

基于端边云协同架构与知识图谱双引擎驱动的糖尿病智能辅助诊断系统设计与实现

吕金萍 张晓梅

宿州学院信息工程学院, 中国·安徽 宿州 234000

摘要: 针对基层医疗机构在糖尿病防治中面临的数据孤岛、专业资源匮乏及隐私安全等痛点, 本研究设计并实现了一套基于“端-边-云”三级协同架构与“数据-知识”双引擎驱动的智能辅助诊断系统。研究创新性地引入边缘计算机制, 在终端侧完成数据清洗与差分隐私脱敏, 有效降低云端负载并保障数据安全; 核心算法采用贝叶斯优化的 Stacking 融合模型与糖尿病专科知识图谱相结合, 既提升了风险预测的鲁棒性, 又通过路径推理实现了并发症预警的可解释性。系统已完成原型开发, 并在公开数据集与脱敏历史数据上完成了离线验证, 同时团队深入社区卫生服务中心开展了实地调研与需求对接。结果表明, 该系统在保持高预测精度的同时, 具备良好的工程适应性与基层落地潜力, 为慢病智能化分级诊疗提供了可行的技术范式。

关键词: 糖尿病辅助诊断; 端边云协同; 医学知识图谱; Stacking 集成学习; 实地调研

Design and Implementation of Intelligent Auxiliary Diagnosis System for Diabetes Driven by End-Edge-Cloud Collaborative Architecture and Knowledge Graph Dual Engine

Lv Jinping, Zhang Xiaomei

School of Information Engineering, Suzhou University, China Anhui Suzhou 234000

Abstract: Aiming at the pain points of data silos, shortage of professional resources and privacy security in diabetes prevention and treatment in primary medical institutions, this study designs and implements an intelligent auxiliary diagnosis system for diabetes based on the three-level collaborative architecture of "End-Edge-Cloud" and driven by "Data-Knowledge" dual engine. The research innovatively introduces the edge computing mechanism to complete data cleaning and differential privacy desensitization at the terminal side, effectively reducing the cloud load and ensuring data security. The core algorithm adopts the Stacking fusion model optimized by Bayesian optimization combined with the specialized knowledge graph of diabetes, which not only improves the robustness of risk prediction, but also realizes the interpretability of complication early warning through path reasoning. The prototype of the system has been developed, and offline verification has been completed on public data sets and desensitized historical data. Meanwhile, the team has carried out field research and demand docking in community health service centers. The results show that the system maintains high prediction accuracy, has good engineering adaptability and grassroots application potential, and provides a feasible technical paradigm for intelligent hierarchical diagnosis and treatment of chronic diseases.

Keywords: Diabetes auxiliary diagnosis; End-edge-cloud collaboration; Medical knowledge graph; Stacking ensemble learning; Field research

1 绪论

1.1 研究背景

“健康中国 2030”战略将慢性病防治列为重点, 而糖尿病作为典型慢病, 其管理高度依赖连续的多维数据监测。然而在基层, 患者数据散落在家用设备、社区档案和不同医院中, 形成了难以打破的“数据孤岛”。加之基层医生经验有限, 面对复杂的并发症早期信号往往难以及时判断, 导致漏诊。

国家卫健委发布的《糖尿病防治行动实施方案(2024—2030年)》明确要求利用信息技术推进早期筛查。这为本研究提供了政策导向: 开发一套能下沉至村卫生室、家庭医生签约服务及巡回医疗点的工具, 解决基层“筛不出、管不住”的难题。

1.2 研究思路与创新

本研究从单纯的“算法验证”转向“系统工程与落地验证”。我们不再死磕单一模型的精度极限, 而是聚焦于架

构的鲁棒性与诊断的可解释性。规划了村卫生室筛查、家庭医生签约、乡镇门诊及巡回医疗四大场景，旨在构建一套适应基层弱网环境、懂医学逻辑的智能辅助系统。

2 系统总体架构设计

2.1 端边云协同机制

初期我们尝试过集中式云计算，但在模拟测试中发现，基层网络波动大，原始数据直传极易拥堵且不安全。因此，我们重构为“终端感知-边缘预处理-云端决策”的三级模式：终端层兼容蓝牙血糖仪、手环等，负责多源数据接入；边缘层作为核心创新，利用社区网关或低成本服务器，在本地完成滑动平均滤波去噪、K近邻插补缺失值，并实施差分隐私处理，实测表明，这使上传数据量减少了约65%，既省带宽又保隐私；云端层专注高算力推理与模型迭代，接收边缘特征向量进行深度分析。

2.2 双引擎驱动的诊断核心

系统诊断能力源于“数据”与“知识”的深度融合。数据驱动方面，采用Stacking融合框架，结合随机森林、XGBoost等算法优势，并引入贝叶斯优化自动调参，提升预测鲁棒性；知识驱动方面，构建糖尿病专科知识图谱。当输入症状时，引擎能搜索病理传导路径（如“高血糖→微血管损伤→视网膜病变”），生成可视化的推理链条。这解决了传统AI“黑盒”难信任的问题，让医生知道系统“为什么这么判”。

3 核心算法原理与关键技术

3.1 基于贝叶斯优化的Stacking融合模型

为了克服单一模型局限，我们构建了双层Stacking框架。第一层用四种异构算法提取特征，第二层用逻辑回归作为元学习器整合结果。引入贝叶斯优化替代传统网格搜索，利用高斯过程引导搜索方向。实验显示，这将超参数搜索时间缩短了约60%，模型AUC值提升了近1.5个百分点。

3.2 知识图谱构建与推理机制

图谱定义了疾病、症状、指标等六大实体及十余种关系。推理采用路径排序算法（PRA），在图谱中搜索从症状到并发症的可行路径，综合路径长度与关系权重计算置信度。系统最终展示得分最高的Top-3路径，不仅给出结论，更解释推导过程，显著提升了诊断透明度。

4 系统原型实现

4.1 开发环境

系统采用轻量化易部署架构：后端基于Python、PyTorch、Neo4j、Flask开发；前端用Vue.js设计简洁Web

界面，适配基层操作；硬件以STM32为主控，搭建便携式血糖监测终端，实现数据自动采集。

4.2 核心功能

结合基层需求，系统设计五大核心功能：1. 双角色数据采集，患者端记录数据、高亮异常，医生端管理患者信息；2. 智能风险评估，实时计算风险评分，可视化展示趋势；3. AI辅助诊断，输出风险等级、并发症预警及诊疗建议，支持人机协同；4. 个性化健康管理，生成饮食建议与营养目标，降低执行难度；5. 群体数据分析，统计患者相关数据，支持报表导出，支撑群体健康管理。

5 实验验证与性能评估

5.1 数据集构建

为科学评估系统，我们构建了混合测试数据集：一是国际通用的PIMA Indians Diabetes Dataset（768条样本）；二是收集整理约500条脱敏历史病历数据。所有数据按7:2:1划分训练集、验证集与测试集，并进行标准化处理。

5.2 模型性能对比

实验结果显示，Stacking融合模型在准确率、精确率、召回率及AUC值上均优于单一模型。特别是召回率达到85.4%，较最优单一模型提升3.1%，这对减少漏诊具有重要意义。模型AUC值达0.892，表明具备良好的区分能力。Stacking模型的优势主要体现在对临界值复杂样本的处理上，通过综合多个基学习器意见做出了更准确判断。

5.3 系统性能测试

模拟并发环境下，系统平均响应时间1.2秒，满足交互需求。边缘预处理使上传数据量减少约65%，显著提升了弱网环境下的可用性。差分隐私保护测试表明，系统能有效防止原始数据还原，符合安全合规要求。

6 总结与展望

本研究立足于基层医疗实际需求，成功设计并实现了一套基于端边云协同与知识图谱融合的糖尿病智能辅助诊断系统原型。通过架构创新解决了数据效率与安全矛盾，通过算法融合实现了高精度与可解释性统一。基于公开数据集与脱敏数据的验证表明，系统在风险识别与并发症预警方面表现优异，且具备良好的工程适应性。项目不仅完成了技术开发，更探索了AI技术下沉基层的可行路径。

未来工作将聚焦于：扩大数据收集范围，开展小范围试点应用；探索联邦学习技术，打破数据孤岛；引入多模态学习，融合影像与文本数据。我们相信，随着技术迭代，此类系统将在基层慢病管理中发挥更大作用。

参考文献:

- [1] 中华医学会糖尿病学分会. 中国 2 型糖尿病防治指南 (2020 年版) [J]. 中华糖尿病杂志, 2021, 13(4): 315-409.
- [2] 张敏, 李娟. 基层糖尿病诊疗现状及智能辅助诊断技术应用进展[J]. 中国基层医药, 2022, 29(7): 1112-1116.
- [3] 王浩, 张丽. 基于 Stacking 集成学习的糖尿病风险预测模型研究[J]. 计算机工程与应用, 2023, 59(12): 234-240.
- [4] 李伟. 基于边缘计算的医疗物联网数据安全研究 [D]. 北京交通大学, 2021.
- [5] 刘敏, 陈静. 医学知识图谱在糖尿病辅助诊断中的应用研究[J]. 中国数字医学, 2022, 17(5): 89-93.
- 基金项目: 项目课题: 宿州学院资助省级大学生创新创业训练计划项目: 基于知识图谱与端边云硬件的糖尿病智能辅助诊断系统; 项目编号: S202510379147。安徽省教育厅自然科学重点项目: 融合注意力与全局上下文信息的图神经网络会话推荐算法研究 (2024AH051810)。
- 作者简介: 吕金萍 (2004-), 女, 汉族, 安徽亳州人, 本科在读, 学生, 研究方向: 计算机应用技术。