ARAŞTIRMA / RESEARCH

Nefroloji Hemşireliği Dergisi,
Journal of Nephrology Nursing. 2025; 20(1): 51-60
DOI:

Reliability of AI Chatbots in Categorising Foods by Oxalate Content

Yapay Zeka Araçlarının Gıdaları Oksalat İçeriğine Göre Sınıflandırmadaki Güvenilirliği

Hüsna KAYA KAÇAR [1]

Abstract

Aim: AI chatbots have shown promise in food classification tasks, but their accuracy in categorising foods based on specific nutritional content, such as oxalates, has not been thoroughly evaluated in the Turkish language. This study assesses the performance of three AI chatbots—ChatGPT 4.0, Gemini, and Microsoft Copilot—in classifying foods according to their oxalate content.

Materials and Methods: A dataset of 63 diverse food items, including commonly consumed Turkish foods, was used to evaluate the chatbots' accuracy across five oxalate categories: little or none, low, moderate, high, and very high. The performance of each model was analysed, and commonly correct and incorrect classifications were identified.

Results: ChatGPT 4.0 demonstrated the highest overall accuracy (69.8%), significantly outperforming Gemini (36.5%) and Microsoft Copilot (26.9%). Foods such as spinach and cocoa were consistently classified correctly, while foods like carrot and walnut were commonly misclassified. Statistical analysis using Cochran's Q test revealed significant differences in accuracy among the chatbots (p-value <0.05).

Conclusion: This study highlights the potential of AI chatbots in dietary management, particularly in supporting clinicians who recommend low oxalate diets for patients with conditions such as hyperoxaluria or kidney disease stones. However, it emphasises the need for further refinement to improve accuracy, especially in classifying foods with regional variations or complex compositions commonly encountered in clinical settings.

Keywords: Artificial Intelligence; Hyperoxaluria; Kidney Stones; Renal Diet; Oxalate

Özet

Amaç: Yapay zeka sohbet botları, gıda sınıflandırma görevlerinde umut vaat etmiştir; ancak oksalat gibi belirli besin içeriklerine göre gıdaları sınıflandırmadaki doğrulukları, Türkçe dilinde yeterince değerlendirilmemiştir. Bu çalışma, üç yapay zeka sohbet botunun—ChatGPT 4.0, Gemini ve Microsoft Copilot—oksalat içeriğine göre gıdaları sınıflandırma performansını değerlendirmektedir.

Gereç ve Yöntem: Yaygın tüketilen Türk gıdalarını içeren 63 farklı gıda maddesinden oluşan bir veri seti, beş oksalat kategorisinde (çok az veya yok, düşük, orta, yüksek ve çok yüksek) sohbet botlarının doğruluğunu değerlendirmek için kullanılmıştır. Her modelin performansı analiz edilmiş ve yaygın doğru ve yanlış sınıflandırmalar belirlenmiştir.

Bulgular: ChatGPT 4.0, genel doğruluk açısından en yüksek performansı (%69,8) sergileyerek Gemini (%36,5) ve Microsoft Copilot'un (%26,9) önüne geçmiştir. Ispanak ve kakao gibi gıdalar tutarlı bir şekilde doğru sınıflandırılırken, havuç ve ceviz gibi gıdalar genellikle yanlış sınıflandırılmıştır. Cochran's Q testi kullanılarak yapılan istatistiksel analiz, sohbet botları arasındaki doğruluk farklılıklarının anlamlı olduğunu göstermiştir (p <0,05).

Sonuç: Bu çalışma, özellikle hiperoksalüri veya böbrek taşı gibi durumları olan hastalar için düşük oksalat diyeti önerilen durumlarda, diyet yönetiminde yapay zeka araçlarının potansiyelini vurgulamaktadır. Ancak, bölgesel farklılıklar veya karmaşık bileşimlere sahip gıdaların sınıflandırılmasında doğruluğun artırılması gerektiğini vurgulamaktadır.

Anahtar Kelimeler: Yapay Zeka; Hiperoksalüri; Böbrek Taşı; Renal Diyet; Oksalat

Geliş Tarihi / Submitted: 12 Aralık / Dec 2024 **Kabul Tarihi / Accepted:** 21 Aralık / Dec 2024

[1] Doktor Öğretim Üyesi, Amasya Üniversitesi, Sağlık Bilimleri Fakültesi, Amasya, Türkiye

İletişim yazarı / Correspondence author: Hüsna KAYA KAÇAR / **E-posta:** husna.kacar@amasya.edu.tr, **Adres:** Amasya Üniversitesi, Sağlık Bilimleri Fakültesi, İpekköy Yerleşkesi, İpekköy/Amasya, Türkiye.

INTRODUCTION

Nephrolithiasis, commonly referred to as kidney stones, ranks among the most prevalent urological disorders, with a notable global rise in both incidence and prevalence—from 3.2% to 11% in recent decades (1-4). Its clinical presentation varies widely, ranging from asymptomatic cases to severe, recurrent episodes that may culminate in compromised kidney function (3, 5, 6). The recurrence of kidney stones is strongly associated with various risk factors and lithogenic conditions, including metabolic alterations influenced by diet and medication (7-9). As a chronic disease, kidney stones significantly diminish patients' quality of life, primarily due to frequent hospitalizations and episodes of renal colic (10-12). These acute, recurrent pain episodes often necessitate surgical intervention and are accompanied by challenges in adhering to preventive measures and attending regular medical follow-ups (13, 14). Beyond the physical impact, kidney stones also contribute to psychological distress, including anxiety over recurrence or disease progression, as well as social challenges, such as isolation and difficulties with intimacy (15, 16).

Building on the significant burden kidney stones impose on patients, calcium oxalate stones, including both monohydrate and dihydrate forms, have been identified as the most prevalent type based on stone composition analyses (17-19). This underscores the critical role of dietary habits in the formation and prevention of oxalate stones, particularly for those prone to calcium oxalate stones (9, 10, 20, 21). Effective dietary management is essential not only to mitigate the risk of recurrence but also to reduce the overall impact of this chronic condition on patients' lives (22). Given the prominence of calcium oxalate stones, dietary oxalate plays a significant role in stone formation (8, 23). Oxalate, a compound naturally present in various foods, particularly plant-based sources, mostly influences urinary oxalate concentrations (24, 25). One key strategy for preventing calcium oxalate kidney stones involves reducing the consumption of oxalate-rich foods while ensuring adequate dietary calcium to bind oxalate in the gastrointestinal tract and minimize its absorption (26). However, clear evidence-based guidelines on optimal daily oxalate intake for kidney stone prevention remain lacking (8, 23, 27). For individuals at risk, including those with hyperoxaluria or kidney stones, healthcare professionals commonly recommend limiting dietary oxalate intake to below 40–50 mg per day (28). This recommendation is critical, as exceeding 50 mg of dietary oxalate daily can lead to a disproportionate increase in its absorption, further elevating urinary oxalate levels (23, 25, 29). Foods such as spinach, avocado, and orange are particularly high in oxalate and can contribute to stone formation in susceptible individuals (30, 31). However, implementing such dietary restrictions requires precise knowledge of oxalate content in foods (32, 33). This highlights the potential for artificial intelligence (AI) to play a transformative role in dietary planning, offering a promising avenue to support patients in making informed choices and adhering to preventative strategies.

In recent years, the exponential growth of technology, particularly AI, has opened new avenues for improving healthcare delivery (34, 35). By integrating chatbots into patient care, healthcare providers can enhance risk assessment, support decision-making, and deliver personalized patient education (36). Research has shown that chatbot interventions can promote healthier lifestyles, including increased physical activity and improved dietary habits (37). For instance, a recent study highlighted the capability of chatbots to assist patients with chronic kidney disease in managing dietary restrictions. Specifically, the chatbots provided accurate information-over 70% accuracy-on the potassium and phosphorus content of foods, showcasing their potential in addressing complex dietary needs (38). This innovation presents an exciting opportunity to apply AI-driven solutions in the dietary management of oxalate for individuals prone to calcium oxalate stones, offering a practical approach to prevention and education.

Advanced generative AI chatbots, including ChatGPT, Microsoft Copilot, and Gemini, exemplify technologies capable of addressing the complexities of dietary planning (39-41). These models leverage extensive datasets to generate contextually relevant information, predict sequences, and provide actionable insights. Despite their potential, these models must be evaluated for accuracy, reliability, and effectiveness before their integration into clinical practice (42). Ensuring their validity is particularly crucial when applied to specialized areas like dietary management for oxalate stones, where precise information can directly impact patient outcomes (43).

This study aims to evaluate the efficacy of three prominent AI models-ChatGPT, Microsoft Copilot, and Gemini-in accurately identifying the oxalate content of various foods in the Turkish language. This focus is especially relevant for individuals seeking to prevent calcium oxalate urolithiasis, where dietary oxalate restriction plays a critical role. As the first study to compare the reliability of these chatbots in categorizing foods by their oxalate content in Turkish, it holds significant potential for clinical integration. The research centres (american english spelling- for consistency) on dietary oxalate management in conditions including hyperoxaluria, calcium oxalate kidney stones, and oxalate nephropathy, using these scenarios as a prototype for evaluating AI chatbot performance. A dataset of diverse food items was employed to assess the accuracy of these chatbots, with a focus on identifying their relative strengths and weaknesses.

MATERIAL AND METHOD

Reference Dataset and Classification Criteria

The oxalate content of 63 different foods was assessed using the Mayo Clinic Oxalate Diet Handbook as the primary reference source. These foods were selected from a total of 539 items based on their simple compositions, defined by a single-ingredient nature or minimal processing, to ensure precise categorisation and reliable analysis. This handbook provides a detailed guide for categorising foods based on their oxalate content per serving, as described elsewhere (32). The foods were classified into five categories: little or none (≤1 mg), low (2–4 mg), moderate (5–8 mg), high (9–11 mg), and very high (≥12 mg) oxalate content. For the purposes of this study, the food names were translated into Turkish to align with the language used in the AI chatbot prompts, ensuring consistency in the dataset.

AI Chatbot Selection and Query Procedure

The study evaluated three widely used AI chatbots: ChatGPT 4.0, Gemini, and Microsoft Copilot (44-46). These chatbots were tasked with classifying each food item into one of the five predefined oxalate categories. To ensure that each chatbot received the same information, a structured querying process was employed. First, each chatbot was provided with a clear definition of the oxalate content categories, as defined above. After receiving the definitions, the following prompt was given to each chatbot for every food item: "Please classify the following food by their oxalate content as little or none, low, moderate, high, or very high: Considering one serving is equal to _." This prompt was consistent across all chatbots to ensure a fair comparison of their performances. The chatbots were asked to categorise each food based on the serving size provided, and their responses were recorded as the final classification for each food.

Accuracy Evaluation

The primary aim of the study was to assess the accuracy of the three chatbots in classifying foods according to their oxalate content. After receiving the responses from the chatbots, these classifications were compared to the reference values from the Mayo Clinic Oxalate Diet Handbook. The accuracy rate of each chatbot was calculated by determining the proportion of correct classifications out of the total number of food items. This allowed for a quantitative comparison of the chatbots' performance in accurately classifying foods based on their oxalate content.

To further evaluate the performance of the chatbots, we identified foods that were consistently classified correctly or incorrectly by all three models. This analysis involved reviewing the classification outputs for each food item across the five oxalate content categories. Foods that achieved unanimous correct classifications across the chatbots were recorded as commonly correct, while those misclassified by all models were noted as commonly incorrect.

Statistical Analysis

To analyse the differences in accuracy across the three AI chatbots, several statistical tests were conducted. Cochran's Q test was used to test for overall differences in the accuracy rates among the three chatbots. This test is particularly suited for evaluating dichotomous outcomes, such as correct vs. incorrect classifications across multiple groups.

A stratified analysis was also carried out, where the accuracy rates of the chatbots were analysed separately for each of the five dietary oxalate content categories (little or none, low, moderate, high, and very high). This stratification allowed for a more detailed examination of how each chatbot performed within specific categories of oxalate content, ensuring that the results were not biased by the distribution of food items across categories.

The statistical significance of all tests was considered at a significance level of $p<0.05$, indicating that any observed differences were likely not due to chance. All statistical analyses were conducted using SPSS Statistics, Version 30.0.0, a widely used software for handling complex data analysis.

RESULTS

The performance of ChatGPT 4.0, Gemini, and Microsoft Copilot in classifying foods according to their oxalate content varied across categories. Overall, ChatGPT 4.0 achieved the highest accuracy, correctly classifying 69.8% (44/63) of the foods, compared to Gemini at 36.5% (23/63) and Microsoft Copilot at 26.9% (17/63).

In the "little or none" oxalate category, ChatGPT 4.0 exhibited perfect classification accuracy (100%, 19/19), significantly outperforming Gemini (36.8%, 7/19) and Microsoft Copilot (52.6%, 10/19). For foods with "low" oxalate content, Gemini demonstrated the highest accuracy (73.3%, 11/15), while ChatGPT 4.0 and Microsoft Copilot achieved accuracies of 40.0% (6/15) and 13.3% (2/15), respectively.

In the "moderate" oxalate category, ChatGPT 4.0 showed higher performance, correctly classifying 85.7% (6/7) of the foods, followed by Gemini (42.8%, 3/7) and Microsoft Copilot (0%). Similarly, ChatGPT 4.0 led in the "high" category, with 42.8% (3/7) correct classifications, whereas Gemini did not classify any foods correctly, and Microsoft Copilot achieved 14.3% (1/7). Finally, in the "very high" category, ChatGPT 4.0 again outperformed the other models, with an accuracy of 66.7% (10/15), while Gemini and Microsoft Copilot demonstrated accuracies of 13.3% (2/15) and 26.7% (4/15), respectively (Table 1, Figure 1).

Table 1. Classification Accuracy of ChatGPT 4.0, Gemini, and Microsoft Copilot Across Different Oxalate Content Categories

Oxalate Content	ChatGPT 4.0	Gemini	Microsoft Copilot
Overall (n=63)	44 (69.8%)	23 (36.5%)	17 (26.9%)
Little or None (n=19)	19 (100%)	7 (36.8%)	10 (52.6%)
Low (n=15)	6 (40.0%)	11 (73.3%)	2 (13.3%)
Moderate (n=7)	6 (85.7%)	3 (42.8%)	0 (0%)
High (n=7)	3 (42.8%)	0 (0%)	1 (14.3%)
Very High (n=15)	10 (66.7%)	2 (13.3%)	4 (26.7%)

* The p (<0.05) from Cochran's Q test indicates significant differences in the accuracy rates among the chatbots across all oxalate content categories.

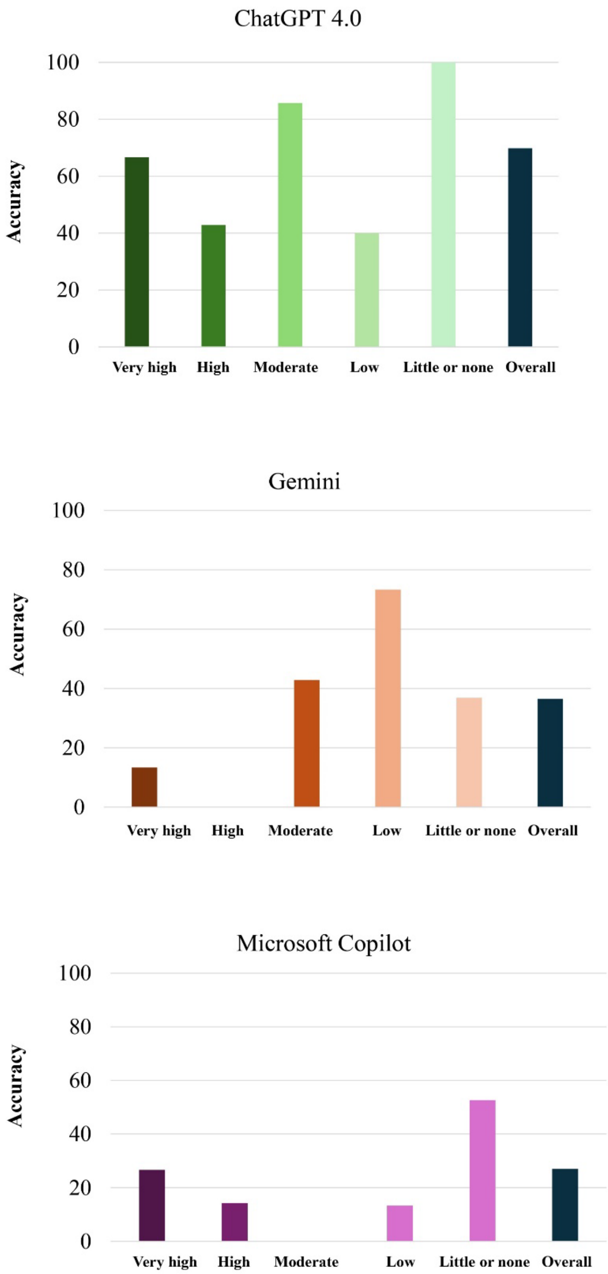

Figure 1. Bar charts illustrating the accuracy rates of ChatGPT 4.0, Gemini, and Microsoft Copilot in classifying foods across different oxalate content categories (little or none, low, moderate, high, very high).

These results underscore ChatGPT 4.0's consistent effectiveness across most oxalate categories, suggesting it is a more reliable tool for food classification tasks related to oxalate content in Turkish language. Statistical analysis of these outcomes revealed significant differences in accuracy among the chatbots ($p < 0.05$), as confirmed by Cochran's Q test, which highlighted significant variability in performance across the models.

The analysis revealed foods that were consistently classified correctly or incorrectly by all three chatbots, highlighting patterns in their performance. Commonly correct classifications included spinach and cocoa in the "very high" oxalate category and watermelon, cucumber, egg, and full-fat milk in the "very low" category. Conversely, several foods were misclassified by all models, particularly in the "very high" category, including carrot, bulgur, whole wheat flour, walnut, and peanut butter. Additionally, certain items in the "high" category, including tangerine, cooked celery, and oat bran, were also frequently misclassified. Foods in the "low" category, including blackberry, raisin, dried apple, and dried apricot, similarly categorised incorrectly by all three chatbots.

DISCUSSION

This study evaluated the accuracy of three advanced chatbot models—ChatGPT 4.0, Gemini, and Microsoft Copilot—in classifying foods according to their oxalate content across five categories: little or none, low, moderate, high, and very high. The findings revealed significant differences in performance, with ChatGPT 4.0 demonstrating consistently higher accuracy rates across most categories. Notably, ChatGPT 4.0 achieved perfect accuracy in the "little or none" category (100%), while Gemini performed best in the "low" category (73.3%). In contrast, Microsoft Copilot struggled in most categories, with particularly low performance in the "moderate" (0%) and "high" (14.3%) categories. Importantly, the foods evaluated in this study were Turkish, reflecting a unique cultural and dietary context that may have influenced the classification challenges faced by the chatbots. These findings highlight the variability in chatbot performance and underscore ChatGPT 4.0's robustness in tasks requiring precise food classification.

The findings of this study reveal both alignments and contrasts with prior research evaluating the accuracy of AI chatbots in food classification tasks, particularly regarding oxalate content. In a study evaluating the reliability of chatbots in categorising foods based on their oxalate content, ChatGPT-3.5, ChatGPT-4, Bard AI, and Bing Chat were assessed for their ability to classify 539 food items into low (<5 mg), moderate (5–8 mg), and high (>8 mg) oxalate content categories (32). Bard AI demonstrated the highest overall accuracy (84%), significantly outperforming Bing (60%), ChatGPT-4 (52%), and ChatGPT-3.5 (49%) ($p < 0.001$) (32). This study highlighted Bard AI's consistency across all oxalate categories, even as accuracy declined with increasing oxalate content. In contrast, our findings identify ChatGPT 4.0 as the most reliable tool, achieving 69.8% accuracy overall in classifying foods with Turkish names into five oxalate content categories. Unlike the prior study, where ChatGPT 4.0 ranked third behind Bard and Bing, ChatGPT 4.0 in our study outperformed both Gemini and Microsoft Copilot across most categories. The difference in rankings could be attributed to the unique cultural context of our dataset, which consists of foods with simple compositions whose names were translated into Turkish. This may have introduced challenges related to language translation nuances and the limited representation of Turkish foods in chatbots' training data, which were not present in the broader dataset of 539 food items evaluated in the earlier study. This variability mirrors the observed trend in the previous study, where higher oxalate levels posed classification challenges for all chatbots. These findings collectively emphasise the potential of AI chatbots in dietary management while also highlighting the impact of food names translated into Turkish and chatbot algorithms on their performance. They also point to the need for further refinement in chatbot capabilities, particularly in handling foods with high oxalate levels and culturally specific items.

Another related study examining the performance of ChatGPT 3.5, ChatGPT 4.0, Bard AI, and Bing Chat in classifying foods based on potassium and phosphorus content for chronic kidney disease patients highlighted similar trends in AI model variability (38). In that study, ChatGPT 4.0 displayed strong performance in categorising potassium content, achieving an 81% overall accuracy and excelling in identifying high potassium foods (99%). However, Bard AI demonstrated highest precision in phosphorus classification, achieving 100% accuracy, surpassing ChatGPT 4.0's 77%. Compared to these findings, where ChatGPT 4.0 excelled in potassium-related tasks but was outperformed by Bard AI for phosphorus, our study positions ChatGPT 4.0 as the most reliable tool for classifying oxalate content in Turkish foods. These differences highlight how AI model performance can vary depending on the nutrient or dietary component being assessed and the context of the dataset. Despite these variations, the consistent reliability of ChatGPT 4.0 across studies demonstrates its potential for diverse applications in dietary management. However, these findings also highlight the need for ongoing refinement of AI algorithms to improve accuracy for nutrient-specific classification tasks and to adapt to diverse cultural and dietary contexts.

The consistent classification patterns observed in this study highlight both the capabilities and the current limitations of AI chatbots in food categorisation tasks. Foods like spinach and cocoa in the "Very High" category and watermelon and cucumber in the "Very Low" category likely benefited from well-established oxalate content data in the chatbots' training datasets. However, the repeated misclassification of certain foods, particularly those in the "High" and "Very High" categories, raises concerns about the adequacy of existing datasets and algorithms. Foods with complex compositions or regional variations, such as bulgur, whole wheat flour, and walnut, may present challenges due to inconsistent representation or conflicting data sources within the training material. Another potential reason for misclassifications is the use of Turkish food names in this study. These findings emphasise the need for a more comprehensive approach to training AI models, ensuring diverse and region-specific foods are accurately incorporated.

The findings of this study show the potential of AI chatbots, particularly ChatGPT 4.0, as supplementary tools in dietary management and food classification tasks. However, none of these tools achieved 100% accuracy in all categories, raising concerns about their reliability in providing precise dietary recommendations. Such inaccuracies can mislead patients, especially those managing conditions like hyperoxaluria or kidney disease, where precise dietary advice is critical (47, 48). These inconsistencies highlight that, although AI chatbots offer promising applications in nutrition science, they are not yet reliable enough to replace human expertise. The human factor remains indispensable, as dietitians and healthcare professionals possess the contextual knowledge, critical thinking, and clinical judgement necessary to address the effectiveness of dietary recommendations. AI tools can serve as valuable aids to streamline certain tasks, including food categorisation or patient education, but their outputs require thorough verification and interpretation by trained professionals. Ensuring patient safety and the accuracy of dietary advice necessitates a collaborative approach, integrating the efficiency of AI with the indispensable oversight of human experts.

This study offers important perspectives on the effectiveness of AI chatbots in categorising foods based on their oxalate content, particularly within the context of the Turkish language. A notable strength of the study is offering a culturally specific perspective that is often underrepresented in AI research. This focus addresses a critical gap, as AI models frequently exhibit varying proficiency across different languages and cultural contexts. However, some limitations must be acknowledged. The use of Turkish food names may have posed additional challenges for chatbots that are not specifically optimised for non-English datasets, potentially affecting their performance. Additionally, the relatively small sample size (63 foods) limits the generalisability of the findings to broader dietary contexts. These limitations emphasise the need for further research to explore the performance of AI models across diverse cultural and linguistic settings and to refine their algorithms for improved accuracy and applicability.

Further studies could explore expanding the datasets to include a broader range of food items from different cultural contexts, which could help refine the models' ability to handle diverse food items. Additionally, examining the performance of AI tools across other nutrients, such as potassium, phosphorus, or vitamin D, could provide a more comprehensive understanding of their potential applications in renal nutrition. Research could also explore the integration of AI models with user-friendly platforms to support healthcare professionals in real-time decision-making, ensuring that the AI outputs are contextually relevant and clinically accurate. Moreover, investigating the impact of these AI tools on patient outcomes, particularly in populations with specific dietary needs (e.g., those with hyperoxaluria or kidney disease), could provide information about their practical utility in healthcare settings.

CONCLUSION

In conclusion, while AI chatbots, particularly ChatGPT 4.0, show promise in classifying foods based on their oxalate content, their performance remains variable, and they are not yet fully reliable for clinical applications without human oversight. Despite ChatGPT 4.0's higher performance compared to other models, the lack of 100% accuracy across different oxalate content categories highlights the need for further refinement in AI technology. These tools can serve as valuable aids for dietitians and healthcare professionals but should not replace their expertise, particularly when advising patients with specific dietary restrictions. The integration of AI into nutrition science holds considerable potential, but it must be accompanied by validation, human interpretation, and continuous improvement to ensure patient safety and clinical efficacy.

ETHICS COMMITTEE APPROVAL

Ethics committee approval is not required for this study.

DECLARATION OF CONFLICT OF INTEREST

Any financial or other interest related to the study there is no conflict.

FINANCIAL SUPPORT

No institution/organization provided financial support related to the study.

PEER REVIEW

External independent, double blind.

ACKNOWLEDGEMENT

I would like to express my sincere gratitude to Ömer Furkan Kaçar for his invaluable assistance in preparing the bar charts used in this study. His contributions were essential in visualising the data and enhancing the clarity of my findings.

AUTHOR CONTRIBUTIONS

Idea, design: HKK
Plan: HKK
Data collection: HKK
Analysis: HKK
Article writing: HKK
Critical review: HKK

References

1. Lang J, Narendrula A, El-Zawahry A, Sindhwani P, Ekwenna O. Global trends in incidence and burden of urolithiasis from 1990 to 2019: an analysis of global burden of disease study data. Eur Urol Suppl. 2022;35:37-46. https://doi.org/10.1016/j.euros.2021.10.008
2. Abufaraj M, Al Karmi J, Yang L. Prevalence and trends of urolithiasis among adults. Curr Opin Urol. 2022;32(4):425-32. https://doi.org/10.1097/mou.0000000000000994
3. Abbas W, Akram M, Sharif A. Nephrolithiasis; prevalence, risk factors and therapeutic strategies: a review. Med J Islam Repub Iran. 2019;3(3):90-5. https://doi.org/10.18689/mjiem-1000120
4. Hill AJ, Basourakos SP, Lewicki P, Wu X, Arenas-Gallo C, Chuang D, et al. Incidence of kidney stones in the United States: the continuous national health and nutrition examination survey. J Urol. 2022;207(4):851-6. https://doi.org/10.1097/ju.0000000000002331
5. Thongprayoon C, Krambeck AE, Rule AD. Determining the true burden of kidney stone disease. Nat Rev Nephrol. 2020;16(12):736-46. https://doi.org/10.1038/s41581-020-0320-7
6. Shastri S, Patel J, Sambandam KK, Lederer ED. Kidney stone pathophysiology, evaluation and management: core curriculum 2023. Am J Kidney Dis. 2023;82(5):617-34. https://doi.org/10.1053/j.ajkd.2023.03.017
7. Khan SR, Pearle MS, Robertson WG, Gambaro G, Canales BK, Doizi S, et al. Kidney stones. Nat Rev Dis Primers. 2016;2(1):1-23. https://doi.org/10.1038/nrdp.2016.8
8. Pearle MS, Goldfarb DS, Assimos DG, Curhan G, Denu-Ciocca CJ, Matlaga BR, et al. Medical management of kidney stones: AUA guideline. J Urol. 2014;192(2):316-24. https://doi.org/10.1016/j.juro.2014.05.006
9. Morgan MS, Pearle MS. Medical management of renal stones. BMJ. 2016;352. https://doi.org/10.1136/bmj.i52
10. Balawender K, Łuszczki E, Mazur A, Wyszyńska J. The Multidisciplinary Approach in the Management of Patients with Kidney Stone Disease—A State-of-the-Art Review. Nutrients. 2024;16(12):1932. https://doi.org/10.3390/nu16121932
11. Eryildirim B, Sahin C, Tuncer M, Sabuncu K, Cetinel C, Tarhan F, et al. Effect of medical expulsive therapy on the health-related quality of life of patients with ureteral stones: a critical evaluation. Int Urol Nephrol. 2015;47:1271-5. https://doi.org/10.1007/s11255-015-1036-7
12. Diniz DH, Blay SL, Schor N. Quality of life of patients with nephrolithiasis and recurrent painful renal colic. Nephrol Dial Transplant. 2007;106(3):c91-c7. https://doi.org/10.1159/000102995

13. Kalaitzidis RG, Damigos D, Siamopoulos KC. Environmental and stressful factors affecting the occurrence of kidney stones and the kidney colic. Int Urol Nephrol. 2014;46:1779-84. https://doi.org/10.1007/s11255-014-0758-2

14. Khopekar F, Nabi S, Shiva M, Stewart M, Rajendran B, Nabi G. Cost-effectiveness of quality improvement intervention to reduce time between CT-detection and ureteroscopic laser fragmentation in acute symptomatic ureteric stones management. World J Urol. 2024;42(1):144. https://doi.org/10.1007/s00345-023-04694-4

15. Neumaier ER. Effects of an integrated behavioral health intervention on the health outcomes and quality of life of patients with kidney stone disease [dissertation]. The University of Wisconsin-Madison;2012.

16. Néill EN, Richards HL, Hennessey D, Ryan EM, Fortune DG. Psychological distress in patients with urolithiasis: a systematic review and meta-analysis. J Urol. 2023;209(1):58-70. https://doi.org/10.1097/ju.0000000000003032

17. Huang Y, Zhang YH, Chi ZP, Huang R, Huang H, Liu G, et al. The handling of oxalate in the body and the origin of oxalate in calcium oxalate stones. Urol Res. 2020;104(3-4):167-76. https://doi.org/10.1159/000504417

18. Singh P, Enders FT, Vaughan LE, Bergstralh EJ, Knoedler JJ, Krambeck AE, et al. Stone composition among first-time symptomatic kidney stone formers in the community Mayo Clin Proc. Elsevier. 2015;90(10):1356-65. https://doi.org/10.1016/j.mayocp.2015.07.016

19. Trinchieri A. Epidemiology of urolithiasis: an update. Clinical cases in mineral and bone metabolism. [Internet]. 2008[cited 2024 Nov 27];5(2):101-106. Available from: https://pmc.ncbi.nlm.nih.gov/articles/PMC2781200/

20. Taylor EN, Curhan GC. Role of nutrition in the formation of calcium-containing kidney stones. Nephrol Dial Transplant. 2004;98(2):55-63. https://doi.org/10.1159/000080265

21. Lin BB, Lin ME, Huang RH, Hong YK, Lin BL, He XJ. Dietary and lifestyle factors for primary prevention of nephrolithiasis: a systematic review and meta-analysis. BMC Nephrol. 2020;21:1-13. https://doi.org/10.1186/s12882-020-01925-3

22. Willett WC, Koplan JP, Nugent R, Dusenbury C, Puska P, Gaziano TA. Prevention of chronic disease by means of diet and lifestyle changes. Dis Control Priorities Dev Ctries [Internet]. 2006[cited 2024 Nov 28];3(2):599-616. Available from: https://www.ncbi.nlm.nih.gov/books/NBK11795/

23. Mitchell T, Kumar P, Reddy T, Wood KD, Knight J, Assimos DG, et al. Dietary oxalate and kidney stone formation. Am J Physiol Renal Physiol. 2019;316(3):F409-F13. https://doi.org/10.1152/ajprenal.00373.2018

24. Franceschi VR, Nakata PA. Calcium oxalate in plants: Formation and function. Annu Rev Plant Biol. 2005;56(1):41-71. https://doi.org/10.1146/annurev.arplant.56.032604.144106

25. Massey LK. Dietary influences on urinary oxalate and risk of kidney stones. Front Biosci. 2003;8:584-94. https://doi.org/10.2741/1082

26. Holmes RP, Knight J, Assimos DG. Lowering urinary oxalate excretion to decrease calcium oxalate stone disease. Urolithiasis. 2016;44(1):27-32. https://doi.org/10.1007/s00240-015-0839-4

27. Lo CYZ, Khor QH, Abdullatif VA, Delgado C, Lu Y, Katz J, et al. Systematic review of pharmacological, complementary and alternative therapies for the prevention of calcium oxalate stones. Asian J Urol. [Article in press]. 2024 [cited 2024 Nov 30] Available from: https://doi.org/10.1016/j.ajur.2024.04.006

28. Siener R, Ernsten C, Welchowski T, Hesse A. Metabolic Profile of Calcium Oxalate Stone Patients with Enteric Hyperoxaluria and Impact of Dietary Intervention. Nutrients. 2024;16(16):2688. https://doi.org/10.3390/nu16162688

29. Taylor EN, Curhan GC. Oxalate intake and the risk for nephrolithiasis. J Am Soc Nephrol. 2007;18(7):2198-204. https://doi.org/10.1681/asn.2007020219

30. Meschi T, Maggiore U, Fiaccadori E, Schianchi T, Bosi S, Adorni G, et al. The effect of fruits and vegetables on urinary stone risk factors. Kidney Int. 2004;66(6):2402-10. https://doi.org/10.1111/j.1523-1755.2004.66029.x

31. Noonan SC, Savage GP. Oxalate content of foods and its effect on humans. Asia Pac J Clin Nutr [Internet]. 1999 [cited 2024 Nov 30];8(1):64-74. Available from: https://pubmed.ncbi.nlm.nih.gov/24393738/

32. Aiumtrakul N, Thongprayoon C, Arayangkool C, Vo KB, Wannaphut C, Suppadungsuk S, et al. Personalized Medicine in Urolithiasis: AI Chatbot-Assisted Dietary Management of Oxalate for Kidney Stone Prevention. J Pers Med. 2024;14(1):107. https://doi.org/10.3390/jpm14010107

33. Attalla K, De S, Monga M. Oxalate content of food: a tangled web. Urology. 2014;84(3):555-60. https://doi.org/10.1016/j.urology. 2014.03.053

34. Udegbe FC, Ebulue OR, Ebulue CC, Ekesiobi CS. The role of artificial intelligence in healthcare: A systematic review of applications and challenges. Int Med Sci Rev J. 2024;4(4):500-8. https://doi.org/10.51594/imsrj.v4i4.1052

35. Poalelungi DG, Musat CL, Fulga A, Neagu M, Neagu AI, Piraianu AI, et al. Advancing patient care: how artificial intelligence is transforming healthcare. Healthcare (Basel). 2023;13(8):1214. https://doi.org/10.3390/jpm13081214

36. Talyshinskii A, Naik N, Hameed BZ, Juliebø-Jones P, Somani BK. Potential of AI-driven chatbots in urology: revolutionizing patient care through artificial intelligence. Curr Urol Rep. 2024;25(1):9-18. https://doi.org/10.1007/s11934-023-01184-3

37. Zhang J, Oh YJ, Lange P, Yu Z, Fukuoka Y. Artificial intelligence chatbot behavior change model for designing artificial intelligence chatbots to promote physical activity and a healthy diet. J Med Internet Res. 2020;22(9):e22845. https://doi.org/10.2196/22845

38. Qarajeh A, Tangpanithandee S, Thongprayoon C, Suppadungsuk S, Krisanapan P, Aiumtrakul N, et al. AI-Powered Renal Diet Support: Performance of ChatGPT, Bard AI, and Bing Chat. J Pers Med. 2023;13(5):1160-72. https://doi.org/10.3390/clinpract 13050104

39. Sharma SK, Gaur S. Optimizing Nutritional Outcomes: The Role of AI in Personalized Diet Planning. International Journal for Research Publication and Seminar. Int J Res Pub Seminar. 2024. https://doi.org/10.36676/jrps.v15.i2.15

40. Alhur A. Redefining healthcare with artificial intelligence (AI): the contributions of ChatGPT, Gemini, and Co-pilot. Curr Opin Clin Nutr Metab Care. 2024;16(4). https://doi.org/10.7759/cureus.57795

41. Oh YJ, Zhang J, Fang M-L, Fukuoka Y. A systematic review of artificial intelligence chatbots for promoting physical activity, healthy diet, and weight loss. Int J Behav Nutr Phys Act. 2021;18:1-25. https://doi.org/10.1186/s12966-021-01224-6

42. Kelly CJ, Karthikesalingam A, Suleyman M, Corrado G, King D. Key challenges for delivering clinical impact with artificial intelligence. Nat Biomed Eng. 2019;17:1-9. https://doi.org/10.1186/s12916-019-1426-2

43. Miao J, Thongprayoon C, Suppadungsuk S, Krisanapan P, Radhakrishnan Y, Cheungpasitporn W. Chain of thought utilization in large language models and application in nephrology. J Med AI Nephrol. 2024;60(1):148. https://doi.org/10.3390/medicina 60010148

44. ChatGPT 4.0 [Internet]. [cited 2024 Nov 13]. Available from: https://chatgpt.com.

45. Gemini [Internet]. [cited 2024 Nov 14]. Available from: https://gemini.google.com/app.

46. Copilot [Internet]. [cited 2024 Nov 15]. Available from: https://copilot.microsoft.com/.

47. Traver MA, Passman CM, LeRoy T, Passmore L, Assimos DG. Is the Internet a reliable source for dietary recommendations for stone formers?. J Endourol. 2009;23(4):715-7. https://doi.org/10.1089/end.2008.0490

48. Lambert K, Mullan J, Mansfield K, Koukomous A, Mesiti L. Evaluation of the quality and health literacy demand of online renal diet information. J Hum Nutr Diet. 2017;30(5):634-45. https://doi.org/10.1111/jhn.12466