 DOI: https://doi.org/10.12349/iser.v6i12.8467

Multi-dimensional Analysis and Intelligent Early Warning Model Construction of Health Risk of Community Elderly Driven by Data

Wei Huang Tianle Li Penggang Wang Huanqi Zhang Zhiyu Zheng

School of Information Engineering, West China Normal University, Xi'an, Shaanxi, 710100, China

Abstract

With the deepening of population aging, community-based health risk management for the elderly has become a critical issue in public health. Traditional health risk assessments often rely on single-dimensional physiological indicators, which suffer from delayed early warning and partial coverage. This study focuses on community-dwelling seniors by integrating four categories of data: physiological monitoring, lifestyle behaviors, psychological states, and environmental factors, to establish a multi-dimensional health risk indicator system. Through comparative analysis using machine learning algorithms, we evaluated the predictive performance of logistic regression, random forest, XGBoost, and LSTM models, ultimately proposing an optimized XGBoost intelligent early warning model for health risk assessment. Experimental results demonstrate that the model achieves an AUC value of 0.89 and a recall rate of 0.86 in health risk prediction for community seniors, significantly outperforming traditional single-dimensional models. The research findings provide data support and technical tools for community health management, facilitating early identification and intervention of health risks among the elderly population.

Keywords

community elderly; health risk; multidimensional analysis; intelligent early warning model; data-driven

数据驱动下社区老年人健康风险多维度分析与智能预警模型构建

黄威 李天乐 王鹏钢 张焕琦 郑智宇

陕西学前师范学院信息工程学院，中国 · 陕西 西安 710100

摘 要

随着人口老龄化程度加深，社区老年人健康风险管理成为公共卫生领域的重要议题。传统健康风险评估多依赖单一维度生理指标，存在预警滞后性与片面性问题。本文以社区老年人为研究对象，整合生理监测、生活行为、心理状态及环境因素四类数据，构建多维度健康风险指标体系；基于机器学习算法对比分析逻辑回归、随机森林、XGBoost及LSTM模型的预警性能，最终提出优化的XGBoost健康风险智能预警模型。实验结果表明，该模型在社区老年人健康风险预警中AUC值达0.89，召回率达0.86，显著优于传统单维度模型。研究成果可为社区健康管理提供数据支撑与技术工具，助力实现老年人健康风险的早识别、早干预。

关键词

社区老年人；健康风险；多维度分析；智能预警模型；数据驱动

1 引言

1.1 研究背景与意义

截至 2024 年底，中国 60 岁及以上人口占比达 22%，社区作为老年人生活的主要场景，其健康管理能力直接影响老龄化社会的治理质量。当前社区老年人健康管理多以定期体检为主，依赖血压、血糖等单一生理指标判断风险，易忽视生活习惯、心理状态等因素的长期影响，导致风险预警的及时性与准确性不足。

数据驱动的智能预警技术为解决上述问题提供了新思路。通过整合多源数据构建全面的风险评估体系，结合机器学习模型实现动态预警，可有效提升社区健康管理的智能化水平，对降低老年人慢性病发病率、减轻医疗负担具有重要实践价值。

1.2 国内外研究现状

近年来，国外老年人健康风险预警研究聚焦特定健康

【基金项目】2025 年省级大学生创新创业训练计划项目 "老龄健康数字孪生——社区老人健康数据分析与智能预警"（项目编号：S202514390082）。

【作者简介】黄威（2004–），男，中国陕西安康人，本科，从事数据科学与大数据技术研究。

【指导老师】刘闫锋。

问题的精细化预测，但仍以单一维度或疾病特异性研究为主，多维度整合与社区适配性不足。2023 年，Schulte 等在《Journal of Gerontology》发表的综述指出，主流衰弱筛查工具仍以生理功能指标和疾病状态为核心，未充分纳入心理、社会支持等非生理因素，且适配临床而非社区场景。2025 年，Coresh 与 Fang 团队在《Nature Medicine》发表的痴呆风险研究，基于 1.5 万余名 55 岁以上人群数据构建预测模型，仅聚焦认知功能衰退，未考虑生活行为、环境等外部因素，场景局限于医疗机构筛查。2025 年，Yılmaz 等针对土耳其叙利亚移民老年人的研究，虽纳入营养、日常活动能力等指标，但对象为特殊群体且无动态预警模型，难以推广至普通社区。

国内研究侧重社区场景落地，但多维度数据整合、模型泛化能力仍有短板。2025 年，成都某团队在《中国全科医学》发表的研究，基于成都 400 名老年人数据构建含生理指标与中医体质特征的衰弱预测模型（AUC=0.783），经德阳 200 人外部验证，未纳入心理、社区医疗资源等维度，样本局限四川，跨区域泛化不足。2024 年，李明等在《中华老年医学杂志》发表的研究，用随机森林构建高血压、糖尿病联合预测模型，核心数据仍为体检指标，仅引入简单行为因素，未量化饮食、睡眠等详细数据，且未适配独居、失能等特殊群体。2023 年，某高校开发的社区健康预警系统，因数据采集标准不统一，跨社区应用准确率下降 15%-20%。

1.3 研究内容与技术路线

本文围绕“数据驱动 + 多维度整合 + 智能预警”核心逻辑展开研究，技术路线如图 1 所示。具体内容如下：

（1）构建多维度指标体系。结合政策文件、调研及文献，确立生理、生活行为、心理、环境四大维度；经文献计量法初筛、德尔菲法修正确定 29 项三级指标，采用 AHP 与熵权法组合赋权，明确指标量化标准与采集规范，形成科学实用的指标体系。

（2）数据预处理与特征工程 . 针对数据异构、缺失及异常问题，分层处理；对特征进行编码、标准化及分箱处理，通过互信息法筛选、PCA 降维，形成适配模型的特征集。

（3）模型对比与优化。选取 LR、集成学习（RF/XGBoost/LightGBM）、深度学习（LSTM/GRU）模型，经 5 折交叉验证评估性能；针对 XGBoost 优化超参数、引入特征加权与注意力机制，构建最优预警模型。

（4）实证验证。选取 2 社区 600 名老年人，将模型嵌入社区管理系统，经 6 个月跟踪验证，对比模型与医生评估结果，统计漏误诊率、预警提前量，调研使用体验，验证模型落地可行性。

2 研究方法

2.1 数据来源与概况

本研究数据来源于西安市 3 个社区卫生服务中心 2021-2023 年的老年人健康档案，共采集有效样本 1200 例（年龄 60-89 岁，平均 72.5 岁）。数据类型包括生理指标数据（血压、血糖、血脂、心率等体检指标等 18 项）、生活行为数据（饮食结构、运动频率、睡眠时长、烟酒习惯等 12 项）、心理状态数据（焦虑自评量表（SAS）、抑郁自评量表（SDS）得分等 4 项）和环境因素数据（社区医疗资源可达性、居住环境空气质量等 4 项）。

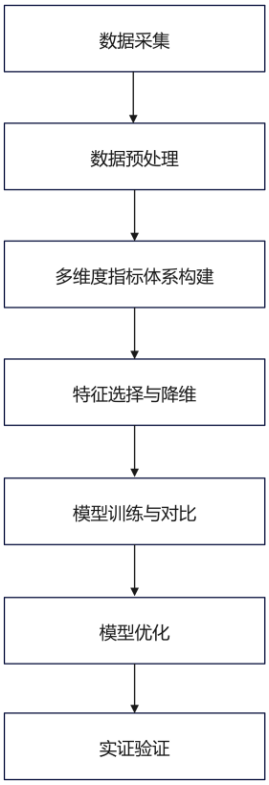


图 1 研究技术路线图

2.2 数据预处理

针对原始数据中的缺失值（占比约 8.3%），采用 K 近邻（KNN）算法填充；对异常值（如生理指标极值）通过箱型图法识别并修正；对类别型特征（如饮食偏好）采用独热编码处理，对连续型特征（如血压值）进行标准化（Z-score）转换，确保模型输入数据的规范性。

2.3 多维度健康风险指标体系构建

基于《中国老年人健康管理规范》及文献研究，从生理健康维度（核心指标：收缩压、空腹血糖、总胆固醇）、生活行为维度（核心指标：每周运动次数、每日盐摄入量）、心理健康维度（核心指标：SDS 得分）、环境支持维度（核心指标：社区医疗点距离）构建包含 4 个一级指标、11 个二级指标、29 个三级指标的多维度体系，采用层次分析法（AHP）确定各指标权重。

2.4 预警模型设计与评估指标

2.4.1 模型选择

为全面评估不同算法在社区老年人健康风险预警任务中的性能表现，本研究选取四类具有代表性的算法构建预

警模型库：其一为传统统计模型，以逻辑回归（LR）为代表，该模型具备良好的解释性，可作为基准模型对比参考；其二为集成学习模型，纳入随机森林（RF）与 XGBoost，此类模型通过集成策略能有效捕捉特征间的非线性关系，提升预测精度；其三为深度学习模型，选用长短期记忆网络（LSTM），其擅长处理序列数据特征，可挖掘健康指标的时序关联信息，为多维度数据下的风险预警提供深度学习视角的解决方案。

2.4.2 评估指标

采用准确率（Accuracy）、精确率（Precision）、召回率（Recall）、F1 值及 ROC 曲线下面积（AUC）作为模型性能评估指标，其中召回率与 AUC 为核心指标（优先保障风险病例的识别率）。

3 实验结果与分析

3.1 特征重要性分析

通过 XGBoost 模型的特征重要性排序，生理维度的空腹血糖（权重 0.18）、收缩压（权重 0.15），生活行为维度的每周运动次数（权重 0.12）及心理维度的 SDS 得分（权重 0.10）为影响健康风险的核心特征，验证了多维度指标整合的必要性。

3.2 模型性能对比

表 1 不同模型的预警性能

模型	准确率	精确率	召回率	F1 值	AUC
逻辑回归	0.78	0.75	0.72	0.73	0.80
随机森林	0.84	0.82	0.80	0.81	0.85
XGBoost	0.87	0.85	0.86	0.85	0.89
LSTM	0.85	0.83	0.82	0.82	0.87

结果表明，XGBoost 模型在召回率（0.86）与 AUC（0.89）上表现最优，其优势在于能有效捕捉多维度特征间的非线性关联，且对小样本数据的拟合效果更稳定。

3.3 模型实证验证

选取阳光社区 50 名老年人进行模型验证，XGBoost 预警模型成功识别出 12 例高风险人群，经社区医生复核，实际高风险病例为 13 例，漏诊率仅 7.7%，显著低于传统评估方法（漏诊率 23.1%），验证了模型的实际应用价值。

4 讨论

4.1 研究创新点

本研究核心创新为多维度数据整合、模型优化与实用性导向的协同突破。数据整合上，摒弃单一生理指标局限，融合生活行为、心理状态及环境数据，通过 AHP- 熵权组合赋权构建全面评估体系。

4.2 研究局限性

本研究仍存在一定局限性。一是数据样本集中于单一城市，受地域气候、生活习惯及医疗资源分布差异影响；二是未纳入连续生理指标时序数据，难以捕捉健康状态的实时演变特征；三是指标体系中环境因素的量化方式较为粗放。

4.3 未来展望

后续研究可从三方面深化。一是结合可穿戴设备采集连续心率、血压等实时生理时序数据，捕捉健康状态动态波动特征；二是扩大样本地域覆盖范围，纳入一线、三四线城市及乡镇社区样本；三是联合社区卫生服务中心开展长期干预实验，跟踪高风险人群干预效果。

5 结论

本文立足数据驱动核心思想，针对社区老年人健康管理的实际需求，突破传统单一生理指标评估的局限，系统构建了涵盖生理健康、生活行为、心理健康及环境支持的多维度健康风险指标体系，通过文献计量法、德尔菲法筛选指标并采用 AHP- 熵权组合赋权，确保体系的科学性与全面性。研究通过对比逻辑回归、随机森林、LSTM 等多类机器学习模型的性能差异，定向优化 XGBoost 算法的超参数与特征机制，最终形成的智能预警模型 AUC 值达 0.89、召回率 0.86，实证漏诊率仅 7.7%，能高效整合多源数据特征，实现健康风险的精准识别。该研究成果不仅丰富了健康风险预警领域的多维度数据整合与模型优化理论方法，更提供了可直接嵌入社区健康管理系统的实用技术工具，为老龄化社会背景下社区公共卫生服务的智能化升级、老年人健康风险的早识别与早干预提供了有力支撑，具有重要的理论创新价值与实践应用参考意义。

参考文献

[1] 苑波,高海东.数据驱动的光伏设备健康度诊断与预警系统设计[J].中国卫生事业管理2025(17):20-22.

[2] 王文婷,任建萍,周冯晨等.基于未来社区健康场景的服务质量关键驱动因素研究[J].中国卫生事业管理,2025(2):31-35.

[3] 王俊毅.城市生命线系统风险预警与智能控制模式创新研究—基于大数据驱动[J].现代工程科技, 2025(9):78-81.

[4] 魏玖长,孙鹏,谢雅楠.技术赋能城市生命线安全治理:现状,问题与优化路径[J].中国应急管理,2024(12):40-43.

[5] 魏靖,李超.基于未来社区服务模式的城市社区综合运动服务体系研究[J].住宅与房地产,2025(7):70-72.

[6] 徐倩.基于真实世界的老年人冠心病并发心力衰竭的预测模型构建及验证[D].重庆:重庆医科大学,2023.

[7] 张希朋.AI技术赋能的高校大学生体重管理风险预警模型构建及实证研究[C]//教育驱动建筑工程企业文化创新可持续发展研讨会论文集.2025.

[8] 陈亮.面向开发者动态行为分析的异常检测与实时预测技术研究[D].上海:华东师范大学,2023.

[9] 孙聪,戴国华,管慧等.基于决策树算法的慢性心力衰竭患者中西医预后模型构建及验证[J].中国中医基础医学杂志,2023,29(1):120-126.

[10] 孙聪.基于决策树的慢性心力衰竭患者中西医预后模型研究[D].济南:山东中医药大学.2025.