Journal of Dentomaxillofacial Radiology, Pathology and Surgery

Summer 2024, Volume 13, Number 4

Review Paper: The Accuracy of ChatGPT in Answering Oromaxillofacial Surgery Questions: A Current and Updated Review

Pedram Hajibagheri[1], Monireh Aghajany-Nasab[2], Pardis Palizban[3*]

1. Department of Oral and Maxillofacial Medicine, Student Research Committee, School of Anzali International Campus, Guilan University of Medical Sciences, Rasht, Iran
2. Cellular & Molecular Research Center, Department of Biochemistry, School of Medicine, Guilan University of Medical Sciences, Rasht, Iran
3. Anzali International Faculty of Dentistry, Guilan University of Medical Sciences, Guilan, Iran

Use your device to scan and read the article online

Citation Hajibagheri P, Aghajany-Nasab M, Palizban P. The Accuracy of ChatGPT in Answering Oromaxillofacial Surgery Questions: A Current and Updated Review. Journal of Dentomaxillofacial Radiology, Pathology and Surgery. 2025; 13(4): 1-8

https://doi.org/10.22034/3DJ.13.4.1

ABSTRACT

Currently, ChatGPT's as an artificial intelligence language model, has been used in various fields of dentistry. Several studies have assessed its accuracy in diagnosing lesions, treatment planning, and surgical procedures. As an updated review, this study provides quick insights into the accuracy of ChatGPT's in the field of oral and maxillofacial surgery. A thorough search was conducted in the PubMed and Web of Science databases. To ensure a clear and comprehensive strategy to literature selection and data synthesis, this review followed the PRISMA guidelines. The study employed the PICO framework, a widely used methodology for structuring clinical research questions and guiding reviews. A total of 30 articles related to the accuracy of ChatGPT's responses in oral and maxillofacial surgery were found. After a review of titles and abstracts followed by a full-text review,10 articles that met the inclusion criteria were assessed in the final review. ChatGPT's might be able to help in responding to oromaxillofacial questions for supporting clinicians, but its role remains supportive rather than replacing professional expertise. Further development is necessary to enhance the model's ability to handle the complexities of clinical practice, where the degrees of patient care require more detailed associated with context-specific knowledge. Therefore, further study of its role for education and clinical decision-making recommended advantageous.

Article info:
Received: 02 Oct 2024
Accepted: 02 Nov 2024
Available Online:09 Nov 2024

Keywords:
*Accuracy
*ChatGPT
*Dentistry
*Oral Surgery

**** Corresponding Authors:***
Pardis Palizban (MD)
Address: *Anzali International Faculty of Dentistry, Guilan University of Medical Sciences, Guilan, Iran.*
Tel: *+989113810694*
E-mail*: pardispalizban@gmail.com*

1. Introduction

Artificial intelligence (AI) is a broad field encompassing various subfields, including machine learning (ML), deep learning, and natural language processing (NLP) (1). Machine learning, a subset of AI, enables computers to learn patterns from data and make predictions or decisions without explicit programming. ChatGPT an advanced Large language learning model (LLM) developed using machine learning techniques, specifically deep learning and transformer-based neural networks, is designed to generate human-like text responses based on the input it receives (2, 3).

Given its ability to process vast amounts of information, ChatGPT has been proposed as a tool for assisting clinicians in decision-making, medical education, and patient communication (4, 5). However, as ChatGPT is a relatively new technology, the accuracy and reliability of its responses remain uncertain (6), particularly in high-stakes medical fields such as surgery. A transformative shift regarding diagnostic, management strategy, communication with patients, and surgical training could be performed by incorporation of AI-powered ChatGPT based on a rapid analysis of existing databases in the field of oral and maxillofacial surgery (7).

Surgical decision-making includes complex, real-time judgments that directly impact patient outcomes. Since surgical procedures are among the most aggressive interventions, clinicians may need to conduct research and evaluate the steps and potential complications before performing the procedure (8). Incorrect or misleading information provided by ChatGPT can lead to severe consequences, including misdiagnoses, and inappropriate surgical planning, potentially leading to life-threatening outcomes for patients (9). On the other hand, surgical procedures are identified as the most stressful dental interventions for patients. One of the effective strategies to reduce patient anxiety is preoperative patient education (10). In the current age, with the advancement of artificial intelligence, patients may seek information through ChatGPT as an easily accessible resource.

The accuracy of AI-generated medical information poses potential risks that require thorough evaluation before widespread implementation in clinical practice. As AI becomes increasingly integrated into healthcare, assessing its reliability is essential, particularly in providing medical guidance to patients and serving as an educational tool for professionals. To address this issue, we conducted a current and updated review to evaluate the accuracy of ChatGPT's responses in oromaxillofacial surgery (OMS). This investigation examined the quality of AI-generated information provided to patients by analyzing responses to frequently asked questions and assessed the accuracy of its educational content using academically standardized questions. By exploring both aspects, this study offers clinicians a comprehensive overview of the reliability of AI-generated medical information in this field, helping them understand its strengths and limitations. These findings will aid healthcare professionals in making informed decisions about integrating AI tools into clinical practice and patient education.

2. Materials and Methods

Review of Literature

As shown in Figure 1, this review followed a systematic search pathway to literature selection and data synthesis. The study employed the PICO framework, a widely used methodology for structuring clinical research questions and guiding reviews. However, since this review did not involve a direct comparative analysis between different interventions or groups, the framework was adapted into the PIO format for the literature search (11). This modification allowed for a more targeted investigation of the accuracy of AI-generated medical information in OMS.

Patient/Problem (P): Individuals seeking information related to OMS, including patients, caregivers, and the general public who rely on AI-based platforms for medical guidance.

Intervention (I): The use of ChatGPT, an AI-driven language model, to generate responses to commonly asked questions in OMS, simulating real-world scenarios in which individuals seek medical advice from AI tools.

Outcome (O): An assessment of the accuracy, reliability, and comprehensiveness of the responses generated by ChatGPT when addressing OMS-

related inquiries, with a focus on the clinical relevance and potential implications for patient education and decision-making.

This study had the following focused question: How accurately can ChatGPT answer dental surgery questions?

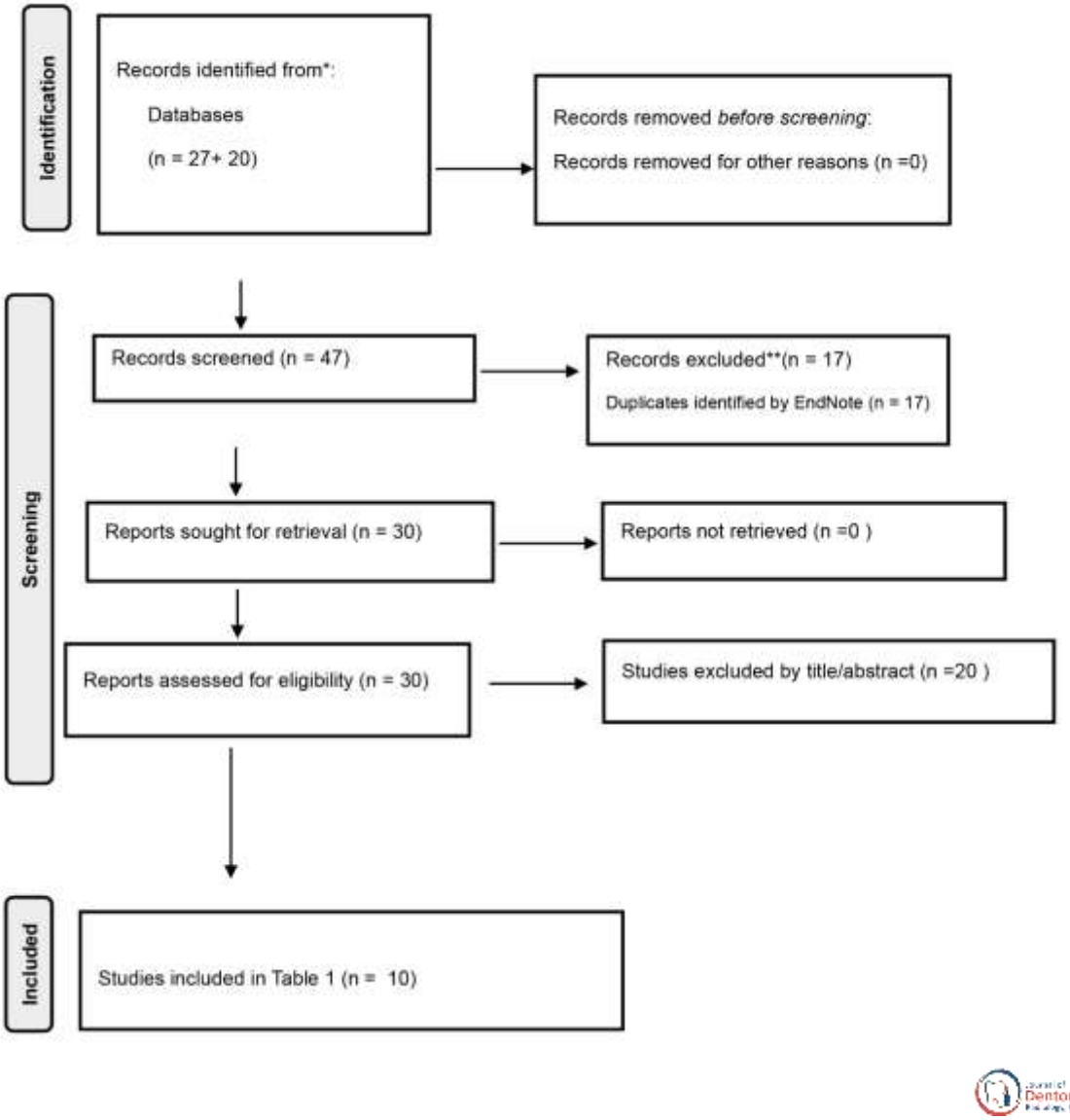

Figure. 1. Strategy of search for review

This study aims to evaluate the reliability, comprehensiveness, and clinical applicability of ChatGPT's responses when addressing various topics within the field of dental surgery. The assessment will consider multiple factors, including the correctness of factual information, alignment with current evidence-based guidelines, clarity of explanations, and potential risks associated with misinformation. Given the increasing reliance on AI-driven platforms for medical information, it is crucial to determine whether ChatGPT can serve as a trustworthy and accurate source for patients and healthcare professionals seeking guidance in dental surgery.

The inclusion criteria for this review were as follows: studies were included if they utilized ChatGPT to answer questions from either patients or clinicians and if they evaluated the accuracy of the responses using a Likert scale or by calculating the percentage of correct answers. Additionally,

only English-language articles were considered, and the questions posed to ChatGPT needed to be in English. These criteria ensured that the studies reviewed were relevant and standardized in terms of language and methods of assessing response accuracy.

Studies were excluded from this review based on several criteria. First, studies that did not utilize ChatGPT for answering patients' questions were excluded, as the focus was specifically on evaluating the accuracy of responses generated by this AI language model. Additionally, non-English articles were not considered, ensuring that only studies published in English were included to maintain consistency in language and interpretation. Studies where questions were posed to ChatGPT in non-English languages were also excluded, as this could introduce language-related variations in response accuracy. Furthermore, studies that included academic questionnaires or board examinations, where the results did not provide a separate statistical analysis for surgery-related questions, were excluded. This criterion ensured that the focus remained on dental surgery-related inquiries, without the inclusion of generalized academic assessments or unrelated topics.

To ensure a comprehensive review of the literature, the authors conducted an advanced search from inception to January 2025, on the PubMed database and a manual search on Web of Science to retrieve relevant studies using the following strategy:

PubMed: ("llms"[Title/Abstract] OR "ChatGPT"[Title/Abstract]) AND ("dental surgery"[Title/Abstract] OR "maxillofacial surgery"[Title/Abstract] OR "molar surgery"[Title/Abstract] OR "tooth extraction"[Title/Abstract] OR "tooth removal"[Title/Abstract])

Two independent reviewers were responsible for examining the titles, introductions, and study designs of all the articles included in the review. This process was conducted using Rayyan, a web-based software designed to streamline the systematic review process by allowing for efficient article screening and collaboration. To mitigate any potential bias in the review process, the “blind on” option in Rayyan was enabled, ensuring that the reviewers were unaware of each other’s decisions and thus reducing the likelihood of influencing each other’s assessments. This approach promoted an objective evaluation of the studies. In cases where the reviewers disagreed on the inclusion or exclusion of an article, they engaged in discussions to thoroughly evaluate the points of contention. Through these discussions, the reviewers worked collaboratively to reach a consensus, ensuring that all decisions were based on a shared understanding and in alignment with the established inclusion criteria. This process aimed to maintain the integrity and accuracy of the review while minimizing any potential biases or errors in the selection of studies.

3. Results

As a result of the comprehensive electronic search conducted across various databases, a total of 27 studies were initially included. Additionally, the manual search, which aimed to capture any relevant studies not found through the electronic search, yielded 20 more studies. After removing duplicate entries from both sources, 30 unique studies remained for further evaluation. Following this, the title and abstract screening process was carried out to assess the relevance of each study based on predefined inclusion criteria. This step led to the inclusion of 10 studies that met the basic requirements for further analysis (12-21).

Four studies were conducted in 2023 (13, 15-17), and six studies were carried out in 2024 (12, 14, 18-21). In three of these studies, academic-standard questions were posed to ChatGPT (12, 14, 18), while one other study involved patient’s frequently asked questions (15), and four studies included questions generated by the researchers themselves (13, 16, 19, 20). One study used both FAQs and questions generated by the researchers (21). One study used the website to select questions (17). Four studies utilized the Likert scale (14, 17, 19, 21), and four studies employed response rate percentages to assess the accuracy of ChatGPT's responses (12, 15, 18, 20). Two studies employed both of these scales to assess the accuracy of the response (13, 16; Table 1).

Table 1. Key characteristics of studies

Author (Year)	Generation of ChatGPT	Sample size	Question's type	Answer's type	Assessment scale
Acar (2023)	ChatGPT-4	20	Questions from website	Explanatory	Likert
Alsayed (2024)	ChatGPT-4	15	generated by the experts	explanatory	5 point-Likert
De Sousa (2023)	ChatGPT-3.5	10	Frequently Asked Questions	-	percentage
Işik G (2024)	ChatGPT Plus	66	academic-standard questions	Explanatory	7 point-Likert
Jacobs (2024)	ChatGPT-3.5	25	Post-operative frequently asked questions	Explanatory	5 point-Likert
Li (2024)	ChatGPT-3.5/4	25	generated by the experts	multiple choice questions	percentage
Mahmoud (2024)	ChatGPT-4o	714	academic-standard questions	multiple choice questions	Percentage
Quah B (2024)	ChatGPT-4	259	academic-standard questions	multiple choice questions	percentage
Suarez (2023)	ChatGPT-4	30	generated by the experts	Explanatory	- 3 point-Likert - percentage
Vaira (2024)	ChatGPT-4	144	generated by the experts	Binary(yes/no) And explanatory	- 6 point-Likert - percentage

4. Discussion

In recent years, numerous studies have mentioned the different applications of ChatGPT in the field of OMS. Several research efforts have demonstrated hopeful results in terms of the AI model's ability to provide scientifically accurate responses to frequently asked questions and clinical scenarios, particularly in the context of surgical procedures.

Recently, De Sousa et al. (15) conducted a study to evaluate the effectiveness of ChatGPT in answering frequently asked questions (FAQs) regarding third-molar tooth extraction, a common dental procedure. The study found that ChatGPT provided scientifically straight responses with a notable precision rate of 90.63%. This level of accuracy indicates that the AI model might be capable of delivering reliable and relevant information, making it a valuable resource for patients seeking information about the procedure, however further study was recommended. The concise and clear nature of the responses also contributes to their usability, as patients might be quickly access straightforward explanations without the need for complex or technical language. The study highlights the possibility of ChatGPT's potential as a user-friendly tool for patient education, particularly in addressing common concerns and providing understandable information on the procedural aspects, post-operative care, and expected outcomes of third-molar tooth extraction. This positive result might be able to give a position to ChatGPT as an accessible, efficient, and easily integrated tool in the patient education process within dental practice.

The results of another study conducted in 2023 by Vaira LA et al. (13) demonstrated a good level of accuracy in ChatGPT's responses. In this study, questions were divided into closed-ended and open-ended questions. The accuracy of responses to closed-ended questions was 84.7%, likewise, the 6-point-LIKERT scale used for open-ended questions indicated that AI's ability to process complex clinical scenarios is favourable (mean 5.2 ± 1.06). The study was suggested that it might be not yet a consistent tool for the decision-making process of clinicians in the field of maxillofacial surgery.

Işik et al. (14) conducted a study to assess the accuracy and quality of ChatGPT Plus's responses to questions based on the Clinical Practice Guide of Ege University, which covers a wide range of clinical scenarios. The researchers employed a 7-point Likert scale to evaluate the responses, with higher scores reflecting greater accuracy and reliability of the information provided. The results of the study highlighted that ChatGPT demonstrated a high level of accuracy, with a median accuracy score of 5 on the 7-point scale. This suggests that ChatGPT might be able to provide relevant and accurate information for most of the questions it addressed, demonstrating its potential as a valuable tool for clinical education and decision-making. However, the study also found that ChatGPT faced challenges in responding to questions that required more in-depth, detailed responses or critical analysis. These questions, which typically involve complex clinical scenarios or nuanced patient-specific factors, led to lower accuracy scores. Despite this limitation, the study concluded that ChatGPT Plus could be shown robust performance overall, particularly in providing concise and

accurate responses to general clinical questions. The findings suggested that while ChatGPT Plus can be a useful resource in clinical practice and education, it may still require refinement to fully address more complex and specialized queries.

In a comparative study, Quah et al. (12) ChatGPT-4 demonstrated greater performance compared to other language learning models (LLMs) when tasked with answering questions from a university's OMS multiple-choice question bank. ChatGPT-4 achieved a mean score of 76.8%, indicating a flattened level of accuracy and proficiency in addressing a range of OMS-related questions. The study found that the model's ability to process and generate relevant answers was particularly better than its counterparts, counting it as a hopeful tool for educational purposes in the field of OMS. Despite these positive results, the researchers concluded that, while LLMs like ChatGPT-4 can be effectively used as complementary tools in education, they are not yet sufficiently advanced to be relied upon for clinical decision-making. The study emphasized that further development is necessary to enhance the model's ability to handle the complexities of clinical practice, where the degrees of patient care require more detailed associated with context-specific knowledge

As such, ChatGPT and similar models should currently be seen as educational aids, assisting students and professionals in understanding concepts and preparing for exams, but not yet suitable for making independent clinical decisions.

In the study by Alsaed et al. (19), a 5-point Likert scale was used to assess the accuracy of ChatGPT's responses to surgical questions. The study found an average accuracy of 3.9, indicating a generally reliable performance. However, the accuracy varied based on the complexity of the surgical questions. For simpler procedures, such as those involving straightforward techniques, the accuracy was higher, with scores of 4/5 or 5/5. In contrast, for more complex surgical scenarios, which required nuanced or detailed responses, the accuracy dropped, with scores of 3/5. This variability underscores ChatGPT's ability to provide dependable answers for basic procedures but also highlights its limitations in addressing intricate surgical questions.

Jacobs et al. (21) conducted a study to assess the response accuracy of ChatGPT to frequently asked questions (FAQ) related to postoperative care following third molar surgery. The study incorporated a variety of patient inquiries, many of which were commonly found through search engines like Google, with additional questions provided by a practicing surgeon to enhance the clinical relevance of the study. Using a 5-point Likert scale to evaluate the responses, the study found that ChatGPT achieved a response accuracy of 4.36, indicating that the AI model demonstrated a great level of precision in addressing questions in this area of oral surgery. This result suggests that ChatGPT might be capable of providing reliable and relevant postoperative information for patients, making it a possibly respected tool for patient education and support. Despite its accuracy, the study also highlighted the importance of ongoing evaluation and development of AI models to ensure they continue to meet the developing needs of both patients and healthcare professionals in complex clinical scenarios.

In September 2023, Acar et al. (17) conducted an evaluation of ChatGPT's response accuracy concerning postoperative complications in oral surgery, utilizing a Likert scale to assess the model's performance. The questions used in the study were sourced from the Quora website, a platform with real-world patient inquiries and concerns related to oral surgery. The study found that ChatGPT consistently provided significantly accurate and relevant responses to these questions, indicating its potential as a reliable source of information for patients. These findings underscore the model's ability to deliver clear and accurate guidance on postoperative care, particularly in addressing common complications and concerns. The results suggest that ChatGPT might be able to play as a valuable role in patient education, offering accessible and trustworthy information to enhance patient understanding and aid in the management of their recovery process. However, the study also emphasizes the need for continuous evaluation and development to ensure the AI model remains accurate and up-to-date with the latest clinical guidelines.

In a 2024 study conducted by Mahmoud et al. (18) ChatGPT demonstrated a response accuracy of 83.7% when answering questions from the OMS

board examination. This high accuracy indicates that ChatGPT-4 has significant potential to serve as an educational aid in the OMS field, particularly in preparing students and practitioners for board exams. The study suggests that the AI tool can provide reliable information and support in the educational process, serving learners improve their knowledge and understanding of complex surgical concepts. Despite its promising performance, the study also notes that further advancements are needed for it to be more effectively integrated into clinical decision-making and practice.

In a study conducted in 2023, Suarez et al. (16) found that ChatGPT achieved a response accuracy of 71.7% when answering questions created by experts in education. The study concluded that while might be able to display possible as an auxiliary assistant in oral surgery, however, it is not suitable to replace clinicians. The results highlight ChatGPT's usefulness in supporting clinicians, but it requires further development before being relied upon for decision-making. Thus, its role remains supportive rather than replacing professional expertise.

In a comparative study by Li et al. in 2024 (20), students demonstrated higher accuracy when using ChatGPT to answer periodontal surgery questions, indicating its potential as a valuable educational tool. The study also found that ChatGPT supported students in their coursework and was helpful to practitioners in drafting clinical letters. While the results were promising, the study emphasized the need for further development for ChatGPT to fully support complex clinical decision-making. This suggests ChatGPT might be able to enhance both academic learning and clinical practice.

5. Conclusion

This review suggested that ChatGPT might be able to demonstrate a suitable appearance in answering oromaxillofacial questions, showing remarkable potential as a supplementary tool by cautious in both education and possibly clinical decision-making. Its ability to generate accurate, evidence-based responses to frequently asked questions and complex clinical scenarios positions as a respected candidate for healthcare professionals, especially in educational settings where it can assist in learning, exam preparation, and reviewing key concepts. Furthermore, the model's capacity to aid clinicians in providing standardized and reliable information can significantly streamline routine clinical tasks such as patient education, preoperative counselling, and postoperative care guidelines. However, its role in clinical decision-making remains in the early stages of development. While ChatGPT can provide useful insights and perhaps careful recommendations, its responses are based on pre-existing data and may not always account for the nuances of individual patient cases or the latest clinical advancements. As such, the model may be most effective when used alongside traditional clinical methods, rather than as an only tool for critical decision-making. In clinical practice, ChatGPT should be viewed as an adjunct, supporting expert judgment rather than replacing it. This collaborative approach could allow healthcare providers to improve the AI's strengths in managing information while still relying on their professional expertise to ensure the best patient outcomes.

Ethical Considerations

Not Applicable

Funding

This study did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors

Authors' Contributions

Pedram Hajibagheri: Conceptualization, Data curation, Methodology, Investigation, Writing-Original draft, Writing-review & editing **Monireh Aghajany-Nasab:** Conceptualization, Project administration, Writing-review & editing **Pardis Palizban:** Investigation, Supervision, Methodology

Conflict of Interests

The authors declare no conflict of interests.

Availability of data and material

Not applicable

Acknowledgments

- The authors used AI-assisted tools, including ChatGPT, for the writing process, grammar checking, and improving the clarity of the text. After using the tool, the authors reviewed and edited the content as necessary, taking full responsibility for the final publication.

- The authors declare that no funding was received for this study.

References

1. Zhou B, Yang G, Shi Z, Ma S. Natural language processing for smart healthcare. IEEE Reviews in Biomedical Engineering. 2022;17:4-18. [DOI: 10.1109/RBME.2022.3210270] [PMID]
2. Liu HY, Alessandri-Bonetti M, Arellano JA, Egro FM. Can ChatGPT be the plastic surgeon's new digital assistant? A bibliometric analysis and scoping review of ChatGPT in plastic surgery literature. Aesthetic Plastic Surgery. 2024;48(8):1644-52. [DOI: 10.1007/s00266-023-03709-0] [PMID]
3. Levin G, Brezinov Y, Meyer R. Exploring the use of ChatGPT in OBGYN: a bibliometric analysis of the first ChatGPT-related publications. Archives of Gynecology and Obstetrics. 2023;308(6):1785-9. [DOI: 10.1007/s00404-023-07081-x] [PMID]
4. Bizzo BC, Almeida RR, Michalski MH, Alkasab TK. Artificial Intelligence and Clinical Decision Support for Radiologists and Referring Providers. J Am Coll Radiol. 2019;16(9): 1351-6. [DOI: 10.1016/j.jacr.2019.06.010] [PMID]
5. Rajjoub R, Arroyave JS, Zaidat B, Ahmed W, Mejia MR, Tang J, et al. ChatGPT and its Role in the Decision-Making for the Diagnosis and Treatment of Lumbar Spinal Stenosis: A Comparative Analysis and Narrative Review. Global Spine J. 2024;14(3):998-1017. [DOI: 10.1177/21925682231195783] [PMID]
6. Helvaci BC, Hepsen S, Candemir B, Boz O, Durantas H, Houssein M, et al. Assessing the accuracy and reliability of ChatGPT's medical responses about thyroid cancer. Int J Med Inform. 2024;191:105593. [DOI: 10.1016/j.ijmedinf.2024.105593] [PMID]
7. Abdalla AQ, Aziz TA. ChatGPT: a game-changer in oral and maxillofacial surgery. Journal of Medicine, Surgery, and Public Health. 2024;2:100078. [DOI:10.1016/j.glmedi.2024.100078]
8. Niekrash C, Goupil MT. Surgical Complications. In: Ferneini EM, Goupil MT, editors. Evidence-Based Oral Surgery: A Clinical Guide for the General Dental Practitioner. Cham: Springer International Publishing. 2019. p. 205-21. [DOI: 10.1007/978-3-319-91361-2]
9. Kozaily E, Geagea M, Akdogan ER, Atkins J, Elshazly MB, Guglin M, et al. Accuracy and consistency of online large language model-based artificial intelligence chat platforms in answering patients' questions about heart failure. Int J Cardiol. 2024;408:132115. [DOI: 10.1016/j.ijcard.2024.132115] [PMID]
10. Ng SX, Wang W, Shen Q, Toh ZA, He HG. The effectiveness of preoperative education interventions on improving perioperative outcomes of adult patients undergoing cardiac surgery: a systematic review and meta-analysis. Eur J Cardiovasc Nurs. 2022;21(6):521-36. [DOI: 10.1093/eurjcn/zvab123] [PMID]
11. Forrest JL, Miller SA. Enhancing your practice through evidence-based decision making. Journal of Evidence Based Dental Practice. 2001;1(1):51-7. [DOI:10.1067/med.2001.116393]
12. Quah B, Yong CW, Lai CW, Islam I. Performance of large language models in oral and maxillofacial surgery examinations. Int J Oral Maxillofac Surg. 2024;53(10):881-6. [DOI: 10.1016/j.ijom.2024.06.003] [PMID]
13. Vaira LA, Lechien JR, Abbate V, Allevi F, Audino G, Beltramini GA, et al. Accuracy of ChatGPT-Generated Information on Head and Neck and Oromaxillofacial Surgery: A Multicenter Collaborative Analysis. Otolaryngol Head Neck Surg. 2024;170(6):1492-503. [DOI: 10.1002/ohn.489] [PMID]
14. Işik G, Kafadar-Gürbüz İ A, Elgün F, Kara RU, Berber B, Özgül S, et al. Is Artificial Intelligence a Useful Tool for Clinical Practice of Oral and Maxillofacial Surgery?. J Craniofac Surg. 2024 ;36(2):558-62. [DOI: 10.1097/SCS.0000000000010686] [PMID]
15. de Sousa RA, Costa SM, Figueiredo PH, Camargos CR, Ribeiro BC, Alves ESMRM. Is ChatGPT a reliable source of scientific information regarding third-molar surgery?. J Am Dent Assoc. 2024;155(3):227-32.e6. [DOI: 10.1016/j.adaj.2023.11.004] [PMID]
16. Suárez A, Jiménez J, de Pedro ML, Andreu-Vázquez C, García VD, Sánchez MG, et al. Beyond the Scalpel: Assessing ChatGPT's potential as an auxiliary intelligent virtual assistant in oral surgery. Comput Struct Biotechnol J. 2024;24:46-52. [DOI: 10.1016/j.csbj.2023.11.058] [PMID]
17. Acar AH. Can natural language processing serve as a consultant in oral surgery? Journal of Stomatology, Oral and Maxillofacial Surgery. 2024;125(3):101724. [DOI: 10.1016/j.jormas.2023.101724] [PMID]
18. Mahmoud R, Shuster A, Kleinman S, Arbel S, Ianculovici C, Peleg O. Evaluating Artificial Intelligence Chatbots in Oral and Maxillofacial Surgery Board Exams: Performance and Potential. J Oral Maxillofac Surg. 2025; 83(3):382-389. [DOI: 10.1016/j.joms.2024.11.007] [PMID]
19. Alsayed AA, Aldajani MB, Aljohani MH, Alamri H, Alwadi MA, Alshammari BZ, et al. Assessing the quality of AI information from ChatGPT regarding oral surgery, preventive dentistry, and oral cancer: An exploration study. Saudi Dent J. 2024;36(11):1483-9. [DOI: 10.1016/j.sdentj.2024.09.009] [PMID]
20. Li C, Zhang J, Abdul-Masih J, Zhang S, Yang J. Performance of ChatGPT and Dental Students on Concepts of Periodontal Surgery. European Journal of Dental Education. 2025;29(1):36-43. [DOI: 10.1111/eje.13047] [PMID]
21. Jacobs T, Shaari A, Gazonas CB, Ziccardi VB. Is ChatGPT an Accurate and readable patient aid for third molar extractions?. Journal of Oral and Maxillofacial Surgery. 2024;82(10):1239-45. [DOI: 10.1016/j.joms.2024.06.177] [PMID]