

“三减三健”政策下基于XGBoost与SHAP分析的可解释性肥胖风险预测研究

曹克苗

上海理工大学管理学院, 上海

收稿日期: 2025年4月29日; 录用日期: 2025年6月13日; 发布日期: 2025年6月20日

摘要

在全球肥胖问题日益严峻及中国“三减三健”政策推进的背景下, 本研究基于墨西哥、秘鲁和哥伦比亚三国2111份多维度健康数据集, 提出一种基于XGBoost与SHAP分析的可解释性肥胖风险预测方法。首先基于XGBoost集成算法构建肥胖风险预测模型(模型准确率达96.22%), 并根据XGBoost内嵌评估指标对模型进行初步特征解释。运用Tree SHAP方法对诊断模型进行特征解释分析, 探究重要特征对肥胖风险的影响关系, 分析特征之间的依赖交互效应, 直观、透明地揭示模型的预测机制。最后利用聚类分析将人群划分为6类, 结合中国“体重管理年”政策提出精准干预策略。研究验证了“饮食-运动”干预路径的跨文化有效性, 为《健康中国2030》的全球化经验输出提供循证依据, 开创了健康政策国际互鉴的新范式。

关键词

三减三健, 肥胖风险预测, XGBoost模型, Tree SHAP, 特征解释

An Explanatory Study on Obesity Risk Prediction Based on XGBoost and SHAP Analysis under the “Three Reductions and Three Healthy Conditions” Policy

Kemiao Cao

Business School, University of Shanghai for Science and Technology, Shanghai

Received: Apr. 29th, 2025; accepted: Jun. 13th, 2025; published: Jun. 20th, 2025

Abstract

Against the background of the increasingly serious global obesity problem and the promotion of China's "Three Reductions and Three Healthy Conditions" policy, this study proposes an interpretable obesity risk prediction method based on XGBoost and SHAP analysis based on 2111 multidimensional health datasets from Mexico, Peru and Colombia. An obesity risk prediction model was first constructed based on the XGBoost integration algorithm (with a model accuracy of 96.22%), and the initial features of the model were interpreted based on the XGBoost embedded assessment metrics. The Tree SHAP method was used to analyze the feature interpretation of the diagnostic model, to explore the relationship of important features on obesity risk, to analyze the dependency interaction effect between features, and to reveal the prediction mechanism of the model intuitively and transparently. Finally, cluster analysis was used to classify the population into 6 categories, and precise intervention strategies were proposed in accordance with China's "Year of Weight Management" policy. The study validates the cross-cultural effectiveness of the "diet-exercise" intervention pathway, provides evidence-based basis for the globalization of the "Healthy China 2030", and creates a new paradigm for the international mutual understanding of health policies.

Keywords

Three Reductions and Three Healthy Conditions, Obesity Risk Prediction, XGBoost Model, Tree SHAP, Feature Interpretation

Copyright © 2025 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

肥胖是一种慢性、易复发、进行性的疾病状态，会对人体健康产生渐进持续性的功能损伤危害[1]。肥胖是多种慢性疾病的重要诱因之一，不仅影响个体健康，还会影响个体的生活质量和劳动生产力，从而对经济社会产生重要影响[2]。

随着全球肥胖流行趋势的加剧，世界卫生组织将肥胖定义为“21 世纪最严峻的公共卫生挑战之一”。据中国居民营养与慢性病状况报告(2020 年)数据显示，中国有超过一半的成年居民超重或肥胖，并且城乡各年龄组居民超重肥胖率仍处于上升态势。预计到 2030 年，全球超重或肥胖人口比例可能接近 60% [3]。

在这一背景下，国务院于 2019 年成立健康中国行动推进委员会，统筹推进《健康中国行动(2019~2030 年)》，明确提出“到 2022 年和 2030 年，成人肥胖增长率持续减缓”的行动目标。2025 年两会期间，国家卫健委提出持续推进“体重管理年”计划，并启动《“体重管理年”活动实施方案》，提出“饮食 - 运动 - 行为”三位一体的干预策略与“三减三健”等健康理念，指导公众正确认识健康体重、科学动态管理体重，消除体重认知误区，提高科学健康知识的传播力和可及性。

本研究通过构建多种模型，分析身体状况、饮食习惯等多种因素对肥胖水平的影响，并揭示其内在机理。同时，结合相关政策，为不同人群提出差异化建议。研究结果可以为具有不同饮食习惯及身体状况的差异化人群提供个性化的体重管理建议，帮助其探寻适配自身特点的体重管理方案以改善自身身体状况，实现人口高质量发展。同时，本文创新性构建跨文化验证框架，验证了政策的普适性，不仅为《健康中国 2030》的全球化经验输出提供循证依据，更开创发展中国家之间健康政策互鉴的新范式，助力构

建人类卫生健康共同体。

2. 文献综述

脂肪是人体的重要组成部分，为人体提供热量，也能够保护内脏器官，缓冲外界对内脏的冲击。而肥胖是指一定程度的明显超重与脂肪层过厚，是体内脂肪积聚过多而导致的一种状态。这主要是由于食物摄入过多或机体代谢的改变而导致体内脂肪积聚过多。多数研究采用体重指数(body mass index, BMI)作为评估身体胖瘦程度及是否健康的标准[1]。

不少学者认为肥胖会产生负面影响。JOSEPH MARFOH 对加纳的 16 个地区进行调研，发现肥胖和超重在全国范围内对个人健康、家庭收入和生活质量均产生负面影响。在个人健康方面，肥胖往往关联着慢性疾病，是糖尿病、不孕症、高血压、心脏病、中风、胆囊疾病、脂肪肝和高胆固醇等疾病的重要原因之一[4]-[6]。在家庭收入方面，肥胖引发的慢性疾病会增加家庭的医疗费用支出，从而使得家庭成员幸福感降低[7]。肥胖不仅影响着个体和家庭生活，还决定着社会的财富。在国家层面，肥胖率的快速增长也会给医疗资源带来更多负担。美国将其年度卫生预算的 70%以上用于与肥胖相关的慢性健康问题[8]。同时，肥胖的增加会减慢人口的活动并影响生产力。

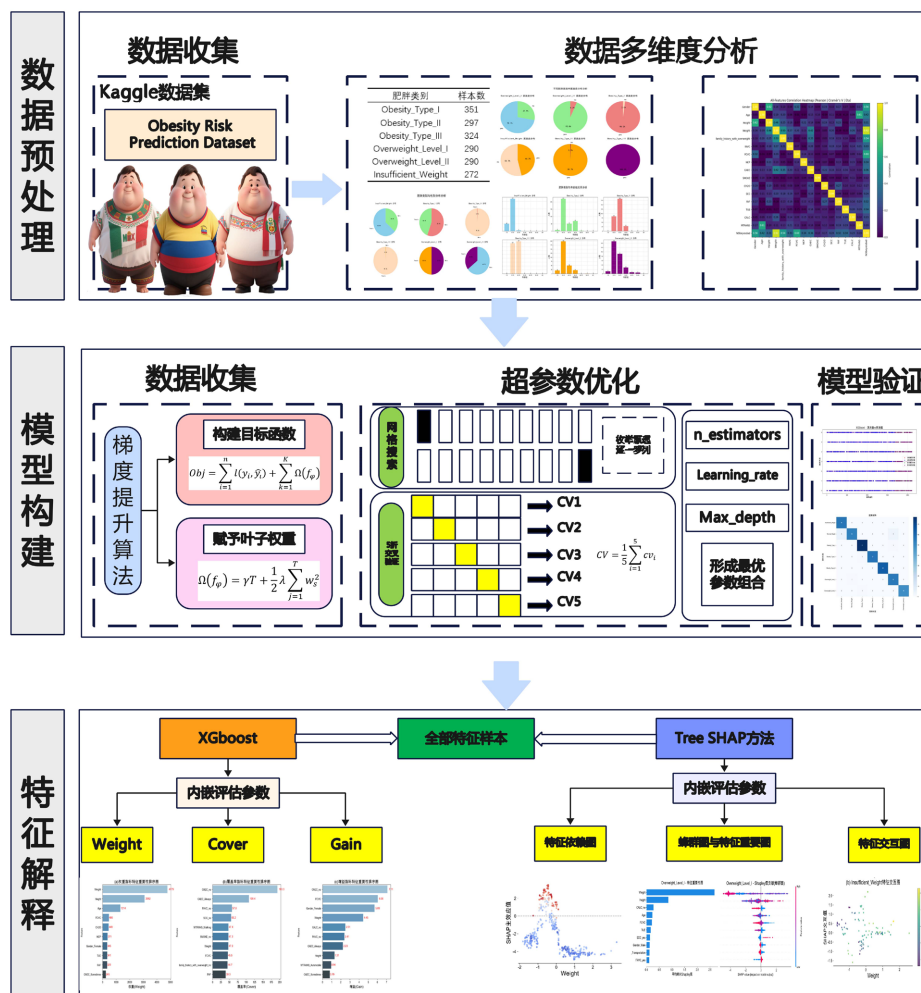


Figure 1. Research framework diagram

图 1. 研究框架图

据世界卫生组织调查报告显示,1990年到2022年,全球5~19岁儿童和青少年的肥胖率从2%增加到8%,而18岁及以上成年人的肥胖率从7%翻了一番到16%。而在我国肥胖问题尤为突出,据2020年中国居民营养与慢性病状况报告显示,我国6~17岁儿童青少年的超重率、肥胖率分别为11.1%和7.9% [9],我国成年人超重肥胖率已超过50%。故为解决肥胖所引起的公共卫生问题,中国正积极采取相关措施。

肥胖是一个由诸多因素共同导致的健康问题。虽然该症状已经存在已久,但目前对肥胖的产生的原因并未有明确定论,主流观点认为肥胖主要受遗传和生活习惯等因素的影响[10][11]。部分人群的肥胖是父母的遗传,尽管生活习惯良好,但仍然产生了肥胖症[12]。而多数人肥胖的产生取决于生活习惯,目前肥胖和超重的激增是由于体育活动的缺乏,导致身体产生运动脂肪[13]。相反,充分体育活动有助于减少体内脂肪和炎性标志物等,从而降低超重肥胖的发生[10]。另外,部分学者认为心理与文化问题也会导致肥胖的产生[14][15]。例如,在非洲某些地区,非洲大多数种族都珍视这些脂肪体,巨大的体型被视为幸福和财富的标志[16]。

尽管目前已有研究探讨肥胖的影响因素,但较少研究结合“体重管理年”政策进行相关分析,并且大多都未针对特定人群提供个性化建议。本研究通过拉美三国数据与中国政策框架的系统对接,既可验证“饮食-运动”基础干预路径的跨文化有效性,又能识别需本土化调整的敏感指标,为《健康中国2030》的全球化经验借鉴提供实证支撑。

3. 分析框架

本文的研究内容主要分为7个部分。论文框架图如图1,第一部分是问题提出:针对国内外研究现状以及目前的研究背景提出研究问题;第二部分是数据来源与预处理:这一部分主要介绍数据来源以及数据预处理;第三部分是肥胖数据多维度分析;第四部分基于XGBoost与SHAP分析的可解释性肥胖风险预测;第五部分结合聚类分析对特定肥胖人群进行对应政策分析。第六部分针对研究结论提出相应的建议;第七部分介绍本文研究创新点、不足之处以及未来研究展望。

4. 数据来源与预处理

4.1. 数据来源

本研究采用的肥胖数据集(Obesity Risk Prediction Dataset)源自Kaggle平台,涵盖了墨西哥、秘鲁和哥伦比亚三国14至61岁人群的肥胖水平估计数据,且样本群体在饮食习惯和身体状况方面存在差异。数据经网络平台匿名收集,由匿名用户填写问卷后,经处理得出,共包含17个属性、2111条记录。其中,与饮食习惯相关的属性包括高热量食物摄入频率(FAVC)、蔬菜摄入频率(FCVC)、主餐次数(NCP)、两餐间食物摄入量(CAEC)、每日饮水量(CH2O)及酒精摄入量(CALC);与身体状况相关的属性涵盖卡路里摄入监测情况(SCC)、体力活动频率(FAF)、使用技术设备时长(TUE)及所使用的交通工具类型(MTRANS)。

4.2. 数据预处理

表1是具体的变量说明,涵盖了从人口统计学信息到生活方式习惯的多个维度,旨在全面描述参与者的背景特征、饮食与生活习惯、身体特征以及家族健康状况,为肥胖等级的分析提供多角度的数据支持。

肥胖数据集主要是数据清洗与特征转换。先对原始数据进行去重、删除空值以及处理异常值。在这部分,本研究将身高低于1.45米和高于1.98米,以及体重低于39公斤和高于165公斤的样本设定为异

常值并进行剔除，经过处理后的肥胖数据集共计 2112 条有效记录。为方便后续分析，对分类变量进行了编码处理。性别(Gender)、家族肥胖史(family_history_with_overweight)、高热量食物消费(FAVC)、吸烟情况(SMOKE)和卡路里监测(SCC)等二分类变量使用二进制编码(0 和 1)；而餐间食物消费(CAEC)、酒精消费(CALC)和交通方式(MTRANS)等多分类变量采用独热编码(One-Hot Encoding)转换为数值特征。

Table 1. Description of variables
表 1. 变量说明

变量分类	变量名称	变量类型	描述
目标变量	NObesdad	分类变量	参与者的肥胖等级
人口统计学变量	Gender	分类变量	参与者的性别
	Age	连续变量	参与者的年龄(岁)
身体特征变量	Height	连续变量	参与者的身高(米)
	Weight	连续变量	参与者的体重(千克)
家族史变量	family_history_with_overweight	分类变量	家族中是否有肥胖成员
饮食习惯变量	FAVC	分类变量	是否经常消费高热量食物
	FCVC	有序分类变量	参与者食用蔬菜的频率
	NCP	有序分类变量	参与者平均每天进餐的次数
	CAEC	有序分类变量	参与者在正餐之外是否额外摄入零食等食物
生活方式变量	SMOKE	分类变量	是否吸烟
	CH2O	有序分类变量	每日饮水量
	SCC	分类变量	是否监测卡路里摄入
	FAF	有序分类变量	参与者进行体育活动的频率
	TUE	有序分类变量	每天使用科技设备(电脑、手机等)的时间
	CALC	有序分类变量	参与者饮酒的频率
	MTRANS	分类变量	参与者通常采用的交通方式

5. 基于 XGBoost 与 SHAP 分析的可解释性肥胖风险预测

5.1. XGBoost 模型构建

XGBoost 是一种基于前向分布算法实现加法模型的集成学习方法，运用迭代思想组合一系列弱学习器逐步形成强学习器，通过计算残差确定每个学习器的权重，利用收缩和正则化技术处理稀疏性和避免过拟合，从而实现高效的分类效果。相较于传统的梯度提升算法，XGBoost 创新地引入二阶导数和正则项，深入考虑模型对真实值的拟合程度以及模型的计算复杂度，进一步提升模型的诊断能力。XGBoost 方法目标函数综合了损失函数和正则项，如式(1)所示，在故障诊断模型构建中，通过梯度提升策略持续优化模型，逐步减小弱学习器的残差，逐步迭代得到最优化诊断模型。

$$\begin{cases} \text{Obj} = \sum_{i=1}^n l(y_i, \hat{y}_i) + \sum_{k=1}^K \Omega(f_\phi) \\ \Omega(f_\phi) = \gamma T + \frac{1}{2} \lambda \sum_{j=1}^T w_s^2 \end{cases} \tag{1}$$

式中：Obj 为 XGBoost 的目标函数，是评估模型性能的理论基准； $l(y_i, \hat{y}_i)$ 为模型训练误差，用于衡量诊

断模型预测结果 $\hat{y}_i$ 和真实结果 y_i 之间的差异； $\Omega(f_\phi)$ 为正则项，用于衡量第 ϕ 棵决策分类树的复杂度，用于控制模型整体复杂度，避免过拟合现象； T 为第 ϕ 棵树中的叶子节点数量； w_s 为第 ϕ 棵树中第 s 个叶子节点的权重； γ 为叶子节点数目的惩罚项； λ 为 L_2 正则惩罚项。

5.2. SHAP 归因解释分析

SHAP 是可解释人工智能技术中一种流行的特征归因机制方法。部分研究者应用该算法成功克服了机器学习的“黑盒”特性，对机器学习模型提供了一致的可解释性。为进一步探究非线性和复杂模型的内在预测机制，SHAP 方法引入了基于 Shapley 值的特征归因计算。Shapley 值计算理论主要包含特征权重和边际贡献两部分，具体如式(2)所示。

$$\Phi_j = \sum_{S \subseteq \{x_1, \dots, x_p\}, \{x_j\}} \frac{|S|!(p-|S|-1)!}{p!} (val(S \cup \{x_j\}) - val(S)) \quad (2)$$

式中： Φ_j 为特征 j 的 Shapley 值； S 为模型中使用的特征子集， $S \subseteq \{x_1, \dots, x_p\}, \{x_j\}$ 为不包括 $\{x_j\}$ 的特征子集， $\{x_1, \dots, x_p\}$ 是输入特征向量的集合； p 为特征的数目， $\frac{|S|!(p-|S|-1)!}{p!}$ 以全排列思想表示子集 S 的权重； $val(S \cup \{x_j\})$ 为包含 $\{x_j\}$ 子集 S 所有特征的边际贡献， $val(S)$ 为不包含 $\{x_j\}$ 的子集 S 所有特征的边际贡献。

模型预测值等价于训练样本平均值与特征 Shapley 值之和， $f(x)$ 为特征变量 Shapley 值的线性函数，可表示为

$$f(x) = \phi_0 + \sum_{j=1}^p \phi_j Z'_j \quad (3)$$

式中： $f(x)$ 为预测值输出； ϕ_0 为训练样本的预测平均值； $Z'_j \in \{0, 1\}^p$ 表示在 p 个特征中，存在的特征为 1，不存在的特征即为 0。

Shapley 值可量化特征在不同故障类别中的相对重要性，特征重要性越高，对故障诊断贡献越大。本文运用 Tree SHAP 方法，通过遍历树结构评估特征对模型预测的重要性。Tree SHAP 优化计算效率，将复杂度从 $O(QL2^M)$ 降至 $O(QLD^2)$ ，其中 Q 为树的数量， L 为最大节点数， M 为特征数， D 为最大深度。这实现了高效特征归因，为模型解释提供有力支持。

5.3. 模型对比

为评估 XGBoost 方法在分类诊断任务中的性能，本文对比决策树(Decision Tree, DT)、随机森林(random forest, RF)、支持向量机(support vector machine, SVM)、K 近邻(K-nearest neighbors, KNN)、朴素贝叶斯(naive Bayes, NB)等常用机器学习方法。所有模型均采用相同的训练集和测试集划分形式，减少数据差异对结果的影响，进一步提高对比结果的可信度和普适性。性能评估指标选取准确率(accuracy)、精确率(precision)、召回率(recall)、F1 值(F1-score)等。如表 2 所示，XGBoost 方法在评估指标上综合表现最佳，表明此方法在肥胖多分类任务中具有优异的性能。

Table 2. Model performance evaluation metrics

表 2. 模型性能评估指标

模型	准确率	精确率	召回率	F1 值
XGBoost	0.9622	0.9636	0.9622	0.9620
Decision Tree	0.9385	0.9389	0.9385	0.9385

续表

random forest	0.9385	0.9391	0.9385	0.9388
Support vector machine	0.8747	0.8788	0.8747	0.8755
K-nearest neighbors	0.8132	0.8112	0.8132	0.8103
naive Bayes	0.5414	0.5165	0.5414	0.4772

5.4. 分析结果

本研究对 Obesity Risk Prediction Dataset 数据集进行了多域特征的量化结果分析, 使用 Scikit-Learn 工具包将数据按 8:2 比例划分为训练集和测试集, 并通过网格搜索和 5 折交叉验证来优化超参数(学习率为 0.1, 最大树深度为 3, 树的数量为 500)。最终, 诊断模型在测试集上的准确率为 96.22%, 表现优异。

混淆矩阵如图 2 所示, 模型对 Insufficient Weight 和 Obesity Type III 两个极端类别的预测最为准确, 均达到了 100%。对 Normal Weight 类别的预测相对较弱, 这可能是由于正常体重与轻度不足体重或轻度超重之间的界限较为模糊。尽管如此, 混淆矩阵结果显示, 模型在分类不同程度的肥胖类别上具有很高的准确性, 尤其是在区分不同类型的肥胖(如 Obesity Type I、Obesity Type II)时表现出色, 表明 XGBoost 算法结合所选超参数在肥胖程度分类任务中非常有效。

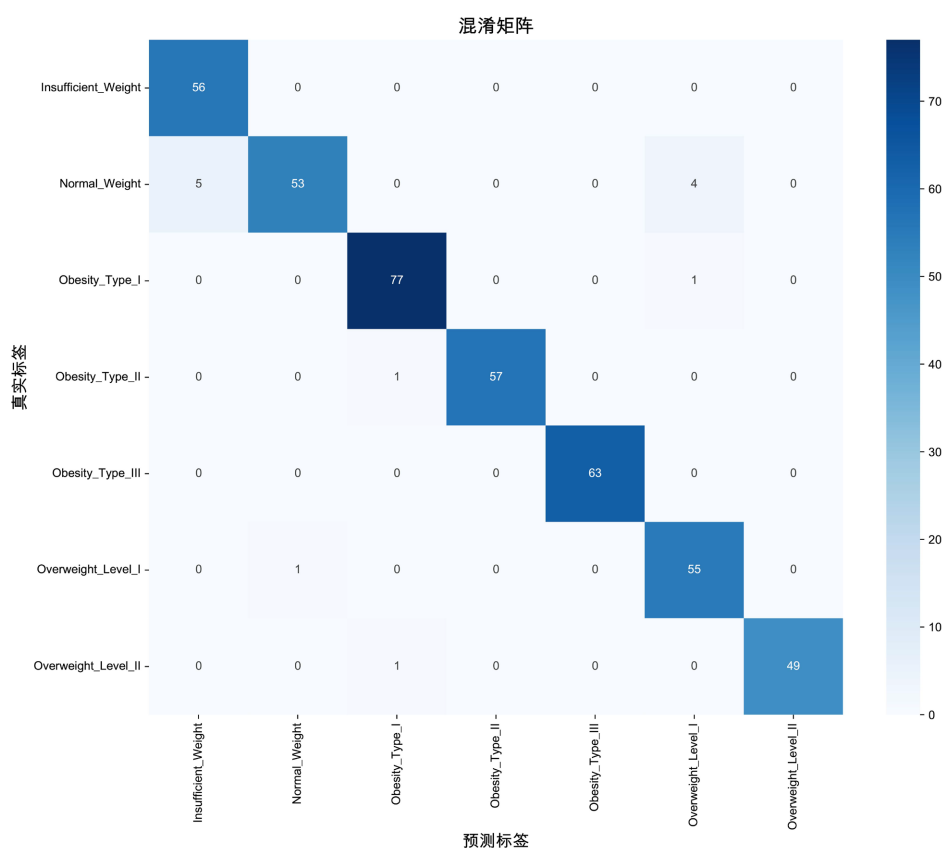


Figure 2. Confusion matrix

图 2. 混淆矩阵

在故障诊断结果分析的基础上, 使用 XGBoost 内嵌评估指标参数可以增强模型特征可解释性。权重、覆盖率和增益三个评估指标分别反映特征的分裂频次、信息增益及影响样本数量, 量化并排序特征的重

要性。结果如图 3 所示，排序图展示前 10 个最重要的特征参数，突出特征在模型中的关键作用。

在权重指标排序中，体重是最重要的特征，其分裂点次数远超其他特征，对分类决策树结构影响显著。身高、年龄和食用蔬菜频率等特征也较为重要，有助于提升分类稳定性。覆盖率指标显示，从不在餐间吃东西覆盖率最高，其次是总是在餐间吃东西，然后是不经常摄入高热量食物。其他重要特征包括不计算卡路里、步行出行和不吸烟。这表明饮食习惯，尤其是餐间进食行为，以及生活方式特征，对样本分类具有广泛影响。增益指标方面，CAEC_no 的信息增益最大，FCVC 次之，女性排名第三，体重排名第四。这说明从信息理论角度，餐间进食习惯、蔬果消费频率和性别对肥胖分类的预测能力最强。综合来看，饮食习惯特征(尤其是 CAEC 和 FAVC)在覆盖率和增益上表现突出，对肥胖分类影响显著。体格特征如 Weight 和 Height 在权重上占优。

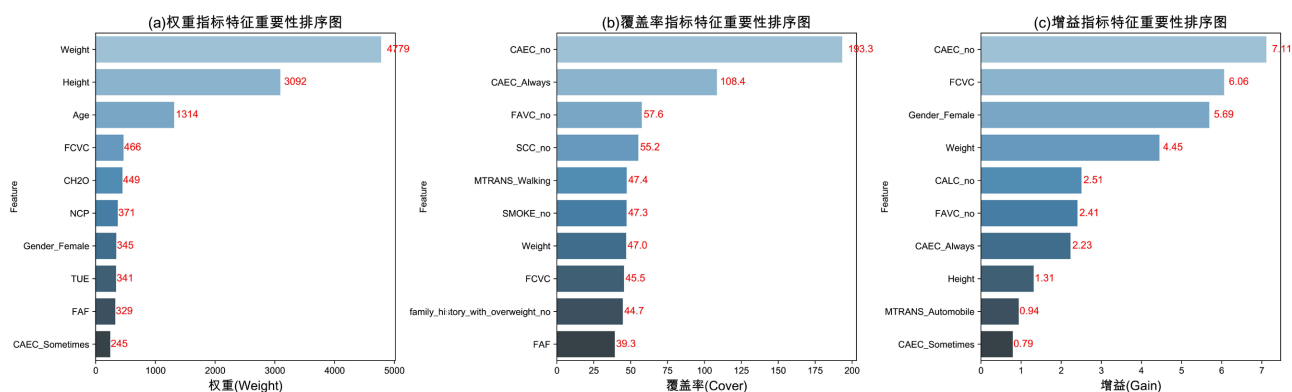


Figure 3. Importance map of XGBoost embedded parameter evaluation metric features
图 3. XGBoost 内嵌参数评估指标特征重要性图

5.5. 特征可解释性

针对分类模型的全局样本特征可解释性，本文采用平均绝对 Shapley 值热图进行可视化分析，由图 4 可以直观地看出，体重和身高始终是模型预测的核心驱动因素，其中体重在所有类别中贡献最大，Insufficient_Weight 类别达到 4.56，Obesity_Type_II 也有 3.51，而身高则在 Obesity_Type_I (1.17) 和 Obesity_Type_III (0.02) 中表现出次要但不可忽视的影响；随着肥胖程度加深，性别、饮水量和年龄等非体重因素的作用日益凸显，尤其在 Obesity_Type_III 中，性别(1.17)与饮水量(1.08)成为关键变量，而在 Overweight_Level_I/II 中，年龄(0.36/0.37)与蔬果消费、零食消费等生活习惯指标也展现出中等影响；相对而言，Insufficient_Weight 类别几乎完全依赖于体重与身高。

为进一步表征输入特征对轴承类别预测的贡献度及作用效果，运用 Tree SHAP 方法分析每种肥胖类别中前 10 个重要特征对故障诊断结果的边际影响。图 5 所示为在 6 种不同异常体重条件下，每类特征对预测结果的影响程度。左侧是特征重要性条形图，展示了平均绝对 Shapley 值，反映各特征对预测“体重不足”类别的重要程度。右侧是 SHAP 值蜂群图，展示了每个特征的 SHAP 值分布，揭示特征值如何影响模型预测。

体重在所有分类中均为最重要的预测变量，其平均绝对 SHAP 值最高，表明对模型输出影响最大。蜂群图中，体重的 SHAP 值分布广泛，高体重多与模型输出正向变化相关，低体重则与负向变化相关。身高在多数分类中位列第二重要特征，其影响在肥胖 I 型和超重 I 级中尤为突出。年龄在肥胖 II 型和超重 II 级中位列第三，显示对特定肥胖类型影响较大。性别在肥胖 III 型中对预测贡献显著，男性 SHAP 值通常较高。饮食习惯如水果、蔬菜、豆类消费频率在肥胖 III 型中位列第三，高热量食物消费频率在超重

II 级中位列第四，表明饮食对肥胖风险的重要作用。饮酒情况在超重 I 级中位列第三，提示饮酒行为与超重状态的关联性。总体而言，体重和身高是肥胖预测的核心变量，年龄、性别、饮食等特征在不同肥胖类型中发挥辅助作用。

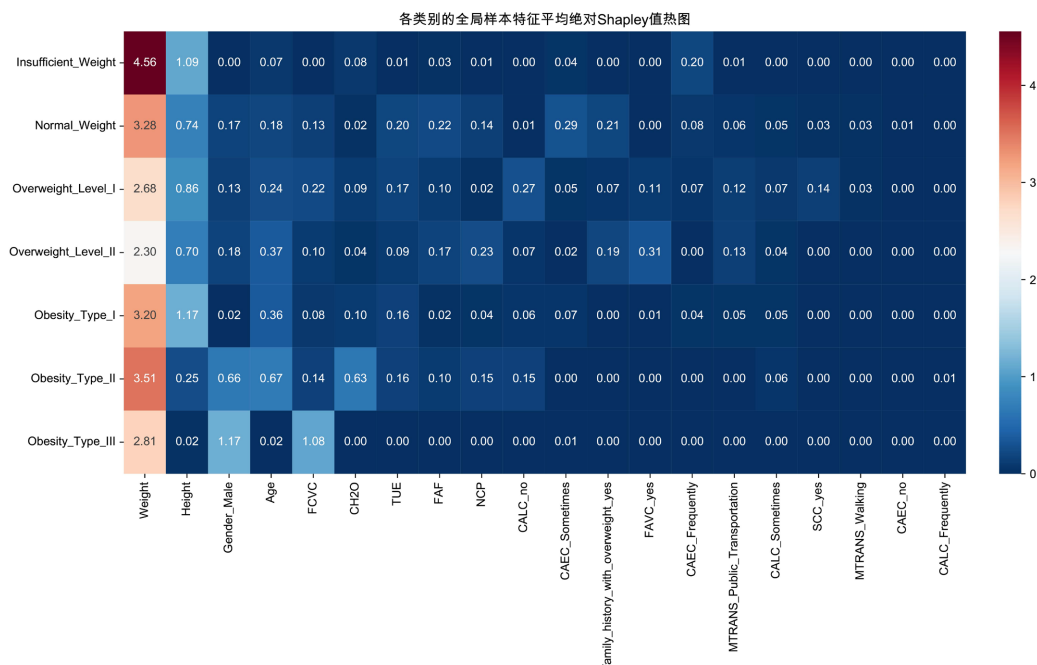
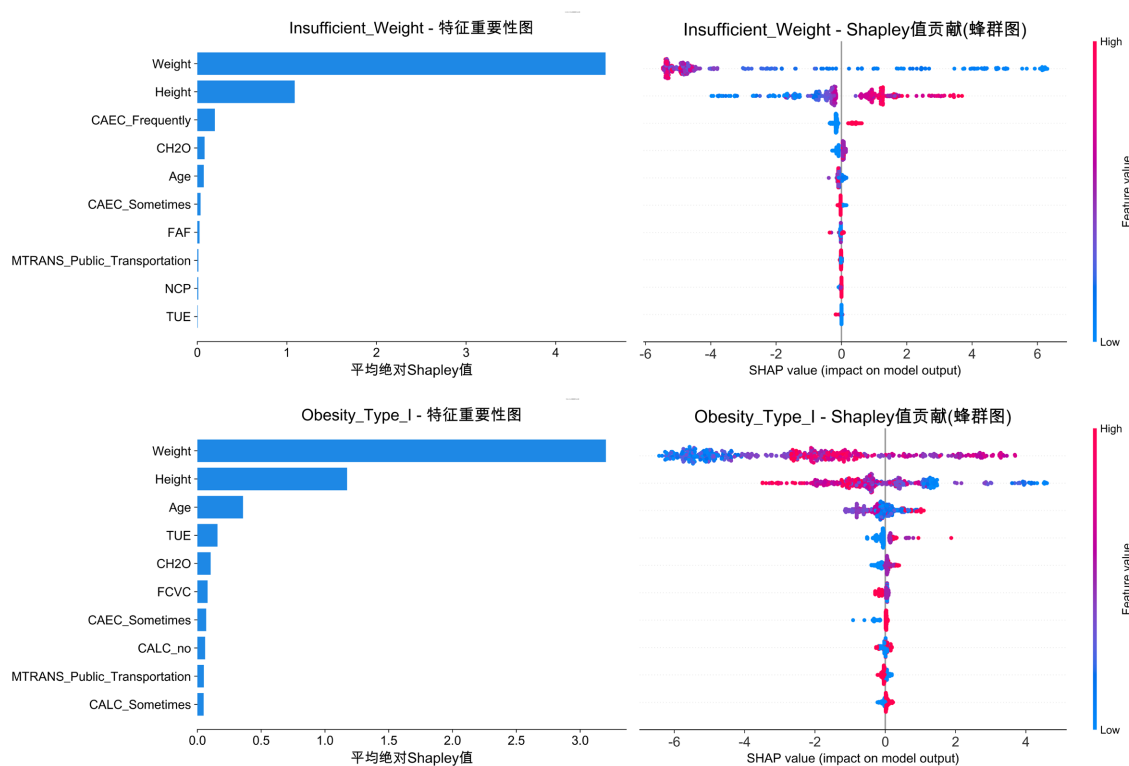


Figure 4. Heatmap of average Shapley values for global sample features

图 4. 全局样本特征的平均 Shapley 值热图



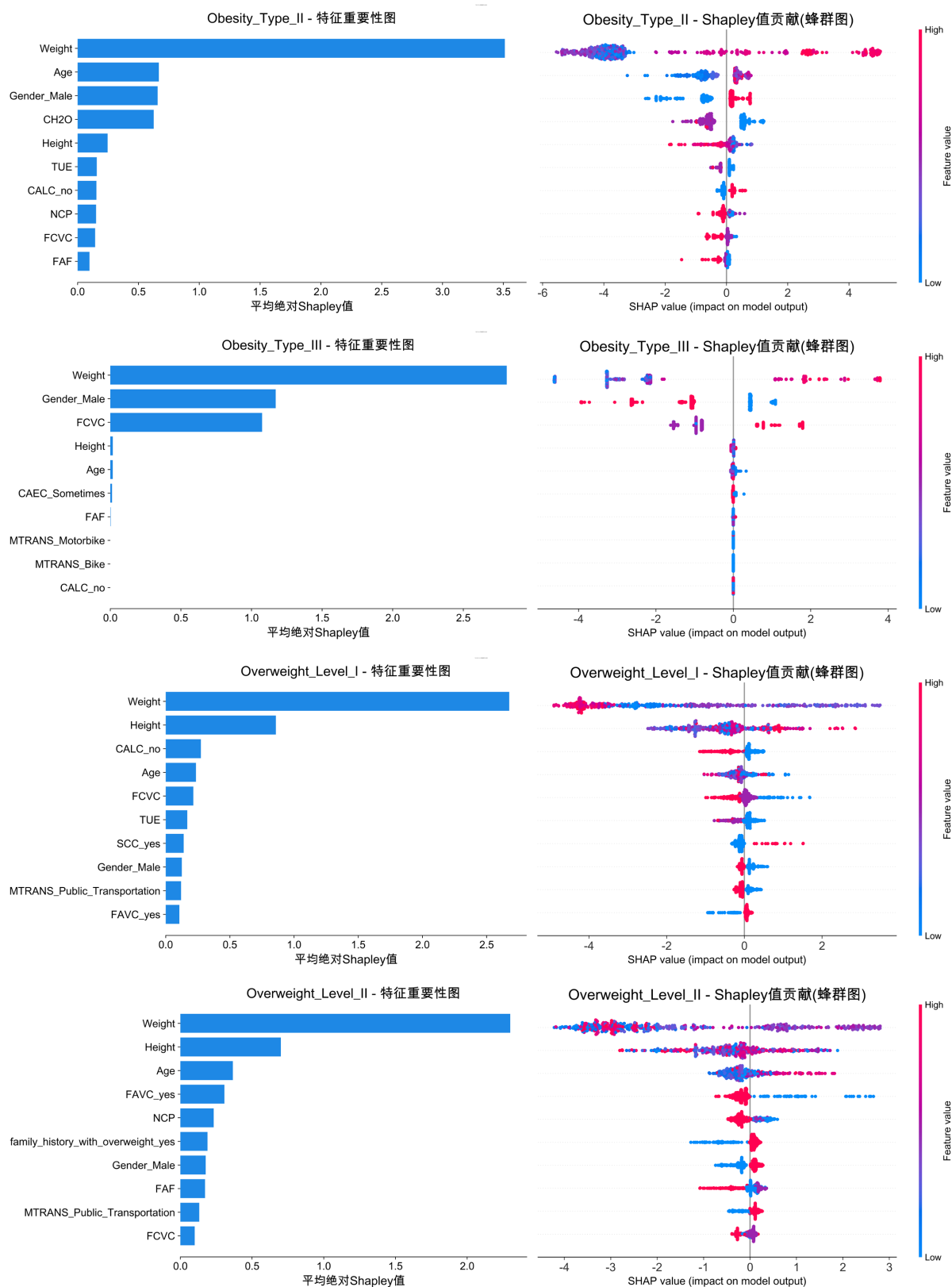


Figure 5. Ranking of the importance of different types of obesity characteristics
图 5. 不同肥胖类型特征重要性排名

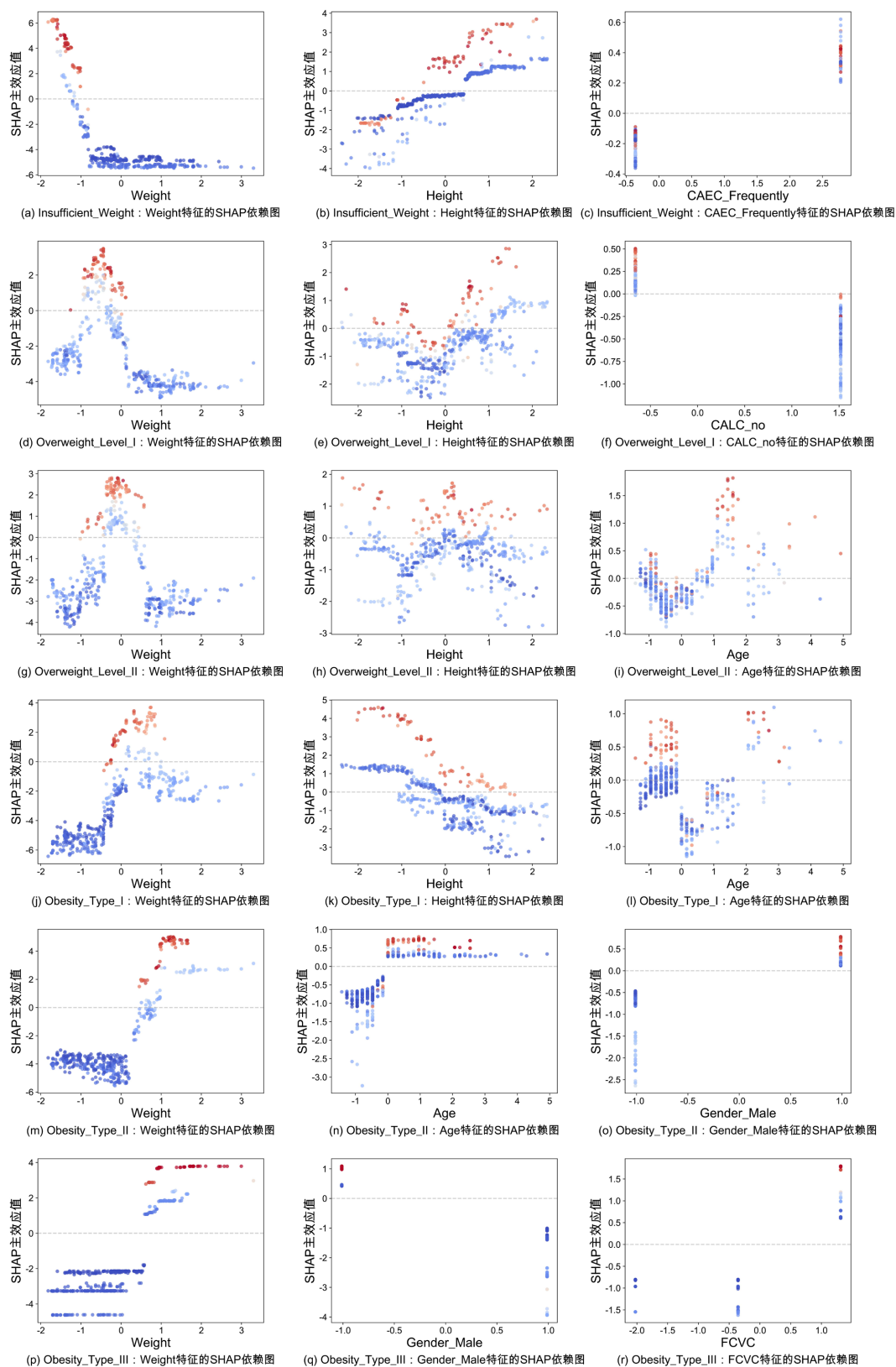


Figure 6. SHAP dependency plots of selected important characteristics for each obesity category
图 6. 各肥胖类别部分重要特征的 SHAP 依赖图

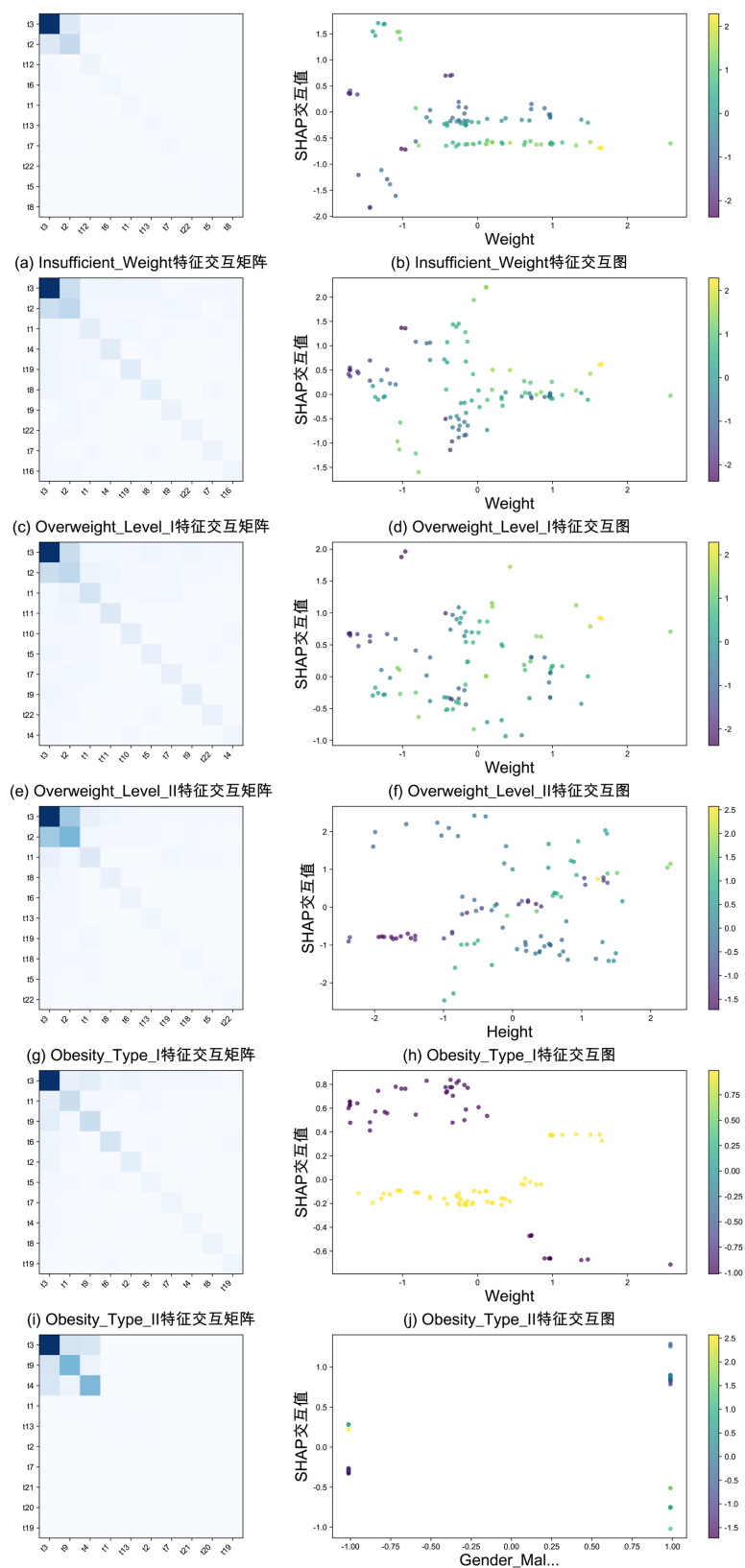


Figure 7. Interaction matrix and interaction plots of SHAP features for each obesity category
图 7. 各肥胖类别 SHAP 特征交互矩阵及交互图

各肥胖类别部分重要特征的 SHAP 依赖图如图 6 所示,在对肥胖不同分类的 SHAP 依赖图进行深入分析时,可以观察到以下几点关键特征与模型输出之间的复杂非线性关系。体重在所有肥胖分类中与 SHAP 值呈显著正相关,体重增加,SHAP 值上升,肥胖风险提高,在超重 II 级、肥胖 I 型和 III 型中边际效应显著。身高总体与 SHAP 值呈负相关,身高增加通常伴随 SHAP 值下降,高身高可能略微抵消肥胖风险,但在体重不足和正常体重分类中影响较弱,在超重和肥胖类型中负相关性更显著。年龄在超重 II 级、肥胖 I 型和 II 型中,随着年龄增长,SHAP 值上升,与肥胖风险增加正相关。这可能与生理变化及生活方式转变有关,提示肥胖干预策略需考虑年龄因素。性别在肥胖 II 型和 III 型中,男性 SHAP 值显著高于女性,可能因生理、代谢和行为因素,男性更易患这些肥胖类型,性别特异性的干预策略应考虑这些差异。饮食习惯,特别是水果、蔬菜和豆类消费频率,在肥胖 III 型中对 SHAP 值有显著负向影响。

SHAP 特征交互矩阵及交互图深入揭示了肥胖不同分类中特征间的复杂交互作用,为理解肥胖多维度成因提供了关键视角。如图 7 所示,体重不足的特征交互矩阵显示体重与身高交互显著,交互图表明体重增加使 SHAP 值上升,身高增加则使其下降,反映能量摄入与消耗不平衡。超重 I 级和 II 级中,体重与身高的交互作用依旧显著,体重正向影响 SHAP 值,身高负向影响,提示身高对肥胖风险的调节作用。肥胖 I 型的矩阵凸显体重与身高的交互,交互图揭示身高与性别特征的交互作用,男性 SHAP 值更高,可能与男性生理结构、代谢特征及行为习惯有关。肥胖 II 型的分析凸显体重与性别的交互作用,两者均正向影响 SHAP 值,强化了性别在肥胖预测中的重要性。肥胖 III 型的矩阵强调性别与体重的交互作用,男性特征值与体重增加共同推动 SHAP 值上升,反映极端肥胖状态下男性的生理和代谢变化。总体而言,体重和身高在多数肥胖分类中交互显著,性别在肥胖 I 型、II 型和 III 型中影响突出,为精准预测肥胖提供了理论依据。

6. 政策适应性分析

本部分将拉美数据映射至中国场景,基于墨西哥、秘鲁和哥伦比亚人群数据进行聚类分析,并通过雷达图(图 8)直观揭示了在身高、体重、技术使用时间等 8 项特征上的差异化分布模式,将人群分为轻度肥胖活跃群体、高大体型重度肥胖和极度肥胖群体等 6 类。

基于聚类分析结果,本研究充分结合我国《健康中国行动(2019~2030 年)》及《“体重管理年”活动实施方案》,针对六大肥胖相关群体提出针对性干预建议:

6.1. 轻度肥胖活跃群体

通过健康行为聚类分析识别出占比 14.7%的轻度肥胖活跃群体(聚类 0),其特征表现为体重略高但具备较高的身体活动水平及显著的技术使用依赖性。针对该群体可以构建“数字环境适配”干预框架,在外卖平台嵌入动态营养提示系统可以实现膳食结构的有效调节。此外,整合物联网技术与智能穿戴设备构建办公场景饮水监测网络,结合卷积神经网络分析个体饮水模式。当系统检测到低饮水量时,触发智能手环震动提醒,通过实时生物与环境数据的动态反馈机制符合自我调节理论中的“监测-反应”闭环,在极大程度上增强健康行为的可持续性。

6.2. 高大体型重度肥胖

聚类 1 为重度肥胖群体,占比 22.5%。其形态特征表现为身高体重显著高于人群均值。该群体虽呈现较高蔬果摄入频率,但存在能量代谢三重失衡机制:隐性高热量摄入、静息代谢率代偿不足及久坐行为占比过高。针对该群体,可以实施“三位一体”的干预方案。构建代谢动态监测平台,开发具备连续动态采集分析体重变化并提供体重管理建议的移动应用程序(APP)和小程序,促进居民体重管理日常化、便

利化、规范化；构建代谢友好型环境，部署社区健康主题公园智能健身环道；同时，积极参与多种形式的全民健身活动。

各聚类特征雷达图

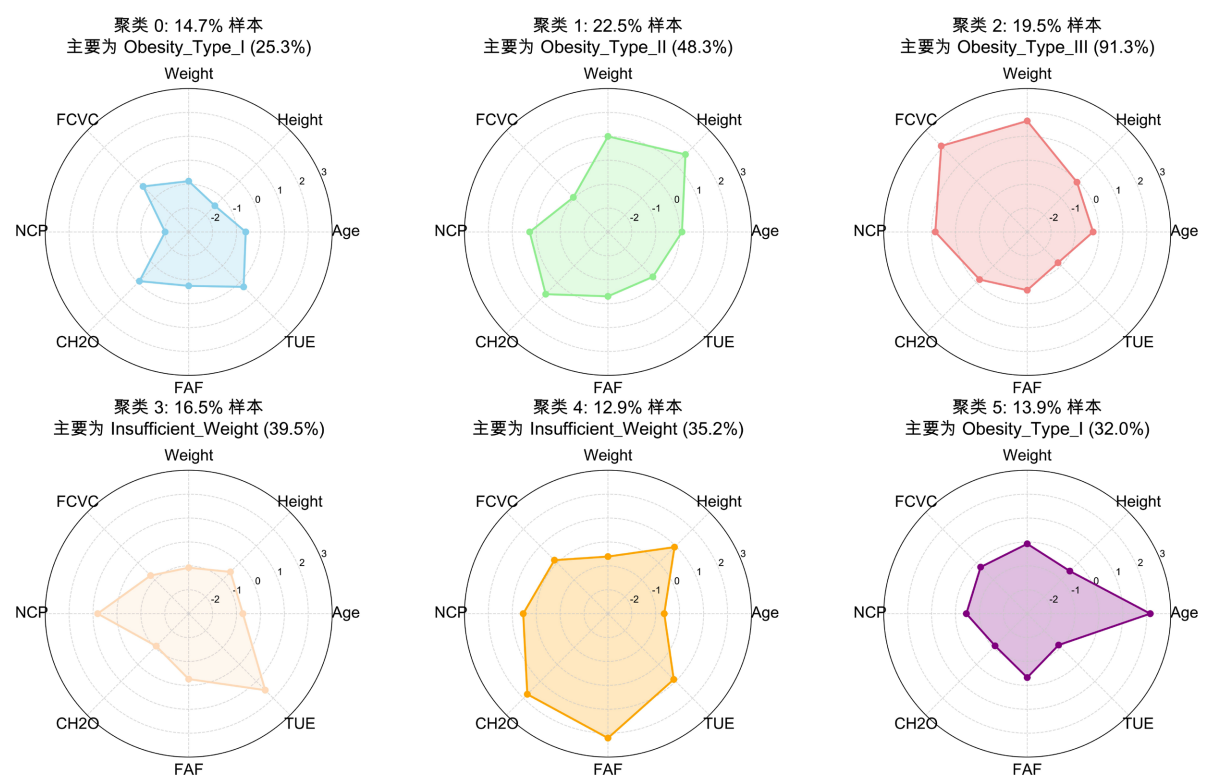


Figure 8. Radar charts of the characteristics of each cluster
图 8. 各聚类特征雷达图

6.3. 极度肥胖群体

聚类 2 为极重度肥胖群体，占比为 19.5%。其代谢特征表现为能量调节系统代偿性失调：尽管蔬果摄入量与饮水量达健康标准，但总热量摄入过高或代谢异常，导致极重度肥胖。该群体可以使用国家药监局认证的智能体脂秤进行代谢监测；实行阶梯式热量管控，通过厨房电子秤和相关 APP 实现每餐食材克重与热量的双重记录。进一步，引入中医体质辨识和包括饮食调养等在内的中医药保健指导服务。

6.4. 体重不足活跃群体

聚类 3 为体重不足活跃群体，占比为 16.5%，其呈现矛盾性代谢特征：高频进食行为与低能量密度饮食模式并存。针对该类群体，可以依据《国民营养计划》实施营养密度跃升计划，实施“333 法则”，即每日 3 次正餐和 3 次加餐，每餐确保含 30 g 优质蛋白；参与抗阻代谢激活训练，利用碎片时间提升肌肉合成效率。

6.5. 高个子体重不足群体

聚类 4 为高个子体重不足群体，占比为 12.9%。其呈现典型形态代谢解离特征：身高显著高于同年龄层，但体重指数远低于健康阈值；尽管水合状态优良，但能量摄入消耗比显著偏离平衡区间。针对这

类群体，可以制定科学增能计划，加强热量摄入；参照国家体育总局《体医融合运动指南》，开展“抗阻训练 + 营养补偿”组合干预；同时，加强自身代谢的动态监控。

6.6. 年长肥胖群体

聚类 5 为年长肥胖群体，占比为 13.9%。其特征表现为年龄普遍高于其他群体，技术设备使用时间极长，可能是久坐办公室工作而导致肥胖的中年群体。针对这类群体，其健康知识可能较为薄弱，需强化其代谢等知识的认知；定期测量体重等指标，开展体重管理综合干预指导，预防肌肉减少及肥胖，减缓机能衰退；实施低冲击运动处方；在社区健身中心配置体医结合监测系统，实时分析关节压力分布数据。

本研究通过多维度聚类分析，为各类人群提供了个性化的建议，为构建智能化的肥胖预警系统提供了跨文化实证基础，助力实现从粗放式健康方案向特征画像驱动的精准确定体重管理方案转型。结果可见，我国的《“体重管理年”实施方案》及《健康中国行动(2019~2030 年)》的政策方案提供了适用于各类人群体重管理的范本方案。

7. 研究结论与政策建议

7.1. 研究结论

本研究基于拉美三国 2111 份样本数据，通过 XGBoost 模型与 SHAP 可解释性分析揭示了肥胖风险的多维驱动机制。研究发现：体重、身高与年龄构成肥胖判定的核心生理指标，其中性别差异在重度肥胖类型中表现显著；饮食习惯呈现非线性影响特征，蔬菜摄入频率(FCVC)每提升 1 单位可降低 15.3% 的肥胖风险，而高频高热量饮食(FAVC)则使超重 II 级概率上升 28.7%；家族史对极端肥胖群体具有决定性作用。通过聚类分析识别出六类特征人群，跨文化验证表明，“饮食 - 运动”基础干预路径在拉美场景中保持 76.8% 的有效性，证实我国体重管理政策具备国际推广的科学基础。

7.2. 政策建议

1) 构建精准的体重管理干预体系

智能设备与动态监测：建议卫生健康部门与科技部门合作，开发一些简单易用的智能设备，像是智能体重秤、健康手环等，实时监测体重、基础代谢率等指标。这些设备可以在家庭和社区中广泛使用，为居民提供个性化的健康反馈。例如，通过手机应用提醒用户每日摄入的热量和消耗的运动量。

外卖平台健康提醒：在外卖平台上增加“健康饮食推荐”功能，对高频外卖用户实施智能提醒，帮助他们选择更健康的餐品。比如，当用户多次选择高热量食物时，系统会推荐一份低脂低糖的健康餐。

2) 全生命周期的健康管理

青少年代谢健康教育：在中小学课程中加入“代谢健康”教育内容，并通过活动激励学生进行健康行为，比如通过运动量兑换积分或奖励。通过这种方式，培养学生从小养成健康的饮食和运动习惯，降低未来肥胖的发生率。

老年人健康管理：考虑到老年人群体的特点，建议社区卫生服务中心为老年人提供专门的体重管理方案，利用智能设备定期监测其体重和基础代谢率，及时发现健康问题。同时，可以在社区健身中心推行“老年人适合的低强度运动”，如轻量抗阻训练和散步等，帮助老年人增强体力，减少肥胖及其相关疾病。

3) 社会共治，健康信息监管

加强健康信息的监管：建议建立一个多方合作的机制，对社交媒体和网络平台上的“反科学减肥法”

进行监管。通过技术手段,拦截不实的健康信息,避免不科学的减肥方法误导公众。

食品营养标签改革:市场监管部门可以推动《食品营养标签通则(2023 版)》的实施,尤其在食品包装上标明每份食物的热量和推荐的运动时长,让消费者在购买时能够直观地了解食物的热量含量,帮助他们做出健康选择。

参考文献

- [1] 中国营养学会肥胖防控分会,中国营养学会临床营养分会,中华预防医学会行为健康分会,等. 中国居民肥胖防治专家共识[J]. 中国预防医学杂志, 2022, 23(5): 321-339.
- [2] 李怡舟, 李有强. 基于体育健康传播视角的肥胖社会传染机制与纾解路径[J]. 体育科学, 2024, 44(9): 49-57, 70.
- [3] Kelly, T., Yang, W., Chen, C., Reynolds, K. and He, J. (2008) Global Burden of Obesity in 2005 and Projections to 2030. *International Journal of Obesity*, **32**, 1431-1437. <https://doi.org/10.1038/ijo.2008.102>
- [4] Dirlam, J. and Zheng, H. (2017) Job Satisfaction Developmental Trajectories and Health: A Life Course Perspective. *Social Science & Medicine*, **178**, 95-103. <https://doi.org/10.1016/j.socscimed.2017.01.040>
- [5] De Fano, D., Schena, R. and Russo, A. (2022) Empowering Plastic Recycling: Empirical Investigation on the Influence of Social Media on Consumer Behavior. *Resources, Conservation and Recycling*, **182**, Article ID: 106269. <https://doi.org/10.1016/j.resconrec.2022.106269>
- [6] Morris, Z.S. and Clarkson, P.J. (2009) Does Social Marketing Provide a Framework for Changing Healthcare Practice? *Health Policy*, **91**, 135-141. <https://doi.org/10.1016/j.healthpol.2008.11.009>
- [7] McCord, G.C., Conley, D. and Sachs, J.D. (2017) Malaria Ecology, Child Mortality & Fertility. *Economics & Human Biology*, **24**, 1-17. <https://doi.org/10.1016/j.ehb.2016.10.011>
- [8] Reyes-Sánchez, F., Basto-Abreu, A., Torres-Alvarez, R., Canto-Orsorio, F., González-Morales, R., Dyer-Leal, D.D., et al. (2022) Fraction of COVID-19 Hospitalizations and Deaths Attributable to Chronic Diseases. *Preventive Medicine*, **155**, Article ID: 106917. <https://doi.org/10.1016/j.ypmed.2021.106917>
- [9] 邓淑珍, 黄鑫, 常利涛, 等. 云南省部分民族 6-18 岁学生超重肥胖现状及影响因素分析[J]. 中国预防医学杂志, 2024, 25(12): 1550-1556.
- [10] Putnis, S.E., Klasan, A., Redgment, J.D., Daniel, M.S., Parker, D.A. and Coolican, M.R.J. (2020) One-Stage Sequential Bilateral Total Knee Arthroplasty: An Effective Treatment for Advanced Bilateral Knee Osteoarthritis Providing High Patient Satisfaction. *The Journal of Arthroplasty*, **35**, 401-406. <https://doi.org/10.1016/j.arth.2019.09.032>
- [11] Gouda, H.N., Charlson, F., Sorsdahl, K., Ahmadzade, S., Ferrari, A.J., Erskine, H., et al. (2019) Burden of Non-Communicable Diseases in Sub-Saharan Africa, 1990-2017: Results from the Global Burden of Disease Study 2017. *The Lancet Global Health*, **7**, e1375-e1387. [https://doi.org/10.1016/s2214-109x\(19\)30374-2](https://doi.org/10.1016/s2214-109x(19)30374-2)
- [12] Al-Bashtawy, M. (2015) Personal Hygiene in School Children Aged 6-12 Years in Jordan. *British Journal of School Nursing*, **10**, 395-398. <https://doi.org/10.12968/bjsn.2015.10.8.395>
- [13] Mehrsafari, A.H., Moghadam Zadeh, A., Jaenes Sánchez, J.C. and Gazerani, P. (2021) Competitive Anxiety or Coronavirus Anxiety? The Psychophysiological Responses of Professional Football Players after Returning to Competition during the COVID-19 Pandemic. *Psychoneuroendocrinology*, **129**, Article ID: 105269. <https://doi.org/10.1016/j.psyneuen.2021.105269>
- [14] Iyengar, K., Jain, V.K. and Vaishya, R. (2020) Pitfalls in Telemedicine Consultations in the Era of COVID 19 and How to Avoid Them. *Diabetes & Metabolic Syndrome: Clinical Research & Reviews*, **14**, 797-799. <https://doi.org/10.1016/j.dsx.2020.06.007>
- [15] Ameyaw Korsah, K. and Ameyaw Domfeh, K. (2020) Research Topic: The Realities of Religious Coping Experiences of Patients with Diabetes Mellitus: Implications for Policy Formulation in Ghana. *International Journal of Africa Nursing Sciences*, **13**, Article ID: 100245. <https://doi.org/10.1016/j.ijans.2020.100245>
- [16] Stern, E., van der Heijden, I. and Dunkle, K. (2020) How People with Disabilities Experience Programs to Prevent Intimate Partner Violence across Four Countries. *Evaluation and Program Planning*, **79**, Article ID: 101770. <https://doi.org/10.1016/j.evalprogplan.2019.101770>