Cumhuriyet Dental Journal, 28(2): 210-215, 2025
DOI: https://doi.org/10.7126/cumudj.1623854

Cumhuriyet Dental Journal

| cdj.cumhuriyet.edu.tr | Founded: 1998 Available online, ISSN: 1302-5805 Publisher: Sivas Cumhuriyet Üniversitesi

Assessing the Success of ChatGPT-4o in Oral Radiology Education and Practice: A Pioneering Research

Fatma Akkoca[1,a], Melih Özdede[1,b], Günnur İlhan[1,c,*], Emre Koyuncu[1,d], Hülya Ellidokuz[2,3,e]

[1]*Department of Dentomaxillofacial Radiology, Faculty of Dentistry, Dokuz Eylul University, Izmir, Turkiye*
[2]*Department of Biostatistics and Medical Informatics, Faculty of Medicine, Dokuz Eylul University, Izmir, Turkiye*
[3]*Department of Preventive Oncology, Institute of Oncology, Dokuz Eylul University, Izmir, Turkiye*
**Corresponding author*

Research Article

History

Received: 21/01/2025
Accepted: 04/04/2025

ABSTRACT

Objectives: This study aims to assess the comprehension and interpretation performance of Chat Generative Pre-Train Omni (GPT-4o) in the context of oral radiology education and practice.

Materials and Methods: Utilizing a set of 99 questions derived from the book "White and Pharoah's Oral Radiology: Principles and Interpretation 8th Edition," this study employed ChatGPT-4o to respond to these questions thrice daily at varying times over 10 days, generating a total of 60 responses for each question. Two oral radiologists independently answered the same questions and verified their answers with the relevant textbook. Responses were compared to those of ChatGPT-4o.

Results: The study revealed that ChatGPT-4o's correct answer rate was 59.4%. Time-based analysis revealed performance differences across specific day periods. Specifically, during noon and evening sessions, the success rate on the first and seventh days was statistically significantly higher ($p = 0.003$ and $p = 0.002$, respectively), while morning performance on those days was significantly lower ($p < 0.05$), indicating that the time and day of the query may influence response accuracy. In contrast, no significant relationship was found between the difficulty level of the questions and the model's accuracy ($p > 0.05$).

Conclusions: Presently, ChatGPT exhibits inadequacies in its application to oral radiology training and clinical practice. Despite this, expectations for platform improvement and expansion in utility persist, particularly with increased data input and advancements in artificial intelligence.

Keywords: Artificial Intelligence, ChatGPT, Dental Education, Oral Radiology

Oral Radyoloji Eğitimi ve Uygulamasında ChatGPT-4o'nun Başarısının Değerlendirilmesi: Öncü Bir Araştırma

Araştırma Makalesi

Süreç

Geliş: 21/01/2025
Kabul: 04/04/2025

ÖZET

Amaç: Bu çalışma, Chat Generative Pre-Train Omni (GPT-4o) platformunun oral radyoloji eğitimi ve pratiği konusunda anlama ve yorumlama performansını değerlendirmeyi amaçlamaktadır.

Gereç ve Yöntemler: Çalışmada, "White and Pharoah's Oral Radiology: Principles and Interpretation 8th Edition" kitabından alınan 99 sorudan oluşan bir set kullanılmıştır. Bu sorular, ChatGPT-4o tarafından günün üç farklı periyodunda, 10 gün boyunca yanıtlanmış ve her soru için toplam 60 yanıt üretilmiştir. Aynı soruları iki ağız, diş ve çene radyolojisi uzmanı da yanıtlamış ve verdikleri cevaplar ilgili ders kitabı ile doğrulanmıştır. ChatGPT-4o'nun cevapları ile uzmanların cevapları karşılaştırılmıştır.

Bulgular: Çalışmada ChatGPT-4o'nun doğru yanıt verme oranı %59,4 olarak bulunmuştur. Zaman temelli analiz, belirli gün ve saat dilimlerinde performans farklılıkları olduğunu ortaya koymuştur. Özellikle öğle ve akşam oturumlarında, birinci ve yedinci günlerdeki başarı oranı istatistiksel olarak anlamlı şekilde daha yüksekken (sırasıyla $p = 0,003$ ve $p = 0,002$), aynı günlerdeki sabah performansı anlamlı derecede daha düşüktü ($p < 0,05$). Bu bulgular, sorgulamanın yapıldığı gün ve saatin yanıt doğruluğunu etkileyebileceğini göstermektedir. Öte yandan, soruların zorluk düzeyi ile modelin doğruluk oranı arasında anlamlı bir ilişki bulunmamıştır ($p > 0,05$).

Sonuç: Mevcut durumda, ChatGPT oral radyoloji eğitimi ve klinik pratiğinde yetersizlikler göstermiştir. Buna rağmen, platformun daha fazla veri girişi ve yapay zekâdaki ilerlemelerle birlikte gelişmesi ve kullanım alanlarının genişlemesi konusunda beklentiler devam etmektedir.

Anahtar Kelimeler: ChatGPT, Diş Hekimliği Eğitimi, Oral Radyoloji, Yapay Zekâ.

[a] *fatma.akkoca@deu.edu.tr* *0000-0002-4522-656X*
[b] *melih.ozdede@deu.edu.tr* *0000-0002-8783-802X*
[c] *gunnur.ilhan@deu.edu.tr* *0009-0001-9587-9059*
[d] *emre.koyuncu@deu.edu.tr* *0009-0007-0114-577X*
[e] *hulya.ellidokuz@deu.edu.tr* *0000-0001-8503-061X*

How to Cite: Akkoca F, Özdede M, İlhan G, Koyuncu E, Ellidokuz H. (2025) Assessing the Success of ChatGPT-4o in Oral Radiology Education and Practice: A Pioneering Research, Cumhuriyet Dental Journal, 28(2): 210-215.

Introduction

Artificial intelligence (AI) signifies the ability of computer systems to emulate cognitive functions and execute tasks with a human-like proficiency.[1] This is achieved through the refinement of algorithms that enable machines to process, learn, and autonomously tackle complex problems.[2] The field of AI has experienced rapid advancements, marked by an increasing prevalence of artificial intelligence-based chatbots and conversational agents.[3]

In 2018, Open AI introduced the Generative Pre-Train (GPT) language model, with subsequent developments leading to the creation of ChatGPT in subsequent years.[3] Operating on the foundation of Large Language Models (LLMs), this software is designed to replicate human speech, comprehend the nuances of language, and generate novel content based on its exposure to training data.[4] ChatGPT finds applications in diverse areas such as idea generation, creative endeavors, coding, text editing, and academic article writing.[5,6]

Launched on May 13, 2024, GPT-4o became a significant upgrade to the GPT series, offering promising features such as multimodal capabilities (text, image, and audio processing), faster performance, expanded multilingual support, a larger context window, and improved accuracy. Launched on March 14, 2023, GPT-4 uses a Transformer-based model, a paradigm that includes pre-training using both public data and "licensed data from third-party providers" to predict the next token.[7] However, ChatGPT's ability to access only a variety of internet data until 2021, combined with its limited access to relevant databases, raises concerns about the timeliness and reliability of the information it produces.[8] Another drawback lies in its tendency to spread false information, which poses a challenge for inexperienced users to distinguish between real and fake information.[6]

While studies in the medical field have showcased ChatGPT's potential as a medical consultation tool, comprehensive examinations are essential to evaluate its performance across different medical specialties.[9,10,11] In the realm of oral radiology, ChatGPT offers swift access to a vast repository of medical information, providing clinical guidelines and current research for various oral conditions and diseases. Dentists and oral radiologists can leverage ChatGPT to enhance diagnostic skills and clinical practice.[12] Multiple studies have been conducted exploring the applicability and effectiveness of ChatGPT in the field of oral radiology, investigating its potential to support diagnostic accuracy, clinical decision-making, and educational purposes.[13-17] Moreover, the software has the potential to augment diagnostic quality by analyzing patient symptoms and relevant data, guiding oral radiologists toward potential diagnoses and imaging options.[13] Beyond this, ChatGPT holds promise for the future of distance consultation and telemedicine.[18]

This study aims to assess the comprehension and interpretation performance of ChatGPT-4o in the context of oral radiology education and practice.

Materials and Methods

Ethical Approval

There is no human as a participant in the study and ethical approval for this research was not required.

Question Design

The questions set are prepared based on 'White and Pharoah's Oral Radiology: Principles and Interpretation 8th Edition'.[19] It was arranged as three questions from each section, totally 99 questions in from 33 sections. According to difficulty levels, questions were classified as easy, medium, and difficult.

Generating Answers in ChatGPT

Two authors (GI, EK), using two different accounts, performed response generation using ChatGPT-4o. ChatGPT responses were captured three times (morning:09.00, noon:12.00, and evening:17.00) by selecting the 'new chat' option for 10 days until 60 responses per question were obtained. This study benefited from the study of Suares *et al.*[20]

Human Expert Answers

Two dentomaxillofacial radiologists with at least five years of experience (FA, MO) independently answered ('yes' or 'no') to 99 questions. In case of inconsistency in expert answers the questions and literature were re-evaluated and a consensus was reached.

Statistical Analysis

All answers were recorded in an Excel spreadsheet and analyzed using the statistical software program of SPSS (Version 29, IBM, Armonk, NY, USA). Kappa and McNemar tests were used for the agreement of two profiles, and the agreement of different days. Chi-square test was performed for the comparison of difficulty levels and correct answers. Cochran-Q test was used for the relationship between different days, also between three periods of the days. The significance level was set to 0.05.

Results

The overall accuracy rate of ChatGPT-4o in providing correct answers was found to be 59.4%. A detailed time-based analysis further delineated the accuracy rates as 59.1% at 09:00, 59.5% at 12:00, and 59.5% at 17:00.

The agreement between profiles was assessed using Kappa and McNemar tests. It was observed that there was perfect agreement for all queries ($p = 1.000$, indicating perfect agreement), and the analysis continued with a single profile.

Sixty ChatGPT-4o queries with expert responses were evaluated using McNemar and Kappa tests (refer to Table 1). There was no significant agreement between ChatGPT-4o responses and expert responses ($p > 0.05$). ChatGPT-4o's performance was evaluated based on the accuracy of its answers to questions from different subjects since only three questions were asked per subject. The obtained accuracy percentages are as follows: while it performed

well in theoretical and structured subjects such as Physics (100%) and Quality Assurance & Infection Control (100%), its accuracy declined in complex diagnostic fields such as Prescribing Diagnostic Imaging (26.7%), Forensics (33.3%), and Cysts (33.3%) (refer to Table 2).

The difficulty level of the questions and the rates of correct and incorrect answers were compared using a chi-square test for a total of 60 queries. No significant relationship was detected between the difficulty level and the correct answer rate for any of these questions ($p > 0.05$). The Kappa values were ranged from 0.091 to 0.863. The relationship between answers and correct answers to questions asked on different days was evaluated using the Cochran-Q test. There was no significant difference between different days during morning interrogations ($p = 0.558$). However, during noon and evening interrogations, the success rate on the first and seventh days was statistically significantly higher (P-values of noon and evening, respectively: $p = 0.003$, $p = 0.002$).

The relationship between periods of the day and the given answers was assessed using the Cochran Q test (refer to Table 3). It was determined that the morning inquiries on the first and seventh days were statistically lower than other day periods ($p < 0.05$).

Table 1. Evaluation of expert response and ChatGPT response of 60 queries

			Chatgpt response					Chatgpt response		
	Day and time	Answer	Yes	No	P value	Day and time	Answer	Yes	No	P value
Expert response	1M	Yes	33 78.60% 50.00%	9 21.40% 27.30%	0.192	6M	Yes	34 81.0% 53.1%	8 19.0% 22.9%	0.265
		No	33 57.90% 50.00%	24 42.10% 72.70%			No	30 52.6% 46.9%	27 47.4% 77.1%	
	1N	Yes	35 83.30% 48.60%	7 16.70% 25.90%	0.168	6N	Yes	34 81.0% 53.1%	8 19.0% 22.9%	0.265
		No	37 64.90% 51.4	20 35.10% 74.10%			No	30 52.6% 46.9%	27 47.4% 77.1%	
	1E	Yes	35 83.30% 48.60%	7 16.70% 25.90%	0.168	6E	Yes	34 81.0% 53.1%	8 19.0% 22.9%	0.265
		No	37 64.90% 51.40%	20 35.10% 74.10%			No	30 52.6% 46.9%	27 47.4% 77.1%	
	2M	Yes	32 76.2% 50.8%	10 23.8% 27.8%	0.205	7M	Yes	33 78.6% 50.0%	9 21.4% 27.3%	0.192
		No	31 54.4% 49.2%	26 45.6% 72.2%			No	33 57.9% 50.0%	24 42.1% 72.7%	
	2N	Yes	34 81.0% 53.1%	8 19.0% 22.9%	0.265	7N	Yes	35 83.3% 48.6%	7 16.7% 25.9%	0.168
		No	30 52.6% 46.9%	27 47.4% 77.1%			No	37 64.9% 51.4%	20 35.1% 74.1%	
	2E	Yes	34 81.0% 53.1%	8 19.0% 22.9%	0.265	7E	Yes	35 83.3% 48.6%	7 16.7% 25.9%	0.168
		No	30 52.6% 46.9%	27 47.4% 77.1%			No	37 64.9% 51.4%	20 35.1% 74.1%	
	3M	Yes	34 81.0% 49.3%	8 19.0% 26.7%	0.18	8M	Yes	32 76.2% 50.8%	10 23.8% 27.8%	0.205
		No	35 61.4% 50.7%	22 38.6% 73.3%			No	31 54.4% 49.2%	26 45.6% 72.2%	

		32	10			34	8		
	Yes	76.2%	23.8%			81.0%	19.0%		
3N		50.8%	27.8%	0.205	8N	53.1%	22.9%	0.265	
		31	26			30	27		
	No	54.4%	45.6%			52.6%	47.4%		
		49.2%	72.2%			46.9%	77.1%		
		32	10			34	8		
	Yes	76.2%	23.8%			81.0%	19.0%		
3E		50.8%	27.8%	0.205	8E	53.1%	22.9%	0.265	
		31	26			30	27		
	No	54.4%	45.6%			52.6%	47.4%		
		49.2%	72.2%			46.9%	77.1%		
		34	8			34	8		
	Yes	81.0%	19.0%			81.0%	19.0%		
4M		53.1%	22.9%	0.265	9M	49.3%	26.7%	0.18	
		30	27			35	22		
	No	52.6%	47.4%			61.4%	38.6%		
		46.9%	77.1%			50.7%	73.3%		
		34	8			32	10		
	Yes	81.0%	19.0%			76.2%	23.8%		
4N		53.1%	22.9%	0.265	9N	50.8%	27.8%	0.205	
		30	27			31	26		
	No	52.6%	47.4%			54.4%	45.6%		
		46.9%	77.1%			49.2%	72.2%		
		34	8			32	10		
	Yes	81.0%	19.0%			76.2%	23.8%		
4E		53.1%	22.9%	0.265	9E	50.8%	27.8%	0.205	
		30	27			31	26		
	No	52.6%	47.4%			54.4%	45.6%		
		46.9%	77.1%			49.2%	72.2%		
		35	7			34	8		
	Yes	83.3%	16.7%			81.0%	19.0%		
5M		52.2%	21.9%	0.252	10M	53.1%	22.9%	0.265	
		32	25			30	27		
	No	56.1%	43.9%			52.6%	47.4%		
		47.8%	78.1%			46.9%	77.1%		
		32	10			34	8		
	Yes	76.2%	23.8%			81.0%	19.0%		
5N		50.8%	27.8%	0.205	10N	53.1%	22.9%	0.265	
		31	26			30	27		
	No	54.4%	45.6%			52.6%	47.4%		
		49.2%	72.2%			46.9%	77.1%		
		32	10			34	8		
	Yes	76.2%	23.8%			81.0%	19.0%		
5E		50.8%	27.8%	0.205	10E	53.1%	22.9%	0.265	
	No	31	26			30	27		
		54.4%	45.6%			52.6%	47.4%		
		49.2%	72.2%			46.9%	77.1%		

M: morning, N: noon, E: evening

Table 2. Percentage of correct answers by ChatGPT-4o for subtopic questions in 'Oral Radiology: Principles and Interpretation, 8th Edition'

	Subjects	*Correct Percentage*
1	Physics	100.0%
2	Quality Assurance and Infection Control	100.0%
3	Dental Caries	91.1%
4	Periodontal Diseases	82.2%
5	Safety and Protection	77.8%
6	Biologic Effects of Ionizing Radiation	73.3%
7	Paranasal Sinus Diseases	68.9%
8	Beyond Three-Dimensional Imaging	66.7%
9	Trauma	66.7%

10	Diseases Affecting the Structure of Bone	66.7%
11	Benign Tumors and Neoplasms	66.7%
12	Soft Tissue Calcifications and Ossifications	66.7%
13	Inflammatory Conditions of the Jaws	66.7%
14	Craniofacial Anomalies	66.7%
15	Intraoral Projections	66.7%
16	Digital Imaging	66.7%
17	Radiographic Anatomy	64.4%
18	Salivary Gland Diseases	60.0%
19	Other Imaging Modalities	57.8%
20	Malignant Neoplasms	55.6%
21	Dental Implants	54.4%
22	Film Imaging	51.1%
23	Principles of Radiographic Interpretation	48.9%
24	Dental Anomalies	48.9%
25	Cone Beam Computed Tomography: Volume Acquisition	42.2%
26	Cephalometric and Skull Imaging	42.2%
27	Projection Geometry	42.2%
28	Temporomandibular Joint Abnormalities	35.6%
29	Cone Beam Computed Tomography: Volume Preparation	35.6%
30	Forensics	33.3%
31	Cysts	33.3%
32	Panoramic Imaging	33.3%
33	Prescribing Diagnostic Imaging	26.7%

Table 3. Assessment of relationship between periods of the day and given answers

Days	Morning (Yes-No)	Noon (Yes-No)	Evening (Yes-No)	P value
D1	66-33	72-27	72-27	**0.011***
D2	63-36	64-35	64-35	0.926
D3	69-30	63-36	63-36	0.050
D4	64-35	64-35	64-35	1.0
D5	67-32	63-36	63-36	0.264
D6	64-35	64-35	64-35	1.0
D7	66-33	72-27	72-27	**0.011***
D8	63-36	64-35	64-35	0.926
D9	69-30	63-36	63-36	0.050
D10	64-35	64-35	64-35	1

Assessment of relationship between periods of the day and given answers using the Cochran Q Test *P < 0.05

Discussion

In recent times, ChatGPT has gained popularity as a valuable research tool, with studies exploring its applications in both clinical and educational contexts within the medical field.[7,9,12,20] Focusing on oral and maxillofacial radiology, a specialized branch of dentistry dealing with image acquisition and interpretation in the maxillofacial region for diagnosis and treatment planning, this study delves into the educational and clinical efficacy of ChatGPT-4o, the latest iteration of the ChatGPT series.[14]

Suarez *et al.* conducted a study assessing the platform's effectiveness in endodontics, achieving a success rate of 57.3%, consistent with the findings of this study.[20] Their study also highlighted consistency across different times of the day. Similarly, Antaki *et al*. utilized the ChatGPT Plus BCSC test set, obtaining 59.4% accuracy in a simulated Ophthalmic Knowledge Assessment Program examination and 49.2% accuracy in the OphthoQuestions test set.[21] The accuracy rates in the field of ophthalmology align closely with the performance of ChatGPT-4o in the realm of oral radiology in this study.

Contrastingly, Bragazzi *et al*. evaluated ChatGPT's diagnostic accuracy in endodontic cases, reporting varied results.[22] While it correctly identified existing endodontic treatments, tooth decay, and dental restorations in certain cases, it displayed limitations in detecting endodontic lesions, resulting in an overall correct interpretation rate of 11%. In the present study, clinical cases were not directly evaluated as images; however, when lesions were described with clinical-radiological features in written form, ChatGPT-4o exhibited more accurate verbal interpretation.

Mago *et al*. explored the use of ChatGPT-3 for radiology report writing and educational purposes in oral and maxillofacial radiology, emphasizing the need for improvement in queries related to anatomical landmarks and radiographic features of pathologies.[15] This study, utilizing ChatGPT-4o, suggests advancements in the platform, showcasing its continuous development and improvement, which bodes well for the future of artificial intelligence technologies.

Bhayana *et al*. evaluated ChatGPT's performance on multiple-choice examination questions in medical radiology without images, achieving a 69% correct response rate.[16] While their results appear more successful than this study, which focused on oral radiology, the differences in question types and content may account for variations in performance.

Ozturk *et al*. conducted a study evaluating ChatGPT's success in oral radiology education, where ChatGPT-4o

demonstrated a high success rate, correctly answering 15 out of 20 multiple-choice questions (75%).[17] However, the smaller question set and focus on undergraduate education may have contributed to a seemingly higher success rate compared to this study. Despite its numerous advantages, ChatGPT raises ethical concerns, persistent misinformation issues, copyright considerations, and legal and regulatory challenges.[15] Addressing these concerns necessitates additional studies and research efforts to ensure this technology's responsible and effective use in various domains.

An analysis of ChatGPT-4o's performance revealed significant variations in accuracy across different subjects. While it performed well in theoretical and structured subjects, such as Physics and Quality Assurance & Infection Control, its accuracy declined in complex diagnostic domains, including Diagnostic Imaging Prescribing, Forensic Medicine, and Cysts. These findings suggest that ChatGPT-4o excels in well-defined, rule-based topics because it can process structured information efficiently. However, the results also indicate that the AI requires further training in clinical diagnosis-based areas, particularly those that rely on visual support. Its lower accuracy in complex diagnostic fields may be attributed to the need for contextual understanding and interpretation of visual data, which remains a challenge for language-based AI models.

Conclusions

This study showed that ChatGPT-4o has limitations in terms of its suitability in oral radiology education and clinical practice due to intra- and inter-day inconsistencies and low correct response rates. It also demonstrates that it is very important and mandatory for the platform to undergo special training that focuses specifically on medical information. However, it is foreseeable that the platform will be further developed, and its scope of use is expected to expand in parallel with increasing data entry and developments in artificial intelligence. Continuous improvement and development of the platform hold the potential to overcome current limitations and make ChatGPT a more robust and effective tool in the field of oral radiology and medical education.

Ethics Approval

Formal ethical approval was not required for this study. The research did not involve human participants and therefore was not subject to ethical standards pertaining to human experimentation.

Acknowledgements

No external funding was received for this study.

Conflicts of Interest Statement

The authors declare that they have no conflicts of interest

References

1. Rodrigues JA, Krois J, Schwendicke F. Demystifying artificial intelligence and deep learning in dentistry. Braz Oral Res 2021;35:e094.
2. Deng L. Artificial intelligence in the rising wave of deep learning: the historical path and futureoutlook. IEEE Signal Process Mag 2018;35(1):180-177.
3. Mogali SR. Initial impressions of ChatGPT for anatomy education. Anat Sci Educ 2024;17(2):444-447.
4. Abd-Alrazaq A, AlSaad R, Alhuwail D, Ahmed A, Healy PM, Latifi S, Aziz S, Damseh R, Alabed Alrazak S, Sheikh J. Large language models in medical education: opportunities, challenges, and future directions. JMIR Med Educ 2023;9:e48291.
5. Cadamuro J, Cabitza F, Debeljak Z, De Bruyne S, Frans G, Perez SM, Ozdemir H, Tolios A, Carobene A, Padoan A. Potentials and pitfalls of ChatGPT and natural-language artificial intelligence models for understanding laboratory medicine test results. Clin Chem Lab Med 2023;61(7):1158-1166.
6. Li H, Moon JT, Purkayastha S, Celi LA, Trivedi H, Gichoya JW. Ethics of large language models in medicine and medical research. Lancet Digit Health 2023;5(6):e333-e335.
7. Gilson A, Safranek CW, Huang T, Socrates V, Chi L, Taylor RA, Chartash D. How does ChatGPT perform on the United States Medical Licensing Examination? The implications of large language models for medical education and knowledge assessment. JMIR Med Educ 2023;9:e45312.
8. Arif TB, Munaf U, Ul-Haque I. The future of medical education and research: is ChatGPT a blessing or blight in disguise? Med Educ Online 2023;28(1):2181052.
9. Kung TH, Cheatham M, Medenilla A, Sillos C, De Leon L, Elepaño C, Madriaga M, Aggabao R, Diaz-Candido G, Maningo J, Tseng V. Performance of ChatGPT on USMLE: potential for AI-assisted medical education using large language models. PLOS Digit Health 2023;2(2):e0000198.
10. Yeo YH, Samaan JS, Ng WH, Ting PS, Trivedi H, Vipani A, Ayoub W, Yang JD, Liran O, Spiegel B, Kuo A. Assessing the performance of ChatGPT in answering questions regarding cirrhosis and hepatocellular carcinoma. Clin Mol Hepatol 2023;29(3):721-732.
11. Samaan JS, Yeo YH, Rajeev N, Hawley L, Abel S, Ng WH, Srinivasan N, Park J, Burch M, Watson R, Liran O, Samakar K. Assessing the Accuracy of Responses by the Language Model ChatGPT to Questions Regarding Bariatric Surgery. Obes Surg 2023;33(6):1790-1796.
12. Fatani B. ChatGPT for future medical and dental research. Cureus 2023;15(4):e37285.
13. de Souza LL, Lopes MA, Santos-Silva AR, Vargas PA. The potential of ChatGPT in oral medicine: a new era of patient care? Oral Surg Oral Med Oral Pathol Oral Radiol 2024;137(1):1-2.
14. Khurana S, Vaddi A. ChatGPT from the perspective of an academic oral and maxillofacial radiologist. Cureus 2023;15(6):e40053.
15. Mago J, Sharma M. The potential usefulness of ChatGPT in oral and maxillofacial radiology. Cureus 2023;15(7):e42133.
16. Bhayana R, Krishna S, Bleakney RR. Performance of ChatGPT on a radiology board-style examination: insights into current strengths and limitations. Radiology 2023;307(5):e230582.
17. Öztürk HP, Avsever İH, Şenel B, Ayran Ş, Özgedik HS, Baysal N. ChatGPT in oral and maxillofacial radiology education. Res Square 2023;3566948.
18. Dave T, Athaluri SA, Singh S. ChatGPT in medicine: an overview of its applications, advantages, limitations, future prospects, and ethical considerations. Front Artif Intell 2023;6:1169595.
19. Mallya S, Lam E. White and Pharoah's oral radiology: principles and interpretation. 8th ed. Elsevier Health Sciences 2018.
20. Suárez A, Díaz-Flores García V, Algar J, Gómez Sánchez M, Llorente de Pedro M, Freire Y. Unveiling the ChatGPT phenomenon: evaluating the consistency and accuracy of endodontic question answers. Int Endod J 2024;57(1):108-113.
21. Antaki F, Touma S, Milad D, El-Khoury J, Duval R. Evaluating the performance of ChatGPT in ophthalmology: an analysis of its successes and shortcomings. Ophthalmol Sci 2023;3(4):100324.
22. Bragazzi NL, Szarpak L, Piccotti F. Assessing ChatGPT's potential in endodontics: preliminary findings from a diagnostic accuracy study. SSRN 2023;4631017.