合, 构建综合性诊断模型。在肿瘤的个性化治疗中, 可以将患者的基因测序结果与 CT 影像、病理诊断报告等信息同时输入深度学习网络, 以跨模态的特征学习来识别潜在的基因变异与肿瘤生长特征对应关系, 为精准医疗决策提供支持。

四、医学检验数据分析的应用案例与趋势

(一) 肿瘤早期筛查与诊断

肿瘤的早期筛查对提高患者生存率具有重要意义。借助人工智能算法对患者的多项检验数据进行综合分析, 可以在病灶尚未形成大规模扩散之前, 对可疑风险患者进行筛选与密切监测。典型的应用案例包括:

乳腺癌筛查: CNN 模型可对大量乳腺 X 光图像进行深度学习训练, 识别组织中是否存在异常钙化或肿瘤影像特征, 协助放射科医生进行更为准确的诊断。

肺癌早期检测: 结合胸部 CT 影像与患者的血清肿瘤标志物检测数据, 构建跨模态学习模型, 大大提升肺部微小结节检测的精准度。

(二) 感染性疾病与公共卫生管理

在流感、COVID-19 等感染性疾病的检验和监测中, 人工智能算法可以根据大规模人口检验数据、核酸检测结果以及流行病学调查信息, 实时监测疫情变化趋势并进行风险预警, 为公共卫生部门的隔离和救治策略提供有效决策支持。

(三) 药物研发与药物反应预测

医学检验数据可以帮助药企或科研人员了解药物在不同患者群体中的药代动力学和药效学差异。借助机器学习或深度学习, 可以对大量临床试验及药物基因组学数据进行综合分析, 预测患者对特定药物的反应, 减少盲目用药和试验成本, 提高新药研发效率。

(四) 个性化医疗与精准医学

在大数据时代下, 个性化医疗与精准医学越来越受到关注。整合患者的多种医学检验数据与

生活方式、环境暴露等信息, 人工智能模型可以对疾病风险进行更加精确的分层, 对治疗方案进行个性化选择, 提高疗效并减少副作用。

五、医学检验数据分析面临的主要问题与对策

(一) 数据匮乏与不均衡

虽然医疗数据总体规模庞大, 但在单个机构或细分疾病领域, 仍可能出现有效样本不足或标签匮乏等问题。此外, 不同疾病亚型或检验结果出现不平衡分布时, 模型训练易受到偏倚, 从而影响预测的准确性。针对该问题, 可以采用迁移学习 (Transfer Learning) 或半监督学习 (Semi-supervised Learning) 等方法, 充分利用其他领域或相似疾病的数据, 并以自适应的方式将已有模型迁移到目标疾病领域。

(二) 可解释性与可靠性

人工智能算法在医学领域应用的一个重要障碍在于: 深度学习模型往往被视为“黑箱”, 难以解释。临床医生和患者通常需要明确知道模型做出某个诊断判断的依据和逻辑。为提升模型的可解释性, 需要开发可视化技术 (如 Grad-CAM、可视化注意力机制) 或将可解释的机器学习方法 (如决策树、注意力可视化) 与深度学习相结合。与此同时, 需要在真实临床环境中进行充分的验证, 确保模型能在更广泛的人群中保持可靠的性能。

六、结束语

基于人工智能的医学检验数据分析在近十年的发展中取得了长足进步, 并在肿瘤筛查、感染性疾病监测、精准医疗和公共卫生管理等方面展现了巨大的应用潜力。机器学习与深度学习为医学检验数据的聚类、关联、分类和预测提供了全新的技术框架, 使研究者和临床医生能够更有效地挖掘海量数据背后的潜在规律并对疾病进行早期干预和精准诊断。然而, 当前在医学检验数据分析领域仍面临数据多源异构、数据标准化不足、模型可解释性与安全合规等多重挑战。为了充分释放人工智能在医学检验中的潜力, 需要进一步加强多学科交叉合作, 在算法优化、数据共享机制以及法规伦理等方面做出系统性改进。随着医疗信息化基础设施的完善以及高性能计算平台与算法的持续演进, 人工智能必将在医疗行业内扮演越来越关键的角色, 不仅提高临床诊断与治疗的效率与准确度, 也为未来医疗模式的革新和健康保障体系的构建奠定坚实基础。

参考文献:

- [1] 张时民. 医学检验领域人工智能技术应用与展望 [J]. 国际检验医学杂志, 2018, 39 (05): 513-516+520.
- [2] 薛真睿, 徐婷, 姚春艳. 医学检验专业人工智能教学现状及改进措施 [J]. 国际检验医学杂志, 2023, 44 (07): 890-893.
- [3] 林花, 余波, 宁熙, 宋如琼, 王勇. 医学生对人工智能的了解与学习意向调查研究 [J]. 高教学刊, 2023, 9 (10): 186-188+192.