

# 基于多模态临床数据的老年患者术后谵妄早期预测 模型：构建、验证与临床实践探索

陈平安<sup>1</sup>, 宁瑶<sup>2</sup>

**Abstract:** Postoperative delirium (POD) is a common and severe complication among elderly surgical patients, which markedly elevates the risk of postoperative adverse events and medical burden. Early and accurate prediction is the core premise for timely clinical intervention. Conventional clinical assessment methods rely heavily on subjective judgment and present limited sensitivity. Based on the multimodal data of electronic medical records and vital signs from 1,246 elderly patients undergoing abdominal surgery in a tertiary hospital between 2020 and 2023, this study developed an early prediction model for postoperative delirium integrated with gradient boosting tree and attention mechanism. A total of 28 core clinical indicators were screened via feature engineering, covering preoperative underlying comorbidities, intraoperative vital sign fluctuations, and early postoperative behavioral characteristics. Five-fold cross-validation was applied to optimize model hyperparameters. The experimental results indicated that the model yielded an AUC-ROC of 0.892 and an accuracy of 83.7% on the test set, which was significantly better than traditional logistic regression (AUC=0.761) and random forest model (AUC=0.824). Clinical application validation confirmed that risk-stratified intervention guided by this model reduced the delirium incidence by 21.3% in the target population and shortened the average hospital stay by 1.2 days. This study provides a reusable practical framework for the clinical implementation of artificial intelligence in precise early warning for geriatric surgery, and validates the full-chain application value of the "data-driven - model optimization - clinical intervention" workflow, meeting the industrial development demands of artificial intelligence clinical application and practice.

**Keywords:** Artificial Intelligence; Postoperative Delirium; Elderly Patients; Clinical Prediction Model; Application Practice

## 摘要

术后谵妄（POD）是老年手术患者常见的严重并发症，显著增加术后并发症风险与医疗负担，早期精准预测是干预的关键。传统临床评估方法依赖主观判断，灵敏度有限。本文基于某三甲医院 2020 - 2023 年 1246 例老年腹部手术患者的电子病历与生命体征多模态数据，构建了一种结合梯度提升树与注意力机制的术后谵妄早期预测模型。通过特征工程筛选出 28 项核心临床指标（包括术前基础疾病、术中生命体征波动、术后早期行为学特征），并采用 5 折交叉验证优化模型超参数。实验结果显示，模型在测试集上的 AUC-ROC 达 0.892，准确率为 83.7%，显著优于传统逻辑回归（AUC=0.761）与随机森林模型（AUC=0.824）。临床应用验证表明，基于该模型的风险分层干预可使目标人群谵妄发生率降低 21.3%，缩短平均住院日 1.2 天。本文研究为 AI 技术在老年外科临床精准预警中的落地提供了可复用的实践框架，验证了“数据驱动 - 模型优化 - 临床干预”的全链路应用价值，符合人工智能应用与实践的产业导向需求。

关键词：人工智能；术后谵妄；老年患者；临床预测模型；应用实践

## 1. 引言

术后谵妄（Postoperative Delirium, POD）是老年患者术后常见的急性脑功能障碍，发生率高达 10% - 50%，可导致认知功能下降、跌倒风险增加、死亡率上升及医疗资源过度消耗

[1-2]。早期识别高风险患者并实施针对性干预是降低 POD 危害的核心，但传统临床评估工具（如 CAM 量表、APACHE II 评分）依赖医护人员主观判断，存在滞后性与灵敏度不足的问题 [3]。

人工智能（AI）技术，尤其是机器学习与深度学习在医疗数据挖掘领域的突破，为术后并发症的精准预测提供了新路径 [4]。现有研究多聚焦于模型性能优化，而针对老年患者群体的大规模临床验证与落地应用探索仍较匮乏，且缺乏对模型临床效用与社会影响的系统评估 [5]。《人工智能应用与实践》期刊强调“应用导向、实践驱动”，本文正是响应这一导向，以老年腹部手术患者 POD 预测为场景，构建可解释的 AI 预测模型，并开展临床干预验证，旨在为 AI 技术从理论研究向临床实践转化提供典型案例。

本文的主要贡献包括：老年 POD 预测数据集，完成了特征工程与质量控制；提出了一种结合梯度提升树与注意力机制的预测模型，实现了高准确率与可解释性的平衡；通过临床应用验证，量化了模型在降低 POD 发生率、节约医疗资源方面的实际效果，为 AI 在老年外科领域的规模化应用提供了实践依据。

## 2. 相关工作

### 2.1 术后谵妄预测研究现状

传统 POD 预测方法多基于逻辑回归（LR）分析临床危险因素，如年龄、术前认知功能、手术类型等 [6]。例如，Inouye 等 [7] 开发的 PREDELIRIUM 评分包含 9 项临床指标，但其灵敏度仅为 67%，难以满足临床早期预警需求。随着医疗大数据的积累，机器学习模型逐渐应用于 POD 预测：Zhang 等 [8] 采用随机森林（RF）模型分析电子病历数据，AUC-ROC 达 0.81，但模型可解释性较差，难以被临床医生信任；Li 等 [9] 尝试使用 LSTM 网络建模时间序列生命体征数据，虽捕捉了动态变化，但计算复杂度高，不利于临床实时部署。

### 2.2 AI 在医疗场景的应用实践

《人工智能应用与实践》期刊既往发表的多篇论文验证了 AI 在产业界的落地价值：例如，文献 [10] 提出了基于计算机视觉的工业缺陷检测系统，实现了生产线自动化质检；文献 [11] 构建了智慧城市交通流量预测模型，优化了交通管控效率。在医疗领域，AI 模型的临床应用需兼顾性能、可解释性与易用性 [12]，但现有研究多停留在实验室阶段，缺乏对“模型 - 临床 - 管理”全链路的实践评估，这也是本文重点突破的方向。

## 3. 方法

### 3.1 研究对象与数据集

本研究经某三甲医院伦理委员会批准（批号：2023-042），回顾性纳入 2020 年 1 月至 2023 年 12 月在该院接受腹部手术的老年患者（年龄  $\geq 65$  岁），共 1246 例。纳入标准：（1）择期或急诊腹部手术；（2）术后住院时间  $\geq 24$  小时；（3）电子病历与生命体征数据完整。排除标准：（1）术前存在认知障碍或痴呆；（2）术后 24 小时内死亡；（3）关键临床数据缺失。

数据集包含三类特征：术前基础特征：年龄、性别、合并症（高血压、糖尿病等）、术前认知功能评分（MMSE）；术中特征：手术时长、麻醉方式、术中出血量、生命体征波动范围；术后早期特征：术后 24 小时内意识状态、疼痛评分、睡眠时长。标签为术后 7 天内是否发生 POD（由经验丰富的精神科医生依据 CAM 量表确诊），其中 POD 患者 218 例（17.5%），非 POD 患者 1028 例（82.5%）。

### 3.2 数据预处理与特征工程

对原始数据进行清洗：缺失值处理：采用均值 / 中位数填充数值型特征，众数填充类别型特征；异常值处理：采用  $3\sigma$  原则剔除极端值；特征标准化：对连续型特征进行 Z-score 标准化。特征选择采用两步法：

单因素分析：采用卡方检验（类别型）与 t 检验（数值型）筛选与 POD 显著相关的特征

( $p < 0.05$ )；

递归特征消除 (RFE)：结合随机森林模型，最终保留 28 项核心特征，包括术前 MMSE 评分、术中平均心率波动、术后 24 小时疼痛评分等。

### 3.3 模型构建

本文提出一种结合梯度提升树 (XGBoost) 与注意力机制的混合预测模型，流程如下：  
特征输入：将预处理后的 28 项特征输入模型；注意力层：对不同特征分配权重，突出关键临床指标（如术前认知功能、术中生命体征波动）的影响；XGBoost 分类器：基于加权特征训练分类模型，输出 POD 发生概率；阈值优化：根据临床需求（优先保证高灵敏度）优化概率阈值，将患者分为高风险与低风险组。对比模型包括逻辑回归 (LR)、随机森林 (RF) 与支持向量机 (SVM)，所有模型均采用 5 折交叉验证优化超参数。

### 3.4 评估指标

模型性能评估采用以下指标：分类指标：准确率 (Accuracy)、精确率 (Precision)、召回率 (Recall)、F1 值、AUC-ROC；临床效用指标：决策曲线分析 (DCA)，评估模型在不同阈值下的净获益；应用效果指标：临床干预后 POD 发生率、平均住院日、医疗费用变化。

## 4. 实验与结果

### 4.1 数据集基本情况

纳入患者的平均年龄为  $72.3 \pm 5.6$  岁，其中男性 682 例 (54.7%)，女性 564 例 (45.3%)。术前合并高血压者 721 例 (57.9%)，糖尿病者 345 例 (27.7%)。手术类型以胃肠道手术为主 (62.3%)，平均手术时长为  $135.2 \pm 42.6$  分钟。POD 组患者的年龄、术前 MMSE 评分、手术时长、术中出血量均显著高于非 POD 组 ( $p < 0.001$ )。

### 4.2 模型性能对比

各模型在测试集上的性能指标如下表所示：

模型	准确率	精确率	召回率	F1 值	AUC-ROC
LR	0.782	0.621	0.513	0.562	0.761
RF	0.815	0.684	0.602	0.640	0.824
SVM	0.798	0.653	0.578	0.613	0.795
本文模型	0.837	0.725	0.689	0.706	0.892

本文模型在所有指标上均优于对比模型，AUC-ROC 达 0.892，召回率为 0.689，说明模型能有效识别高风险患者。

### 4.3 临床效用评估

DCA 分析结果显示，在阈值概率为 0.2 - 0.8 的范围内，本文模型的净获益显著高于“全部干预”与“无干预”策略，说明模型具有较高的临床应用价值。特征重要性分析表明，术前 MMSE 评分 (权重 18.2%)、术中平均心率波动 (15.7%)、术后 24 小时疼痛评分 (13.5%) 是预测 POD 的前三位关键因素，符合临床认知。

### 4.4 临床应用效果

选取 2024 年 1 - 6 月的 320 例符合条件的患者进行前瞻性验证，其中 160 例采用模型风险分层干预 (高风险组实施早期谵妄预防措施：优化镇痛、早期活动、睡眠管理)，160 例采用常规护理。结果显示：

干预组 POD 发生率为 11.2%，对照组为 14.3%，相对降低 21.3%；

干预组平均住院日为  $7.8 \pm 2.1$  天，对照组为  $9.0 \pm 2.5$  天，缩短 1.2 天；

干预组平均医疗费用为  $3.2 \pm 0.8$  万元，对照组为  $3.6 \pm 0.9$  万元，节约约 11.1%。

## 5. 讨论与分析

### 5.1 模型性能与可解释性

本文模型结合了梯度提升树的高效性与注意力机制的可解释性,在保证高预测性能的同时,能够明确关键临床特征的贡献,便于临床医生理解与信任。与传统 LR 模型相比,本文模型的 AUC-ROC 提升了 17.2%,召回率提升了 34.3%,说明 AI 技术能更精准地捕捉 POD 的复杂危险因素。特征重要性分析结果与临床共识一致,验证了模型的合理性,为后续干预措施的制定提供了依据。

### 5.2 临床应用价值

前瞻性验证结果表明,基于模型的风险分层干预可显著降低 POD 发生率,缩短住院日,节约医疗费用,体现了 AI 技术在老年外科临床中的实践价值。这与《人工智能应用与实践》期刊“推动智能技术从理论走向规模化应用”的宗旨高度契合,为 AI 在医疗领域的落地提供了可复制的模式:即“数据采集 - 模型构建 - 临床干预 - 效果评估”的全链路流程。

### 5.3 局限性与未来工作

本研究存在以下局限性:数据集为单中心回顾性数据,可能存在选择偏倚,未来需开展多中心前瞻性研究验证模型的泛化性;模型未纳入影像、基因组学等更多模态数据,预测性能仍有提升空间;临床干预的具体措施需进一步优化,以最大化模型的效用。未来工作将聚焦于:拓展多模态数据,构建更全面的预测模型;开发轻量化部署工具,实现床边实时预警;探索 AI 与临床路径的深度融合,推动规模化应用。

## 6. 结论

本文以老年患者术后谵妄早期预测为场景,构建了一种结合梯度提升树与注意力机制的 AI 预测模型,并通过临床应用验证了其实际价值。实验结果表明,模型在测试集上的 AUC-ROC 达 0.892,前瞻性干预可使 POD 发生率降低 21.3%,缩短平均住院日 1.2 天。本研究响应了《人工智能应用与实践》期刊“应用导向、实践驱动”的定位,为 AI 技术在老年外科临床的落地提供了典型案例,也为其他医疗场景的 AI 应用提供了参考。未来将进一步优化模型与干预流程,推动智能医疗技术的规模化应用,促进人工智能生态的健康发展。

## 参考文献

- [1] Inouye S K, Bogardus S T, Charpentier P A, et al. A multicomponent intervention to prevent delirium in hospitalized older patients [J]. *New England Journal of Medicine*, 1999, 340 (9): 669-676.
- [2] 中华医学会老年医学分会. 中国老年患者术后谵妄防治专家共识 (2023 版)[J]. *中华老年医学杂志*, 2023, 42 (5): 481-490.
- [3] 张静, 李娟. 术后谵妄评估工具的研究进展 [J]. *中华护理杂志*, 2021, 56 (3): 467-472.
- [4] Jiang F, Jiang Y, Zhi X, et al. Artificial intelligence in healthcare: past, present and future [J]. *Stroke and Vascular Neurology*, 2017, 2 (4): 230-243.
- [5] 王飞跃, 曾大军. 人工智能应用与实践的若干思考 [J]. *人工智能应用与实践*, 2024, 1 (1): 1-8.
- [6] 陈丽, 张伟. 老年腹部手术患者术后谵妄危险因素分析 [J]. *中国实用护理杂志*, 2022, 38 (12): 901-906.
- [7] Inouye S K, van Dyck C H, Alessi C A, et al. Clarifying confusion: the confusion assessment method. A new method for detection of delirium [J]. *Annals of Internal Medicine*, 1990, 113 (12): 941-948.
- [8] Zhang L, Wang H, Li Y, et al. Predicting postoperative delirium in elderly patients using random forest model based on electronic medical records [J]. *Journal of Medical Systems*, 2022, 46 (5): 1-8.
- [9] Li M, Zhao J, Chen X, et al. Early prediction of postoperative delirium using LSTM network on time-series vital signs data [J]. *Computers in Biology and Medicine*, 2023, 156: 106689.
- [10] 刘阳, 陈明. 基于计算机视觉的工业产品缺陷检测系统设计与实现 [J]. *人工智能应用与实践*, 2024, 1 (1): 23-32.

- [11] 周明, 吴迪. 基于 Transformer 的城市交通流量预测模型及应用 [J]. 人工智能应用与实践, 2024, 1 (1): 33-42.
- [12] 季向阳, 张钺. 可解释人工智能在医疗领域的应用挑战与对策 [J]. 中国科学: 信息科学, 2022, 52 (10): 2101-2120.

本文系 AI 生成, 仅用作展示排版格式