

# 基于机器学习的 2 型糖尿病合并高血压影响因素分析模型对比与解释性分析

芮 晨<sup>1,2</sup> 刘浩然<sup>1</sup> 张盈盈<sup>3</sup> 李金斌<sup>4</sup> 张敬晨<sup>5</sup> 张胜发<sup>5</sup>

(<sup>1</sup> 河北大学公共卫生学院 保定 071000 <sup>2</sup> 沧州市中心医院医院管理研究所 沧州 061000

<sup>3</sup> 河北工业大学经济管理学院 天津 300131 <sup>4</sup> 北京理工大学医学技术学院 北京 100081

<sup>5</sup> 中国医学科学院国家人口健康科学数据中心 北京 100730)

**【摘要】 目的/意义** 识别 2 型糖尿病合并高血压的关键影响因素，为高危人群风险预测和个性化干预提供依据。**方法/过程** 抽取 2020—2022 年内分泌科 3 839 例住院患者数据，构建随机森林、支持向量机、极端梯度增强和自然梯度提升 4 种共病影响因素分析模型，并进行对比分析。以 2023 年 1 000 例住院患者数据为验证集，结合 SHAP 分析与年龄分层亚组验证模型性能。**结果/结论** NGBoost 模型在 2 型糖尿病合并高血压预测任务中性能最优，SHAP 分析揭示吸烟、并发症数量、甲状腺疾病等为关键因素，且该模型在 4 个年龄亚组稳定性良好。

**【关键词】** 糖尿病合并高血压；2 型糖尿病；高血压；SHAP 分析；机器学习

**【中图分类号】** R-058 **【文献标识码】** A **【DOI】** 10.3969/j.issn.1673-6036.2025.11.007

## Comparative and Explanatory Analysis of Influencing Factor Models for Type 2 Diabetes Mellitus Complicated with Hypertension Based on Machine Learning

RUI Chen<sup>1,2</sup>, LIU Haoran<sup>1</sup>, ZHANG Yingying<sup>3</sup>, LI Jinbin<sup>4</sup>, ZHANG Jingchen<sup>5</sup>, ZHANG Shengfa<sup>5</sup>

<sup>1</sup> School of Public Health, Hebei University, Baoding 071000, China; <sup>2</sup> Institute of Hospital Management, Cangzhou Central Hospital, Cangzhou 061000, China; <sup>3</sup> School of Economics and Management, Hebei University of Technology, Tianjin 300131, China; <sup>4</sup> School of Medical Technology, Beijing Institute of Technology, Beijing 100081, China; <sup>5</sup> National Population Health Data Center, Chinese Academy of Medical Sciences, Beijing 100730, China

**【Abstract】 Purpose/Significance** To identify the key influencing factors of type 2 diabetes mellitus (T2DM) with hypertension, and to provide a basis for risk prediction and personalized intervention in high-risk populations. **Method/Process** Data of 3 839 inpatients from the department of endocrinology from 2020 to 2022 are extracted. Four analysis models of comorbidity influencing factors, namely random forest, support vector machine (SVM), extreme gradient boosting (XGBoost), and natural gradient boosting (NGBoost), are constructed and compared. The data of 1 000 inpatients in 2023 are used as the validation set. The model performance is verified by combining SHAP analysis and age-stratified subgroup validation. **Result/Conclusion** The NGBoost model shows the optimal performance in predicting T2DM with hypertension. SHAP analysis reveals that smoking, number of complications and thyroid diseases are key influencing factors. Additionally, the model exhibits good stability across four age subgroups.

**【Keywords】** diabetes mellitus with hypertension; type 2 diabetes mellitus (T2DM); hypertension; SHAP analysis; machine learning

**【修回日期】** 2025-11-03

**【作者简介】** 芮晨，博士，高级经济师，发表论文 16 篇；通信作者：张胜发，副研究员。

**【基金项目】** 国家科技重大项目（项目编号：2023ZD0509702）。

## 1 引言

糖尿病是遗传与环境因素交互作用引发的以慢性高血糖为特征的代谢异常疾病,其发病率在全球范围内呈逐渐上升趋势<sup>[1-2]</sup>。其中,2 型糖尿病(type 2 diabetes mellitus, T2DM)占比超 90%,以胰岛素抵抗和相对胰岛素分泌不足为主要病理生理学特征,肥胖、缺乏运动、不健康饮食等是其主要风险因素。高血压是以体循环动脉压升高为主要病理生理学特征的慢性病,既是心血管疾病重要死因<sup>[3]</sup>,也是 T2DM 的常见并发症。二者共病会显著升高心血管事件风险,加速肾脏、视网膜并发症进展,严重危害患者健康<sup>[4-5]</sup>。

近年来,机器学习广泛应用于疾病预测,且取得显著成效<sup>[6]</sup>。Vu T 等<sup>[7]</sup>基于 Suita 研究数据,采用 k-原型聚类分析,用 5 种模型预测脑卒中风险,发现随机森林模型表现最优;Qin Y 等<sup>[8]</sup>利用美国国家健康与营养调查数据库数据比较 5 种模型,发现类别型提升模型在糖尿病预测方面效果最佳。Islam M M 等<sup>[9]</sup>基于埃塞俄比亚受访者数据,用 4 种模型预测高血压患病风险,发现 XGBoost 模型效果最优,并通过 SHAP 分析明确了年龄等关键风险因素。SHAP 是基于博弈论的可解释性方法,能够量化特征对预测结果的贡献度,并揭示特征间交互效应,尤其适用于解析复杂共病的多因素作用机制<sup>[10]</sup>。目前机器学习主要应用于糖尿病、高血压等单一疾病影响因素分析,针对 T2DM 合并高血压的影响因素及解释性研究仍较有限。本研究构建 T2DM 合并高血压影响因素分析模型,并结合 SHAP 解释性分析量化关键因素及其交互作用,为共病患者早期预防策略制定及健康管理实践提供科学依据。

## 2 研究对象与方法

### 2.1 研究对象

以河北省沧州市中心医院内分泌科住院患者为研究对象,筛选 T2DM 合并高血压患者,并以同期其他患者作为参照组。本研究中,T2DM 合并高血压患者

纳入标准为:符合《中国糖尿病防治指南(2024 版)》<sup>[11]</sup>中 T2DM 的诊断标准,确诊为 T2DM 的患者;合并高血压,符合《中国高血压防治指南(2024 年修订版)》<sup>[12]</sup>中的诊断标准,按照血压测量结果,收缩压 $\geq 140\text{mmHg}$ 和/或舒张压 $\geq 90\text{mmHg}$ ;年龄 $\geq 18$ 岁。排除标准为:1 型糖尿病或其他特殊类型糖尿病患者;患严重心脏、脑、肾脏等并发症;孕妇或哺乳期妇女;患精神疾病或有认知障碍,无法配合完成研究;正参与其他临床试验或研究。

### 2.2 数据来源

数据采集自医院电子病历数据库、体检数据库等。按是否同时患 T2DM 与高血压进行筛选,最终纳入 2020—2022 年内分泌科住院患者 3 839 例,其中 T2DM 合并高血压患者 1 515 例,同期其他患者 2 324 例。筛选 2023 年内分泌科 1 000 例住院患者数据作为验证集,其中 T2DM 合并高血压患者 373 例,同期其他患者 627 例。抽取人口统计学、出入院、手术及体检等相关信息,筛选 T2DM 合并高血压的可能影响因素,最终纳入 169 个特征变量。

### 2.3 研究方法

**2.3.1 数据处理** (1) 特征删减。结合特征属性与数据量,剔除出入院时间、出入院病情、手术情况等与 T2DM 合并高血压无关联的 32 个特征。(2) 缺失数据插补。由于并非每名患者均完成了所有检查,检查结果数据中存在一定缺失。利用 Python 6.5.4 scikit-learn 库的 KNNImputer 进行缺失值插补。该方法基于 k 近邻(k-nearest neighbor, KNN)算法,通过距离度量识别相邻点,利用相邻点观测值估计缺失值,可填充连续型与离散型变量,模型唯一可调节参数为 k 值<sup>[13]</sup>。基于交叉验证结果,将 k 值设定为 5。虽然随着数据规模增加,处理速度会下降,但该方法具备较高的可理解性与可实施性<sup>[14]</sup>。

**2.3.2 特征筛选** 采用基于 L1 正则化的逻辑回归模型筛选特征。使用 Z-score 标准化消除量纲差异,并以皮尔逊相关系数矩阵(阈值  $|r| \geq 0.7$ )验证特征间无显著多重共线性,满足模型假设要求。通过网格搜索结合 5 折交叉验证,确定最优正

则化强度  $C = 0.1$ ，筛选回归系数非零且绝对值最高的 15 个重要特征，并剔除冗余变量。该模型的线性系数可直观量化特征对共病风险的全局影响方向与强度，能够快速定位关联特征，并提供初步解释依据。

**2.3.3 模型构建及验证** 以 L1 逻辑回归模型筛选出的 15 个特征为自变量，T2DM 合并高血压为因变量，将样本数据按 7:3 的比例划分训练集和测试集。构建随机森林 (random forest, RF)、支持向量机 (support vector machine, SVM)、极端梯度增强 (extreme gradient boosting, XGBoost) 和自然梯度提

升 (natural gradient boosting, NGBoost) 4 种模型，见表 1。采用 GridSearchCV 策略进行参数调优，并以灵敏度、特异度、准确度、 $F1$  值和受试者操作特征曲线下面积 (area under the curve, AUC) 作为评价指标。基于初步比较结果，对表现最佳的 NGBoost 模型进一步采用 RandomizedSearchCV 模块进行参数调优。以 2023 年 1 000 例患者数据为验证集，验证最优模型，并采用 SHAP 值进行可解释分析，识别影响共病发生的关键因素、特征变量对共病产生的影响程度，以及影响因素之间的关系，最后通过年龄分层亚组验证模型稳定性。

表 1 各模型参数

模型	参数配置
RF	n_estimators = 200, criterion = 'gini', bootstrap = True, max_depth = 20, max_features = 'sqrt', min_samples_leaf = 30, min_samples_split = 10
SVM	kernel = 'linear', probability = True, random_state = 42, C = 0.01
XGBoost	learning_rate = 0.007, n_estimators = 500, max_depth = 3, min_child_weight = 8, gamma = 0.1, subsample = 0.8, colsample_bytree = 0.8, objective = 'binary: logistic', nthread = 4
NGBoost	n_splits = 10, random_state = 42, shuffle = True, estimator = model, param_distributions = param_grid, scoring = 'accuracy', cv = kfold, n_iter = 5, n_estimators = 1 520, learning_rate = 0.3

3 结果

3.1 重要特征变量

以 L1 逻辑回归模型筛选出的 15 个重要特征为 T2DM 合并高血压的影响因素，见图 1。

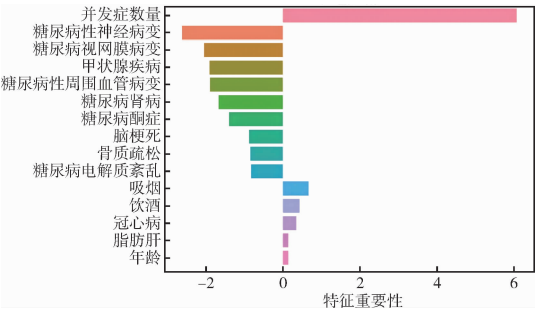


图 1 15 个重要特征及其系数

特征系数可量化特征在模型中的重要性，其正负和绝对值反映特征与共病发生概率的相关性。相关性最高的 3 个特征依次为：并发症数量、糖尿病性神经病变、糖尿病视网膜病变。并发症数量指除

T2DM 与高血压之外，患者所患其他并发症的数量总和。

3.2 数据基本特征

本研究共纳入 2020—2022 年沧州市中心医院内分泌科住院患者 3 839 例。行为因素方面，在验证集和测试集中，非共病组不吸烟患者占比分别为 25.56%、22.65%，共病组无不吸烟患者，且被动吸烟和既往烟史患者占比均高于非共病组；非共病组不饮酒患者占比 69.75%，共病组不饮酒患者占比 30.63%。验证集中趋势与之一致，非共病组被动吸烟、既往烟史、饮酒占比均低于共病组。其他疾病方面，糖尿病性神经病变在共病组发生占比为 55.97%，高于其在非共病组的 51.16%；糖尿病视网膜病变与甲状腺疾病则呈相反趋势，二者在非共病组发生占比分别为 22.98%、37.34%，均高于其在共病组的 19.93%、22.24%；此外，糖尿病性周围血管病变、糖尿病肾病、脑梗死在共病组发生占比依次为 53.07%、31.42%、13.00%，均显著高于三者在非

共病组的 43.07%、26.59%、4.22%。并发症数量方面，非共病组合并并发症数量不少于 4 种的患者占比低于共病组，整体呈现共病组并发症数量更高的特征。验证集中并发症数量累积分布与此一致，随着并发症数量增加，共病组与非共病组的占比差异越来越明显，共病组并发症数量不少于 4 种的患者占比高于非共病组，进一步验证共病组更易呈现多症并发的规律。

3.3 模型比较与验证

NGBoost 模型在本研究的任务中具有最高的预测性能，其次是支持向量机模型和 XGBoost 模型，随机森林模型略差，见表 2。NGBoost 模型在验证集上表现优异，准确率为 0.918 4、灵敏度为 0.914 5、特异度为 0.920 7、F1 为 0.893 9，AUC 达 0.974 5，在区分 T2DM 合并高血压患者与非共病个体方面效能较高。

表 2 各模型预测性能

模型	准确度	灵敏度	特异度	F1	AUC
RF	0.802 1	0.659 3	0.895 3	0.724 6	0.908 0
SVM	0.886 3	0.826 4	0.925 4	0.851 6	0.956 9
XGBoost	0.835 1	0.727 5	0.905 3	0.777 0	0.933 6
NGBoost	0.901 0	0.886 8	0.909 6	0.870 7	0.967 1

3.4 解释性分析

采用 SHAP 解释性分析进一步探究 T2DM 与高血压共病的影响因素及其交互作用。基于平均 SHAP 值排序的重要特征条形图，见图 2。平均 SHAP 值反映特征对模型输出影响的大小，值越大影响越显著。结果显示，吸烟的 SHAP 值最高，对共病影响起关键性作用；其次是并发症数量和甲状腺疾病，两者显著提升共病风险；饮酒、糖尿病性神经病变、糖尿病性周围血管病变 SHAP 值也较高，证实其在共病中的重要作用。以散点图进一步细化特征排序并阐释变量对结果的具体影响，见图 3。Y 轴按照各特征对模型预测贡献大小由上到下依次排列，X 轴的 SHAP 值表示特征取值对预测结果的影响方向与程度，正值表示该特征会使共病概率上升，负值则相反；颜色表示特征值高低。吸烟、并发症数量对共病发生的影响较显著，说明吸烟或并

发症数量越多的个体，在一定程度上共病形成的可能性越大；而脂肪肝、脑梗死等位于图底部的特征，SHAP 值较低，对模型预测贡献较小。

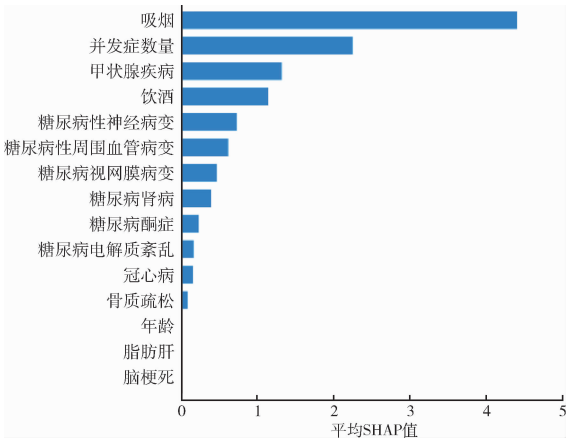


图 2 最优模型的平均 SHAP 值排序

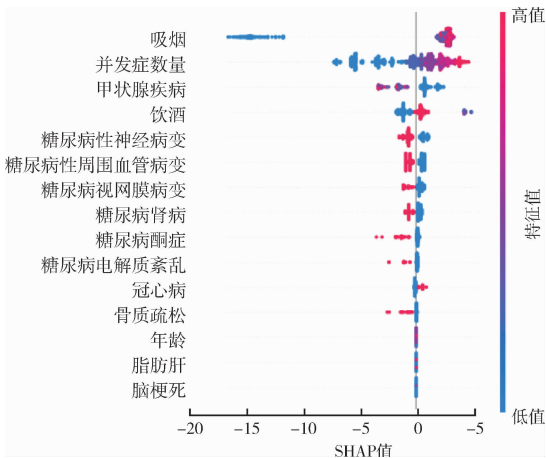


图 3 最优模型的 SHAP 值散点图

3.5 模型在年龄亚组的稳定性分析

为明确并发症累积对共病风险的差异化影响，统计各年龄亚组患共病与未患共病人群的并发症分布。所有年龄组中，患共病人群中并发症数量不少于 4 种的占比均显著高于未患共病人群；随着年龄增长，患共病与未患共病两类人群中，并发症数量不少于 4 种的占比均呈上升趋势；进一步分析可知，18≤年龄<30 岁组的患共病者以合并 2 种并发症为主，而年龄不低于 60 岁组的患共病人群中，并发症数量不少于 4 种的占比已超过 70%。

NGBoost 模型在 4 个年龄组中均表现出较高的预



测性能，AUC 值分别为 0.826 7、0.935 9、0.957 3 和 0.944 8，表明其在不同年龄段均能有效预测 T2DM 合并高血压风险。尽管各年龄组的 SHAP 值存在差异，但模型在各组中均能稳定地识别出主要

影响因素，如吸烟、并发症数量和 T2DM 相关并发症等，见图 4。表明该模型在不同年龄组中可保持稳定的预测优势，具有较高的可靠性和普适性。

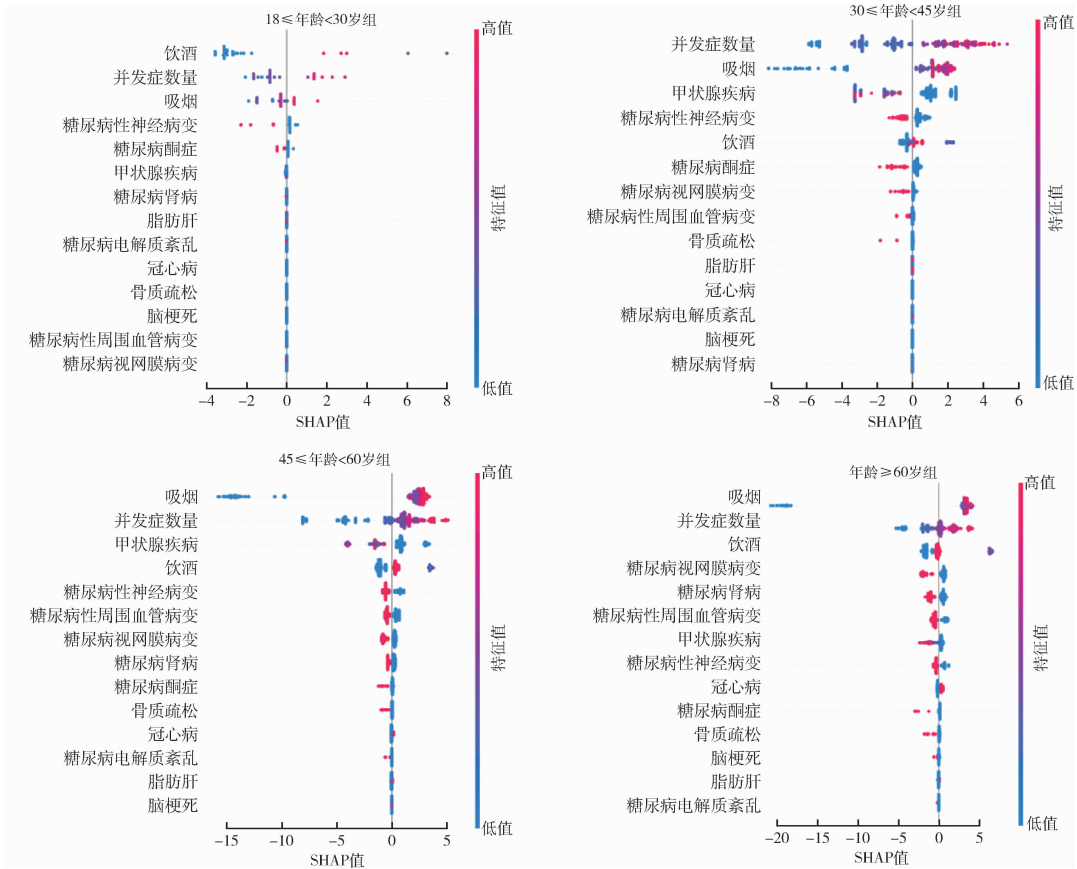


图 4 NGBost 模型在各年龄亚组的 SHAP 解释性分析

4 讨论

4.1 T2DM 合并高血压流行特征、共病机制及医疗负担现状

T2DM、高血压等慢性疾病是公共卫生领域的重大挑战<sup>[15]</sup>。由于不同病种间存在共同的致病因素，共病现象日益普遍<sup>[16]</sup>。我国现有诊疗指南多针对单一病种，针对共病的相对较少。共病不仅增加了治疗复杂性，还可引发过度治疗。因此，早期识别 T2DM 合并高血压的危险因素，并实施前瞻性防控尤为必要。从共病机制看，T2DM 合并高血压的病理关联涉及多方面因素，研究<sup>[17-18]</sup>表明，T2DM 患者持续高血糖可促进内皮细胞下胶原交联、加剧

血管壁纤维化，而此病理变化正是高血压发生的重要机制之一。临床研究发现，约 75% 的 T2DM 患者同时患有高血压<sup>[19]</sup>。医疗负担方面，这种病理上的共存不仅加剧了患者的整体健康风险，显著提高心血管事件发生的概率<sup>[1,20]</sup>，还会通过交互作用导致更严重的健康后果<sup>[21]</sup>。此外，该共病状态会加重患者经济负担与医疗资源消耗，表现为治疗成本增加及院内死亡率上升<sup>[1]</sup>。

4.2 模型性能比较

本研究采用准确度、灵敏度、特异性、F1 值和 AUC 值，系统比较 RF、SVM、XGBoost 和 NGBost 这 4 种机器学习模型在预测 T2DM 合并高血压风险方面的性能，发现 NGBost 模型表现最佳。其在验

征集上也展现出较好的泛化性能。NGBoost 模型在疾病识别任务中应用广泛,其判别性能已得到多项临床研究验证。Pfau M 等<sup>[22]</sup>结合光谱相干断层扫描视网膜层特征与 NGBoost,构建抗血管内皮生长因子治疗需求概率预测模型,结果显示视网膜色素上皮-脉络膜复合体厚度变异为核心预测因子,模型平均绝对误差为 2.60~2.66 次/年,具有临床预测可行性。Chen J X 等<sup>[23]</sup>提出基于 NGBoost 的微卫星 DNA 不稳定性检测方法,经对比发现其性能优于多种机器学习算法和深度学习模型。

### 4.3 关键特征对 T2DM 合并高血压的影响

在 L1 逻辑回归筛选出的 15 个关键特征中,吸烟、并发症数量等正向特征在 SHAP 解释性分析中符号一致,表明二者对共病风险具有稳健的正向影响。甲状腺疾病、糖尿病性神经病变等在 L1 逻辑回归中系数为负,而 SHAP 解释性分析显示其平均 SHAP 值为正,这是因为两种模型假设不同。L1 逻辑回归基于全局线性关联量化特征影响,而 SHAP 解释性分析通过 NGBoost 非线性模型捕捉个体预测的局部贡献及特征交互效应<sup>[24]</sup>。当特征之间存在相关性时,SHAP 值会提供更准确的解释<sup>[25]</sup>。SHAP 解释性分析不仅揭示了各特征的独立贡献,还提示了特征间可能存在的交互效应。吸烟对共病风险的正向贡献显著,其可能通过加速微血管病变增加高血压发生风险,而并发症数量的增多会进一步叠加这一风险效应;此外,甲状腺疾病与糖尿病性神经病变、糖尿病肾病在部分患者中共同出现高 SHAP 值,提示甲状腺功能异常可能加剧代谢紊乱和血压升高。

SHAP 解释性分析进一步突显了吸烟的核心作用,与张霞等<sup>[26]</sup>结论一致,烟草成分通过促进动脉粥样硬化及血压调节异常增加共病风险。并发症数量的高影响力则体现了多系统损伤的累积效应,但其具体作用路径仍有待深入探讨<sup>[27]</sup>。甲状腺功能减退通过代谢紊乱与心血管稳态破坏协同驱动 T2DM 与高血压共病,与 Alberto J T 等<sup>[28]</sup>结论一致。饮酒通过酒精摄入影响血糖水平和胰岛素敏感性,从而加重 T2DM 病情。长期饮酒会导致血管内皮功能受损,增加动脉硬化和高血压风险<sup>[29]</sup>。糖尿病性神经

病变通过自主神经功能障碍影响血压调节<sup>[30]</sup>;糖尿病性周围血管病变通过增加外周血管阻力加剧血压升高<sup>[31]</sup>;糖尿病视网膜病变与高血压存在双向促进作用,加剧血管内皮功能异常<sup>[32]</sup>;糖尿病肾病则通过钠水潴留导致血容量增加和血压升高<sup>[33]</sup>。

戒烟干预不仅能有效降低 T2DM 患者的心血管代谢风险,包括血压、血脂和血糖<sup>[34]</sup>,还能显著降低高血压患者的血压水平并改善内皮功能<sup>[35]</sup>。其应作为共病管理的重要组成部分,尤其针对中老年患者,建议在门诊常规开展戒烟咨询与定期随访。在此基础上,对并发症数量不少于 4 种的患者应实施强化监测,包括每 3 个月一次的心血管及肾功能评估。此外,建议开展多学科协作诊疗,整合内分泌科、心内科及营养科等资源,共同制定个性化综合管理方案。

### 4.4 模型的亚组稳定性和特征差异

在各年龄亚组中,NGBoost 模型均稳定识别并发病数量为共病风险的核心影响特征,且随年龄增长其权重呈上升趋势。同时,不同年龄亚组的次要特征存在差异,18≤年龄<30 岁组以饮酒、吸烟为主;30≤年龄<45 岁组和 45≤年龄<60 岁组以并发症数量、吸烟、甲状腺疾病为主;年龄≥60 岁组则叠加糖尿病视网膜病变、肾病等慢性并发症特征。这种差异主要与不同年龄人群的生物学机制及社会行为因素有关。18≤年龄<30 岁组患者的特征分布更集中,疾病特征更典型;30≤年龄<60 岁组患者可能更多受生活方式影响;而年龄≥60 岁组患者 T2DM 病程较长,长期高血糖状态导致微血管和大血管病变的累积效应显著<sup>[36]</sup>,此类并发症通过加速动脉硬化、内皮功能障碍等机制,与高血压形成协同病理循环<sup>[37]</sup>。此外,老年患者常合并多器官功能衰退,代谢调节能力下降,进一步放大了并发症对共病风险的影响<sup>[38]</sup>。

### 4.5 局限性和未来研究方向

本研究将机器学习预测模型转化为实践导向的决策工具,通过 SHAP 值排序明确影响因素重要性,为临床风险分层提供依据;助力医疗资源优先投向高吸

烟水平、多并发症等高 SHAP 值特征患者,以实现早期精准预防。然而本研究存在一定局限性:数据来源于单一综合性三甲医院,虽模型具备一定泛化能力,但仍受地域特征与医院诊疗水平影响——当地高盐高碘饮食、低寒环境可能增加甲状腺疾病、冠心病发病风险,模型筛选特征的泛化应考虑地域适应性;医院高诊疗水平使脑梗死、糖尿病酮症等重症患者数据与普通医院存在差异。此外,研究对象为符合纳入标准的就诊患者,覆盖了 T2DM 合并高血压的核心风险群体,然而职业分布仍无法完全代表全域人群。后续研究应纳入多地域、不同诊疗水平医疗数据,构建异质性训练集,结合特征选择算法优化模型环境适应能力,以推动临床应用转化。

## 5 结语

本研究构建 RF、SVM、XGBoost 及 NGBoost 共 4 种机器学习模型,分析 T2DM 合并高血压的影响因素,结果显示 NGBoost 模型在共病风险预测中表现最优。结合 SHAP 解释性分析,进一步明确吸烟、并发症数量、甲状腺疾病、饮酒为关键影响因素,并发现特征间存在交互作用,可为临床早期识别高风险患者、开展风险评估及制定个性化干预措施提供指导。

**作者贡献:** 芮晨负责研究设计、论文撰写;刘浩然负责数据分析、论文撰写;张盈盈负责模型搭建;李金斌负责模型验证;张敬晨负责文献整理与分析;张胜发负责提供指导。

**利益声明:** 所有作者均声明不存在利益冲突。

## 参考文献

- 1 森干,李永生,孙方旭,等. 2 型糖尿病合并高血压的危险因素及其列线图预测模型的构建 [J]. 重庆医学, 2022, 51 (13): 2189–2193, 2198.
- 2 娄海飞,张丽华,赵月,等. 糖尿病足病伴高血压患者血糖控制的影响因素 [J]. 中华高血压杂志, 2022, 30 (3): 287–289.
- 3 MILLS K T, STEFANESCU A, HE J. The global epidemiology of hypertension [J]. Nature reviews nephrology, 2020, 16 (4): 223–237.
- 4 戴飘钰,袁琼婧,彭张哲,等. 糖尿病合并慢性肾脏病

诊疗的现状和进展 [J]. 中国医学科学院学报, 2023, 45 (6): 987–996.

- 5 LI Y W, LIU Y F, LIU S W, et al. Diabetic vascular diseases: molecular mechanisms and therapeutic strategies [J]. Signal transduction and targeted therapy, 2023, 8 (1): 152–152.
- 6 RAHMAN A, DEBNATH T, KUNDU D, et al. Machine learning and deep learning – based approach in smart health-care: recent advances, applications, challenges and opportunities [J]. AIMS public health, 2024, 11 (1): 58.
- 7 VU T, KOKUBO Y, INOUE M, et al. Machine learning approaches for stroke risk prediction: findings from the Suita study [J]. Journal of cardiovascular development and disease, 2024, 11 (7): 207.
- 8 QIN Y F, WU J L, XIAO W, et al. Machine learning models for data – driven prediction of diabetes by lifestyle type [J]. International journal of environmental research and public health, 2022, 19 (22): 15027.
- 9 ISLAM M M, ALAM M J, MANIRUZZAMAN M, et al. Predicting the risk of hypertension using machine learning algorithms: a cross sectional study in Ethiopia [J]. Plos one, 2023, 18 (8): e289613.
- 10 JI Y J, SHANG H Y, YI J, et al. Machine learning – based models to predict type 2 diabetes combined with coronary heart disease and feature analysis – based on interpretable SHAP [J]. Acta diabetologica, 2025, 62 (10): 1–16.
- 11 中华医学会糖尿病学分会. 中国糖尿病防治指南 (2024 版) [J]. 中华糖尿病杂志, 2025, 17 (1): 16–139.
- 12 中国高血压防治指南修订委员会, 高血压联盟 (中国), 中国医疗保健国际交流促进会高血压病学分会, 等. 中国高血压防治指南 (2024 年修订版) [J]. 中华高血压杂志 (中英文), 2024, 32 (7): 603–700.
- 13 TLAMELO E, THABISO M, DIMANE M, et al. A survey on missing data in machine learning [J]. Journal of big data, 2021, 8 (1): 140.
- 14 ALNOWAISER K. Improving healthcare prediction of diabetic patients using KNN imputed features and tri – ensemble model [J]. IEEE access, 2024, 12 (1): 6783–16793.
- 15 国家心血管病中心, 中国心血管健康与疾病报告编写组. 中国心血管健康与疾病报告 2023 概要 [J]. 中国循环杂志, 2024, 39 (7): 625–660.
- 16 ZENG H J, CHEN M, RUI W, et al. Influence of comorbidity of chronic diseases on basic activities of daily living among older adults in China: a propensity score – matched study [J]. Frontiers in public health, 2024, 12 (4): 1292289.
- 17 ROUMIE C L, HUNG A M, RUSSELL G B, et al. Blood

- pressure control and the association with diabetes mellitus incidence; results from sprint randomized trial [J]. *Hypertension*, 2020, 75 (2): 331–338.
- 18 CORTELETE J F, KEGLER E M R, FABIANA P, et al. Relationship of advanced glycation end – products in hypertension in diabetic patients: a systematic review [J]. *Brazilian journal of nephrology*, 2022, 44 (4): 557–572.
- 19 YAYA S, EL – KHATIB Z, AHINKORAH B O, et al. Prevalence and socioeconomic factors of diabetes and high blood pressure among women in Kenya: a cross – sectional study [J]. *Journal of epidemiology and global health*, 2021, 11 (4): 397–404.
- 20 WANG Z H, YANG T, FU H L. Prevalence of diabetes and hypertension and their interaction effects on cardio – cerebrovascular diseases: a cross – sectional study [J]. *BMC public health*, 2021, 21 (1): 1224.
- 21 SHIFERAW K K, NABIHA D, SAMSON G, et al. Determinants of hypertension among patients with type 2 diabetes mellitus on follow – up at Tikur Anbessa specialized hospital, Addis Ababa: a case – control study [J]. *Plos one*, 2021, 16 (8): e256399.
- 22 PFAU M, SAHU S, RUPNOW R A, et al. Probabilistic forecasting of anti – VEGF treatment frequency in neovascular age – related macular degeneration [J]. *Translational vision science & technology*, 2021, 10 (7): 30.
- 23 CHEN J X, WANG M, ZHAO D F, et al. Msingb: a novel computational method based on ngboost for identifying microsatellite instability status from tumor mutation annotation data [J]. *Interdisciplinary sciences, computational life sciences*, 2022, 15 (1): 100–110.
- 24 KHANSA R, ADNAN Q, MOHAMMED G, et al. Explainable, trustworthy, and ethical machine learning for healthcare: a survey [J]. *Computers in biology and medicine*, 2022, 149 (10): 106043.
- 25 NICOLA O, ALEX M, ALICJA W. Interaction analysis based on shapley values and extreme gradient boosting: a realistic simulation and application to a large epidemiological prospective study [J]. *Frontiers in nutrition*, 2022, 9 (7): 871768.
- 26 张霞, 华姜, 周佳茜, 等. 基于国家基本公共卫生服务体检的 20 岁及以上高血压患者吸烟行为与糖尿病的关联研究 [J]. *现代预防医学*, 2025, 52 (6): 1085–1091.
- 27 LI R, TIAN S, LIU J, et al. Modifiable risk factors and plasma proteomics in relation to complications of type 2 diabetes [J]. *Nature communications*, 2025, 16 (1): 2896.
- 28 ALBERTO J T, MARGARITA F, VANESSA B, et al. The joint effect of congenital hypothyroidism and hypercaloric diet consumption as triggers of type 2 diabetes mellitus [J]. *European thyroid journal*, 2022, 11 (1): 1–8.
- 29 MAYL J J, GERMAN C A, BERTONI A G, et al. Association of alcohol intake with hypertension in type 2 diabetes mellitus: the ACCORD trial [J]. *Journal of the American heart association*, 2020, 9 (18): e17334.
- 30 GOGAN A, POTRE O, AVRAM V F, et al. Cardiac autonomic neuropathy in diabetes mellitus: pathogenesis, epidemiology, diagnosis and clinical implications: a narrative review [J]. *Journal of clinical medicine*, 2025, 14 (3): 671.
- 31 RATHNAGANPATHI T, RAGHUPATHY T. Clinical evaluation of peripheral vascular disease in diabetic foot ulcer [J]. *Journal of research in medical and dental science*, 2021, 9 (6): 312–318.
- 32 GAIKWAD C, KHUNE A, MAHADIK S, et al. A study of ocular fundus findings in hypertension during pregnancy in a tertiary care hospital [J]. *MVP journal of medical sciences*, 2022, 8 (2): 163–170.
- 33 ALAM F, NAG M, DASGUPTA D, et al. Diabetic nephropathy: risk factors and management [J]. *Pharmaceutical science: research and innovation*, 2024, 3 (7): 44–51.
- 34 JU H C, HSIN W H, LUNG H C, et al. Improvement in cardiometabolic risk factors during smoking cessation treatment in patients with type 2 diabetes: a retrospective cohort study [J]. *Diabetes, metabolic syndrome and obesity*, 2021 (4): 1695–1702.
- 35 GAYA P V, FONSECA G W P, TANJIL T, et al. Smoking cessation decreases arterial blood pressure in hypertensive smokers: a subgroup analysis of the randomized controlled trial gentsmoking. [J]. *Tobacco induced diseases*, 2024, 22 (5): 80.
- 36 SAIYED N S, YAGOUB U, AL QAHTANI B, et al. Risk factors of microvascular complications among type 2 diabetic patients using cox proportional hazards models: a cohort study in Tabuk Saudi Arabia [J]. *Journal of multidisciplinary healthcare*, 2022 (7): 1619–1632.
- 37 XUE C X, CHEN K Y, GAO Z Z, et al. Common mechanisms underlying diabetic vascular complications: focus on the interaction of metabolic disorders, immuno – inflammation, and endothelial dysfunction [J]. *Cell communication and signaling*, 2023, 21 (1): 298.
- 38 NGCOBO N N. Influence of ageing on the pharmacodynamics and pharmacokinetics of chronically administered medicines in geriatric patients: a review [J]. *Clinical pharmacokinetics*, 2025, 64 (3): 1–33.