

ChatGPT辅助大学生中医体质辨别的可行性研究

王 茜^{1*}, 张 娟^{2#}

¹云南医药健康职业学院基础与药学院, 云南 昆明

²云南医药健康职业学院临床医学院, 云南 昆明

收稿日期: 2026年7月13日; 录用日期: 2026年8月6日; 发布日期: 2026年8月14日

摘 要

目的: 基于云南医药健康职业学院500名大学生中医体质分布数据, 评估ChatGPT-4.0模拟并复现《中医体质分类与判定表》判别逻辑的可行性与准确性。方法: 以《中医体质分类与判定表》对500名大学生进行传统量表测评, 计算9种体质转化分并判定主体质及兼夹体质, 并将该量表测评结果作为参照标准; 同时将量表症状描述转换为结构化提示词输入ChatGPT-4.0, 分析其输出结果与传统量表转化分的相关性及判别效能。结果: 500名大学生中偏颇体质者占71.20%, 主倾向体质前三位为阳虚质(29.00%)、阴虚质(20.80%)、湿热质(18.80%), 兼夹体质者占总人群50.00%; 女性阳虚质、阴虚质、气虚质、血瘀质、特禀质、气郁质转化分均高于男性, 平和质转化分低于男性($P < 0.05$); ChatGPT-4.0总体判别准确率为81.20% (406/500), $Kappa = 0.687$ ($P < 0.001$); 灵敏度居前三位为阳虚质(87.23%)、气虚质(85.42%)、平和质(83.33%), 特异度居前三位为特禀质(95.89%)、血瘀质(94.12%)、痰湿质(92.94%); 亚组分析中, 无兼夹体质组判别准确率为90.40%, 有兼夹体质组为72.00%, 非明确偏颇组为84.39%, 明确偏颇组为78.33%, 两组间差异均有统计学意义($\chi^2 = 27.589, 5.927, P < 0.05$)。结论: ChatGPT-4.0在复现《中医体质分类与判定表》的判别逻辑方面具有中等一致性(准确率81.20%), 对单一偏颇体质识别效能较好, 但在兼夹体质判别中准确率存在下降。研究表明, 大语言模型可作为高校中医体质量表化初步筛查的辅助工具, 但鉴于量表本身作为参照标准存在固有局限性, 目前尚不能替代专业中医师的综合判断。

关键词

中医体质, 体质辨别, ChatGPT, 大学生

Feasibility of ChatGPT-Assisted TCM Constitution Classification for College Students

*第一作者。

#通讯作者。

文章引用: 王茜, 张娟. ChatGPT 辅助大学生中医体质辨别的可行性研究[J]. 临床医学进展, 2026, 16(8): 1001-1008.
DOI: 10.12677/acm.2026.1682874

Xi Wang^{1*}, Juan Zhang^{2#}

¹School of Basic Medical Sciences and Pharmacy, Yunnan Medical Health College, Kunming Yunnan

²School of Clinical Medicine, Yunnan Medical Health College, Kunming Yunnan

Received: July 13, 2026; accepted: August 6, 2026; published: August 14, 2026

Abstract

Objective: To evaluate the feasibility and accuracy of ChatGPT-4.0 in simulating and replicating the discriminative logic of the *Classification and Determination of Constitution in Traditional Chinese Medicine Scale*, based on the distribution data of TCM constitution types among 500 students at Yunnan Medical Health College. **Methods:** A total of 500 college students were assessed using the *Classification and Determination of Constitution in Traditional Chinese Medicine Scale*. Transformation scores for the nine constitution types were calculated to identify the dominant constitution and coexisting constitution types for each participant, and these scale-based classification results were adopted as the reference standard. Additionally, symptom descriptions from the scale were converted into structured prompts and fed into ChatGPT-4.0. The correlation between ChatGPT's outputs and the transformation scores of the traditional scale, as well as its discriminative performance, were statistically analyzed. **Results:** Among the 500 college students, 71.20% presented with biased constitutions. The top three dominant types were yang-deficiency constitution (29.00%), yin-deficiency constitution (20.80%), and dampness-heat constitution (18.80%), with coexisting constitutions observed in 50.00% of all subjects. Females had significantly higher transformation scores for yang-deficiency constitution, yin-deficiency constitution, qi-deficiency constitution, blood-stasis constitution, special diathesis constitution, and qi-depression constitution, but a significantly lower score for balanced constitution compared with males ($P < 0.05$). The overall identification accuracy of ChatGPT-4.0 reached 81.20% (406/500), with a Kappa coefficient of 0.687 ($P < 0.001$). The top three types in terms of sensitivity were yang-deficiency constitution (87.23%), qi-deficiency constitution (85.42%), and balanced constitution (83.33%). The top three in specificity were special diathesis constitution (95.89%), blood-stasis constitution (94.12%), and phlegm-dampness constitution (92.94%). Subgroup analysis revealed that the identification accuracy was 90.40% in the group without coexisting constitutions, 72.00% in the group with coexisting constitutions, 84.39% in the non-obvious biased group, and 78.33% in the obvious biased group. The differences between the two subgroup pairs were statistically significant ($\chi^2 = 27.589$ and 5.927 respectively; both $P < 0.05$). **Conclusion:** ChatGPT-4.0 demonstrated moderate agreement (overall accuracy: 81.20%) in replicating the discriminative logic of the *Classification and Determination of Constitution in Traditional Chinese Medicine Scale*, with satisfactory performance in identifying single biased constitutions but decreased accuracy in cases involving coexisting constitutions. This study suggests that large language models may serve as an auxiliary tool for scale-based preliminary screening of TCM constitutions in university settings; however, given the inherent limitations of the scale itself as a reference standard, they are not yet capable of replacing the comprehensive judgment of professional TCM practitioners.

Keywords

TCM Constitution, Constitution Classification, ChatGPT, College Students

Copyright © 2026 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

中医体质学说属于中医药理论体系的重要组成部分, 为疾病的预防、诊断与治疗提供个体化的理论依据。流行病学调查显示, 偏颇体质占比达 62.19%, 其中兼夹体质占 40.23% [1]; 另有研究发现, 气虚质(48.14%)、阴虚质(46.58%)及阳虚质(37.38%)是主要的偏颇体质类型[2]。上述研究表明, 大学生群体的体质健康状况不容乐观, 亟须建立便捷、高效的体质辨别与健康管理途径。

当前中医体质辨别主要依赖《中医体质分类与判定表》[3]评测及医师四诊合参, 虽有较好临床适用性, 但在大规模人群筛查场景下存在效率较低、主观性较强等局限性; 而随着人工智能技术的不断发展, 其在中医药领域的应用也日益受到关注, 以 ChatGPT 为代表的大语言模型在中医诊断辅助方面的潜力开始被探索[4]。然而, 《中医体质分类与判定表》[3]虽为目前最广泛应用的标准化工具, 但其作为参照标准存在固有局限: 量表依赖个体主观自评, 缺失舌脉等客观诊察信息, 且对兼夹体质采用简化阈值判定, 难以全面反映中医体质辨识所需的全息信息。因此, 以量表结果评估 AI 效能, 实质上检验的是 AI 对该量表标准化判读逻辑的复现能力, 而非其真实临床诊断能力。

目前关于大语言模型在中医体质辨别这一具体场景中的应用研究仍十分有限, 其在大规模人群中的判别效能与可行性尚缺乏实证依据。基于此背景, 本研究以 500 名大学生为研究对象, 以《中医体质分类与判定表》[3]测评结果为参照标准, 旨在严格评估 ChatGPT-4.0 模拟并复现该量表判别逻辑的性能, 而非泛化评估其临床中医体质辨别能力, 以期客观审视大语言模型在高校中医量表化初步筛查中的适用边界提供参考依据。未来研究可进一步引入专业中医师判读结果, 构建量表 - ChatGPT - 医师三方对比框架, 以更全面评估其临床适用性。

2. 资料与方法

2.1. 研究对象

采用便利抽样法, 于 2025 年 12 月至 2026 年 1 月选取 500 例我校在校大学生为研究对象。样本量估算依据: 参考既往研究[5]中偏颇体质检出率约 63.65%~71.27%, 设定容许误差为 5%, 置信水平为 95%, 按横断面研究样本量公式 $n = Z^2 * p \times (1 - p) / d^2$ 计算, 所需最小样本量为 369 例; 同时, 考虑 10% 无效应答率, 最终确定样本量为 500 例。纳入标准: (1) 年龄 ≥ 18 岁; (2) 自愿参与本研究并签署知情同意书; (3) 能够独立完成问卷填写。排除标准: (1) 患有严重心、脑、肝、肾等系统原发性疾病者; (2) 存在精神或认知障碍, 无法配合调查者; (3) 调查期间因故休学或长期不在校者。500 例中男性 194 例(38.80%), 女性 306 例(61.20%); 平均年龄(19.10 ± 1.20)岁。

2.2. 中医体质判定标准

依据中华中医药学会 2009 年发布的《中医体质分类与判定》标准[6], 将中医体质分为平和质、气虚质、阳虚质、阴虚质、痰湿质、湿热质、血瘀质、气郁质、特禀质 9 种基本类型, 采用《中医体质分类与判定表》[3]进行体质测评, 该量表包含 9 个亚量表、共计 60 余个条目, 各条目按“没有、很少、有时、经常、总是”5 级 Likert 尺度计分; 各亚量表原始分经公式“转化分 = (原始分 - 条目数) / (条目数 $\times$ 4) $\times$ 100”换算为 0~100 分的转化分。平和质转化分 ≥ 60 分且其余 8 种偏颇体质转化分均 < 30 分者判定为平和质; 偏颇体质转化分 ≥ 40 分者判定为该体质类型, 转化分 30~39 分者判定为有该体质倾向(本研究将最终体质类型判定为“是”的个体纳入偏颇体质统计, 将“偏颇倾向”者单独归类分析)。

在亚组分析中, 划分标准与分组明细如下, (1) “明确偏颇组”: 至少一种偏颇体质转化分 ≥ 40 分, 且其余偏颇体质转化分均 < 30 分(无任何偏颇倾向); (2) “非明确偏颇组”: 包括平和质、基本平和质,

以及虽存在偏颇体质转化分 ≥ 40 分但同时兼有其他偏颇体质转化分 30~39 分者(两组划分依据为是否存在 30~39 分的体质倾向, 即体质诊断界限是否清晰); 偏颇倾向者指: 虽存在至少一种偏颇体质转化分 ≥ 40 分, 但同时兼有其他偏颇体质转化分处于 30~39 分者。量表判读与 ChatGPT 测试由两名研究者分别独立完成, 双方互盲, 待全部测试结束后由第三方合并数据并进行一致性分析。

2.3. 资料收集方法

采用问卷调查法收集资料, 包括一般资料(性别、年龄、年级、专业等)与《中医体质分类与判定表》[3]; 调查前向研究对象说明研究目的与填写要求, 征得知情同意后发放问卷, 当场填写并回收(本研究共发放问卷 530 份, 回收有效问卷 500 份, 有效回收率为 94.5%)。

2.4. ChatGPT 模拟体质判别测试设计

将量表中各体质亚量表中核心症状描述(如阳虚质条目“您手脚发凉吗”“您胃脘部、背部或腰膝部怕冷吗”, 气虚质条目“您容易疲乏吗”“您容易气短吗”等)转换为结构化问诊提示词, 输入 ChatGPT-4.0 模型, 提示词模板统一要求模型“模拟中医医师进行问诊, 基于以下症状信息判断该个体的 9 种中医体质转化分(0~100 分)”, 为确保输出结果的确定性与可重复性, 本研究通过 API 调用方式访问模型, 并将温度参数固定设置为 0, 以消除模型响应的随机性, 使同一输入在不同时间调用下输出结果一致。由两名研究者分别独立记录 ChatGPT 输出的 9 种体质得分, 取其平均值作为最终判别结果, 为计算判别准确率, 本研究参照《中医体质分类与判定》标准[6]转化分阈值, 对 ChatGPT 输出的 9 种模拟转化分进行相同的体质类型判定: 偏颇体质转化分 ≥ 40 分判定为该体质类型“是”, 30~39 分判定为“倾向”; 平和质判定标准同量表(平和质转化分 ≥ 60 分且其余 8 种偏颇体质转化分均 < 30 分), 在该判定规则下, 将 ChatGPT 输出的体质类型判定结果与金标准(量表判定结果)进行逐例比对, 以计算各体质类型的灵敏度、特异度及总体准确率, 其中, 当金标准判定为“倾向”(30~39 分)时, 因倾向本身不属于最终体质类型, 准确率分析中将“倾向”单独归为“非明确偏颇”类别, 不纳入“是/否”二分类的灵敏度/特异度计算。以《中医体质分类与判定表》[3]量表判定结果(转化分及体质类型)为金标准, 计算 ChatGPT 判别结果与金标准的一致性(评估指标包括: 总体准确率、Kappa 一致性系数、各体质类型的灵敏度与特异度)。

2.5. 统计学方法

采用 Excel 2016 建立数据库, SPSS 22.0 统计软件进行数据分析。计量资料先检验方差齐性(Levene's 检验)与正态性(Shapiro-Wilk 检验), 符合检验要求者以 $(\bar{x} \pm s)$ 表示, 组间比较采用独立样本 t 检验; 不符合检验要求者以 $[M(Q_1, Q_3)]$ 表示, 组间比较采用 Mann-Whitney U 检验。计数资料以 n 与 $[n(\%)]$ 表示, 计数资料组间比较采用 χ^2 检验。Kappa 一致性系数用于评估 ChatGPT 判别结果与金标准之间一致性程度。以 $P < 0.05$ 为差异具有统计学意义。

3. 结果

3.1. 大学生中医体质分布特征

500 名大学生中医体质分布特征及兼夹体质频次分布见表 1、表 2。

3.2. 不同性别体质分布比较

女性阳虚质、阴虚质、气虚质、血瘀质、特禀质、气郁质转化分均高于男性, 平和质转化分低于男性 ($P < 0.05$), 见表 3。

Table 1. Distribution characteristics of TCM constitutions among college students [n (%)]**表 1.** 大学生中医体质分布特征[n (%)]

类型	人数(例)	占比(%)
主体质		
非平和质	356	71.20
平和质	90	18.00
基本平和质	54	10.80
主倾向体质		
阳虚质	145	29.00
阴虚质	104	20.80
湿热质	94	18.80
气虚质	58	11.60
气郁质	41	8.20
特禀质	23	4.60
痰湿质	20	4.00
血瘀质	15	3.00
兼夹总体情况		
有兼夹体质	250	50.00
无兼夹体质	250	50.00
偏颇倾向者	93	18.60

Table 2. Frequency distribution of coexisting constitutions**表 2.** 兼夹体质频次分布

类型	出现频次(次)	占兼夹总频次比例(%)
气虚质	164	17.60
阴虚质	139	14.90
阳虚质	121	13.00
痰湿质	121	13.00
气郁质	105	11.30
血瘀质	103	11.10
湿热质	91	9.80
特禀质	87	9.30

Table 3. Comparison of transformation scores of each constitution type between different sexes [$M(Q_1, Q_3)$, scores]**表 3.** 不同性别各体质转化分比较[$M(Q_1, Q_3)$, 分]

体质类型	男性($n = 194$)	女性($n = 306$)	Z 值	P 值
阳虚质	25.00 (14.29, 46.43)	42.86 (28.57, 57.14)	7.3807	<0.001
阴虚质	31.25 (15.62, 46.88)	37.50 (25.00, 50.00)	3.58	<0.001

续表

气虚质	28.12 (12.50, 50.00)	37.50 (25.00, 53.12)	4.4912	<0.001
痰湿质	25.00 (6.25, 42.97)	25.00 (9.38, 40.62)	1.3819	0.160
湿热质	27.08 (8.33, 45.83)	29.17 (13.54, 45.83)	1.3047	0.215
血瘀质	17.86 (3.57, 32.14)	25.00 (11.61, 42.86)	4.0932	<0.001
特禀质	21.43 (3.57, 32.14)	25.00 (10.71, 39.29)	3.0464	0.004
气郁质	25.00 (3.57, 42.86)	28.57 (14.29, 50.00)	3.4764	<0.001
平和质	65.62 (50.00, 78.12)	59.38 (50.00, 68.75)	3.8585	<0.001

3.3. ChatGPT 判别效能分析

以量表判定为金标准, ChatGPT-4.0 总体判别准确率为 81.20% (406/500), $Kappa = 0.687$ ($P < 0.001$), 与金标准具有中等一致性。灵敏度居前三位为阳虚质(87.23%)、气虚质(85.42%)与平和质(83.33%); 特异度居前三位为特禀质(95.89%)、血瘀质(94.12%)与痰湿质(92.94%), 见表 4。

Table 4. Discriminative performance analysis of ChatGPT
表 4. ChatGPT 判别效能分析

体质类型	灵敏度(%)	特异度(%)
阳虚质	87.23	88.67
阴虚质	79.69	90.37
气虚质	85.42	89.38
痰湿质	70.59	92.94
湿热质	74.58	90.48
血瘀质	71.43	94.12
特禀质	61.54	95.89
气郁质	75.00	91.38
平和质	83.33	89.51

3.4. ChatGPT 判别效能亚组分析

在有无兼夹体质的亚组分析中, ChatGPT-4.0 在无兼夹体质组判别准确率为 90.40%, 在存在兼夹体质组判别准确率为 72.00%, 两组间差异具有统计学意义($\chi^2 = 27.589$, $P < 0.001$); 在偏颇程度亚组分析中, 明确偏颇组判别准确率为 78.33%, 非明确偏颇组判别准确率为 84.39%, 两组间差异具有统计学意义($\chi^2 = 5.927$, $P = 0.015$)。详见表 5。

Table 5. Subgroup analysis of the discriminative performance of ChatGPT-4.0
表 5. ChatGPT-4.0 判别效能的亚组分析

分组	例数	判别正确例数	准确率(%)	χ^2 值	P 值
无兼夹体质	250	226	90.40	27.589	<0.001
有兼夹体质	250	180	72.00		
非明确偏颇	237	200	84.39	5.927	0.015
明确偏颇	263	206	78.33		

4. 讨论

中医体质学说认为, 体质是人体生命过程中在先天禀赋与后天获得基础上所形成的固有特质, 其在疾病的发生、发展与转归中发挥着重要作用, 随着人工智能技术的快速发展, 以大语言模型为代表的智能工具在医疗健康领域的应用日益受到关注。目前关于大语言模型在中医体质辨别这一具体场景中的应用研究仍较为有限。本研究基于 500 名大学生的中医体质调查数据, 系统评估了 ChatGPT-4.0 在中医体质判别中的效能, 以期为人工智能技术在大规模人群中中医健康管理中的应用提供参考依据。

本研究结果显示, 500 名大学生中偏颇体质者占比高达 71.20%, 其中位居主倾向体质前三位分别是阳虚质(29.00%)、阴虚质(20.80%)及湿热质(18.80%), 兼夹体质者占 50.00%。这一结果与既往多项研究基本一致, 邵颖等[7]调查发现, 偏颇体质在大学生群体中较为突出; 一项针对 1568 名大学生的调查亦显示偏颇体质占比为 62.9%, 且呈现显著的性别特征[8]。上述研究均发现偏颇体质在大学生群体中具有高检出率, 由此提示其体质健康状况不容乐观, 亟须建立高效的体质辨别与健康管理途径, 但不同研究中大学生体质类型分布存在一定差异, 这可能与研究地域、样本特征及调查时间的差异有关。进一步发现, 不同性别大学生在体质分布上存在显著差异, 女性在阳虚质、阴虚质、气虚质、血瘀质、特禀质及气郁质方面转化分均高于男性, 而平和质转化分缺较低。分析产生这一差异的原因, 从中医理论角度分析, 《灵枢·阴阳二十五人》指出“妇女之生, 有余于气, 不足于血”, 女性以血为本, 经、带、胎、产等生理活动易耗伤阴血, 故阴虚质、血瘀质较为多见; 而男性以气为用, 生活工作中承受的压力及烟酒等不良习惯易致阳气受损、湿热内蕴; 而从现代医学机制层面来看, 性别差异与性激素水平的调控密切相关, 雌激素可通过调节免疫细胞功能、影响能量代谢及线粒体功能等途径参与体质的形成与维持[9]; 同时, 有研究发现, 阳虚质及寒证个体表现出较低的能量代谢水平, 表现为基础代谢率及细胞 ATP、丙二醛水平的降低, 线粒体酶活性亦有所下降[10]。

在 ChatGPT 判别效能方面, 本研究发现 ChatGPT-4.0 总体判别准确率为 81.20%, 与金标准具有中等一致性, 该结果提示大语言模型在复现标准化量表判读逻辑方面的确存在潜在应用价值; Liu 等[11]研究表明, GPT-4.0 等模型在中医诊断领域的表现与专业针灸医师具有较高的一致性, 其结果与本研究一致。然而, 必须审慎解读这一准确率, 由于本研究参考标准本身是结构化的自评量表, 其输入信息与输出规则均具有明确规范性与封闭性, ChatGPT-4.0 的高准确率更准确地说是其自然语言处理能力与规则匹配能力的体现, 而非其具备真正意义上的中医临床思辨能力, 这一区分对于界定 AI 在中医药领域的应用边界至关重要。亚组分析显示, 在存在兼夹体质个案中, 判别准确率从 90.40%降至 72.00%, 提示兼夹体质的复杂性可能影响模型的判别效能, 而这一发现与大语言模型在临床决策中的已知局限性相呼应; 同时研究提醒, 当前最先进大语言模型并不能在所有病理条件下准确诊断患者, 其表现劣于医师, 且不能遵循诊断或治疗指南, 也无法正确解读实验室结果[12]; 可见, 在存在多种症状、体征相互交织的复杂临床情境中, 大语言模型的信息整合与推理能力仍存在明显不足。就中医体质辨别而言, 兼夹体质涉及多种体质类型的症状重叠与交互影响, 模型难以准确区分主次与权重, 从而导致判别准确率下降; 此外, 大语言模型的判别效果高度依赖于输入信息的质量与结构, 而本研究将量表条目转化为结构化提示词的方式虽已尽量规范化, 但仍可能因语义表述的偏差影响模型的理解与输出。

本研究存在以下局限性: (1) 以《中医体质分类与判定表》[3]为参照标准, 该量表依赖个体主观自评, 缺失舌脉等客观诊察信息, 且对兼夹体质采用简化阈值判定, 故本研究的准确率结果实质反映 AI 对量表判读逻辑的复现能力, 而非临床辨别能力; (2) 样本仅来源于单所高校, 大学生群体在年龄、生活环境及健康认知方面具有较高同质性, 研究结果的外部有效性有待多中心、多样本人群验证; (3) 提示词设计虽经规范化, 但不同语义表述仍可能影响输出, 且未评估测试-复测信度。未来展望方面, 为增强该技术

在真实场景中的可用性与可靠性, 建议后续开发基于标准化问卷的前端应用程序, 由用户直接点选症状条目, 后台自动生成结构化高质量提示词并调用 API 接口, 从而规避普通用户自行编写提示词的困难与偏差; 同时可引入专业中医师判读结果, 构建将量表、模型与医师结合的三方对比评估体系, 以全面检验其临床适用边界。

综上, ChatGPT 在大学生中医体质判别中具有一定判别效能, 对单一偏颇体质的识别具有较高灵敏度与特异度, 但在面对复合兼夹体质的判别时, 其准确率存在下降; 临床中, 大语言模型可作为高校中医健康管理中体质初步筛查的辅助工具, 但其临床应用仍需结合专业中医师的四诊合参与综合判断, 目前不建议替代医师的专业决策。

基金项目

云南省教育厅科学研究基金项目资助(2024J2138)。

参考文献

- [1] 何汶霞, 王俊杰, 刘泽, 等. 郴州市某高校大学生中医体质兼夹规律调查[J]. 湘南学院学报(医学版), 2023, 25(4): 52-55.
- [2] 边文彦, 徐学权. 广州大学生行为方式与中医体质相关性研究[J]. 光明中医, 2025, 40(5): 861-864.
- [3] 中华中医药学会. 中医体质分类与判定[S]. 北京: 中国中医药出版社, 2009.
- [4] 潘晨浩, 周明. ChatGPT 在中医药领域中的应用前景[J]. 医学信息, 2023, 36(18): 15-20.
- [5] 王琦, 朱燕波. 中国一般人群中中医体质流行病学调查——基于全国 9 省市 21948 例流行病学调查数据[J]. 中华中医药杂志, 2009, 24(1): 7-12.
- [6] 中医体质分类与判定(ZYYXH/T157-2009) [J]. 世界中西医结合杂志, 2009, 4(4): 303-304.
- [7] <https://d.wanfangdata.com.cn/periodical/CiBQZXJpb2RpY2FsQ0hJU29scjkyMDI2MDcyNzAzMTgxMhINZ3N6eTl-wMjMxMTAyMRoIbTNma2s3cWQ%3D>
- [8] 李邱诺, 杨柯, 杨春红, 等. 1568 例大学生中医体质调查与调护——以贵州中医药大学为例[J]. 贵州中医药大学学报, 2024, 46(2): 86-90.
- [9] Hu, H., Yang, Q., Hu, J., Fan, Y., Zhang, Z., Yang, K., et al. (2026) ESRRA-ATG5-Mediated Mitophagy Enhances Arginine Metabolism to Alleviate Diabetic Kidney Disease. *Autophagy*, **22**, 666-690. <https://doi.org/10.1080/15548627.2025.2601874>
- [10] Ma, G., Li, C., Ji, P., Chen, Y., Li, A., Hu, Q., et al. (2023) Association of Traditional Chinese Medicine Body Constitution and Cold Syndrome with Leukocyte Mitochondrial Functions: An Observational Study. *Medicine*, **102**, e32694. <https://doi.org/10.1097/md.00000000000032694>
- [11] Liu, Y., Yuan, Y., Yan, K., Li, Y., Sacca, V., Hodges, S., et al. (2025) Evaluating the Role of Large Language Models in Traditional Chinese Medicine Diagnosis and Treatment Recommendations. *npj Digital Medicine*, **8**, Article No. 466. <https://doi.org/10.1038/s41746-025-01845-2>
- [12] Hager, P., Jungmann, F., Holland, R., Bhagat, K., Hubrecht, I., Knauer, M., et al. (2024) Evaluation and Mitigation of the Limitations of Large Language Models in Clinical Decision-Making. *Nature Medicine*, **30**, 2613-2622. <https://doi.org/10.1038/s41591-024-03097-1>