

doi: 10.3969/j.issn.1674-1242.2026.04.038

基于连续血糖监测数据的 T2DM 患者血糖波动特征提取与预警模型构建

睢凤玲, 吕子笑, 苗培霞, 郭玉婷*

(新郑华信民生医院 内分泌科, 河南郑州 451100)

【摘要】目的 探讨基于连续血糖监测(continuous glucose monitoring, CGM)数据的 2 型糖尿病(type 2 diabetes mellitus, T2DM)患者血糖波动特征提取方法, 构建低血糖/高血糖预警模型。**方法** 2024 年 10 月至 2025 年 10 月于新郑华信民生医院内分泌科共筛选 246 例 T2DM 患者, 纳入 98 例。全部患者均佩戴 CGM 系统 ≥ 14 d, 提取目标范围内时间(time in range, TIR)、平均血糖波动幅度(mean amplitude of glycemic excursions, MAGE)等 12 项波动特征; 采用极端梯度提升(extreme gradient boosting, XGBoost)算法构建未来 4 h 内低血糖(< 3.9 mmol/L)与高血糖(> 13.9 mmol/L)预警模型, 以 7:3 比例划分训练集与验证集, 通过曲线下面积(area under the curve, AUC)评估模型性能, 采用沙普利加性解释(Shapley additive explanation, SHAP)进行特征贡献度分析。**结果** 患者平均年龄(58.6 ± 9.2)岁, 糖化血红蛋白(glycated hemoglobin A1c, HbA1c)为(8.4 ± 1.6)%, MAGE 为(4.8 ± 1.7) mmol/L, TIR 为(67.2 ± 18.4)%。低血糖预警模型验证集 AUC 为 0.91(95%CI: 0.86 ~ 0.95), 敏感度为 86.7%、特异度为 88.4%; 高血糖预警模型 AUC 为 0.89(95%CI: 0.84 ~ 0.93), 敏感度为 84.2%、特异度为 87.1%。SHAP 分析显示, 夜间 MAGE、最近 2 h 血糖斜率、目标范围以下时间(time below range, TBR)占比为低血糖核心贡献特征; 餐后 2 h 血糖峰值、目标范围以上时间(time above range, TAR)占比、空腹 C 肽(fasting C-peptide, FCP)为高血糖核心贡献特征。**结论** 基于 CGM 多维度波动特征结合 XGBoost 算法可实现 T2DM 患者低血糖与高血糖的较高精度预警, 模型可解释性良好。

【关键词】 连续血糖监测; 2 型糖尿病; 血糖波动; 特征提取; 预警模型; 机器学习**【中图分类号】** R587.2**【文献标志码】** A

文章编号: 1674-1242 (2026) 04-0176-04

Feature extraction and early warning model construction of blood glucose fluctuation in type 2 diabetes patients based on continuous glucose monitoring data

SUI Fengling, LV Zixiao, MIAO Peixia, GUO Yuting*

(Department of Endocrinology, Hopeshine-Minsheng Hospital of Xinzheng, Zhengzhou 451100, Henan, China)

【Abstract】Objective To explore a method for extracting features of blood glucose fluctuations in patients with type 2 diabetes mellitus (T2DM) based on continuous glucose monitoring (CGM) data, and to construct a hypoglycemia/hyperglycemia early warning model. **Methods** From October 2024 to October 2025, a total of 246 patients with T2DM were screened in the Department of Endocrinology of Hopeshine-Minsheng Hospital of Xinzheng, and 98 of them were included. All patients wore the CGM system for a duration of at least 14 days, and 12 fluctuation characteristics such as time in range (TIR) and mean amplitude of glycemic excursions (MAGE) were extracted. The extreme gradient boosting (XGBoost) algorithm was used to construct early warning models for hypoglycemia (< 3.9 mmol/L) and hyperglycemia (> 13.9 mmol/L) within the next 4 h, and the training set and the validation set were divided in a 7 vs. 3 ratio. Furthermore, model performance was evaluated by area under the curve (AUC), and feature contribution analysis was performed using Shapley additive explanation (SHAP). **Results** The mean age of patients was (58.6 ± 9.2) years, glycated hemoglobin A1c (HbA1c) was (8.4 ± 1.6)%, MAGE was (4.8 ± 1.7) mmol/L, and TIR was (67.2 ± 18.4)%. The AUC of the hypoglycemia early warning model validation set was 0.91 (95%CI: 0.86–0.95), the sensitivity was 86.7%, and the specificity was 88.4%. The AUC of the hyperglycemia early warning model was 0.89 (95%CI: 0.84–0.93), the sensitivity was 84.2%, and the specificity was 87.1%. The SHAP analysis showed that nocturnal MAGE, the most recent 2-hour blood glucose slope, and the proportion of time below range (TBR) were the core contributing features of hypoglycemia. Meanwhile, the peak 2-h postprandial blood glucose and the proportion of time

收稿日期: 2025-11-26。

作者简介: 睢凤玲, 本科, 副主任医师, 研究方向: 糖尿病。E-mail: 15038001399@163.com。

通信作者: 郭玉婷, 本科, 副主任医师, 研究方向: 糖尿病。E-mail: 15136707991@163.com。

above range (TAR) were characteristics of the core contribution of fasting C-peptide (FCP) to hyperglycemia. **Conclusion** Based on the multi-dimensional fluctuation characteristics of CGM combined with XGBoost algorithm, a high-precision early warning of hypoglycemia and hyperglycemia in T2DM patients can be achieved, and the model has good interpretability.

【**Key words**】 Continuous glucose monitoring; Type 2 diabetes mellitus; Blood glucose fluctuations; Feature extraction; Early warning models; Machine learning

2 型糖尿病 (type 2 diabetes mellitus, T2DM) 发病率持续攀升, 传统指尖血糖监测信息有限, 而血糖波动是独立于糖化血红蛋白的并发症危险因素^[1-2]。连续血糖监测 (continuous glucose monitoring, CGM) 技术为血糖波动分析与风险预警提供了数据基础^[3]。目前, CGM 衍生参数与并发症的相关性已有报道^[4-5], 但基于多参数融合的 T2DM 预警模型仍有限。现有研究多聚焦单一风险或依赖国外人群数据^[6-12]。本研究以 98 例 T2DM 患者为对象, 提取 CGM 波动特征, 构建极端梯度提升 (extreme gradient boosting, XGBoost) 算法双向预警模型并进行沙普利加性解释 (Shapley additive explanation, SHAP) 分析, 旨在为基层医院血糖管理提供参考方案。

1 资料与方法

1.1 一般资料

本研究为回顾性研究设计。2024 年 10 月至 2025 年 10 月于新郑华信民生医院内分泌科筛选 246 例 T2DM 患者, 最终纳入 98 例。排除原因如下: CGM 数据不足 (68 例)、严重肝肾功能不全 (32 例)、资料不完整 (24 例)、拒绝参与 (16 例) 及其他 (8 例)。排除组糖化血红蛋白 (glycated hemoglobin A1c, HbA1c) 高于纳入组, 提示存在选择偏倚。纳入患者男 52 例, 女 46 例; 年龄 (58.6 ± 9.2) 岁; 中位病程 6.0 年; HbA1c (8.4 ± 1.6)%。纳入标准: ① 符合 T2DM 诊断标准; ② 年龄 18 ~ 80 岁; ③ CGM 佩戴 ≥ 14 d 且有效数据 $\geq 70\%$; ④ 临床资料完整。排除标准: ① 1 型糖尿病; ② 急性并发症史; ③ 估算肾小球滤过率 (estimated glomerular filtration rate, eGFR) $< 30 \text{ ml} \cdot \text{min}^{-1} \cdot 1.73 \text{ m}^{-2}$; ④ 妊娠或哺乳期、联用影响血糖药物。本研究经新郑华信民生医院医学伦理委员会批准 (批件号: XZYX-2024-LW-098)。

1.2 CGM 数据采集

采用硅基动感 GS1 系统 (采样间隔 15 min), 配套软件导出数据后剔除异常值, 线性插值填补缺失。

1.3 血糖波动特征提取

提取以下 4 类 12 项特征: 动态葡萄糖图谱

(ambulatory glucose profile, AGP) 参数 [目标范围内时间 (time in range, TIR)、目标范围以上时间 (time above range, TAR)、目标范围以下时间 (time below range, TBR)]、波动幅度 [平均血糖波动幅度 (mean amplitude of glycemic excursions, MAGE)、日间血糖平均绝对差 (mean of daily differences, MODD)、标准差 (standard deviation, SD)、变异系数 (coefficient of variation, CV)]、动态趋势 (日内最大血糖斜率、夜间 MAGE、最近 2 h 血糖斜率) 及衍生指标 [血糖管理指标 (glycemic management indicator, GMI)、低血糖风险指数 (low blood glucose index, LBGi)、高血糖风险指数 (high blood glucose index, HBGi)]。MAGE 采用 Service 法, 最近 2 h 血糖斜率为最小二乘线性回归斜率。

1.4 预警标签定义与模型构建

以未来 4 h 低血糖 ($< 3.9 \text{ mmol/L}$) 或高血糖 ($> 13.9 \text{ mmol/L}$) 为正样本, 按 HbA1c 分层后以 7:3 划分训练集 (69 例) 与验证集 (29 例)。采用 XGBoost 算法, 输入 12 项特征 + 年龄 + HbA1c + 病程 + 空腹 C 肽 (fasting C-peptide, FCP), 设置 scale_pos_weight ≈ 11.2 处理不平衡。低血糖阳性预测值 (positive predictive value, PPV) = 72.3%、阴性预测值 (negative predictive value, NPV) = 94.8%; 高血糖 PPV = 68.5%、NPV = 93.1%。超参数经 5 折交叉验证确定 (学习率 0.05、最大树深 6、子树数 300), 受试者操作特征 (receiver operating characteristic, ROC) 曲线的曲线下面积 (area under the curve, AUC) 均值为 (0.902 ± 0.015)。同时构建逻辑回归 (logistic regression, LR) 与随机森林 (random forest, RF) 作为对照。

1.5 可解释性分析

采用 SHAP 方法计算各特征平均绝对贡献值, 用于解析 XGBoost 模型的全局预测逻辑; 本研究仅汇报 SHAP 特征贡献度排名前 6 位的特征, 未绘制 SHAP 蜂群图及 SHAP 依赖图。

1.6 统计学方法

采用 Python 3.10 及 R 4.3 软件进行数据分析。

符合正态分布的计量资料以均数 $\pm$ 标准差($\bar{x}\pm s$)表示,组间比较采用 t 检验;不符合者以 $M(Q_1, Q_3)$ 表示,组间比较采用 Mann-Whitney U 检验。模型性能以 AUC(95%CI)、敏感度、特异度、准确度、 F_1 评价, DeLong 法比较 AUC 差异。 $P<0.05$ 为差异有统计学意义。

2 结果

2.1 患者基线资料与 CGM 波动参数

98 例 T2DM 患者 CGM 佩戴依从性良好,中位有效数据占比 92.4%(85.6%~97.1%),基线及 CGM 核心参数如下:(1)基线资料:年龄(58.6 ± 9.2)岁;男性 52 例(53.1%);病程 6.0(3.0, 10.0)年;HbA1c(8.4 ± 1.6)%;FCP(1.8 ± 0.9)ng/ml。(2)CGM 波动参数:TIR(67.2 ± 18.4)(3.9~10.0 mmol/L,%);TAR(8.1 ± 10.3)

(>13.9 mmol/L,%);TBR(4.3 ± 6.8)(<3.9 mmol/L,%);MAGE(4.8 ± 1.7)mmol/L;MODD(1.9 ± 0.6)mmol/L;CV(28.4 ± 7.1)%;LBGI(2.1 ± 1.8);HBGI(6.7 ± 4.2)。

2.2 预警事件分布与模型整体性能

98 例患者共生成有效样本 4 704 条(低血糖正样本 387 条,高血糖正样本 421 条)。表 1 显示三种算法在验证集的性能对比。DeLong 检验显示 XGBoost 与 RF、LR 的 AUC 差异均有统计学意义(均 $P<0.05$)。Brier 评分低血糖 XGBoost 模型 0.089、高血糖 XGBoost 模型 0.102,优于 LR(0.134、0.147);校准曲线在 0.1~0.9 区间与实际频率基本吻合(Hosmer-Lemeshow $P=0.342$),极端区间存在轻度高估。图 1 为验证集 ROC 曲线。

表 1 不同算法模型在验证集的预警性能比较

预警类型	模型	AUC (95%CI)	准确度 (%)	敏感度 (%)	特异度 (%)	F_1 (%)
低血糖	XGBoost	0.91 (0.86~0.95)	87.5	86.7	88.4	85.1
	RF	0.88 (0.83~0.93)	84.2	82.3	85.6	81.8
	LR	0.82 (0.76~0.88)	79.6	78.1	81.0	77.3
高血糖	XGBoost	0.89 (0.84~0.93)	85.6	84.2	87.1	83.7
	RF	0.86 (0.81~0.91)	82.8	80.5	84.6	81
	LR	0.80 (0.74~0.86)	77.4	75.8	79.2	74.6

注: AUC: 曲线下面积; XGBoost: 极端梯度提升; RF: 随机森林; LR: 逻辑回归。

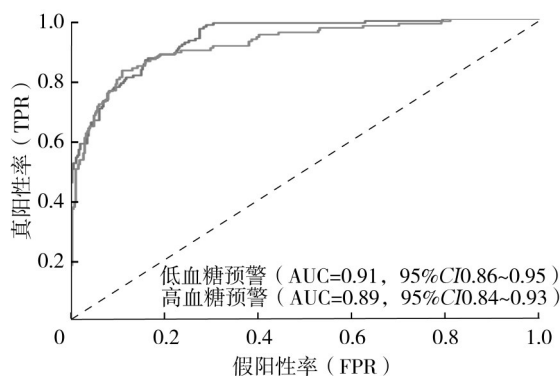


图 1 XGBoost 预警模型验证集 ROC 曲线

注: XGBoost: 极端梯度提升; ROC: 受试者操作特征; AUC: 曲线下面积。

2.3 亚组分析与特征贡献

按 HbA1c 分层后, XGBoost 模型在 HbA1c $>9.0\%$ 组低血糖 AUC 达 0.94, 在 7.0%~9.0% 组为 0.90, $<7.0\%$ 组为 0.87; 高血糖预警仅 $>9.0\%$ 组 AUC 为 0.92(另两组事件稀疏未纳入)。SHAP 分析显示, 夜间 MAGE 越大、最近 2 h 血糖斜率为负时, 低血糖风险显著升高; 餐后 2 h 血糖峰值越高、TAR 占比越

大, 高血糖风险越高。提示模型对血糖控制较差人群预警效能更优(表 2)。

表 2 SHAP 特征贡献度 Top 6 (SHAP 平均值)

排名	低血糖预警	高血糖预警
1	夜间 MAGE (0.38)	餐后 2 h 血糖峰值 (0.41)
2	最近 2 h 血糖斜率 (0.31)	TAR 占比 (0.36)
3	TBR 占比 (0.27)	FCP (0.29)
4	HbA1c (0.24)	HbA1c (0.26)
5	LBGI (0.21)	HBGI (0.23)
6	年龄 (0.14)	病程 (0.17)

注: SHAP: 沙普利加性解释; MAGE: 平均血糖波动幅度; TAR: 目标范围以上时间; TBR: 目标范围以下时间; FCP: 空腹 C 肽; HbA1c: 糖化血红蛋白; LBGI: 低血糖风险指数; HBGI: 高血糖风险指数。

3 讨论

本研究基于 98 例 T2DM 患者 14 d CGM 数据, 提取 12 维波动特征, 构建 XGBoost 双向预警模型, 验证集 AUC 分别达 0.91 与 0.89, 并通过 SHAP 实现临床可解释性。与国际研究相比, Lutsker 等^[7] 的 CGM 基础模型泛化能力强但难以迁移至小样本场景; Carletti 等^[8] 多模态 AI 未专门区分 T2DM 亚群;

Metwally 等^[9]为本研究“特征-风险”映射思路提供了参照。国内层面,汪俊等^[5]提示 CGM 参数可作为并发症“代理标志物”,本研究将其拓展至实时预警;程梅等^[6]的管理效果验证可与本研究预警模型互补。模型在 HbA1c>9.0% 亚组 AUC 最高,提示对血糖失控人群预警收益最大^[13]。未来临床落地路径:① 医院信息系统接口自动化接入 CGM 数据;② 每日定时推理输出风险概率;③ 护士站可视化分级预警;④ 闭环反馈用于模型迭代。拟开展多中心 cluster RCT 验证临床效用,主要结局为严重低血糖发生率,假报率容忍上限设为 30%。

本研究局限性:① 单中心、样本量偏小($n=98$),需多中心验证;② 14 d 佩戴期未捕捉长期变化;③ 4 h 预警窗口为经验性设定,更短时窗样本稀疏问题突出;④ 未融合饮食、运动等多模态数据。

综上,基于 CGM 多维度波动特征的 XGBoost 双向预警模型在 T2DM 管理中具备较高精度与可解释性,尤其适用于血糖波动大、HbA1c 控制不佳人群,后续需多中心前瞻性验证。

参考文献

- [1] 中华医学会糖尿病学分会. 中国血糖监测临床应用指南(2021 年版)[J]. 中华糖尿病杂志, 2021, 13 (10): 936-948.
- [2] 中华护理学会糖尿病护理专业委员会, 国家高性能医疗器械创新中心, 张明霞, 等. 动态葡萄糖图谱报告护理应用专家共识(2025 版)[J]. 中华糖尿病杂志, 2025, 17 (3): 311-321.
- [3] 邹大进, 张征, 纪立农. 缓解 2 型糖尿病中国专家共识[J]. 中国糖尿病杂志, 2021, 29 (9): 641-652.
- [4] 张苗苗, 宋锦, 朱晓秋, 等. 连续血糖监测技术在血糖管理中的研究进展[J]. 中华危重症医学杂志电子版, 2024, 17 (1): 62-67.
- [5] 汪俊, 刘德翠, 孙艳雷. 动态血糖监测参数与 2 型糖尿病周围神经病变患者神经传导、炎症因子及病情严重程度的关系[J]. 检验医学与临床, 2025, 22 (22): 3090-3094.
- [6] 程梅, 赵进东, 窦丽娜, 等. 持续动态血糖监测联合健康教练技术对 2 型糖尿病患者的影响研究[J]. 安徽医学专学报, 2026, 25 (1): 125-128.
- [7] LUTSKER G, SAPIR G, SHILO S, *et al.* A foundation model for continuous glucose monitoring data[J]. Nature, 2026, 650 (8103):978-986.
- [8] CARLETTI M, PANDIT J, GADALETA M, *et al.* Multimodal AI correlates of glucose spikes in people with normal glucose regulation, pre-diabetes and type 2 diabetes[J]. Nat Med, 2025, 31(9):3121-3127.
- [9] METWALLY A A, PERELMAN D, PARK H, *et al.* Prediction of metabolic subphenotypes of type 2 diabetes via continuous glucose monitoring and machine learning[J]. Nat Biomed Eng, 2025, 9(8):1222-1239.
- [10] SHAO J, PAN Y, KOU W B, *et al.* Generalization of a deep learning model for continuous glucose monitoring-based hypoglycemia prediction: algorithm development and validation study[J]. JMIR Med Inform, 2024, 12:e56909.
- [11] SUGIMOTO H, HIRONAKA K, NAKAMURA T, *et al.* Improved detection of decreased glucose handling capacities via continuous glucose monitoring-derived indices[J]. Commun Med, 2025, 5(1):103.
- [12] HALEEM M S, KATSAROU D, GEORGA E I, *et al.* A multimodal deep learning architecture for predicting interstitial glucose for effective type 2 diabetes management[J]. Sci Rep, 2025, 15(1):27625.
- [13] SERVICE F J, MOLNAR G D, ROSEVEAR J W, *et al.* Mean Amplitude of Glycemic Excursions, a Measure of Diabetic Instability[J]. Diabetes, 1970, 19(9):644-655.