

· 论著 ·

DOI: 10.12464/j.issn.0253-9802.2025-0086

两种人工智能模型对于原发性骨质疏松症诊疗指南及临床骨密度检查解读的对比分析

胡朝峰, 郑松涛, 陈瑞强, 李尚福✉

(中山大学附属第三医院脊柱外科, 广东 广州 510630)

【摘要】 目的 为人工智能模型应用于骨质疏松症临床诊疗的效用提供实证依据。**方法** 选取原发性骨质疏松症诊疗指南核心章节, 由 10 位骨科医师从准确性、全面性、可解释性、临床应用性 4 个维度对 ChatGPT 和 DeepSeek 的回答内容与指南标准对比评分, 并结合临床骨密度检测结果, 评估模型决策的准确性与科学性。**结果** 10 位骨科医师评分一致性较高。DeepSeek 准确性评分高于 ChatGPT, 在药物管理及参数解析方面更精准契合指南更新要点。2 种模型全面性及可解释性评分虽无统计学差异, 但 DeepSeek 通过整合药物假期管理、FLS 协作模式及 OSTA 筛查工具构建多维评估框架, 并采用图表提升信息呈现效率。DeepSeek 临床应用性评分优于 ChatGPT, 其建立的决策路径与全周期监测体系有效缩短临床决策时耗。临床骨密度检测分析显示, ChatGPT 仅提供基础诊疗框架, 而 DeepSeek 具有更完整的指南依从性和临床适用性。**结论** AI 能提升骨质疏松症临床决策效率和质量, DeepSeek 解读的表现比 ChatGPT 更优异。

【关键词】 人工智能模型; 骨质疏松症; DeepSeek; ChatGPT; 精准医疗

Comparative analysis of two artificial intelligence models for the diagnosis and treatment guidelines of primary osteoporosis and the interpretation of clinical bone mineral density examination

HU Chaofeng, ZHENG Songtao, CHEN Ruiqiang, LI Shangfu✉

(Department of Spinal Surgery, the Third Affiliated Hospital of Sun Yat-sen University, Guangzhou 510630, China)

Corresponding author: LI Shangfu, E-mail: lishangfu@mail.sysu.edu.cn

【Abstract】 Objective To provide empirical evidence for the application of artificial intelligence model in the clinical diagnosis and treatment of osteoporosis. **Methods** The core chapters of the guidelines for the diagnosis and treatment of primary osteoporosis were selected, and 10 orthopedic surgeons compared the answers of ChatGPT and DeepSeek with the guidelines from the four dimensions of accuracy, comprehensiveness, interpretability, and clinical applicability. Combined with the results of clinical bone mineral density test, the accuracy and scientificity of the model decision-making were evaluated. **Results** The scores of 10 orthopedic surgeons were consistent. The accuracy score of DeepSeek was significantly higher than that of ChatGPT, and DeepSeek was more accurate in drug management and parameter analysis, which conformed to the key points of guideline update. Although there was no statistical difference in comprehensiveness and interpretability score, DeepSeek built a multidimensional evaluation framework by integrating drug leave management, FLS collaboration mode and OSTA screening tools, and used charts to improve the efficiency of information presentation. The clinical application score of DeepSeek was better than that of ChatGPT, and its decision-making path and full-cycle monitoring system effectively shortened the time consumption of clinical decision-making. Clinical bone mineral density analysis showed that ChatGPT only provided a basic diagnosis and treatment framework, while DeepSeek showed a more complete guideline compliance and clinical applicability. **Conclusion** AI can improve the efficiency and quality of clinical decision-making in osteoporosis, and the performance of DeepSeek interpretation is better than that of ChatGPT.

【Key words】 Artificial intelligence model; Osteoporosis; DeepSeek; ChatGPT; Precision medicine

收稿日期: 2025-03-10

基金项目: 广东省自然科学基金面上项目 (2021A1515010316)

作者简介: 胡朝峰, 硕士研究生, 研究方向: 骨质疏松症诊疗, E-mail: hcf1161212741@163.com; 李尚福, 通信作者, 副主任医师, 研究方向: 骨质疏松症诊疗, E-mail: lishangfu@mail.sysu.edu.cn

随着全球人口老龄化的加剧,原发性骨质疏松症作为全球性公共卫生问题挑战日益严峻,据估计,骨质疏松全球患病率约为19.7%^[1],已成为患者致残及死亡的重要因素^[2]。由于诊疗指南不断更新^[3],诊疗标准也逐步向多方面风险评估、精准化药物选择及动态监测方向发展。然而,由于指南内容复杂且涉及多学科交叉^[4-5],相关媒体及网络的骨质疏松信息良莠不齐,如何高效准确解读指南及临床骨密度检查结果已成为临床实践及大众科普的迫切需求。

近年来,人工智能(artificial intelligence, AI)在医学相关领域中的应用也越来越广泛^[6-7]。在牙科放射学领域, AI影像系统可精准识别龋齿及牙科植入物,推动个性化口腔健康管理模式的革新^[8]。在高等教育领域, AI驱动的教学评估体系已成为跨学科知识转化的例子^[9]。在骨质疏松领域, AI应用已涵盖抗骨质疏松药物研发的分子模拟^[10]、电子健康记录的骨折风险预测模型^[11]及骨质疏松性椎体骨折识别等方面^[12-13],其高灵敏度和低假阴性率的特性有效优化了临床工作流程^[14]。AI通过结构化知识提取与动态循证更新机制,为骨质疏松症标准化诊疗路径的优化提供解决方案^[15]。目前, ChatGPT 4.0 与 DeepSeek-R1 作为代表性模型展示了各自的技术特征。ChatGPT 依托 GPT-4 架构实现多语言通用性,但其知识时效性受到训练数据截止周期的限制,在动态医学知识整合方面存在一定局限^[16-17]。DeepSeek 作为中文领域优化的大语言模型,其私有化部署特性与实时知识更新机制,在医疗数据安全与临床场景适配方面具有显著优势^[18]。尽管可解释性 AI 评估框架已在医学其他领域建立了较为完善的方法论体系^[8, 19-20],但针对临床骨质疏松症诊疗的准确性和科学性评价方面仍属空白。

本研究拟对比 ChatGPT 和 DeepSeek 2 种模型在原发性骨质疏松症诊疗指南解读中的适用性,并利用临床骨密度检查结果验证 AI 对骨质疏松症诊疗的准确性与科学性。同时,探讨 AI 技术如何融入临床骨质疏松症诊疗及决策,旨在为构建“DNA”(Doctor-Nurse-AI)模式,通过整合医生、护士和 AI 技术的优势,优化医疗流程、提升诊疗效率和患者护理质量,推动骨质疏松症分级诊疗体系智能化转型提供理论依据。

1 对象与方法

1.1 研究对象

本研究以 ChatGPT 4.0 和 DeepSeek-R1 2 种人工智能模型针对《原发性骨质疏松症诊疗指南(2022)》(以下简称指南)生成的问答内容为研究对象。

1.2 研究方法

针对指南,向 2 种人工智能模型提问,选取 10 名经验丰富且熟悉指南的骨科医师,从 4 个维度对 2 种模型回答结果与指南进行系统性对比分析:准确性、全面性、可解释性、临床应用性。再通过临床实例验证其回答,并与临床骨密度检查结果进行对比。研究方法及流程见图 1。

1.3 问题选择

本研究围绕指南的核心临床实践需求,确保问题集兼具指南依从性、临床实用性与知识挑战性,从骨质疏松症的诊断标准、防治措施、骨折危险因素及风险评估、鉴别诊断和抗骨质疏松症药物 5 个方面进行提问。对 2 种人工智能模型赋予具有丰富临床经验并了解指南的骨科医师的角色。

问题 1: 帮我解读一下原发性骨质疏松症诊疗指南(2022)中关于骨质疏松症诊断标准。

问题 2: 帮我解读一下原发性骨质疏松症诊疗指南(2022)中关于骨质疏松症防治措施。

问题 3: 帮我解读一下原发性骨质疏松症诊疗指南(2022)中关于骨质疏松症骨折危险因素及风险评估。

问题 4: 帮我解读一下原发性骨质疏松症诊疗指南(2022)中关于骨质疏松症的鉴别诊断。

问题 5: 帮我解读一下原发性骨质疏松症诊疗指南(2022)中关于抗骨质疏松症药物。

1.4 统计学方法

采用 SPSS 27.0 进行数据分析,针对 10 位骨科医师对 ChatGPT 和 DeepSeek 模型在 5 个临床问题上的评分结果(每位医师对同一问题的 2 种模型评分形成 50 对配对数据),采用 Wilcoxon 符号秩和检验进行组间比较。由于某些评分者的评分分布极端,如部分评分为 1 或 5,常规的评分者间一致性检验不适用^[21],因此本研究将评分者间的评分一致性分成 3 个等级:完全一致(评分为 5),基本一致(评分为 2、3、4),完全不一致(评分

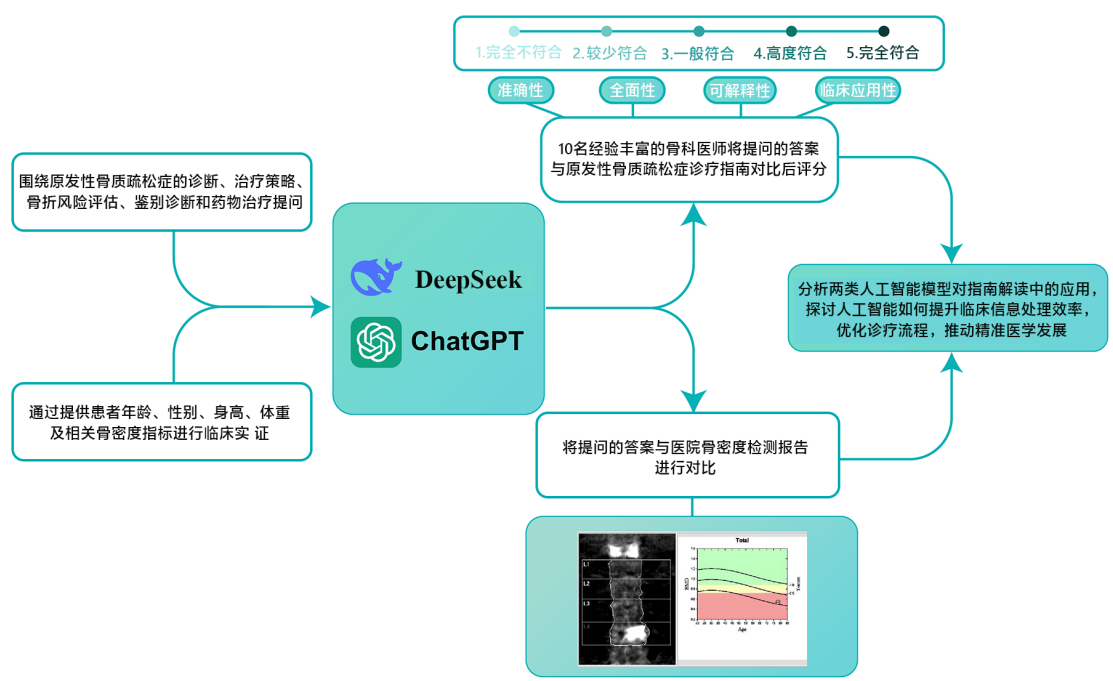


图 1 人工智能模型对于原发性骨质疏松症诊疗指南及临床骨密度结果解读的评估流程图

Figure 1 Flow chart of interpretation of artificial intelligence model for diagnosis and treatment guidelines of primary osteoporosis and the results of clinical bone mineral density

为 1)，通过计算百分率评估一致性情况。所有检验以双侧 $P < 0.05$ 为差异有统计学意义。

2 结 果

2.1 10 位骨科医师基本情况及评估结果的一致性

10 名骨科医师均来自三甲医院，具有博士学位和较丰富的临床诊疗经验，见表 1。10 位骨科医师对 2 种模型回答情况的评分见表 2。骨科医师之间的评分一致性均较高，见图 2。

2.2 2 种模型回答结果对比分析

2.2.1 准确性

ChatGPT 与 DeepSeek 在回答原发性骨质疏松症诊疗指南相关问题时，均展现了较高的准确性，但两者的准确性评分存在差异 ($Z = -3.62$, $P = 0.001$)，见图 3A，骨科医师对 DeepSeek 的回答的准确性认可度更高。具体而言，ChatGPT 虽然能覆盖骨密度阈值判定等基础诊断标准，并对抗骨质疏松症药物进行基本分类，但在一些关键细节的表述上仍存在模糊性。相比之下，DeepSeek 在关键细节的处理上更贴合指南精细化要求。此外，DeepSeek 还明确了中国人群 FRAX® 骨折风险干预阈值的地域化调整，即主要骨折概率达到 7% 或以

表 1 10 位骨科医师的基本情况 & 工作经验

Table 1 Basic information and working experience of 10 orthopedic surgeons

医 师	年 龄 / 岁	职 称	学 历	医 院 等级	工 作 年 限 / 年
1	40	副主任医师	医学博士	三甲	11
2	57	主任医师	医学博士	三甲	32
3	29	主治医师	医学博士	三甲	5
4	38	副主任医师	医学博士	三甲	8
5	53	主任医师	医学博士	三甲	28
6	32	副主任医师	医学博士	三甲	7
7	28	主治医师	医学博士	三甲	5
8	46	副主任医师	医学博士	三甲	18
9	45	副主任医师	医学博士	三甲	15
10	59	主任医师	医学博士	三甲	34

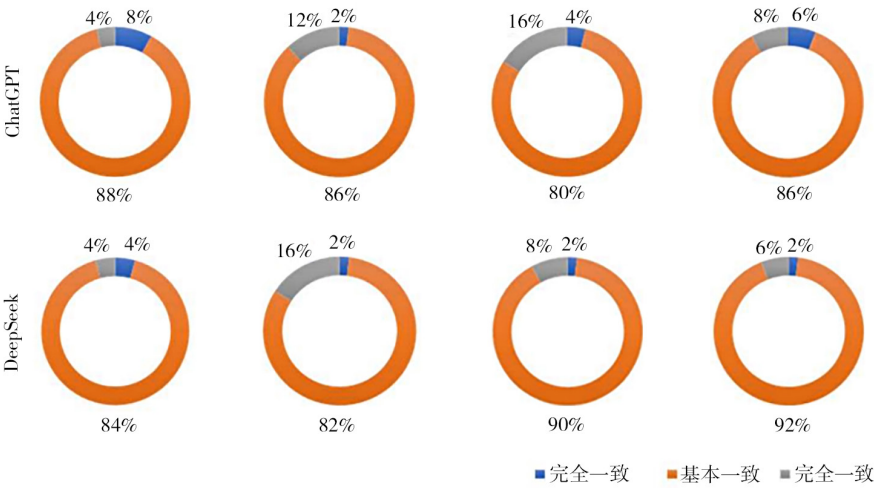
上时需采取干预措施。在抗骨质疏松症药物分类方面，DeepSeek 不仅详细阐述了罗莫珠单抗在临床应用中 对心血管事件风险的监测要求，还规范了停药后双膦酸盐的序贯治疗方案，与指南的更新要点保持高度一致。对回答结果进一步分析发现，DeepSeek 在复杂参数解析和动态决策支持方面具有一定优势，特别是在风险分层工具的本土化应用以及药物序贯治疗逻辑的精准匹配方面更为突出。准确度更高的 AI 模型能够帮助临床医师快速获取最新指南的核心标准，减少因信息遗漏

表 2 10 位骨科医师对 2 种人工智能模型在指南五大关键领域回答内容的评分

Table 2 Scores of 10 orthopedic surgeons on the answers of the two AI models in the five key areas of the guide

医 师	问题 1		问题 2		问题 3		问题 4		问题 5	
	ChatGPT	DeepSeek	ChatGPT	DeepSeek	ChatGPT	DeepSeek	ChatGPT	DeepSeek	ChatGPT	DeepSeek
1	2/2/3/3	4/2/3/4	3/2/3/3	4/2/3/5	2/2/3/5	3/1/2/3	3/3/4/1	3/2/4/4	3/3/2/2	4/3/4/4
2	3/3/4/3	3/3/3/3	2/2/3/3	4/2/3/4	3/4/3/5	4/3/2/4	2/3/4/3	5/2/4/4	3/3/4/3	4/2/4/4
3	3/4/3/5	2/5/4/4	3/1/1/4	4/2/2/4	4/2/2/5	3/3/3/4	3/2/3/3	4/4/3/4	2/5/2/2	4/4/3/5
4	1/4/4/2	1/4/3/3	4/5/3/2	3/5/4/3	1/3/3/3	4/3/2/3	4/3/2/1	3/3/2/4	3/2/2/3	3/3/3/4
5	3/2/4/2	4/3/4/3	3/2/2/3	3/5/2/2	3/2/3/3	4/2/3/4	1/2/2/3	4/4/5/3	2/4/5/2	4/4/5/3
6	3/4/3/2	4/4/4/3	1/4/4/2	4/4/4/3	2/5/5/4	5/4/5/3	3/3/4/3	5/3/4/4	3/2/3/2	2/3/4/3
7	3/2/2/2	4/2/2/3	3/2/3/2	3/2/4/4	5/3/3/2	1/3/3/1	2/3/3/1	4/3/4/4	4/4/5/3	5/5/3/3
8	2/2/3/2	4/3/3/4	2/3/3/3	4/2/2/3	3/3/5/2	3/3/4/3	4/5/2/4	4/5/3/3	3/2/4/3	4/3/4/2
9	3/3/1/4	3/3/1/3	4/2/2/4	4/3/3/3	4/4/2/4	4/4/2/3	3/3/4/4	3/5/4/4	5/4/5/2	5/3/4/3
10	4/2/5/2	5/2/5/3	3/4/5/2	3/4/4/3	4/5/5/2	4/4/3/3	4/3/3/2	4/5/3/3	3/5/3/4	3/5/3/5

注：1、2、3、4、5 分别对应完全不符合、较少符合、一般符合、高度符合、完全符合。每个问题的评分顺序依次为准确性、全面性、可解释性和临床应用性。



注：从左往右分别是准确性、全面性、可解释性和临床应用性的评价结果一致性情况

图 2 10 位骨科医师评价结果的一致性

Figure 2 Consistency of evaluation among 10 orthopedic surgeons

导致的诊疗偏差，尤其在涉及 Z 值、FRAX® 骨折风险阈值等复杂参数的应用时，能够提供更精准的临床决策支持。

2.2.2 全面性

2 种模型在指南内容覆盖全面性方面的评分差异没有统计学意义 ($Z = -1.65, P = 0.092$), 见图 3B。ChatGPT 虽然能够概述个体化治疗原则，但在关键内容的深度解析上仍存在不足。例如，其对 FLS 流程、OSTA 与 FRAX® 联合筛查的阶梯式策略缺乏详细说明，这可能影响指南在临床实践中的有效落地。而 DeepSeek 在全面性覆盖方面仍存在一定的优势。例如，在指南的系统性整合方面，DeepSeek 表现更为突出，其防治措施描述

不仅涵盖基础干预，如钙剂补充和运动处方，还完整纳入了药物假期管理，包括双膦酸盐 5 年疗程后的风险评估。此外，该模型对骨折联络服务 (fracture liaison service, FLS) 这一多学科协作模式进行了详细解析，并覆盖了基层医疗机构的分级诊疗路径，使指南应用更加系统化和具有可操作性。在风险评估维度，DeepSeek 不仅提及了 FRAX® 模型，还补充了亚洲人骨质疏松自我筛查工具 (Osteoporosis Self-Assessment Tool for Asians, OSTA) 的临床应用，以及骨转换标志物动态监测的价值。此外，该模型还能识别中国人群的特异性危险因素，如握力 <18 kg 和慢性腰痛史，从而形成更为全面的多维评估框架。

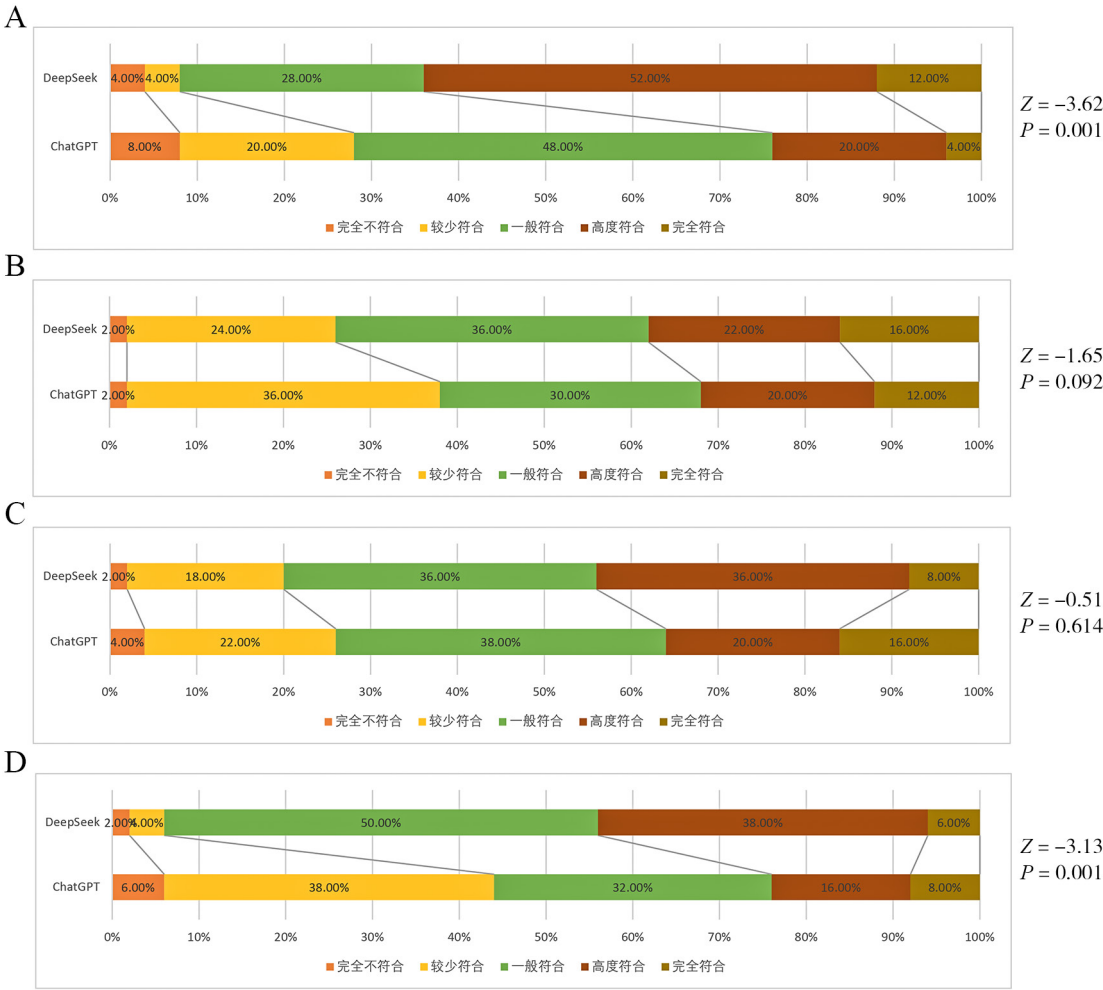
2.2.3 可解释性

ChatGPT 与 DeepSeek 在指南解读的可解释性方面的表现无统计学差异 ($Z = -0.51$, $P = 0.614$), 见图 3C。但在结构化信息的呈现方式上, 二者展现出不同的特征。ChatGPT 虽然能够系统列举鉴别诊断条目, 并能分点描述 7 种继发性骨质疏松病因, 但在信息提炼上存在不足。同时, 在部分术语解释上更倾向于病理机制理论, 而非临床实践。与之相比, DeepSeek 在解读指南时更倾向于通过临床场景化工具提升可读性。此外, DeepSeek 构建了流程图来直观展示诊断路径, 如基于骨密度 T 值和脆性骨折史的风险分级流程, 从而优化了信息提取效率。在术语解释方面, DeepSeek 的描述更贴合临床决策语境。例如, 其对序贯治疗的定义明确指向药物转换的最佳时机, 而迫在眉睫的

骨折风险也给出了量化标准, 使临床应用更具指导性。

2.2.4 临床应用性

临床应用性是临床医师最为关心的一个指标, AI 技术不仅需要具备高准确性、全面性和可解释性, 更重要的是能够顺畅融入临床工作流程, 切实提升医师的诊疗效率并确保医疗决策的安全性和可靠性。ChatGPT 与 DeepSeek 在临床应用性的评分存在差异 ($Z = -3.13$, $P = 0.001$), 见图 3D, 骨科医师对 DeepSeek 的临床实践价值认可度更高, 这种差异主要体现在临床场景适配性和动态管理支持方面。ChatGPT 虽然能够系统分类抗骨质疏松药物, 但在特殊人群剂量调整方面缺乏针对性指导。相比之下, DeepSeek 不仅能够将指南建议转化为可操作的决策路径, 例如在肾功能不全患者



注: A~D 分别为 2 种模型准确性、全面性、可解释性、临床应用性评分结果对比。

图 3 2 种人工智能模型准确性、全面性、可解释性和临床应用性评分的统计结果

Figure 3 Statistical results of the accuracy, comprehensiveness, interpretability and clinical applicability scores of the two AI models

中优先推荐唑来膦酸而不是地舒单抗，还能细化围手术期药物的管理方案，包括双膦酸盐停药时机及术后抗骨质疏松治疗的衔接策略。此外，该模型构建了涵盖一线用药选择、极高风险强化治疗以及长期监测调整的全周期管理框架。在动态

监测方面，DeepSeek 建议每年进行骨密度复查，并结合骨转换标志物检测，同时强调了地舒单抗停药后椎体骨折反跳的风险及相应的预防策略，与指南的精细化要求高度一致。

2 个模型具体比较结果详见表 3。

表 3 ChatGPT 和 DeepSeek 在原发性骨质疏松症诊疗指南解读准确性、全面性、可解释性和临床应用性的比较
Table 3 Comparison of the accuracy, comprehensiveness, interpretability, and clinical application of ChatGPT and DeepSeek in the interpretation of primary osteoporosis diagnosis and treatment guidelines

模 型	准确性	全面性	可解释性	临床应用性
ChatGPT	①覆盖骨密度判定等诊断标准 ②抗骨质疏松药物分类描述完整	①概述个体化治疗原则 ②提供基础骨质疏松风险因素	①列举骨质疏松鉴别诊断条目 ②分点描述骨质疏松继发性病因	①对骨质疏松药物系统分类 ②提供基础治疗建议
DeepSeek	①精准区分 T 值与 Z 值的适用场景 ②明确中国人群 FRAX® 阈值 ③规范骨质疏松药物序贯治疗方案	①整合骨质疏松药物假期管理 ②解析 FLS 协作模式及 OSTA 筛查工具 ③纳入中国特异性骨质疏松风险因素	①使用表格详细对比骨质疏松症的鉴别诊断 ②骨质疏松症诊断流程图 ③临床术语解释更易理解	①将骨质疏松症诊疗指南转换为可操作的决策路径 ②骨质疏松症全周期管理框架 ③强调每年进行骨密度复查和相应预防策略

2.3 临床实例验证

提问内容：女性患者，年龄 75 岁，身高 153 cm，体重 53 kg；腰椎 L1 骨密度为 0.600 g/cm²，腰椎 L2 骨密度为 0.668 g/cm²，腰椎 L3 骨密度为 0.634 g/cm²，腰椎 L4 骨密度为 0.604 g/cm²，请问这位患者的诊断是什么？有什么治疗和预防的建议？

在骨质疏松症临床诊断的规范性方面，ChatGPT 依据于世界卫生组织推荐的 T-score 标准（诊断阈值≤-2.5），计算出平均 T 值 -3.12，符合基本诊断要求。然而，该模型的分析存在两方面不足。首先，未纳入 Z 值评估体系，忽略了不同年龄群体在骨密度判定上的差异。其次，未提及继发性骨质疏松的鉴别诊断，与 2022 版《原发性骨质疏松症诊疗指南》的要求存在一定差距。指南明确指出，当 Z 值≤-2.0 时，应进一步排查内分泌和代谢性疾病，例如甲状腺功能亢进或维生素 D 缺乏，以鉴别继发性骨质疏松的可能性。与之对比，DeepSeek 展现出更完整的诊断逻辑，不仅系统计算了各椎体的 Z 值，例如 L3 和 L4 均≤-2.0，还针对性地建议开展血清钙、甲状旁腺激素和甲状腺功能等实验室检验，以进一步完善鉴别诊断。这种诊断流程与指南规范高度一致，提升了对潜在继发性因素的识别能力。此外，DeepSeek 严格遵循指南分层管理原则，强调严重骨质疏松症的诊断需结合患者的脆性骨折史进行病情分级，这一做法对于制定阶梯化治疗方案具有重要的临床指

导价值，能够确保高风险患者获得更精准的干预措施。

在治疗策略的完整性方面，虽然 ChatGPT 正确推荐了双膦酸盐、地舒单抗等一线抗骨质疏松药物和基础的营养补充方案，但其分析中存在 3 个关键遗漏。首先，未明确药物使用的周期，例如未指出双膦酸盐的标准疗程应为 3~5 年。其次，未界定特立帕肽的适用条件，未提及该药物应仅用于严重骨质疏松或骨折高风险患者。最后，缺乏对药物禁忌证的详细说明。反观 DeepSeek 的治疗建议更加符合循证医学的原则。该模型明确指出，双膦酸盐的使用需评估患者的肾功能和胃肠道耐受性，并推荐将地舒单抗作为肾功能不全患者的首选药物。对于特立帕肽，DeepSeek 严格限制其用于严重病例，并推荐其治疗疗程不超过 2 年。此外，在营养干预方面，DeepSeek 特别强调动态监测血清 1,25（OH）₂D₃ 水平，并建议目标值应≥30 ng/mL，符合骨质疏松症个体化治疗的要求。DeepSeek 还在诊疗流程中引入了 Z 值异常时优先排查继发性骨质疏松的临床路径，进一步确保了治疗的全面性和针对性，并且有助于为患者提供更为精准和个性化的治疗方案。

在疾病管理的科学性方面，虽然 ChatGPT 提出的基础预防措施具有一定合理性，但未纳入国际通用的 FRAX® 骨折风险评估工具，也未强调指南推荐的高危人群筛查策略，例如 65 岁以上

女性应常规进行骨密度检测等建议。相比之下, DeepSeek 的防控体系更为系统化, 采用 FRAX[®] 工具量化 10 年骨折风险, 严格执行年龄分层筛查标准, 并对跌倒预防措施进行细化, 包括居家环境改造和辅助器具的使用。此外, DeepSeek 特别提醒在长期管理过程中, 如地舒单抗需要持续给药, 避免因患者自主停药而导致的骨量丢失反弹, 这一提醒对提高患者依从性具有重要的临床意义。

3 讨论

本研究对比了 2 种代表性的人工智能模型在原发性骨质疏松症诊疗指南解读和临床实例验证中的表现, 观察两者对于临床实际工作的应用价值。结果显示, 两者都能通过高效的信息整合提升骨质疏松症的诊疗效率, 人工智能模型解读与临床骨密度检查的实际诊断报告的结果亦基本吻合。在指南解读方面, 10 位骨科医师在 4 个维度的评分一致性高, DeepSeek 在准确性和临床应用性上比 ChatGPT 更优。而全面性和可解释性方面无统计学差异, 这可能与样本量限制或评分尺度趋中性有关, 还可能受评分者对可视化工具偏好差异或部分场景下文本解释等效性的影响。两者对解读指南结果的差异主要可能是因为其机制不同, DeepSeek 可结合患者的数据进行精准分析, 还能够直接调用医学数据库等; 而 ChatGPT 主要依赖已有的训练数据, 更擅长语言理解和对话等。本研究通过临床实例验证的结果显示, 相较于 ChatGPT 的基础性诊疗框架, DeepSeek 展现出更为完整的指南依从性和临床适用性。这种系统性、全周期的疾病管理框架, 有助于提高骨科医师临床决策的精准性和效率, 从而优化骨质疏松症患者的长期治疗效果。因此, AI 能够辅助医师诊治骨质疏松症, 提升临床效率和决策质量, DeepSeek 在原发性骨质疏松症诊疗指南及临床骨密度检查中解读的表现比 ChatGPT 更优异。

人工智能目前已广泛应用于骨质疏松症诊疗研究领域。Ferizi 等^[22]的研究结果显示, AI 通过基于 MRI 影像特征的机器学习预测模型可以对 MRI 数据进行分析后预测脆性骨折。本研究通过 AI 对骨密度、FRAX[®] 骨折风险预测和 OSTA 初筛进行了更全面的评估。此外, Kruse 等^[20]的研究表明机器通过学习可以预测髌部骨折的风险, 其研究价值主要体现为特定解剖部位的风险预警。本

研究亦提示 DeepSeek 在预测骨质疏松性骨折方面有更好的表现, 为骨质疏松症的预测及相关预防措施的开展提供了新途径。与 Scanlan 等^[23]的研究相比, 本研究对临床应用性方面进行了验证, 而 Scanlan 等介绍的方法可能仅适合作为骨质疏松症早期诊断的人群筛查。

然而, AI 医疗化进程仍面临多重挑战。首先, 尽管 AI 能够根据现有的骨质疏松症指南提供决策支持, 但由于指南不断发展, 要求模型也要持续迭代^[24]。若 AI 更新滞后可能会影响其临床适用性^[25]。其次, AI 模型的有效性依赖于大量个人健康数据的收集与处理, 数据依赖性可能会引发患者隐私风险^[26]。在医疗全流程中, 必须确保数据的安全性和患者的隐私得到充分保护, 防止数据泄露或滥用^[27-29]。随着 AI 在骨质疏松症临床中的深入应用, AI 系统的临床决策失误的责任界定也需要符合相关伦理和法律^[30]。因此, 未来当 AI 模型融入骨质疏松症诊疗时, AI 生成的建议虽能辅助诊疗, 但仍然无法完全替代医师的专业判断, 最终决策还是需要依赖医师的临床经验与综合分析。

未来若能将 AI 模型嵌入骨质疏松症医疗信息系统来构建预警网络, 例如能够在电子病历实时监测骨代谢指标异常波动, 通过可穿戴设备的数据来预判跌倒风险, 并在区域医疗网络中建立分级诊疗决策树^[27], 将显著提高骨质疏松症的早期检出率并为患者提供动态健康管理的支持。而 AI 模型经过训练后, 在基层医师接诊骨质疏松患者时, AI 可同时完成风险评估、禁忌筛查和治疗推荐; 三级医院专家则能通过 AI 预筛的复杂病例开展深度会诊^[31], 能够大大提高医疗效率^[32]。目前, AI 在医学领域呈现出专业化与泛化能力的互补态势。DeepSeek 模型在精准医疗领域展现出独特优势^[33], 其阶梯式筛查算法能融合 FRAX[®] 骨折风险预测与 OSTA 初筛特性, 在保证骨质疏松症诊断准确性的同时提升筛查效率。相较而言, ChatGPT 凭借其强大的自然语言处理能力, 能够成为骨质疏松症患者对话式的健康助手^[34], 能动态解析骨质疏松症相关的专业术语并为不同文化程度患者生成科普内容^[35-36]。

本研究亦存在一定的局限性, 首先, 研究只选取了 10 位骨科医师对 AI 模型回答的问题进行评分, 评分的人数可能稍显不足。另外, 研究还应评估将 AI 模型整合到骨质疏松症诊疗过程中的

准确性以及对工作效率提升的作用。

综上所述,本研究结果表明,DeepSeek 比 ChatGPT 在原发性骨质疏松症诊疗指南及临床骨密度检查结果的解读中表现更为优异,有望在未来成为一种辅助骨质疏松症诊疗的工具。当 AI 模型融入骨质疏松症诊疗中时,应注重 AI 算力和骨科医师的临床经验,始终坚持以人为主导的决策机制,将骨科医师的临床经验作为诊疗决策的核心依据,通过构建“DNA”模式,为患者提供更加个性化的诊疗服务及提升临床的诊疗效率。

利益冲突声明: 本研究未受到企业、公司等第三方资助,不存在潜在利益冲突。

参 考 文 献

- [1] LIU Y, HUANG X, TANG K, et al. Prevalence of osteoporosis and associated factors among Chinese adults: a systematic review and modelling study[J]. J Glob Health, 2025, 15: 04009. DOI: 10.7189/jogh.15.04009.
- [2] WANG H S, KARNIK S J, MARGETTS T J, et al. Mind gaps and bone snaps: exploring the connection between Alzheimer's disease and osteoporosis[J]. Curr Osteoporos Rep, 2024, 22(5): 483-494. DOI: 10.1007/s11914-023-00851-1.
- [3] 中华医学会骨质疏松和骨矿盐疾病分会. 原发性骨质疏松症诊疗指南(2022)[J]. 中国全科医学, 2023, 26(14): 1671-91. DOI: 10.12114/j.issn.1007-9572.2023.0121. Chinese Society of Osteoporosis and Bone Mineral Research. Guidelines for the Diagnosis and Treatment of Primary Osteoporosis (2022)[J]. Chin Gen Prac, 2023, 26(14): 1671-1691. DOI: 10.12114/j.issn.1007-9572.2023.0121.
- [4] SONG M, ELSON J, BASTOLA D. Digital age transformation in patient-physician communication: 25-year narrative review (1999-2023)[J]. J Med Internet Res, 2025, 27: e60512. DOI: 10.2196/60512.
- [5] KARAM L, PACCOU J. Management of adverse skeletal effects following bariatric surgery procedures in people living with obesity[J]. Curr Osteoporos Rep, 2025, 23(1): 11. DOI: 10.1007/s11914-025-00902-9.
- [6] CONROY G, MALLAPATY S. How China created AI model DeepSeek and shocked the world[J]. Nature, 2025, 638(8050): 300-301. DOI: 10.1038/d41586-025-00259-0.
- [7] GIBNEY E. China's cheap, open AI model DeepSeek thrills scientists[J]. Nature, 2025, 638(8049): 13-14. DOI: 10.1038/d41586-025-00229-6.
- [8] PUTRA R H, DOI C, YODA N, et al. Current applications and development of artificial intelligence for digital dental radiography[J]. Dentomaxillofac Radiol, 2022, 51(1): 20210197. DOI: 10.1259/dmfr.20210197.
- [9] SYMEOU L, LOUCA L, KAVADELLA A, et al. Development of evidence-based guidelines for the integration of generative AI in university education through a multidisciplinary, consensus-based approach[J]. Eur J Dent Educ, 2025. DOI: 10.1111/eje.13069.
- [10] LEI C, SONG J H, LI S, et al. Advances in materials-based therapeutic strategies against osteoporosis[J]. Biomaterials, 2023, 296: 122066. DOI: 10.1016/j.biomaterials.2023.122066.
- [11] KHANNA V V, CHADAGA K, SAMPATHILA N, et al. A decision support system for osteoporosis risk prediction using machine learning and explainable artificial intelligence[J]. Heliyon, 2023, 9(12): e22456. DOI: 10.1016/j.heliyon.2023.e22456.
- [12] SHEN L, GAO C, HU S, et al. Using artificial intelligence to diagnose osteoporotic vertebral fractures on plain radiographs[J]. J Bone Miner Res, 2023, 38(9): 1278-1287. DOI: 10.1002/jbmr.4879.
- [13] TSAI D J, LIN C, LIN C S, et al. Artificial intelligence-enabled chest X-ray classifies osteoporosis and identifies mortality risk[J]. J Med Syst, 2024, 48(1): 12. DOI: 10.1007/s10916-023-02030-2.
- [14] FERIZI U, HONIG S, CHANG G. Artificial intelligence, osteoporosis and fragility fractures[J]. Curr Opin Rheumatol, 2019, 31(4): 368-375. DOI: 10.1097/BOR.0000000000000607.
- [15] GATINEAU G, SHEVROJA E, VENDRAMI C, et al. Development and reporting of artificial intelligence in osteoporosis management[J]. J Bone Miner Res, 2024, 39(11): 1553-1573. DOI: 10.1093/jbmr/zjae131.
- [16] ERDEN Y, TEMEL M H, BAĞCIER F. Artificial intelligence insights into osteoporosis: assessing ChatGPT's information quality and readability[J]. Arch Osteoporos, 2024, 19(1): 17. DOI: 10.1007/s11657-024-01376-5.
- [17] CHOUDHURY A, SHAMSZARE H. The impact of performance expectancy, workload, risk, and satisfaction on trust in ChatGPT: cross-sectional survey analysis[J]. JMIR Hum Factors, 2024, 11: e55399. DOI: 10.2196/55399.
- [18] KAYAALP M E, PRILL R, SEZGIN E A, et al. DeepSeek versus ChatGPT: multimodal artificial intelligence revolutionizing scientific discovery. From language editing to autonomous content generation-redefining innovation in research and practice[J]. Knee Surg Sports Traumatol Arthrosc, 2025. DOI: 10.1002/ksa.12628.
- [19] JIN W, LI X, FATEHI M, et al. Guidelines and evaluation of clinical explainable AI in medical image analysis[J]. Med Image Anal, 2023, 84: 102684. DOI: 10.1016/j.media.2022.102684.
- [20] KRUSE C, EIKEN P, VESTERGAARD P. Machine learning principles can improve hip fracture prediction[J]. Calcif Tissue Int, 2017, 100(4): 348-360. DOI: 10.1007/s00223-017-0238-7.
- [21] 黎超, 陈优美, 段亚妮, 等. 生成式人工智能在生成影像学报告方面的表现评估[J]. 新医学, 2024, 55(11): 853-860. DOI: 10.3969/j.issn.0253-9802.2024.11.001.

- LI C, CHEN Y M, DUAN Y N, et al. Evaluation of the performance of generative artificial intelligence in generating radiology reports [J]. *J New Med*, 2024, 55 (11): 853-860. DOI: 10.3969/j.issn.0253-9802.2024.11.001.
- [22] FERIZI U, BESSER H, HYSI P, et al. Artificial intelligence applied to osteoporosis: a performance comparison of machine learning algorithms in predicting fragility fractures from MRI data [J]. *J Magn Reson Imaging*, 2019, 49 (4): 1029-1038. DOI: 10.1002/jmri.26280.
- [23] SCANLAN J, LI F F, UMNNOVA O, et al. Detection of osteoporosis from percussion responses using an electronic stethoscope and machine learning [J]. *Bioengineering*, 2018, 5 (4): 107. DOI: 10.3390/bioengineering5040107.
- [24] LAI H, GE L, SUN M, et al. Assessing the risk of bias in randomized clinical trials with large language models [J]. *JAMA Netw Open*, 2024, 7 (5): e2412687. DOI: 10.1001/jamanetworkopen.2024.12687.
- [25] KANJEE Z, CROWE B, RODMAN A. Accuracy of a generative artificial intelligence model in a complex diagnostic challenge [J]. *JAMA*, 2023, 330 (1): 78-80. DOI: 10.1001/jama.2023.8288.
- [26] ZHU L, LAI Y, MOU W, et al. ChatGPT's ability to generate realistic experimental images poses a new challenge to academic integrity [J]. *J Hematol Oncol*, 2024, 17 (1): 27. DOI: 10.1186/s13045-024-01543-8.
- [27] PENG Y, MALIN B A, ROUSSEAU J F, et al. From GPT to DeepSeek: significant gaps remain in realizing AI in healthcare [J]. *J Biomed Inform*, 2025, 163: 104791. DOI: 10.1016/j.jbi.2025.104791.
- [28] JEYARAMAN M, BALAJI S, JEYARAMAN N, et al. Unraveling the ethical enigma: artificial intelligence in healthcare [J]. *Cureus*, 2023, 15 (8): e43262. DOI: 10.7759/cureus.43262.
- [29] PETERS V, BAUMGARTNER M, FROESE S, et al. Risk and potential of ChatGPT in scientific publishing [J]. *J Inher Metab Dis*, 2023, 46 (6): 1005-1006. DOI: 10.1002/jimd.12666.
- [30] MELLO M M, GUHA N. ChatGPT and physicians' malpractice risk [J]. *JAMA Health Forum*, 2023, 4 (5): e231938. DOI: 10.1001/jamahealthforum.2023.1938.
- [31] XU T, WENG H, LIU F, et al. Current status of ChatGPT use in medical education: potentials, challenges, and strategies [J]. *J Med Internet Res*, 2024, 26: e57896. DOI: 10.2196/57896.
- [32] 祁冬, 姚传顺, 胡淑敏, 等. 人工智能在冠状动脉 CT 血管成像图像后处理和冠状动脉狭窄诊断中的应用 [J]. *江苏大学学报 (医学版)*, 2023, 33 (4): 323-327, 332. DOI: 10.13312/j.issn.1671-7783.y220147.
- Qi D, YAO C S, HU S M, et al. The value of artificial intelligence in post-processing coronary CTA images and diagnosing coronary artery stenosis [J]. *J Jiangsu Univ (Med Ed)*, 2023, 33 (4): 323-327, 332. DOI: 10.13312/j.issn.1671-7783.y220147.
- [33] NORMILE D. Chinese firm's large language model makes a splash [J]. *Science*, 2025, 387 (6731): 238. DOI: 10.1126/science.adv9836.
- [34] LIM Z W, PUSHPANATHAN K, YE W S M E, et al. Benchmarking large language models' performances for myopia care: a comparative analysis of ChatGPT-3.5, ChatGPT-4.0, and Google Bard [J]. *EBioMedicine*, 2023, 95: 104770. DOI: 10.1016/j.ebiom.2023.104770.
- [35] PREIKSAITIS C, ROSE C. Opportunities, challenges, and future directions of generative artificial intelligence in medical education: scoping review [J]. *JMIR Med Educ*, 2023, 9: e48785. DOI: 10.2196/48785.
- [36] GAN W, OUYANG J, LI H, et al. Integrating ChatGPT in orthopedic education for medical undergraduates: randomized controlled trial [J]. *J Med Internet Res*, 2024, 26: e57037. DOI: 10.2196/57037.

(责任编辑:郑巧兰)