（八）人工智能在用药依从性中应用

用药依从性是指个体在服药时间、剂量和频率方面遵循处方方案的程度，是确保治疗疗效和长期健康结果的关键决定因素。然而，用药不依从在临床实践和公共卫生中仍是一个重大问题，尤其对于糖尿病、高血压和心血管疾病等慢性病患者，可导致健康状况恶化、住院率上升及沉重的经济负担。现有监测方法主要依赖患者自我报告、药房续药记录或照护者观察，这些方法主观性强、易受回忆偏差影响，且多为间接测量。在此背景下，人工智能技术的快速发展为用药依从性评估开辟了新路径。近年来，两类技术路线尤为值得关注：一是以大语言模型为代表的模拟方法，通过在虚拟环境中测试心理测量工具来预判真实世界依从模式；二是以计算机视觉为代表的目标检测方法，通过对患者拍摄的药品图像进行自动识别与计数来实现客观评估。二者虽技术路径迥异，但共同指向了从主观报告向客观、自动化监测转型的学术趋势。

摩洛哥 Hassan 第一大学 Maryem Arraji 团队开展的一项探索性研究是利用 ChatGPT-4 生成 11 个摩洛哥二型糖尿病患者的虚拟档案，并施以已在摩洛哥阿拉伯方言中经过心理测量验证的一般用药依从性量表，以此评估在模拟环境中应用该量表的可行性。研究选取的变量包括年龄、性别、居住地（城市或农村）、教育水平、职业、健康保险状况、治疗类型（仅口服抗糖尿病药、仅胰岛素或联合治疗）、糖尿病病程及常见合并症（高血压、血脂异常），旨在涵盖摩洛哥临床实践中遇到的多样化情境。用药依从性量表由 11 个项目组成，分为三个维度：无意不依从（遗忘或无意中服药困难）、有意停药（自愿中断或调整治疗）和经济限制（与治疗费用相关的困难）。该量表的摩洛哥阿拉伯方言版本于 2022 年完成验证，显示出良好的内部一致性（Cronbach α ≈0.80）和因子效度（验证性因子分析>0.95）。专家评估结果显示，模拟档案在临床可信度（3.6±0.4）、文化/语言相关性（3.8±0.3）和内部一致性（3.7±0.5）三个维度上均获得较高评分（4 分制）。依从性评分结果显示，4 个档案（36.4%）呈高依从性（评分≥27），6 个（54.5%）呈中度依从性（17—26 分），1 个（9.1%）呈低依从性（<17 分）。交叉分析揭示出与文献一致的合理趋势：

仅接受口服抗糖尿病药物治疗的档案高依从性比例最高（75%），而仅使用胰岛素治疗者无一达到高依从性；有医保覆盖者高依从性比例为 57.1%，无医保者仅 25%；病程≤5 年者高依从性占 75%，而>10 年者仅 14.3%。按用药依从性量表维度分析显示，无意遗忘维度得分较高（10.8±2.1/12 分），经济限制维度则存在显著差异——无医保者（11.1±2.3）高于有医保者（7.8±2.0），农村档案（10.4±2.5）高于城市档案（8.6±2.4），胰岛素治疗者（10.9±2.6）高于仅口服药者（8.7±2.2）。该研究的价值在于方法学层面的创新——据研究者所述，这是用药依从性量表首次在 AI 模拟环境中应用于摩洛哥阿拉伯方言。通过大语言模型（Large Language Models，LLM）生成虚拟患者档案并在模拟环境中施测经过验证的心理测量工具，该方法可在真实世界调查前识别潜在偏差、优化方案并降低成本。然而，该研究存在明显局限：模拟样本仅 11 例，无法进行推断性统计分析；AI 生成的行为数据不能替代真实患者体验，且可能强化训练数据中的既有偏差；此外，模拟结果中高依从性比例（36.4%）远低于摩洛哥真实世界调查中通常超过 90%的水平，这虽与研究有意纳入“高风险”画像的方法学决策有关，但也提示模拟与真实之间仍存在显著差距。

与上述模拟方法不同，美国史蒂文斯理工学院 Haozhe Liu 团队从客观监测角度出发，开发了一个多模型计算机视觉框架，用于在多种真实拍摄条件下自动检测和计数药片，以支持剂量水平的用药依从性客观评估。该框架借用“专家混合”概念，整合了两种目标检测模型——YOLOv12（单阶段检测器）和 Faster R-CNN（两阶段检测器）——以及一个分割模型 Segment Anything Model（SAM）。框架分为两个阶段：第一阶段由两个检测器并行识别感兴趣区域；第二阶段以检测框为提示，通过 SAM 进行像素级分割细化，并辅以冗余去除和几何形状验证（确认分割区域符合药片的圆形轮廓），以抑制背景伪影造成的假阳性。研究使用公开 Kaggle 数据集（152 张图像）评估检测与计数性能，同时自行构建了模拟 20 天每日三剂服药场景的纵向依从性数据集（60 张图像，处方剂量为每张 4 片，其中 30 例依从、30 例非依从）。结果显示，所提框架在药片计数层面达到平均绝对误差（MAE）0.167，剂量层面准确率 0.933。在纵向依从性评估中，框架总体准确率 91.67%，检测依从病例的灵敏度 93.33%，识别非依从病例的特异性 90.00%。消融研究表明，单一检测器（YOLOv12 或 Faster R-CNN）的 MAE 为 0.300、准确率 0.733—0.767；加入 SAM 细化和冗余去除后 MAE 降至 0.200、准确率升至 0.900；完整框架加入几何形状验证后进一步优化至 MAE 0.167、准确率 0.933。在家庭、办公室、户外和餐厅等多样化场景中，框架均保持稳定的检测性能。该研究的意义在于填补了用药依从性监测中的一项重要空白

——现有方法大多依赖间接信号（如开瓶事件、续药记录），无法直接验证可见药片数量是否匹配处方剂量。本研究通过静态图像中的药片检测与计数，提供了客观、可扩展的剂量验证手段。然而，局限同样显著：依从性场景为模拟设定，未涉及真实患者服药行为；数据集规模有限；框架评估的是可见药片数量而非实际摄入情况。

上述两项研究代表了人工智能在用药依从性领域的两条差异化技术路径。摩洛哥团队的 LLM 模拟研究侧重于方法学预验证——在真实调查前通过 AI 生成虚拟患者并施测心理量表，以识别潜在偏差、优化研究方案；美国团队的计算机视觉研究则侧重于客观行为监测——通过自动识别患者拍摄的药品图像来验证剂量是否正确。二者均体现了 AI 技术在用药依从性评估中从“依赖主观报告”向“客观、自动化”转型的努力，但侧重点不同：前者服务于研究设计阶段的方法学准备，后者服务于临床监测阶段的行为验证。但是这两项研究都存在一定局限。其一，均基于模拟环境——LLM 研究生成的是虚拟患者档案而非真实行为数据，计算机视觉研究使用的是模拟服药场景而非真实患者队列，二者均缺乏与真实世界患者数据的直接比较。其二，样本量均有限——11 个虚拟档案和 60 张模拟图像均不足以支持统计推断或外部验证。其三，LLM 研究面临提示工程依赖、模型版本更迭影响可重复性以及“幻觉”风险等挑战；计算机视觉研究则受限于光照变化、药品重叠、背景杂乱等现实成像条件下的可靠性问题。

未来研究应着重推进以下方向：一是开展大样本真实世界队列研究，将 AI 生成或检测的结果与实际患者数据进行直接比较，以验证外部效度；二是将两类方法进行融合——例如，利用计算机视觉客观记录服药行为，再结合 LLM 模拟分析影响依从性的社会心理因素，构建多模态评估体系；三是在更多文化背景和慢性病种中验证方法的可迁移性；四是关注 AI 辅助依从性评估中的伦理问题，包括数据隐私、算法透明性和避免自动化临床决策。在人口老龄化与慢病负担持续加重的背景下，人工智能为用药依从性评估提供了从主观走向客观、从回顾走向实时的技术可能，但其从方法学探索走向临床常规应用，仍需经过严谨的验证与转化过程。

（徐健编辑）

译文一：

在摩洛哥利用 ChatGPT 和经过验证的用药依从性通用量表进行人工智能评估用药依从性评估

Maryem Arraji, Mohamed Khalis, Mohamed Chahboune

来源：Osong Public Health Res Perspect.

时间：2026 年 8 月

链接：https://doi.org/10.24171/j.phrp.2025.0396.

1. 简介

服药依从性是慢性病管理的关键，尤其是在二型糖尿病（T2D）中。患者对处方治疗不依从与发病率和死亡率升高以及医疗费用显著增加密切相关。尽管已有有效的治疗方法，但其在实际临床中的疗效在很大程度上取决于患者对治疗方案的依从性，而这种依从性在常规临床实践中仍处于较低水平。

在摩洛哥，T2D 的负担正在迅速增加。2023 年，国际糖尿病联合会估计成年人糖尿病患病率约为 12.4%，这一数字在过去二十年中持续上升。这一趋势由快速的营养转变、日益久坐的生活方式、人口老龄化以及传统医疗实践的持续性所推动。在此背景下，用药依从性受到多种障碍影响，包括经济限制、文化因素、认知限制以及获取医疗服务的困难。

现有多种工具可用于测量用药依从性。最初，巴基斯坦开发了一般用药依从性量表（GMAS），评估三个维度的依从性：非故意不依从、故意停药和经济限制。其阿拉伯语版本于 2022 年在摩洛哥被翻译、被文化认同并验证，显示出在二型糖尿病患者中表现出优异的心理测量特性（Cronbach α ≈0.80;confirmatory factor analysis >0.95;Kaiser-Meyer-Olkin [KMO] index >0.70）。因此，该工具在摩洛哥的用药依从性评估中尤为适用。在实地应用前，在模拟环境中测试此类工具，可作为方法学上的预验证步骤，帮助识别潜在偏差、优化方案，并降低实际调查的成本和时间。

与此同时，人工智能（AI）的兴起，尤其是大型语言模型（LLM），如 ChatGPT，为健康研究开辟了新方向。多项研究表明，这些模型在生成临床病例、支持医学培训以及协助科学写作方面具有实用性。然而，这些工具还存在一定的局限性，包括偏差、潜在错误和可靠性问题。

近年来，研究探讨了 ChatGPT 在模拟临床情境和在教育或探索环境中可信互动的能力。然而，迄今尚无研究评估其在特定文化背景中模拟患者特征，以应用验证心理测量工具的潜力。这一差距在中东和北非地区尤为重要，该地在医疗人工智能整合方面的科学贡献有限，且文化和语言语境化是数字方法有效性的核心。据我们所知，这是 GMAS 首次在基于 AI 模拟环境中应用于摩洛哥阿拉伯语方言。

在此背景下，本探索性试点模拟研究探讨了利用 ChatGPT 模拟摩洛哥二型糖尿病患者特征，以应用于本地验证的 GMAS 版本的可行性。本研究提出了一种方法论方法，结合经验类工具的心理测量优势与语言模型的生成灵活性，作为迈向现实世界用药依从性调查的准备步骤。

2. 材料与方法

2.1 研究设计

这项方法学探索性研究评估了使用大型语言模型（特别是 ChatGPT）生成摩洛哥二型糖尿病患者的虚拟画像的可行性。该方法旨在评估经过验证的摩洛哥阿拉伯语方言版 GMAS 在相应文化背景的模拟环境中应用。

选择该探索性设计旨在探讨 LLM 生成合理行为数据的能力，评估结果与文献研究的趋势一致，并完善量表给药方案和数据解读程序，以支持未来摩洛哥二型糖尿病患者在真实世界里的调查设计。

2.2 用药依从性评估工具

表 1 GMAS 的规模

一般用药依从性评分量表	描述
无意的不遵守	药物服用时的遗忘或无意中困难
有意停产	自愿中断或调整治疗
财务限制	与治疗费用或负担能力相关的困难

用药依从性使用经验类的摩洛哥阿拉伯方言版 GMAS 进行评估。GMAS 最初在巴基斯坦开发，由 11 个项目组成，分为三个主要维度。表 1 总结了这些维度。

本研究采用的翻译和文化适应版本是专门为摩洛哥二型糖尿病患者开发的，，并于 2022 年通过心理测量验证。该验证研究显示内部令人满意（Cronbach α ≈ 0.80），确认的因子效度（确认因子分析>0.95），以及充分的 KMO 抽样充足指数（>0.70）。这些心理测量特性支持 GMAS 在摩洛哥用药依从性测量中的相关性和稳健性。

2.3 使用 ChatGPT 的模拟过程

使用 OpenAI 开发的大型语言模型 ChatGPT-4 生成了 11 个摩洛哥二型糖尿病患者的虚拟档案。这些模拟画像基于摩洛哥背景下选定社会人口和临床参数构建，基于现有的本地流行病学数据及作者在 2D 管理方面的专业经验。所选变量包括年龄、性别、居住地（城市或农村）、教育水平、职业、健康保险状况、治疗类型（仅口服抗糖尿病药、仅胰岛素或联合治疗）、糖尿病持续时间及常见并存的疾病，如高血压和血脂异常。目标不是获取具有统计代表性的样本，而是涵盖摩洛哥临床实践中遇到的现实且多样化的情境。

最初提供给 ChatGPT 的提示包含详细说明，描述每个模拟患者的预期社会人口、临床和行为特征，包括年龄、性别、教育水平、健康保险状况、治疗类型和疾病持续时间。模拟在单次会话中进行，采用迭代方法对每个配置文件进行操作，以支持响应的内部一致性和协议的可重复性。所有数据均于 2024 年 8 月 1 日生成，该日期的报告是为了确保透明度和可追溯性。

模拟说明旨在重现代表摩洛哥人口的社会人口统计情境，包括年龄、性别、教育水平、职业、居住地和健康保险覆盖情况。现实的治疗模式，包括国家临床实践中主要的治疗方式，如仅口服抗糖尿病药物、单用胰岛素或联合治疗，以及对用药依从性态度的不同，从最佳到差的依从性不等，且与合理的情境因素相关。

每个模拟档案随后通过文本交互，按照面对面调查相同的管理程序，对经验类的摩洛哥阿拉伯语方言版 GMAS 中的 11 个项目进行管理。

ChatGPT 生成的回答被完整记录，按照官方 GMAS 评分系统编码，并被解读，以确定依从性为高、中或低。

图 1 展示了方法学步骤：定义代表性画像、GMAS 的执行、响应的标准化编码，以及使用 IBM SPSS 23.0 版（IBM 公司）进行描述性分析。

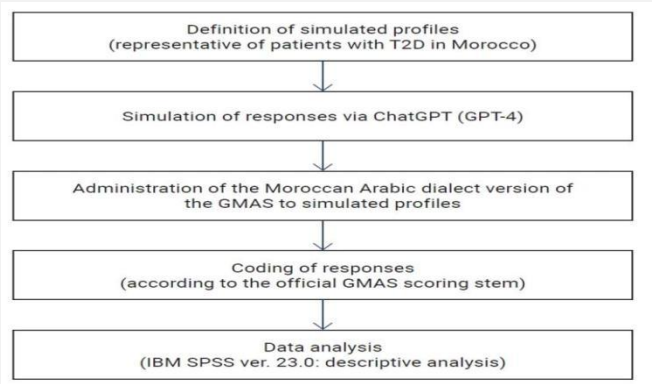

图 1 仿真协议模式

2.3 样本选择与分析的合理性

利用 ChatGPT（GPT-4）生成了 11 个二型糖尿病患者的虚拟档案，模拟摩洛哥医疗环境中遇到的多样社会人口和临床情境。选择档案数量（n=11）是为了平衡两项要求：确保有足够的异质性以体现影响用药依从性的关键因素，包括年龄、性别、教育水平、社会经济地位、健康保险类型、治疗方案、合并症以及居住地，以及保持可管理的样本量以进行探索性方法学研究。每个档案均基于国家流行病学数据及摩洛哥以往关于二型糖尿病的研究中报告的现实世界分布，进行迭代提示创建。由内分泌学家、糖尿病教育护士和生物统计学家组成的专家小组审查了生成特征的合理性，临床支持性和文化相关性。经过专家审核，这 11 份资料采用了摩洛哥阿拉伯方言版的 GMAS。数据进行了描述性分析，并与摩洛哥及国际场所已发表的依从性研究进行比较，以评估其一致性和方法学可行性。

模拟样本数量被有意限制，因为本研究设计为探索性试点模拟，旨在评估方法学可行性，而非统计代表性。类似的试点和基于仿真的方法学研究通常使用小样本来评估其合理性、方案一致性和操作可行性，然后再进行大规模实施。未来使用更大模拟数据集（如 50 – 100 个画像）的研究，可能有助于评估该方法的可靠性、可重复性和可扩展性。

2.4 基于专家的定性验证

三位独立专家评估了 ChatGPT 生成的回答对 11 个模拟档案的合理性和文化相关性。该小组由一名内分泌学家、一名糖尿病护理教育者和一名方法学家兼生物统计学家组成。

这些画像基于预先定义的社会人口和临床参数构建，包括年龄、性别、居住地、教育水平、健康保险覆盖范围、治疗类型——口服抗糖尿病药物、胰岛素或联合治疗——以及可能的并存的疾病。然而，这些模拟参数并未向专家披露。每位专家仅获得了模拟画像的完整叙述描述及其对经验证的摩洛哥阿拉伯语方言版 GMAS 的回应。

每位专家评估了三个主要维度：回答的临床可信度、文化和语言相关性，以及画像特征与相应回答之间的内部一致性。

此步骤旨在确定模拟样本是否可信、文化适应性强，并符合摩洛哥实际情况，并进行初步分析。专家们对具体的提示指令和模拟生成过程一无所知。每位专家在参与共识讨论前，独立审查了模拟画像及对应的 GMAS 回答，以比较不同观点并确认专家判断在三个评估领域中的一致性。对于小规模分歧，最终判断通过集体讨论达成，直到达成一致。鉴于该试点模拟研究的探索性质和样本量较小，专家评估旨在通过定性方法验证，而非正式的心理测量可靠性分析。ChatGPT 生成的原始回复未做任何修改。所有专家评估均有记录并进行描述性分析。

2.5 数据处理

模拟过程中生成的数据通过 IBM SPSS 23.0 版录入和处理。

分析包括对经验证的摩洛哥阿拉伯语方言 GMAS 中 11 个题目的回答的详细描述，并根据官方评分系统，对用药依从性进行总体分类：高依从性，得分≥27;中度依从性，得分 17–26;低依从性，得分<17。

鉴于研究具有探索性质且模拟样本量较小（n=11），未进行推断性统计分析。其目的是检验模拟画像的内部一致性和可信度，作为未来真实世界研究的准备。

2.6 伦理考量

本研究未涉及真实患者数据。这些配置文件及其回复完全由人工智能使用 ChatGPT 4 生成。摩洛哥阿拉伯方言版的 GMAS 基于摩洛哥二型糖尿病患者的心理测验验证。在这项方法学研究中，仅咨询了独立专家——内分泌学家、糖尿病教育护士和生物统计学家——以评估模拟画像的合理性和文化相关性。这些专家在参与前提供了口头知情同意书，研究团队对此进行了记录。除了程序性考虑外，使用人工智能模拟患者行为也引发了关于透明度、问责制及潜在滥用等重要伦理问题。AI 生成的档案不应被解读为真实患者体验或临床判断的替代品。模拟输出也可能强化训练数据中的现有偏差，或过度简化复杂的心理社会依从性决定因素。因此，这种方法应仅被视为一种补充性探索研究工具。该研究遵循了医疗中负责任人工智能使用的一般原则，包括方法学透明度、承认局限性以及避免自动临床决策。

因此，这项研究为未来实地调查提供了方法学上的准备，既要最大限度地降低伦理风险，又要遵守《赫尔辛基宣言》关于涉及人类参与者的研究原则。

3. 结果

3.1 模拟样本的定性验证

专家们用 4 分评级标准评估了三个标准：1，差;2，普通;3，良好;4，优秀。该尺度反映了各维度的可信度和充分性。

表 2 总结了这些评估结果。高的平均分表明模拟画像总体上被认为可信、符合文化，并且符合其社会人口和临床特征。文化和语言相关性的平均得分最高（3.8±0.3），反映了模拟的语言和语境充分性。临床可信度（3.6±0.4）和内部一致性（3.7±0.5）也获得了很高的专家评级，支持了 ChatGPT 生成的画像与摩洛哥背景的一致性。没有档案在任何标准下得分低于 3，表明正面评价整体均质性。

表 2 11 个模拟档案中的标准均值专家评估

评估标准	Mean±SD	Mix	Max
临床可信度	3.6±0.4	3	4
文化/语言相关性	3.8±0.3	3	4
内部一致性	3.7±0.5	3	4

评分量表：1，差;2，普通;3，良好;4，非常好。

3.2 模拟配置文件的特征

表 3 展示了 11 个摩洛哥二型糖尿病患者的虚拟档案的社会人口和临床特征。模拟样本旨在从社会人口和医学角度重现摩洛哥真实临床实践中遇到的不同病例。

表 3 模拟画像的社会人口和临床特征

病例	年龄（y）	性	居住地	教育水平	职业	保险	处理类型	二型糖尿病持续时间（年）	并存疾病
P1	45	女性	城市	中学	失业	AMO	仅限 OAD 信号	2	没有
P2	60	男性	农村	文盲	农民	没有	仅注射胰岛素	15	HTN
P3	52	男性	城市	小学	车手	CNSS	OAD+胰岛素	8	HTN+ 血脂异常
P4	38	女性	城市	大学	在职经历	CNOPS	仅限 OAD 信号	3	没有
P5	70	男性	农村	文盲	退役	AMO	仅注射胰岛素	20	HTN+其他
P6	55	女性	城市	高中时期	裁缝	没有	两个 OAD 系统	10	血脂异常
P7	62	女性	农村	小学	家庭主妇	CNOPS	OAD+胰岛素	12	HTN
第 8 页	49	男性	城市	大学	公务员生涯	CNSS	仅限 OAD 信号	5	没有
P9	59	男性	农村	中学	店主	没有	仅注射胰岛素	17	HTN+ 血脂异常
P10	67	女性	城市	文盲	退役	AMO	OAD+胰岛素	11	HTN
P11	42	女性	城市	高中时期	护士	CNOPS	仅限 OAD 信号	6	没有

T2D，二型糖尿病；AMO，强制健康保险；OAD，口服抗糖尿病药物;HTN,高血压;CNSS，国家社会保障基金;CNOPS，国家社会福利组织基金。

这些档案的年龄范围在 38 至 70 岁之间，性别分布几乎均衡，6 名女性和 5 名男性。教育水平从文盲到大学教育不等，职业身份包括就业、失业、退休和非正规部门。居住地分布在城市和农村地区，可反映该国的领土和社会经济多样性（见表 3）。

健康保险的分配反映了摩洛哥的医疗覆盖模式，包括强制性健康保险（AMO）、国家社会保障基金（CNSS）、国家社会福利组织基金（CNOPS）以及无社会保障的保险体系。模拟治疗方案包括口服抗糖尿病药物单独、胰岛素单独及联合治疗。疾病持续时间从 2 年到 20 年不等，涵盖了近期病例和长期疾病（见表 3）。

若干病例还包括与二型糖尿病常见相关的合并症，尤其是高血压和血脂异常，进一步增强了模拟面板的临床合理性和相关性（见表 3）。

3.3 根据 GMAS 的用药依从性评分

表 4 展示了 11 个模拟档案的 GMAS 评分及其对应的依从性水平。得分范围为 16 至 33 分，满分 33 分。

表 4 模拟档案的 GMAS 评分和用药依从性水平

病例	GMAS 评分（满分 33 分）	依从性水平
P1	30	高
P2	18	中等
P3	24	中等
P4	32	高
P5	17	中等
P6	25	中等
P7	20	中等
P8	31	高
P9	16	低
P10	22	中等
P11	33	高

在这些档案中，4 个档案（P1、P4、P8 和 P11）表现出高依从性（得分≥27），6 个显示中等依从性（得分 17–26），1 个档案（P9）显示低依从性（得分<17）。

该分布表明行为多样性合理，符合模拟变量，包括教育水平、健康覆盖率、治疗复杂性和并存的疾病。这些发现表明，在所采用的提示策略下，ChatGPT 能够生成反映合理用药依从态度的差异化画像，从而使 GMAS 在探索性语境中能够连贯应用。

3.4 根据潜在解释变量对依从性进行交叉分析

表 5 对三个选定变量——治疗类型、健康覆盖和糖尿病持续时间——进行了用药依从性的描述性交叉分析。观察到几个趋势。仅接受口服抗糖尿病药物治疗的病例依从性最高（75%），而单用胰岛素治疗的病例未达到此水平。疾病持续时间为≤5 年，75%病例与高依从性相符，而>10 年患者为 14.3%。在健康保障方面，57.1%的

有保险档案（CNSS、CNOPS 或 AMO）表现出较高的依从性，而无保险档案中为 25%。其中一例无保险档案依从性较低，而有保险档案中无低依从性病例。

表 5 根据选定特征对依从性的描述性交叉分析

变量	高依从性（%）	中度依从性（%）	低依从性（%）
治疗类型：仅限 OAD（孤注一体）	75.0	25.0	0.0
治疗类型：OAD+胰岛素	33.3	66.7	0.0
治疗类型：仅注射胰岛素	0.0	66.7	33.3
保险：覆盖（CNSS/CNOPS/AMO）	57.1	42.9	0.0
保险：无保险	25.0	50.0	25.0
持续时间：≤5 年	75.0	25.0	0.0
持续时间：6–10 年	50.0	50.0	0.0
持续时间：>10 年	14.3	57.1	28.6

虽然基于一个小样本（n=11），但这些趋势与文献中报道的结果一致。他们认为，治疗简单性、通过医疗保险获得的制度支持以及疾病持续时间更短，与更好的用药依从性相关。由于模拟样本量有限，未进行统计检验。相反，目标是识别可行的趋势，以便在真实世界调查中进一步探索。

3.5 GMAS 的维度分析

表 6 根据选定的特征按维度划分的一般用药依从性量表平均得分

测量项	平均得分±SD	最小	最大	没有保险	有保险	农村	城市	仅限 OAD	胰岛素
无意的遗忘	10.8±2.1	6	12	–	–	–	–	–	–
自愿停药	11.2±3.4	5	13	–	–	–	–	–	–
财务限制	9.6±2.7	4	11	11.1 ±2.3	7.8 ± 2.0	10.4 ± 2.5	8.6± 2.4	8.7 ± 2.2	10.9 ± 2.6

按 GMAS 维度检查平均分数显示各领域间存在明显趋势（见表 6）。无意遗忘的平均得分较高（10.8±2.1 分（满分 12 分），反映了大多数模拟画像中该成分的良好控制。这种一致性表明，无论社会人口特征如何，提醒或服药习惯在模拟画像中都得到了良好整合。

相比之下，自愿停药维度的变异性更大（13 个中 11.2±3.4 个），反映了对治疗态度的差异，这些差异无法仅用经济因素来解释。这些差异可能反映了个体对治疗效果、治疗耐受性或长期动机认知的不同。

财务约束维度的总体平均分为 9.6±2.7 分（满分 11 分），表明整体经济负担适中，但存在显著差异。没有健康保险的档案得分高于有保险的（7.8± ±2.0），显示保险具有保护作用。居住地也显得相关：农村人口档案（10.4±2.5）往往比城市人口更易暴露（8.6±2.4），这很可能是因为补贴医疗服务的可及性更有限且间接成本更高。治疗类型也成为区分因素：接受胰岛素治疗的模拟患者（10.9±2.6）面临的经济压力大于仅口服抗糖尿病药物的患者（8.7±2.2），反映胰岛素及相关用品的较高成本。

总体而言，这些结果表明，尽管模拟样本中无意遗忘的控制相对较好，但自愿停药尤其是经济限制可能成为量身定制干预的优先目标。特别应关注无医疗保险患者、农村地区患者以及需要胰岛素治疗的患者，以减少对用药依从性经济的影响。

3.5.1 依从性等级的图形可视化

图 2 显示，超过一半的模拟画像具有中等依从性，表明在建模管理情景下仍有改进空间。低依从性较少见，且观察到的特征是与依从性较差相关因素，包括缺乏医疗保障和仅使用胰岛素治疗。

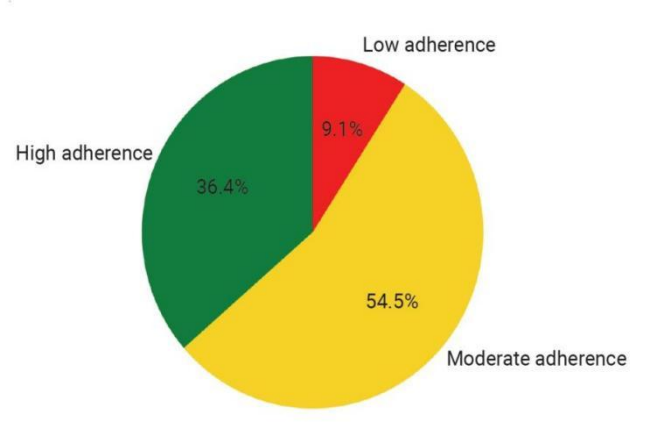

图 2 按用药依从性水平分布模拟画像

3.5 2 模拟与现实研究的比较

表 7 比较了模拟中获得的用药依从性水平与摩洛哥及国际实地研究报告的数据。模拟的依从性水平明显低于摩洛哥临床调查中观察到的高依从性比例通常超过 90%。

表 7 模拟画像与摩洛哥及国际研究的依从性水平比较

来源/背景	高（%）	中（%）	低（%）
ChatGPT 模拟（当前研究）	36.4	54.5	9.1
Maryem 等 [14] （2023）;摩洛哥（*n*=284，GMAS）	95.8	3.9	0.3
多中心研究[15];摩洛哥，2025 年（*n*=584，GMAS）	91.2	7.5	1.3
GMAS 验证[17];约旦（*n*=428）	83.4	12.1	4.5
GMAS 验证[16];沙特阿拉伯（*n*=502）	80.7	15.3	4.0

在模拟中，只有 36.4%的画像表现出高依从性，而 2023 年摩洛哥验证研究为 95.8%，2025 年多中心研究为 91.2%。相反，超过一半的模拟依从性显示中（54.5%），9.1%显示低依从性。这些水平在摩洛哥研究中很少被报告，但与一些国际场合中观察到的比例更为一致，包括约旦和沙特阿拉伯。

这些差异可能有多种解释：故意采用异质建模方法，旨在测试模拟协议的灵敏度，包括引入具有较低依从性的“高风险”画像；缺乏积极的人为因素，如医患关系、家庭支持或治疗教育，这些通常能增强真实调查中的依从性；以及方法论上的决定，将模拟定位为更多样化的案例，而非仅仅复制摩洛哥报道的高依从率，从而突出 ChatGPT 生成差异化画像的能力。

4. 讨论

这项探索性研究评估了利用 ChatGPT（GPT-4）模拟摩洛哥二型糖尿病患者特征并应用 GMAS 的可行性，这些方法已在摩洛哥阿拉伯方言中得到验证。这种方法使心理测量工具在模拟且在相应文化背景中能够测试其一致性和相关性，同时识别出用药依从性的合理趋势。

对生成的画像分析揭示出明显模式：在仅接受口服抗糖尿病药物治疗、有健康保险覆盖者以及病程≤5 年的患者中观察到最高依从性。相比之下，最低依从性则在无健康保险、仅接受胰岛素治疗者以及病程>10 年者中被建模。这些发现表明，即使在模拟环境中，社会经济、治疗和临床决定因素也可能塑造人工智能所建模的行为。

这些模拟模式不应被解读为观察到的患者行为。相反，它们应被视为 AI 生成的回应，旨在反映基于假设和提示策略的合理依从性情境。

这些结果与现实环境中的研究一致。Maryem 等人在 284 名二型糖尿病患者中验证了摩洛哥阿拉伯语方言的 GMAS 版本，展现出优异的心理测量特性。在使用该版本的实地调查中，95.8%的患者报告了高用药依从性。统计分析显示治疗类型与依从性有显著关联：接受单一口服抗糖尿病药物的患者依从性优于接受胰岛素治疗的患者。一项涵盖 584 名患者的多中心研究报告了疾病持续时间和治疗复杂性与依从性的关联。在约旦、沙特阿拉伯和苏丹进行的国际验证也强调了该工具的跨文化稳健性。

这些发现对邻近的阿拉伯语语境同样具有相关意义。在约旦、沙特阿拉伯和苏丹进行的验证研究表明，阿拉伯语版本的 GMAS 保持良好的心理测量特性，仍适合评估糖尿病或其他慢性病患者的用药依从性。这些证据支持 GMAS 在阿拉伯语群体间的跨文化适应性，同时强调了针对各地方语境进行语言和文化适应的重要性。在这方面，摩洛哥阿拉伯方言版具有重要贡献，因为它考虑到了摩洛哥二型糖尿病患者的具体语言、文化和医疗现实。

模拟档案的社会人口多样性反映了摩洛哥人口中观察到的关键变异，包括年龄、性别、教育水平、职业地位和社会覆盖率。按维度分析显示，低教育水平或日常工作量较重的个体中，无意遗忘更为常见；自愿停药常与治疗的负面看法或对其效用缺乏信心有关；财务约束则与保险状况相关联进行建模。这些动态在模拟响应中部分重现，展示了 ChatGPT 在预定义模拟条件下建模合理药物依从行为的能力。

从方法论角度看，这项工作在将先进语言模型整合到心理测量评估过程中，且在现实世界之外具有创新性。尽管以往研究已证明 ChatGPT 能够生成临床病例或通过体检，但据我们所知，尚无研究专门探讨在特定文化背景下使用人工智能进行经过验证的量表。

需要承认若干局限性。模拟样本量较小（n=11）、缺乏与真实患者数据的比较以及缺乏推断性统计分析，意味着结果不可推广，仅作为方法学可行性的证据来解释。

重要的是，本研究应被解读为概念性和方法论评估，而非对药物依从行为的实证验证。模拟样本未直接与真实患者级数据集进行比较。未来研究应包括将 AI 生成的

反应与从摩洛哥二型糖尿病患者收集的真实 GMAS 数据集进行直接比较，以评估外部效度和行为准确性。

增加模拟样本数量可以产生更大的行为变异性，并提高模式分析的可靠性。未来涉及 50 – 100 个或更多虚拟配置文件的大规模模拟，可能实现更稳定的趋势探索，并更好地评估不同提示策略和模型配置的重复性。

此外，尽管 AI 生成的结果是一致的，但它们未能考虑可能影响真实依从性的情感和心理因素，以及人类行为的复杂性，尤其是在社会和关系环境中。需要通过比较模拟与真实配置文件进行交叉验证，以确认该方法的外部有效性。

5. 基于 LLM 仿真中的偏倚与可重复性

使用 LLM 进行行为模拟带来了若干方法学挑战。首先，生成的回答高度依赖提示结构、措辞和上下文框架;小幅的即时调整可能影响依从模式和个人资料一致性。其次，可重复性可能因模型版本、参数配置以及 AI 提供者随时间引入的更新而有所不同。第三，尽管努力确保真实性，LLM 仍易产生幻觉，可能生成看似合理的回应，但不反映真实的患者推理过程。最后，确定性与概率生成机制之间的平衡可能影响重复模拟中响应的稳定性。这些局限性凸显了未来 AI 辅助健康研究中透明即时设计和标准化模拟协议的重要性。

尽管存在这些局限性，研究结果仍暗示了若干潜在应用：探索性研究可以在实际实施前识别脆弱特征；通过完善协议准备实际调查;以及通过患者文化和社会经济特征调整的模拟进行医治教育。

这种方法可以在其他阿拉伯语语环境中复制，并有可能推广到其他慢性病，从而促进人工智能在公共卫生研究中的伦理和有意识地整合。

这种方法也可能具有教学价值。它可以作为医疗专业人员和学生的培训工具，帮助他们快速识别有依从性不佳风险的患者。通过模拟多样的临床和社会经济情景，该方法能够支持分析和决策技能的发展，同时避免真实患者暴露于风险之中。将继续教育项目整合进继续教育项目还能提升医疗团队对心理测量工具的熟悉度，并帮助标准化评估实践。

从公共卫生角度看，如果通过大规模比较研究得到验证，这种方法论可以通过早期识别关键情形，从而减少准备研究阶段的时间和成本。该协议在其他语言及其他慢性疾病中（包括阿尔及利亚、突尼斯和埃及等阿拉伯语系语域）的可重复性，也为更广泛的文化和语言适应创造了机会，并可能增强该方法的国际适用性。

6. 结论

这项探索性研究凸显了人工智能，特别是 ChatGPT，在模拟真实患者档案和应用经过验证的心理测量工具于特定文化背景中的潜力。研究结果表明，即使在模拟环境中，社会经济、治疗和临床决定因素也可能影响用药依从性模式，支持模拟框架内生成档案的内部一致性。

尽管存在一些局限性，包括模拟样本量较小且无法直接与现实世界数据进行比较，但这种方法提出了若干潜在应用。它可以在实地研究前用于识别易感患者特征、完善研究方案，并支持个性化治疗教育。它也可能作为培训医疗专业人员和学生的教学工具，使他们能够练习分析多样化的病历，同时避免真实患者面临风险。

除了方法论贡献外，这项研究还因其新颖性而著称。据我们所知，它是 GMAS 在基于 AI 的模拟环境中最早的应用之一，并以摩洛哥阿拉伯方言进行验证。该应用凸显了将文化语境化与技术创新结合的重要性。

总之，尽管实证验证仍是心理测量研究的基石，这种方法代表了人工智能在公共卫生研究中伦理且有意识地整合的第一步。其潜在扩展至其他慢性疾病和语言语境，可能支持更广泛的文化和教学适应，并加强其国际适用性。

译文二：

基于计算机视觉框架的自动药片检测技术在用药依从性客观评估中的应用：开发与评价研究

Haozhe Liu, Yaguang Zheng, Bei Wu, Yu Gan

来源：JMIR AI .

时间：2026 年 8 月

链接：https://doi.org/10.2196/92528.

1. 简介

1.1 背景

用药依从性是指个体在服药时间、剂量和频率方面遵循处方药方案的程度。治疗方案的依从性是治疗效果和保持长期健康结果的关键决定因素。药物不依从在临床实践和公共卫生中仍是一个重大问题，尤其是对于患有糖尿病、高血压和心血管等慢性疾病的患者。不遵守处方药方案会导致健康状况下降、住院率上升以及患者和医疗系统面临沉重的经济负担。既往研究估计，未按时服药治疗可导致每年约 12.5 万可预防死亡病例，且当包括直接和间接成本时，总金额达 1000 亿至 3000 亿美元。

鉴于依从性在治疗效果中起着关键作用，准确评估用药依从性至关重要。然而，当前的监测方法主要基于患者自我报告、药房续药数据或护理人员观察。这些方法通常不可靠、主观且容易产生回忆偏差。这一挑战因认知障碍老人数量的增加而进一步加剧，他们可能在管理复杂的药物方案时更加困难。因此，迫切需要客观、自动化的解决方案来准确评估用药行为。

1.2 人工智能在用药依从性中的应用

人工智能是一个很有前途的客观评估用药依从性的方法。现已有多项研究是采用机器学习方法（如监督学习、深度学习），通过自我报告的依从性问卷来分类依从状态，但是此类方法仍高度依赖自我报告输入，因此报告不准确性会对结果产生显著影响。近年来，分析患者提供的图像和视频数据（如药物照片或短视频）的计

算机视觉模型日益受到关注。这些模型在非传染性疾病的管理中尤为重要，如癌症、心血管疾病、慢性呼吸系统疾病和糖尿病。通过自动检测和计数药品、识别药物容器或追踪服药行为，这些方法试图提供更客观且可扩展的依从性验证方法。在这些任务中，药品检测和计数是最重要的任务之一，因为它反映了所服用药物的剂量。与患者自报方法相比，基于视觉的深度学习技术减少了对患者自我报告的依赖，从而在现实环境中能更准确、及时和持续地评估用药依从性。

目前，研究人员已识别出基于人工智能的药品计数应用存在若干关键缺陷，这些缺陷限制了其可靠性和现实世界的可用性。这些缺陷包括对光线变化的敏感、背景杂乱、药品重叠和朝向等。这些限制可能导致药品计数不准确，并降低基于视力的方法在不受限制的真实环境中用药依从性评估的有效性。

1.3 本研究目标

为解决这些局限性，本研究旨在开发一个可靠且准确的计算机视觉框架，用于在多种真实影像条件下检测和计数，以支持客观的用药依从性评估。具体来说，我们评估了集成互补对象探测器、基于分割的细化、冗余消除和几何验证的多模型设计，是否比单个探测基线，更能提升药品计数准确性和剂量依从性分类。

1.4 相关工作

1.4.1 用药依从性监测技术

现已开发出多种技术来监测用药依从性，包括自我报告问卷、药品计数、药房续药记录、照护者观察以及电子用药监测设备。虽然这些方法在临床实践和研究中很有用，但许多方法只能间接或代理测量服药行为，例如，自我报告受回忆和社会期望偏差影响；药房续药记录确认的是获取而非正确剂量；电子监控系统、智能药瓶和智能用药依从性产品可以记录开瓶事件、提醒、配药记录或警报，但这些信号并不能验证被取出、计数或服用的药品数量是否正确。因此，当目标是直接评估符合剂量水平的药物摄入时，现有的依从监测方法仍然有限。

1.4.2 基于人工智能的药品识别、检测与计数

计算机视觉已被广泛用于药品识别、检测和验证。公众难题和深度学习使得消费者药品图片与参考图片的匹配成为可能，并提升了从移动手机或消费者照片中药品识别能力。这些研究展示了基于人工智能的药片图像分析可行性，但大多数研究侧重于药片身份识别，而非剂量依从性评估。其他研究则涉及药品定位、验证或计数，适用于更结构化的环境，如医院药房工作流程、基于器械的药品核查以及一般药品计数应用。虽然与药品检测和药物安全性相关，但用于剂量依从性评估的日常生活药品计数仍未被充分研究。本研究重点关注患者拍摄的静态图像中可见药片数量是否符合现实日常条件下的处方剂量。

2. 方法

2.1 开发了一个计算机视觉模型

我们构建了一个多模态框架，整合了两个成熟的物体检测模型和一个分割模型，其灵感来源于“专家混合”（MoE）的概念。如图 1 所示，三个模型——你只看一次模型（YOLOv12）、更快基于区域卷积神经网络（Faster R-CNN）和分段任意模型（SAM）——作为决策的三个专家。该框架包含两个阶段：第一阶段大致识别出人类手部中的感兴趣区域；第二阶段通过逐像素分割，在候选区域内对药片数量进行精细化处理。在第一阶段，两个独立的检测器旨在减少对单一检测范式的依赖，并提升对异构视觉模式的覆盖能力。在第二阶段，采用检测引导的分割策略，利用训练于大规模真实图像数据的基础模型 SAM 的泛化能力，实现跨不同成像条件下的鲁棒分割，而无需针对特定任务进行额外训练。这种模块化设计可在复杂的现实环境中实现可靠性能，并便于适应不同的服药场景。第一阶段识别出的潜在感兴趣区域被用作提示，生成药品掩码，从而显著降低 SAM 的计算成本，同时实现高效且准确的分割。该掩码生成步骤能够精确划分物体边界，在复杂现实拍摄条件下实现稳健的药片检测，并支持实时推理。这两个阶段通过在物体检测后去除冗余检测、在药品分割后进行形状平滑性检查而得到优化。当两个分割掩码之间的交并比（IoU）超过 0.8 时，对应的边界框检测被视为表示同一药品实例，并合并为单一检测，保留置信度更高且几何合理性更强的掩码。为了进一步抑制由背景伪影引起的假阳性，开发了一个形状平滑性检查模块，用于确认分割区域符合医用药品的圆形轮廓。

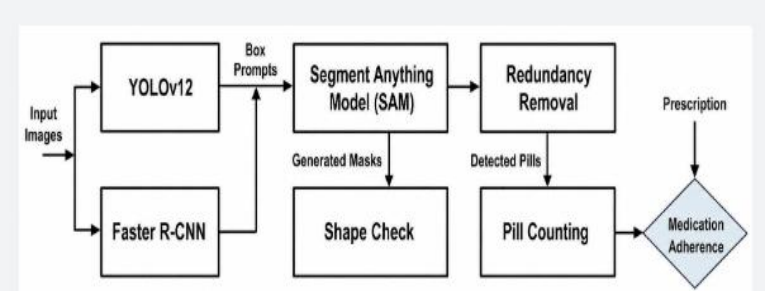

图 1　拟议用药依从性框架

具体来说，YOLOv12 和 Faster R-CNN 是并行且独立地应用于每张输入图像，它们的候选边界框被合并成统一的候选区域集，而不是顺序处理。每个候选框随后作为提示，SAM 生成相应的分割掩模，切片后通过掩膜层冗余消除和几何形状验证解决两个检测器间的矛盾，防止同一药品被多次计数。这种由检测器提示的 SAM 策略使该框架能够在像素层面细化候选药品区域，同时保持一级和两级探测器的互补检测模式。

2.2 数据收集

我们使用公开数据集评估药品检测和计数的表现，使用自己收集的数据集来评价用药依从性评估的表现。我们共从公开的 Kaggle 数据集中获得了 152 张图片。数据集被随机分为 10：2：3，用于训练、验证和测试。我们还进一步评估了在纵向用药依从性情境下的框架。具体来说，用药依从性的情景涵盖了四种具有代表性的场境（即家中、办公室、餐厅和户外）。精心整理的纵向用药依从性数据集包含 60 张模拟每日三剂服药 20 天情景的图像。处方剂量是每张图像四片。这些图像中，30 例是依从性病例，正好有四颗可见药片，30 例为非依从性病例，药片数量与处方剂量不同，包括 18 例剂量不足（1-3 片）、1 例漏服（0 片）和 11 例过量服药（5 片或以上）。非依从病例的真实药品计数为 0（n=1）、1（n=4）、2（n=9）、3（n=5）、5（n=5）、6（n=2）、7（n=1）、8（n=1）、9（n=1）和 10（n=1）。该分布用于评估框架是否能区分正确剂量、过量和过量剂量错误。

2.3 评估指标

我们直接评估药品计数和检测在药品层面和剂量（即影像）层面的表现。在药品层面，我们使用平均绝对误差（MAE）评估药片计数性能，MAE 定义为预测药片数与参考药片数的绝对差值。在剂量层面，我们评估是否能正确识别整张图像的准确

性。在这两个层级上，我们都比较了传统 AI 模型 Faster R-CNN 和 YOLOv12 的性能。纵向研究中，我们侧重于评估依从性与不依从性。具体来说，我们评估敏感性和特异性。敏感度以正确识别依从性次数的百分比来衡量。相比之下，特异性通过正确识别不依从性的次数来衡量（例如，如果服用五颗药片而非四颗，我们的人工智能模型能识别出药品数量为五颗，则正确归类为不依从）。对于这两种评估，最真实的都是对图像中药品数量的人工注释。

2.4 统计分析

由于所提出的框架和基线模型（YOLOv12 和 Faster R-CNN）在同一图像上进行了评估，因此采用了配对统计检验，所有比较分别针对公开的 Kaggle 药片检测数据集和纵向药物依从数据集进行。对于剂量级（图像级）准确性，精确的 McNemar 检验应用于所提出框架和每个基线的配对正确与错误分类结果，因此只有不一致的图像对，即一种方法正确分类但另一种方法未正确分类的图像对，才对检验统计量有贡献;由于不一致的图像对数量较少，采用了麦克内马检验的精确二项形式，而非 x^2 近似。对于药品计数，每张图像的绝对计数误差（定义为预测药品计数与参考药品计数的绝对差）使用 Wilcoxon 符号秩检验（一种不假设正态分布差异的非参数配对检验）在所提出的框架与每个基线之间进行比较。同一配对测试程序也应用于纵向用药依从性数据集中的依从性分类结果。所有检验均为双向，统计显著性定义为 $P<0.05$。

2.5 伦理考量

本研究使用了公开的药片图像数据集和一个为原型评估创建的策划模拟用药依从性数据集。由于该研究未涉及患者参与者、可识别个人、受保护的健康信息、临床记录或真实的用药行为，因此无需伦理审批。这些精选图像仅用于模拟药品计数和剂量依从性情景，用于对拟议计算机视觉框架的技术评估。

3. 结果

3.1 药品检测和计数的整体表现

本研究提出的多模态框架在剂量层面的准确率为 0.933，药品级 MAE 为 0.167，如表 1 所示。与同一图像集单模型基线相比，Faster R-CNN 在 MAE 为 0.300 的情况下实现了 0.733 的准确率，而 YOLOv12 在相同 MAE 为 0.300 的情况下实现了 0.767 的准确率。所提出的药品检测框架比对照模型实现了更高的准确性。具体来说，它显著优于 Faster R-CNN，McNemar 检验显示（*P*=.02），Wilcoxon 签名秩检验显示（*P*=0.03）。它也显著优于 YOLOv12，McNemar 检验显示（*P*=.02），Wilcoxon 签名排名检验显示（*P*=.03）。由于所有报告的 *P* 值均低于 0.05，这些结果表明所提框架在两个基线模型上的性能提升具有统计学显著性。我们框架获得的最高准确率和最低 MAE 表明其在药品计数中优于现有单一 AI 模型。我们还进行了消融研究，以评估各组分的贡献，如图所示表 1.单个探测器的平均平均能量（MAE）为 0.300，剂量水平准确率为 0.733-0.767。结合基于 SAM 的精细化和掩膜级冗余去除，将 MAE 降至 0.200，精度提升至 0.900;在完整框架中加入几何形状验证，MAE 进一步降至 0.167，精度提升至 0.933。

表 1 Kaggle 测试组的药片检测性能和消融、报告药品计数 MAE 及剂量准确性的研究

Method	SAM refinement	Redundancy removal	Shape check	MAE	Accuracy
YOLOv12	No	No	No	0.300	0.767
Faster R-CNN	No	No	No	0.300	0.733
YOLOv12 & Faster R-CNN+ SAM + Redundancy removal	Yes	Yes	No	0.200	0.900
Full proposed method	Yes	Yes	Yes	0.167	0.933

[a]MAE：绝对误差。

[b]SAM：分段任何模型。

[c]YOLOv12：你只看一次，第 12 版。

[d]R-CNN：基于区域卷积神经网络。

Faster R-CNN、YOLOv12 及所提方法在测试数据集上的详细药品计数误差分布如图 2。我们发现大多数误判案件只偏离了极少数的药品±1。

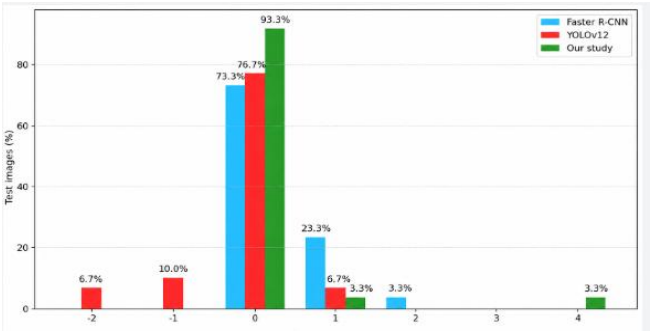

图 2 评估该方法在测试数据集上与 Faster R-CNN 和 YOLOv12 的性能比较

代表性结果显示于图 3.更快的 R-CNN 经常将看起来像药品的背景区域误判为药品，有时甚至无法检测到场景中的所有药品（图 3A).YOLOv12 减少了背景假阳性，但当药品密集聚集时，往往会为同一药品生成多个重叠的边界框，导致误差过度计数，如图所示图 3B.图 3C 展示了 SAM 使用检测边界框作为提示时产生的分割结果，这些框提供了更好的边界划定，但并不能直接解决冗余或误分类问题。通过整合 YOLOv12 和 Faster R-CNN 的单个检测数据，结合基于 SAM 的细化和几何验证，提出了该框架（图 3D）有效抑制冗余检测并去除虚假背景区域，从而实现更准确的药品定位和计数。

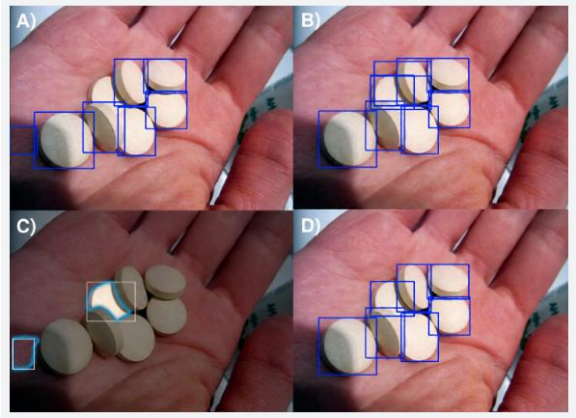

图 3 （**A**） 更快的 R-CNN 检测结果;（**B**） YOLOv12 检测结果;（**C**） 使用边界框作为提示的 SAM 分割结果;（**D**）我们改进的多模型框架的检测结果。R-CNN：基于区域的快速卷积神经网络;SAM：分段任何模型;YOLOv12：你只看一次，第 12 版

3.2 纵向用药依从性评估

依从性评估评估是否能自动判断受试者是否遵循处方（即每剂四片）。如图 2 所示，所提出的多模型框架总体准确率为 91.67%，检测依附病例的灵敏度为 93.33%，识别非依从病例的特异性为 90.00%。相比之下，我们观察到 YOLOv12 和 Faster R-

CNN 在真实世界数据集上的依从性模拟测试中普遍适用性有限。这一观察与先前的发现一致，即标准 Faster R-CNN 探测器在现实世界域变化下常常会出现显著的性能下降。通过采用基于大规模真实世界数据训练的 SAM 基础模型，所提出的框架展示了在多种真实世界成像条件下的鲁棒性。值得注意的是，Faster R-CNN 在我们策划数据集中的表现远低于我们整理的数据集(表 1），因为它在识别在户外、办公室、餐厅等各种场景下获得的药品具有有限的泛化性。这项试点研究表明，所提框架能够通过准确区分现实影像条件下依从和不合规服药行为，强有力地支持自动用药依从性评估。

表 2 在精选数据集上的依从性评估表现

指标	本研究	YOLOv12	更快的 R-CNN
准确性	0.917	0.600	0.183
灵敏度	0.933	0.467	0.067
特异性	0.900	0.733	0.300

我们还评估了所提出的框架在常见服药环境中的应用，这些环境代表典型的服药场景。药品图像采集于日常生活中常见的各种场所，如家中、办公室、户外和餐厅。这些图片涵盖了黑暗场景（图 4A）和明亮（图 4B).胶囊中也包含药品的形状（图 4A），圆 （图 4A），三角形（图 4B），以及长方形（图 4C，D).如图 4 我们的框架正确识别了这些情景中的药品数量。这些实验展示了在受控测试条件下真实环境变化下的稳定性。结果还显示了检测被封闭药品的能力。所提框架在这些场景下始终保持了可靠的检测和计数性能。

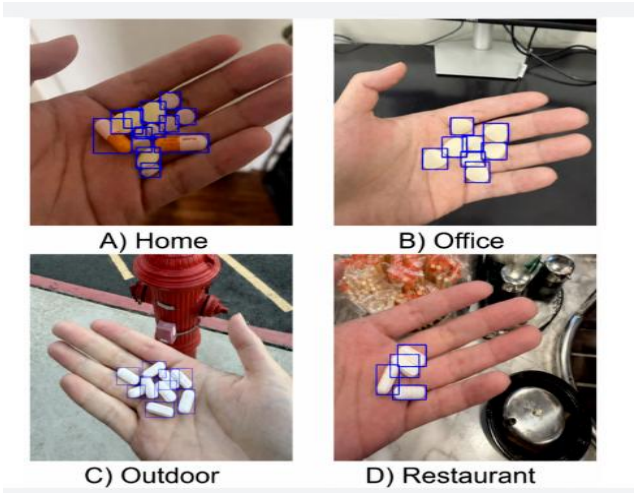

图 4 药品检测的代表性例子出现在常见的现实服药环境中，包括（A）家庭、（B）办公室、（C）户外和 D）餐厅环境

4. 讨论

4.1 摘要

这项概念验证研究开发并评估了一个多模型计算机视觉框架，用于药片检测和计数，以支持客观的用药依从性评估。所提框架在公开药品数据集和精心策划的纵向依从性模拟中均实现了低计药误差和高剂量级分类精度。这些发现表明，结合YOLOv12、Faster R-CNN、基于 SAM 的分割细化、冗余去除和几何验证，可以提升日常生活成像条件下基于静态图像的药品计数。

该方法弥补了用药依从性监测中的一项重要空白。许多现有的依从性监测方法依赖间接或代理信号，如自我报告、药房续药记录、护理人员观察、电子监测设备和智能药瓶。虽然这些方法有用，但可能无法直接验证可见的药品计数或剂量的正确性。在本研究中，所提出的框架优于评估的 YOLOv12 和 Faster R-CNN 基线，显示其在检测低剂量和过量情景下的潜力。

4.2 局限性

这项研究存在若干局限性。首先，依从性情景是模拟的，没有涉及真实患者服药行为。其次，策划的数据集规模适中，包含有限的药品外观和成像环境。第三，当前框架评估的是可见的药品数量，但不验证实际摄入情况。未来研究应在更大规模的真实世界人受试者队列中验证这一方法，并探索与纵向、保护隐私的依从监测工作流程的整合。

4.3 结论

本研究提出了一个多模态的计算机视觉框架，用于客观、基于图像的用药依从性评估。结果表明，基于静态图像的药片检测和计数可能成为可扩展的数字健康工具中用于剂量依从监测的有用组成部分。需要更大规模的真实世界研究来评估普适性、可用性和临床整合性。

*注：原文和译文版权分属作者和译者所有，若转载、引用或发表，请标明出处。