

基于机器学习的 2 型糖尿病发病因素分析与预测

成路平¹ 吴思扬² 路 波¹

(¹ 陕西省中医医院 西安 710004 ² 西北工业大学管理学院 西安 710129)

〔摘要〕 **目的/意义** 构建多维数据挖掘预测框架,提升 2 型糖尿病风险预测准确率与临床决策效率。**方法/过程** 基于 Pima 数据集,通过单、双、多变量分析筛选核心因素,采用逻辑回归、随机森林、支持向量机、极限梯度提升和轻量梯度提升机 5 种机器学习模型建模,经网格搜索与交叉验证优化参数。**结果/结论** 识别出血糖水平、身体质量指数、年龄等核心风险因素,与传统循证结论相符;随机森林预测准确率达 0.870 1,整体性能最优。通过数据挖掘与特征筛选,降低了数据采集成本,缩短了风险因素识别周期,揭示了变量间非线性交互机制,为社区高危人群普筛提供了高效工具。

〔关键词〕 2 型糖尿病; 风险预测; 多维数据挖掘; 机器学习

〔中图分类号〕 R-058 **〔文献标识码〕** A **〔DOI〕** 10.3969/j.issn.1673-6036.2025.09.003

Analysis and Prediction of Risk Factors of Type 2 Diabetes Mellitus Based on Machine Learning

CHENG Luping¹, WU Siyang², LU Bo¹

¹ Shaanxi Provincial Hospital of Traditional Chinese Medicine, Xi'an 710004, China; ² School of Management, Northwestern Polytechnical University, Xi'an 710129, China

〔Abstract〕 **Purpose/Significance** To construct a multidimensional data mining prediction framework, and to enhance the accuracy of risk prediction of type 2 diabetes mellitus (T2DM) and the efficiency of clinical decision-making. **Method/Process** Based on the Pima dataset, univariate, bivariate, and multivariate analyses are conducted to screen core risk factors. Five machine learning models, namely logistic regression, random forest, support vector machine, extreme gradient boosting and light gradient boosting machine, are employed for modeling. Hyperparameter optimization is performed using grid search and cross-validation. **Result/Conclusion** The identified key risk factors such as blood glucose level, body mass index, and age are consistent with conclusions from traditional evidence-based medicine. The prediction accuracy of random forest reaches 0.870 1, and the overall performance is the best. By data mining and feature selection, the cost of data collection is reduced, the cycle of risk factor identification is shortened, and the nonlinear interaction mechanism among variables is revealed, providing an efficient tool for the general screening of high-risk groups in the community.

〔Keywords〕 type 2 diabetes mellitus (T2DM); risk prediction; multidimensional data mining; machine learning

〔修回日期〕 2025-08-06

〔作者简介〕 成路平, 主治医师, 发表论文 6 篇; 通信作者: 路波。

〔基金项目〕 陕西省高水平中医药重点学科(中医内分泌病学)(项目编号: SX2YY2DXK-2024006); 陕西省卫生健康科研创新能力提升计划(项目编号: 2025YF-41)。

1 引言

以糖代谢紊乱为特征的 2 型糖尿病已成为全球重大公共卫生问题^[1]。近年来,受生活方式和饮食习惯变化影响,中国 2 型糖尿病患病率持续上升,严重影响患者生活质量,也给社会经济与医疗体系造成沉重负担^[2-3]。

2 型糖尿病的发生受年龄、生活方式、血糖水平、遗传背景、身体质量指数等多种生物学和环境因素影响^[4]。现有研究多聚焦单一变量影响,或者基于传统统计方法探索变量间线性关系,如血糖^[5]、血压^[6-8]、家族遗传史^[9]、运动量^[10-12]与患病的直接联系。然而,2 型糖尿病致病机制高度复杂,因素间存在非线性关系与交互效应,传统方法难以揭示此类高维复杂关联^[13-17]。

近年来数据挖掘技术在医疗领域广泛应用,为高维数据分析与疾病机制探索提供了新路径^[18-20],尤其在 2 型糖尿病预测建模、变量筛选与机制揭示中展现出显著潜力^[21-24]。但现有研究仍存在不足:一是多数研究未全面捕捉多变量交互,疾病机制解释力弱^[25-27];二是部分研究在数据预处理、模型构建和结果验证等环节规范性欠缺,影响结论可靠性与应用价值^[28-29]。

本研究基于公开 2 型糖尿病数据集,挖掘关键影响因素及其交互关系,筛选核心患病特征。借助逻辑回归(logistic regression, LR)、随机森林(random forest, RF)、支持向量机(support vector machine, SVM)、极限梯度提升(extreme gradient boosting, XGBoost)和轻量梯度提升机(light gradient boosting machine, LightGBM) 5 种机器学习模型,探索特征筛选方法的有效性,比较患病预测算法的准确性,通过多变量交互分析弥补传统研究局限,为 2 型糖尿病早期诊断、个性化预防及慢性病预测建模提供科学支持与方法论参考。

2 数据与方法

2.1 数据来源

采用 Pima 印第安人糖尿病数据集。该数据集由

美国国家糖尿病、消化和肾脏疾病研究所构建,包含 768 名 21 岁以上女性的临床数据。受生活方式、饮食习惯及遗传等因素影响,该群体 2 型糖尿病发病率较高,是研究多因素交互作用的典型样本。

数据集涵盖妊娠次数、血糖水平等 8 项生理指标自变量和一个二分类结果变量(是否患 2 型糖尿病),见表 1。其中糖尿病谱系函数通过家族患病情况及遗传关系等量化个体遗传风险。

表 1 变量定义

变量	具体内容	单位
妊娠次数	妊娠次数的整数值	次
血糖水平	血糖水平换算后的浮点数值 (原整数值除以 18)	mmol/L
舒张压	舒张血压的整数值	mmHg
皮肤褶皱厚度	皮肤褶皱厚度的整数值	mm
胰岛素水平	饭后 2 小时血清胰岛素的整数值	μU/mL
身体质量指数	身体质量指数的浮点数值	kg/m ²
糖尿病谱系函数	反映糖尿病家族遗传倾向的浮点数值	无
年龄	年龄的整数值	岁
是否患 2 型糖尿病	0 表示未患病, 1 表示患病	无

2.2 分析方法 (图 1)

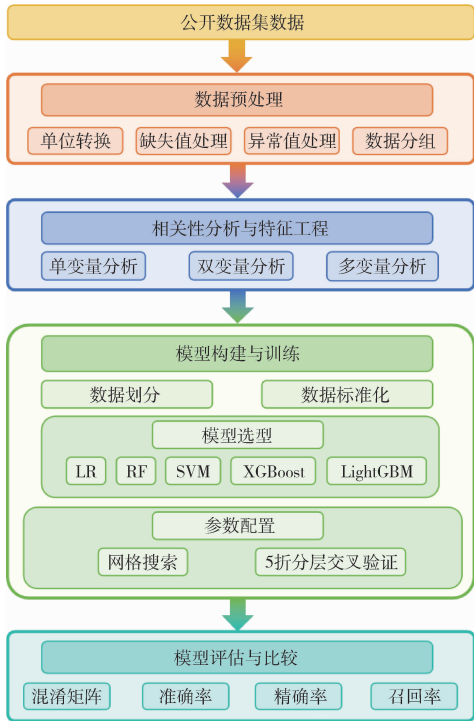


图 1 2 型糖尿病预测研究技术路线

2.2.1 数据预处理 为保障数据质量，对原始数据进行单位统一以及缺失值、重复值、异常值等处理。血糖水平依据 $\text{mmol/L} = \text{mg/dL}/18$ 转换为国际单位 mmol/L 。经核查，数据无缺失值和重复样本，具备完整性与独立性。异常值处理分两步：先以中位数填补法修正生理学不合理的极端 0 值，减少其干扰；再计算各变量四分位距（ $\text{inter} - \text{quartile range}$, IQR）。其中， Q_1 为下四分位数， Q_3 为上四分位数。随后，依据统计学常用的 1.5 倍 IQR 准则定义正常取值范围，即样本值在 $[Q_1 - 1.5 \times \text{IQR}, Q_3 + 1.5 \times \text{IQR}]$ 区间内时为正常，超出该范围的样本判定为异常值并剔除，以此保证数据分布合理。

IQR = Q₃ - Q₁ (1)

为便于后续分析与模型解释，依据临床标准及

数据分布特征，对主要连续变量分段处理，见表 2。核心分组逻辑如下：妊娠次数按临床常见区间划分；血糖水平结合临床诊断标准与数据特征界定；舒张压参照《美国高血压预防、检测、评估及治疗联合委员会第七次报告》分组；皮肤褶皱厚度按临床常见范围划分；胰岛素水平依据临床评估分组，其中胰岛素抵抗指组织对胰岛素敏感性降低，高胰岛素血症通常为其代偿性结果；身体质量指数遵循世界卫生组织标准分组；糖尿病谱系函数参照临床研究共识划分风险等级；年龄分组以临床常规习惯为基础，结合本数据集特征微调，与世界卫生组织标准（世界卫生组织定义青年 <44 岁、中年 45 ~ 59 岁）略有差异，以适配本数据集多变量交互分析需求。

表 2 变量分组标准

变量	分组标准
妊娠次数	0 次；1 ~ 2 次；3 ~ 5 次；> 5 次
血糖水平	<3.9 mmol/L（正常）；3.9 ~ 7.8 mmol/L（较高血糖）；>7.8 mmol/L（高血糖）
舒张压	<80 mmHg（正常）；80 ~ 90 mmHg（高血压前期）；>90 mmHg（高血压）
皮肤褶皱厚度	<10 mm（较薄皮褶）；10 ~ 30 mm（中等皮褶）；>30 mm（较厚皮褶）
胰岛素水平	<50 μU/mL（正常）；50 ~ 100 μU/mL（胰岛素抵抗）；>100 μU/mL（高胰岛素血症）
身体质量指数	<18.5 kg/m ² （体重过低）；18.5 ~ 24.9 kg/m ² （正常，含 24.9 kg/m ² ）；24.9 ~ 29.9 kg/m ² （超重，不含 24.9 kg/m ² ）；>29.9 kg/m ² （肥胖）
糖尿病谱系函数	<0.5（低风险）；0.5 ~ 1.5（中等风险）；>1.5（高风险）
年龄	<30 岁（青年）；30 ~ 45 岁（中青年，含 45 岁）；45 ~ 60 岁（中年，不含 45 岁）；>60 岁（老年）

2.2.2 相关性分析 通过皮尔逊相关系数、卡方检验等统计方法量化变量间的关联强度与协同效应，剔除冗余变量、降低多重共线性影响，筛选显著独立影响是否患病的核心特征，提升预测准确性与临床解释力。首先，基于皮尔逊相关系数开展单变量相关性分析，评估各变量与糖尿病患病状态之间的线性关联强度：

r(X_i, Y) = Cov(X_i, Y) / (σ_{X_i}σ_Y) (2)

其中， X_i 代表各候选变量， Y 为患病标签， Cov 为协方差， σ 为标准差。以 $|r| > 0.2$ 为阈值筛选显著相关变量，初步识别出血糖、身体质量指数、年龄等 6 个关键变量。其次，双变量交互分析，通过构建交叉表和计算卡方统计量，探究变量间的协同效应：

χ² = ∑ (O_{i,j} - E_{i,j})² / E_{i,j} (3)

其中， $O_{i,j}$ 表示第 i 行第 j 列的实际观测频数， $E_{i,j}$ 表示第 i 行第 j 列的理论期望频数， $\sum$ 表示对交叉表中所有单元格求和。该公式衡量实际与理论值差异程度，差异越大关联越显著。最后，多变量分析，借助三维可视化，通过绘制所有变量对（ $C_s^2 = 28$ ）与患病状态的三维散点图，展示变量间的交互效应：

Z = f(X_i, X_j, Y) (4)

其中， X_i 、 X_j 为自变量对， Y 为患病状态。可视化结果与前期筛选一致，最终确定妊娠次数、血糖水平、身体质量指数、年龄、胰岛素水平与皮肤褶皱厚度为建模变量。

2.2.3 模型构建与训练 为准确预测 2 型糖尿病，采用 LR、RF、SVM、XGBoost 和 LightGBM 这 5 种机器学习模型建模。先将数据集划分为特征变量 X (6 个核心自变量) 和目标变量 Y (患病状态)。接着采用 $Z - score$ 标准化方法消除变量间量纲差异。

$$x_{new} = \frac{x - \mu}{\sigma} \tag{5}$$

其中， μ 为样本均值， σ 为样本标准差，确保标准化后特征服从均值为 0、标准差为 1 的正态分布。采用网格搜索结合 5 折分层交叉验证的方法对 5 种模型进行超参数优化。为每个模型定义多维参数搜索空间，LR 重点优化正则化系数 C 、正则化类型及求解器；RF 关注树数量、最大深度、节点分裂阈值等参数；SVM 以径向基核函数为基础，搜索惩罚系数 C 与核系数 γ ；XGBoost 和 LightGBM 则围绕学习率、树深度、迭代次数和正则化参数展开优化。交叉验证过程采用 StratifiedKFold 进行分层抽样，将数据集划分为 5 个子集，其中 1 个子集作为验证集、4 个子集作为训练集。对每个参数组合执行训练及验证循环，每次使用不同随机种子划分的分层抽样子集。

2.2.4 模型评估与比较 选取测试集上准确率最高、泛化能力最强的参数组合参与模型比较。依据混淆矩阵，计算准确率 (accuracy)、精确率 (precision)、召回率 (recall) 作为评估指标。其中混淆矩阵通过真实标签与模型预测标签的交叉分类，量化模型预测结果的分布。将患病状态定义为“1”，未患病状态定义为“0”。真阳性 (true positive, TP) 是真实标签为“1”、模型预测标签也为“1”的样本数量，即模型正确识别的患病者数；真阴性 (true negative, TN) 是真实标签为“0”、模型预测标签也为“0”的样本数量，即模型正确识别的未患病者数；假阳性 (false positive, FP) 是真实标签为“0”、模型预测标签为“1”的样本数量，即模型误判为患病者的未患病者数，也称误诊；假阴性 (false negative, FN) 是真实标签为“1”、模型预测标签为“0”的样本数量，即模型误判为未患病者的患病者数，也称漏诊。准确率是模型预测正确的样本数占总样本数的比例，反映模型整体预测的正确性。精确率是

模型预测为“1”的样本中真实标签为“1”的样本数，反映模型预患病结果的可靠性。召回率是真实为“1”的样本中，被模型正确预测的样本数占比，反映模型识别患病者的全面性。

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{6}$$

$$Precision = \frac{TP}{TP + FP} \tag{7}$$

$$Recall = \frac{TP}{TP + FN} \tag{8}$$

3 结果分析

3.1 相关性分析

3.1.1 单变量与患病相关性分析 通过皮尔逊相关系数量化各变量与患病状态的线性关联强度，并借助热力图可视化呈现变量间相关关系，见图 2。结果显示，血糖水平与患病状态的相关系数达 0.49，呈显著正相关，提示血糖升高会增加患病风险；舒张压与患病状态的相关系数仅 0.17，关联较弱。进一步基于相关系数阈值筛选，考虑自变量临床意义与共线性影响，部分自变量存在弱线性关联，未达到强共线性干扰程度，最终识别出妊娠次数、血糖水平、皮肤褶皱厚度、胰岛素水平、身体质量指数、年龄共 6 个关键变量，为后续筛选核心预测因子提供依据。同时，绘制各变量分组患病比例，见图 3，以观察不同分组患病比例差异。

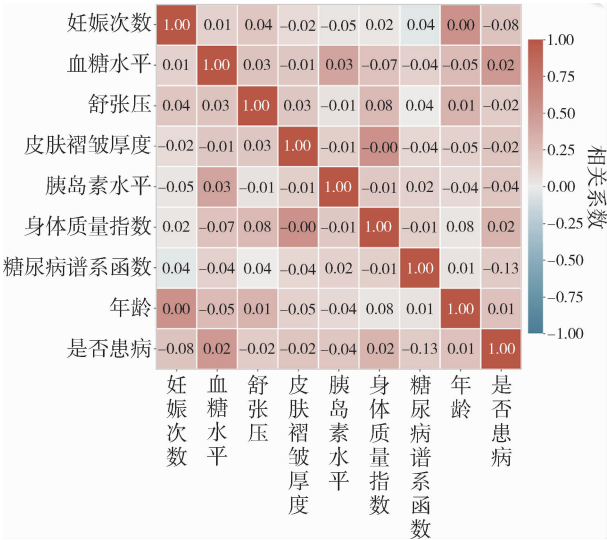


图2 变量间皮尔逊相关系数矩阵

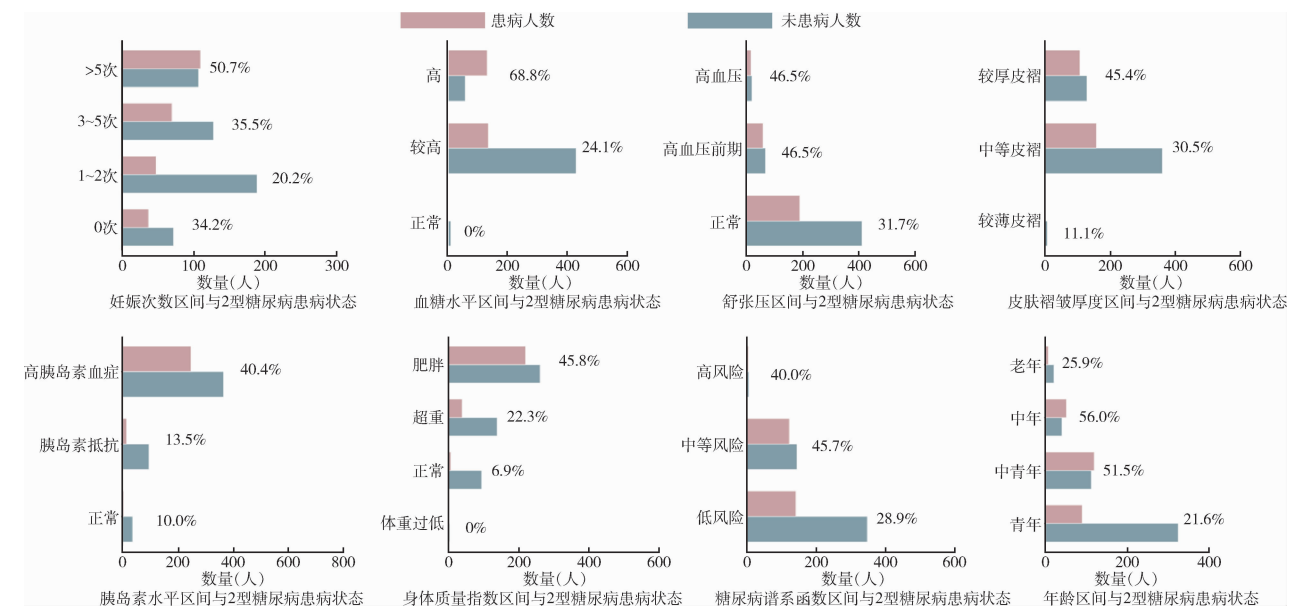


图 3 各变量分组患病比例

以妊娠次数为例，随着妊娠次数增加，2 型糖尿病患病率总体上升，其中妊娠次数大于 5 次的群体患病率显著高于其他组别，印证妊娠次数是重要影响因素。分析结果表明，多项变量与患病风险密切相关且分布趋势具有临床意义：血糖水平为最显著预测因子，正常血糖时患病率为 0，高血糖时近 70%；身体质量指数与糖尿病强关联，肥胖人群患病率达 45.8%；年龄维度中，中青年及中年人群患病率均超 50%，显著高于青年组。此外，胰岛素水平、皮肤褶皱厚度、糖尿病谱系函数与患病风险在不同水平上呈正相关；舒张压关联相对较弱，但仍呈一定正相关。变量分组分析结果与相关系数分析结果一致，印证了关键变量筛选的有效性。

3.1.2 双变量相关性分析 对预先分组的危险因素进行双变量分析，以系统探究其关联以及与 2 型糖尿病患病风险的关联程度。基于卡方检验与交互分布分析，在 28 对变量中，妊娠次数分组 - 年龄分组等 12 对呈高度显著关联 ($P < 0.001$)，妊娠次数分组 - 糖尿病谱系函数分组等 7 对呈显著关联 ($0.001 \leq P < 0.01$)，胰岛素水平分组 - 身体质量指数分组等 3 对呈边缘显著关联 ($0.01 \leq P < 0.05$)，余下 6 对无显著关联 ($P \geq 0.05$)。多组变量通过代

谢、衰老及病理网络产生协同作用，老年群体中舒张压异常者患病率更高，体现衰老与血压异常的协同作用。而高血糖、肥胖是糖尿病发病核心表型，身体质量指数与皮肤褶皱厚度关联，提示肥胖通过脂肪分布、炎症微环境加剧胰岛素抵抗。血糖与胰岛素分组分布的协同，契合高血糖到高胰岛素再到胰岛素抵抗的恶性循环；老年群体中，舒张压异常与糖尿病遗传易感性协同，损伤内皮功能，放大衰老代谢损伤，加速糖尿病进程。

3.1.3 多因素相关性分析 以多变量交互患病率为干预组、单变量患病率为基线组，探究 8 个核心生理变量与 2 型糖尿病患病状态的关系。结果显示，血糖、身体质量指数和年龄是主要风险因素且存在协同作用。高血糖个体患病风险显著升高，若同时合并肥胖，风险较单一高血糖个体提高约 80%，表明肥胖可能加剧胰岛素抵抗，放大血糖异常危害；中老年肥胖人群患病风险较青年提高近 3 倍，提示年龄相关代谢退化（如胰岛素敏感性下降、线粒体功能减退）在发病中起关键作用。此外，糖尿病谱系函数升高者在高血糖状态下患病率明显高于单纯高血糖人群，印证遗传易感性与环境因素协同致险。已有研究^[30]指出 TCF7L2 等基因变异与胰岛素分泌障碍密切相关，与本研究结果一

致。胰岛素水平与皮肤褶皱厚度正相关，支持胰岛素抵抗通过影响脂肪分布参与代谢紊乱的机制。部分研究^[31]表明，脂肪组织受 PPAR- γ 通路调控异常时易发生结构性扩张，本研究数据支持该机制。值得注意的是，高血糖、超重和高血压兼具的个体 2 型糖尿病患病率近 90%，远超单一因素风险；高龄（老年组）合并高血压者患病风险较年轻且血压正常者高约 45%，暗示血管老化和胰岛素信号障碍可能存在潜在关联。上述发现表明，2 型糖尿病的发生是代谢状态、年龄及遗传背景等多因素联合作用的结果，所选变量有效捕捉了核心风险因素及其交互作用。

3.2 预测性能分析

3.2.1 模型参数优化过程 为确保 5 种模型性能稳定且最优，采用网格搜索结合 5 折分层交叉验证进行参数优化，核心参数搜索空间及最优结果如下。LR 搜索空间涵盖正则化系数 $C = [0.001, 0.01, 0.1, 1, 10, 100]$ 、正则化类型 $\text{penalty} = ['l1', 'l2', 'elasticnet']$ 等。经迭代验证，最优参数为 $C = 0.1$ 、 $\text{penalty} = 'l2'$ 、 $\text{solver} = 'liblinear'$ ，交叉验证准确率达 0.863 6。RF 搜索空间包括树数量 $n_estimators = [50, 100, 150]$ 、最大深度 $\text{max_depth} = [None, 5, 10, 15]$ 等。最优参数为 $n_estimators = 50$ 、 $\text{max_depth} = 5$ 、 $\text{min_samples_leaf} = 1$ 、 $\text{min_samples_split} = 5$ 、 $\text{criterion} = 'gini'$ 、 $\text{max_features} = 'sqrt'$ ，交叉验证准确率达 0.870 1。SVM 搜索空间为惩罚系数 $C = [0.01, 0.1, 1, 10, 100]$ 、核系数 $\gamma = ['scale', 'auto', 0.1, 1]$ 等。最优参数为 $C = 1$ 、 $\gamma = 'scale'$ 、 $\text{kernel} = 'rbf'$ ，交叉验证准确率达 0.850 6。XGBoost 搜索空间涵盖树深度 $\text{max_depth} = [3, 5, 7, 9]$ 、学习率 $\text{learning_rate} = [0.01, 0.05, 0.1, 0.2]$ 等。最优参数为 $\text{max_depth} = 3$ 、 $\text{learning_rate} = 0.1$ 、 $\text{col_sample_bytree} = 0.8$ 、 $\text{min_child_weight} = 1$ 、 $n_estimators = 50$ 、 $\text{reg_alpha} = 0.1$ 、 $\text{reg_lambda} = 1$ 、 $\text{subsample} = 0.8$ 、 $\text{objective} = 'binary: logistic'$ ，交叉验证准确率为 0.844 2。LightGBM 搜索空间包括

$\text{num_leaves} = [10, 31, 63]$ 、 $\text{subsample} = [0.6, 0.8, 1.0]$ 等。最优参数为 $\text{num_leaves} = 10$ 、 $\text{subsample} = 0.6$ 、 $\text{max_depth} = 3$ 、 $\text{learning_rate} = 0.1$ 、 $n_estimators = 50$ 、 $\text{colsample_bytree} = 1.0$ 、 $\text{reg_alpha} = 1$ 、 $\text{reg_lambda} = 0.1$ ，交叉验证准确率达 0.867 1。

上述优化过程中，各模型均通过 10 次随机种子重复实验消除随机性，最终选取泛化能力最优的参数组合，以随机种子数 6 115 情况下的实验结果用于后续性能评估。

3.2.2 最优参数模型的预测性能分析 基于预处理后的数据及 6 项核心特征，采用各模型最优参数，评估 5 种模型在 2 型糖尿病预测中的表现，根据混淆矩阵计算准确率、精确率和召回率。不同模型对非患病（类别 0）与患病（类别 1）样本分类能力不同。其中，RF 的 TN 为 96 例、TP 为 38 例，两类样本识别均衡性最优；XGBoost 的 FN 仅 10 例，漏诊风险最低，见图 4。



图 4 各模型混淆矩阵对比

各模型性能雷达图，见图 5。各模型准确率均处于较高水平，RF 以 0.870 1 的准确率领先，整体预测性能最优。由于数据类别不平衡，5 种模型健康样本识别的效能普遍优于患病样本。实际应用中应根据临床场景选择模型，结合混淆矩阵计算召回率（召回率 = 正确识别数/实际总数），类别 0 上 SVM、LightGBM 的召回率更高，分别为 0.934 9 和 0.943 4，可有效降低初筛误诊带来的医疗负担；类别 1 上 XGBoost 的召回率最优，达 0.795 9，漏诊风险最低，更适配精准干预场景。

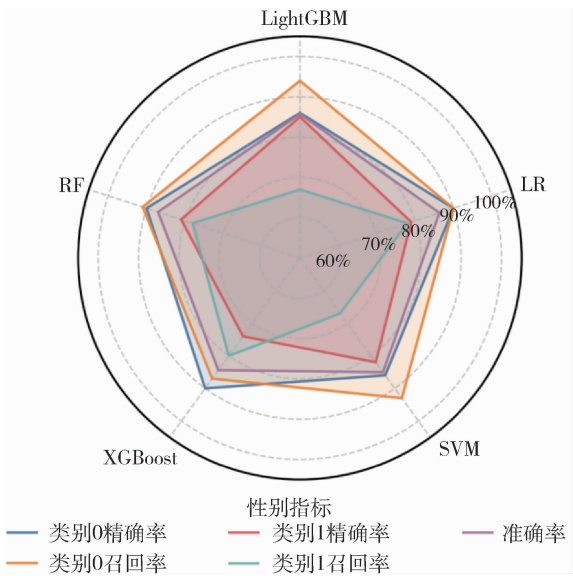


图 5 各模型性能雷达图

上述结果印证了前期特征筛选的有效性，利用筛选后特征建立的模型均能高效预测是否患有 2 型糖尿病。其中，RF 性能均衡优异，整体准确率最高，达 0.870 1；0 类精确率达 0.897 2，召回率达 0.914 3；1 类精确率达 0.808 5，召回率达 0.775 5。

4 结论

本研究通过机器学习驱动的多维数据挖掘框架，系统解析 2 型糖尿病核心风险因素及其非线性交互机制，构建高性能预测模型，为疾病早期预警、精准干预以及临床决策提供科学依据。主要结论如下。一是血糖水平、身体质量指数和年龄是 2 型糖尿病风险预测的 3 大核心变量。高血糖对患病风险的独立影响显著，中老年与肥胖协同效应可使患病风险提升数倍，突显代谢综合征累积危害；胰岛素水平与皮肤褶皱厚度正相关，揭示脂肪分布异常与胰岛素抵抗潜在联系；高遗传风险与高龄（老年组）交互，揭示遗传在代谢衰老中的放大机制。二是相较于传统循证医学，本研究提出的单变量、双变量、多变量渐进式分析路径，在复现经典医学结论的同时，仅用 6 项核心指标，降低了数据采集成本，缩短了风险识别周期。基于筛选特征构建的 LR、RF、SVM、XGBoost、LightGBM 5 种机器学习

模型中，RF 综合性能最优，整体准确率达 0.870 1，两类标签的各项指标优异且均衡；SVM、LightGBM 排除非患者效能突出，但对患病样本敏感度不足；XGBoost 展现低漏诊特性，假负例最少，可作为辅助诊断工具。受数据集标签非均衡性限制，本研究对患病状态的预测精度仍有提升空间。后续可引入先进深度学习框架，结合多组学数据，进一步提升预测准确性与稳定性，推进临床转化应用。

作者贡献：成路平负责研究设计、论文撰写；吴思扬负责数据分析、模型搭建与验证；路波负责数据处理、论文修订。

利益声明：所有作者均声明不存在利益冲突。

参考文献

- 1 卿雪莲, 陈小玉, 廖娟, 等. 回授法联合微信全程健康教育对糖尿病患者自我效能与医疗依从性的影响 [J]. 中华全科医学, 2025, 23 (5): 814 - 818.
- 2 杨予青, 陈荃, 何琪乐, 等. 北京市东城区老年糖尿病患者自我管理水平及其影响因素研究 [J]. 医学信息学杂志, 2024, 45 (1): 59 - 63, 88.
- 3 傅乐斌, 时程遥, 徐佳玲, 等. 2 型糖尿病从气耗阴伤论治的研究概况 [J]. 江西中医药, 2025, 56 (5): 61 - 66.
- 4 HE Q, CHEN X, ZEH P. Promoting self - care and management for patients with type 2 diabetes based on lifestyle changes under the concept of person - centered care: a review [J]. Alternative therapies in health and medicine, 2023, 29 (5): 302 - 307.
- 5 GAO X, ZHANG X, CAI R, et al. Advanced maternal age, overweight and obese positively correlate to the abnormal plasma glucose among gestational diabetes mellitus women even with physical exercise >90 min/day: a prospective cohort study in Shanghai [J]. Scientific reports, 2025, 15 (1): 21191.
- 6 LI C, YU H, ZHU Z, et al. Association of blood pressure with incident diabetic microvascular complications among diabetic patients: longitudinal findings from the UK Biobank [J]. Journal of global health, 2023, 13 (3): 04027.
- 7 ALAMRO M N, ALAHMARI N A, BATAIS A M, et al. Social inequalities in hypertension, dyslipidemia, and cardiovascular events among adults with type 2 diabetes: a cross - sectional study from Saudi Arabia [J]. Healthcare,

- 2025, 13 (13): 1480.
- 8 GROSSMAN E, MESSERLI H F. Intensive blood - pressure control in patients with type 2 diabetes [J]. The New England journal of medicine, 2025, 392 (12): 1155 - 1167.
- 9 倪蕾, 陈振龙, 戴霞云. 职业噪声与 2 型糖尿病相关的遗传变异对糖尿病前期的交互作用 [J]. 现代预防医学, 2024, 51 (10): 1760 - 1765.
- 10 COX R E, GAJANAND T, KEATING E S, et al. Effect of low - volume combined aerobic and resistance high - intensity interval training on vascular health in people with type 2 diabetes: a randomised controlled trial [J]. European journal of applied physiology, 2024, 124 (9): 2819 - 2833.
- 11 HAN Y, HAN X Y, HUANG F, et al. The relationship between abdominal fat and sleep quality after combined exercise in patients with type 2 diabetes mellitus [J]. Frontiers in endocrinology, 2025, 16 (6): 1471608.
- 12 WRENCH E, RATTLEY K, LAMBERT J E, et al. There is no dose - response relationship between the amount of exercise and improvement in HbA1c in interventions over 12 weeks in patients with type 2 diabetes: a meta - analysis and meta - regression [J]. Acta diabetologica, 2022, 59 (11): 1399 - 1415.
- 13 李书楠, 张诗妍, 邓亚楠, 等. 1990—2021 年中国 2 型糖尿病肾病的疾病负担变化趋势及预测研究 [J]. 中国全科医学, 2025, 28 (33): 4214 - 4226.
- 14 杨洋, 杜丽娟, 崔燕燕, 等. 基于肠道菌群防治 2 型糖尿病伴抑郁、焦虑研究进展 [J]. 陕西医学杂志, 2024, 53 (9): 1294 - 1297.
- 15 王一凡, 石超君, 马晓洁, 等. 2 型糖尿病并发动脉粥样硬化风险预测模型比较 [J]. 医学信息学杂志, 2024, 45 (7): 74 - 80.
- 16 孟继娴, 刘蕾, 甄紫伊, 等. 2 型糖尿病患者糖尿病肾病发生风险预测模型的研究进展 [J]. 沈阳医学院学报, 2023, 25 (5): 525 - 528, 534.
- 17 曹慧颖, 冯磊, 唐灵通, 等. 遗传风险评分在预测 2 型糖尿病中的进展 [J]. 中华全科医学, 2023, 21 (8): 1383 - 1387.
- 18 彭赛, 卢丛兰, 孙中洋. 人工智能在骨质疏松症诊疗中的应用进展 [J]. 陕西医学杂志, 2024, 53 (3): 425 - 427, 432.
- 19 于立恒, 林锡祥, 陈煦, 等. 人工智能技术在心脏超声常规参数测量及左室舒张性慢性心力衰竭诊断中的应用 [J]. 陕西医学杂志, 2023, 52 (7): 826 - 830.
- 20 李瑞瑶, 许婧怡, 戴浩宇, 等. 基于深度置信网络的 2 型糖尿病微血管并发症预测 [J]. 医学信息学杂志, 2024, 45 (7): 68 - 73.
- 21 吴挺超, 何茗莛, 冯皓月, 等. 基于数据挖掘探讨中药复方治疗气阴两虚型 2 型糖尿病的用药规律 [J]. 中草药, 2024, 55 (14): 4811 - 4823.
- 22 乔岩, 郑利涛, 李金林. 基于数据挖掘的疾病并发症发现方法研究: 以 2 型糖尿病为例 [J]. 数学的实践与认识, 2024, 54 (6): 23 - 30.
- 23 周婧雅, 吕仁和, 陶经纬, 等. 基于数据挖掘和网络药理分析吕仁和教授治疗 2 型糖尿病的用药规律与作用机制 [J]. 现代中医临床, 2024, 31 (2): 109 - 116.
- 24 黄湘宁, 姚心怡, 谭艳, 等. 基于数据挖掘的 2 型糖尿病动物模型应用分析 [J]. 中国实验方剂学杂志, 2023, 29 (2): 159 - 165.
- 25 YANG Z. Application of machine learning in diabetes prediction based on electronic health record data analysis [C]. Online: 2nd International Conference on Data Science, Advanced Algorithm and Intelligent Computing, 2024.
- 26 魏锦轩, 李荣, 杨素珍, 等. 基于数据挖掘探讨高血压合并射血分数保留型心力衰竭的用药规律 [J]. 医学信息学杂志, 2024, 45 (1): 52 - 58.
- 27 ZARCH E M, MASOUD S. Application of machine learning in affordable and accessible insulin management for type 1 and 2 diabetes: a comprehensive review [J]. Artificial intelligence in medicine, 2024, 151 (8): 102868.
- 28 LUO P, HUANG C. Causal associations between type 2 diabetes mellitus, glycemic traits, dietary habits and the risk of pressure ulcers: univariable, bidirectional and multivariable Mendelian randomization [J]. Frontiers in nutrition, 2024, 11 (10): 1375179.
- 29 向俊杰. 机器学习在预测糖尿病中的应用 [J]. 中国科技信息, 2025 (14): 77 - 80.
- 30 郑晓雅, 任伟, 张素华, 等. TCF7L2 基因 rs7903146 和 rs11196218 单核苷酸多态性与新诊断 2 型糖尿病患者早相胰岛素分泌的相关性分析 [J]. 解放军医学杂志, 2011, 36 (3): 269 - 272.
- 31 李秀凉, 吕莹, 孙庆申. PPAR - γ 及其调节剂与脂肪代谢的关系研究进展 [J]. 食品安全质量检测学报, 2019, 10 (17): 5722 - 5727.