



ChatGPT-4.0 in Oral and Maxillofacial Radiology: Prediction of Anatomical and Pathological Conditions from Radiographic Images

Shila Kahalian¹, Marieh Rajabzadeh¹, Melisa Öçbe², Mahmut Sabri Medisoglu³

¹ Faculty of Dentistry, Kocaeli Health and Technology University, Kocaeli, Türkiye

² Department of Oral and Maxillofacial Radiology, Faculty of Dentistry, Kocaeli Health and Technology University, Kocaeli, Türkiye

³ Department of Anatomy, Faculty of Dentistry, Kocaeli Health and Technology University, Kocaeli, Türkiye

Corresponding author: Melisa Öçbe, Department of Oral and Maxillofacial Radiology, Faculty of Dentistry, Kocaeli Health and Technology University, 29 Yeniköy Mahallesi Ilıca Caddesi, 21275 Başiskele, Kocaeli, Türkiye; Email: melisabozkurt@windowslive.com, Tel.: +905332946233

Received: 26 August 2024 ♦ **Accepted:** 26 November 2024 ♦ **Published:** 31 December 2024

Citation: Kahalian S, Rajabzadeh M, Öçbe M, Medisoglu MS. ChatGPT-4.0 in oral and maxillofacial radiology: prediction of anatomical and pathological conditions from radiographic images. *Folia Med (Plovdiv)* 2024;66(6):863-868. doi: 10.3897/folmed.66.e135584.

Abstract

Introduction: ChatGPT has the ability to generate human-like text, analyze and understand medical images using natural Language processing (NLP) algorithms. It can generate real-time diagnosis and recognize patterns and learn from previous cases to improve accuracy by combining patient history, symptoms, and image characteristics. It has been used recently for learning about maxillofacial diseases, writing and translating radiology reports, and identifying anatomical landmarks, among other things.

Materials and methods: In this study, 52 radiographic images were queried on the OpenAI application ChatGPT-4.0. The responses were evaluated with and without using clues for specific radiographs to see if adding clues during prompting improved diagnostic accuracy.

Results: The true prediagnosis rate without any clue was 30.7%. By adding one clue this rate significantly increased to 56.9%. There was not a significant difference in accurate diagnosis of anatomical landmarks, cysts, and tumors ($p>0.05$). However, including internal structure information improved the diagnostic accuracy ($p<0.05$).

Conclusion: GPT-4.0 showed a tendency to misdiagnose closely located anatomical structures and by adding additional clues its performance showed improvement, while its ability to recognize diverse differential diagnoses remains limited.

Keywords

anatomical landmarks, artificial intelligence, ChatGPT, oral radiology

INTRODUCTION

ChatGPT (Chat Generative Pre-Trained Transformer) has lately become the most popular artificial intelligence (AI) chatbot developed by OpenAI (AIBY Inc., Miami, Florida).^[1] This model's core ability to generate human-like text by leveraging its understanding of contextual clues in a con-

versation. Since its introduction, the model has had updates with additional features and enhancements. ChatGPT-4.0, the latest version, has features such as voice interaction and image-based conversations. The image-based conversation feature can create image descriptions from hyperlinks when provided with specific prompts, which could be useful in the field of radiology.^[2,3]

ChatGPT has the ability to analyze and understand medical images using Natural Language Processing (NLP) algorithms, which is a major advantage in radiology.^[1-3] In radiology, this process involves recognizing key anatomical landmarks, identifying potential pathologies, and suggesting possible diagnoses.^[4-7] ChatGPT achieves this by associating visual patterns in the image with textual descriptions, allowing it to generate preliminary assessments.^[6,7] As it is exposed to more diverse cases and datasets, its accuracy in differentiating anatomical structures from abnormalities improves, making it a useful tool for supporting radiological evaluations.^[4] However, its performance in radiology depends on high-quality, diverse data for accurate diagnosis.^[7-9] Data privacy, limited access, and diverse protocols are some of the challenges while biased or limited data can have an impact on accuracy.^[10-12]

Given the rapid adoption of AI in healthcare, exploring ChatGPT's diagnostic potential in dentomaxillofacial radiology is essential.^[3] Understanding how well it may identify anatomical landmarks and pathologies not only helps evaluate its current capabilities but also informs future improvements needed. Previous studies demonstrated various benefits of using ChatGPT in the field of radiology including image reporting, text-based radiology exams, image interpretation, diagnostic performance.^[3-9] These possibilities are also increasing as the AI applications develop. Investigating these possibilities in radiology can provide important insights and establish its role in this field.

AIM

The aim of this study was to evaluate the performance of ChatGPT-4.0, specifically utilizing its image-based conversation feature. The study focuses on assessing the tool's ability to accurately identify anatomical landmarks, cysts, and tumors by highlighting these structures with arrows and providing relevant diagnostic clues.

MATERIALS AND METHODS

In this study, a total of 52 radiographic images (panoramic radiographies, periapical radiographies, cone-beam computed tomography sections) were obtained from the archive of the Department of Oral and Maxillofacial Radiology at the Health and Technology University in Kocaeli and analyzed using the OpenAI ChatGPT-4.0 application. Images depicting anatomical structures were selected randomly, while those used for cyst/tumor detection were chosen from cases with confirmed histopathological diagnoses. All images were collected from the archive between May and July 2024. The inclusion criteria required suboptimal image quality. The radiographs were classified based on 3 categories; anatomical landmarks (n=18), odontogenic and non-odontogenic cysts (n=11), and tumors (n=23).

Protocol for prompting and clue selection

The process for choosing clues was systematic, aimed at emulating clinical context by providing ChatGPT with pertinent anatomical and structural information to improve its interpretative capability. The clues were selected based on key diagnostic criteria commonly used by radiologists and were categorized into four types:

1. Location: Each radiograph included a clue about the anatomical location (e.g., “mandible,” “maxilla,” “anterior,” or “posterior”) to contextualize the image and aid ChatGPT in narrowing down diagnostic possibilities based on spatial orientation.

2. Internal structure: Clues about the internal characteristics of the lesion or structure were provided, describing radiopacity or radiolucency, the presence of an impacted tooth, ground glass appearance, or textural qualities (heterogeneous or homogeneous) within the region of interest. These descriptions were intended to guide ChatGPT's interpretation toward specific pathologies by highlighting relevant radiographic features.

3. Peripheral structure: Information on the borders of the lesion, such as “well-defined sclerotic,” “well-defined corticated,” or “ill-defined,” was added to indicate lesion boundary characteristics. This clue type helped distinguish between benign and malignant pathologies, as border definitions can be critical in radiological assessment.

4. Adjacent structures: Clues on proximity to critical anatomical landmarks (e.g., “close to mandibular canal,” “adjacent to maxillary sinus floor”) were included to provide context on nearby structures. This aimed to enhance ChatGPT's understanding of the lesion's spatial relationships, which are essential for accurate localization and diagnosis.

Each image underwent a standardized prompting sequence: initially presented without clues, followed by a second presentation incorporating the designated clues as outlined above.

After gathering all the answers provided by ChatGPT-4.0, one oral and maxillofacial radiologist and one anatomy specialist evaluated the answers as true or false. During the data collection process, responses were evaluated both with and without the inclusion of clues for certain specific radiographs to determine if adding clues during prompting resulted in improved diagnostic accuracy. The clues given to ChatGPT included location, internal structure (radiopacity/radiolucency, impacted tooth, ground glass appearance, heterogenous or homogenous), peripheral structure, and adjacent structures. Each anatomical structure or pathological condition image was presented twice to the ChatGPT: once without any clue and once with a clue.

Scoring and evaluation

After collecting ChatGPT's responses for each prompted image, an oral and maxillofacial radiologist and an anatomo-

my specialist independently evaluated the answers, determining their correctness based on the expected diagnosis. Responses were scored on a scale from 0 to 2 (Fig. 1). If ChatGPT provided multiple pre-diagnoses, 2 points were awarded if the correct diagnosis appeared as the first option, supporting the analysis of ChatGPT's primary interpretation accuracy. Scoring criteria can be found in Fig. 1.

Data were analyzed using IBM Statistical Package for the Social Sciences (SPSS) Statistics 22 (SPSS Inc., Chicago, IL, USA). Descriptive statistical methods for average prompt scoring were used to evaluate the data. The chi-square test was used to compare qualitative data. Statistical significance was set at $p < 0.05$.

Prompt Scoring Rubric (without clue)

The prediagnosis of the anatomical or pathological condition was not in the pre-diagnosis list	0
The prediagnosis of the anatomical or pathological condition was in the pre-diagnosis list	1
The anatomical or pathological structure prediction was true or in the first place in the list of pre-diagnosis	2

Figure 1. The prompt scoring rubric is depicted. Responses to the prompts were evaluated on a scale from 0 to 2. If multiple pre-diagnoses were suggested, 2 points were given if the correct diagnosis appeared first on the list.

RESULTS

The accuracy of the prediagnosis was measured, and the results were analyzed statistically. The true prediagnosis rate without any clue was 30.7%. When at least one clue was included, the incidence of true answers increased significantly to 56.9%. Average prompt score was found 9.3 for anatomical structures, 4.1 for odontogenic and non-odontogenic cysts, and 7.6 for odontogenic and non-odontogenic tumor pre-diagnosis. There was no statistically significant difference found between anatomical structure prediction, odontogenic and non-odontogenic cyst, and tumor pre-diagnosis ($p > 0.05$).

A statistical analysis was performed to determine the significance of the different types of clues. There was no statistically significant correlation found with the inclusion of a location, peripheral structure, radiopacity/radiolucency, and adjacent structures clues to the prompt ($p > 0.05$). However, mentioning the internal structure of the anatomical or pathological condition was found to be significant for achieving a true prediagnosis, indicating that this type of information is particularly effective in enhancing diagnostic accuracy ($p < 0.05$).

The analysis of false predictions made by ChatGPT revealed that errors often involved confusing adjacent anatomical structures. Specifically, nutrient canals of mandibular molars were frequently misdiagnosed as mandibular molar roots, the anterior nasal spine was misdiagnosed as

the incisive canal, and the greater palatine foramen was mistaken for the maxillary sinus. These misdiagnoses highlight the AI's challenge in distinguishing between closely situated structures, suggesting a need for improved contextual understanding and differentiation capabilities.

Additionally, the study found that querying about lesions associated with impacted teeth did not consistently lead to the accurate prediagnosis of a dentigerous cyst. However, when the presence of an impacted tooth was explicitly mentioned in the prompt, ChatGPT correctly suggested a dentigerous cyst but did not identify other potential pathologies such as adenomatoid odontogenic tumor, odontogenic myxoma, or ameloblastoma. This indicates that while the AI can recognize specific pathologies when given explicit cues, it may overlook other relevant differential diagnoses in the context of impacted teeth. Ameloblastoma, tooth socket, external oblique ridge, median palatal suture, genial tubercle, mandibular canal was detected by ChatGPT without giving any location, peripheral, internal or adjacent structure information. **Table 1** provides a summary of the questions and the false pre-diagnoses.

DISCUSSION

In this study, we evaluated the capability of ChatGPT-4.0 to diagnose dental radiographs using its image-based conversation feature, focusing on anatomical landmarks and lesions such as tumors and odontogenic and non-odontogenic cysts. Initially, without providing any clues, the prediagnosis rate was 30.7%. Subsequently, when we offered at least one clue, the rate increased to 56.9%. With additional hints, the average prompt scores increased.

In previous studies, Silva et al. evaluated ChatGPT-3.5's ability to describe radiolucent lesions in panoramic radiographs and establish differential diagnoses, finding that it correctly identified the diagnosis as the first hypothesis in 25% of cases, within the first two hypotheses in 57.14% of cases, and within the first three in 67.85% of cases, with 46% of responses including contraindications, indicating a reliance on general patterns rather than detailed evaluations.^[13] Likewise, Yani et al. assessed ChatGPT's potential for making diagnoses based on chief complaints and cone beam computed tomography radiologic findings. It achieved an overall accuracy score of 3.7 across 102 complex oral and maxillofacial cases, and its performance in generating pathological diagnoses for neoplastic/cystic diseases was less satisfactory.^[14]

Despite its improved performance with additional hints, the analysis of ChatGPT's diagnostic performance revealed several common misdiagnoses, often involving the confusion of closely situated anatomical structures.^[15,16] For instance, misidentification of nutrient canals of mandibular molars as molars, the anterior nasal spine as the incisive canal, and the greater palatine foramen as the maxillary sinus. So, this indicates a challenge in differentiating closely situated structures, emphasizing the need for better con-

Table 1. Structures or lesions queried without any location, internal structure, peripheral structure and adjacent structure information to ChatGPT and the pre-diagnosis

Anatomical structure or lesion	Pre-diagnosis
Nutrient canals (Mandible posterior)	Mandibular molar roots
Nutrient canals (Mandible anterior)	Mandibular fracture
Internal oblique ridge	Mandibular canal
Anterior nasal spine	Incisive canal
Greater palatine foramen	Maxillary sinus
Mandibular incisive foramen	Periapical lesion and a Stafne's bone cavity
Dentigerous cyst	Periapical lesion
Odontogenic keratocyst	Traumatic bone cyst and ameloblastoma
Buccal bifurcation cyst	Radicular cyst and dentigerous cyst
Radicular cyst (Maxilla anterior region)	Nasopalatine canal cyst
Odontoma	Dentigerous cyst
Odontogenic myxoma	Odontogenic keratocyst and osteoradionecrosis
Hypersementosis	Internal root resorption
Osteoma	Sialolithiasis and mandibular torus
Cementoblastoma	Periapical cemento-osseous dysplasia
Osteoid osteoma	Cemento-osseous dysplasia and ameloblastic fibro-odontoma
Osteoblastoma	Ameloblastic fibro-odontoma
Arteriovenous malformation	Stafne bone cavity
Idiopathic osteosclerosis	Condensing osteitis
Torus maxillaris	Internal root resorption and maxillary sinus floor
Calcifying epithelial odontogenic tumor	Dentigerous cyst and odontogenic keratocyst

textual understanding. Furthermore, while ChatGPT could accurately identify a dentigerous cyst when an impacted tooth was mentioned, it failed to consider other potential pathologies like adenomatoid odontogenic tumor or ameloblastoma, indicating a limitation in recognizing diverse differential diagnoses.

When comparing ChatGPT versions 3, 3.5, and 4.0 in predicting preliminary diagnoses based on radiological images, notable differences in performance emerge. ChatGPT-3, while capable of basic interpretation and clinical decision-making tasks, often lacked the nuance needed for accurate radiological predictions.^[15,17] ChatGPT-3.5 showed some improvement, but it tended to take a maximalist approach, frequently recommending more imaging modalities than necessary and failing to recognize cases where imaging would not be beneficial.^[17-19]

In the present study, ChatGPT-4.0's diagnostic performance was significantly influenced by the inclusion of specific clues. The true pre-diagnosis rate increased from 30.7% without clues to 56.9% with at least one clue, emphasizing the importance of contextual information in improving AI accuracy. Specifically, clues related to internal structure were found to be statistically significant ($p < 0.05$), underscoring their critical role in enhancing diagnostic precision. Albagieh et al. did not include diagnostic clues in their study design and found lower overall accuracy among large language models (LLMs), with a median accuracy score of 50%.^[19] Notably, LLMs failed entirely to identify complex cases, such as ectodermal dysplasia, which were correctly diagnosed by 85% of senior residents ($p = 0.011$). This suggests that AI models, including ChatGPT, may

struggle with complex, less common pathologies unless provided with additional context, reinforcing the current study's finding that diagnostic clues are essential for AI performance.

The current study highlighted ChatGPT-4.0's tendency to misdiagnose adjacent anatomical structures, revealing challenges in differentiating closely situated elements. For example, nutrient canals were mistaken for mandibular molar roots, while the anterior nasal spine was confused with the incisive canal. This points to the model's limitations in spatial recognition, suggesting a need for more advanced contextual understanding to improve AI accuracy in complex anatomical regions. Albagieh et al., on the other hand, reported moderate agreement among LLMs ($\kappa = 0.622$), which was higher than the weak agreement observed among senior residents ($\kappa = 0.396$). This suggests that while LLMs may offer more consistent responses across cases, they still lack the nuanced understanding that human clinicians can apply to diverse clinical scenarios.^[19] Additionally, the kappa values between resident and LLM responses were consistently low (around 0.4), indicating discrepancies between AI-generated and clinician-generated diagnoses.

The present study found that ChatGPT-4.0's diagnostic accuracy varied across different case types, with an average prompt score of 9.3 for anatomical structures, 4.1 for odontogenic cysts, and 7.6 for tumors. This suggests that while the model performs relatively well in anatomical identification, it has limitations in diagnosing pathologies, particularly when dealing with complex cases like odontogenic cysts and tumors. Similarly, Albagieh et al. observed signif-

icant performance divergence in specific cases, such as the management of a large painful ulcer in a post-kidney transplant patient, where LLMs uniformly favored topical corticosteroids, contrasting with 70% of residents who chose intralesional injections ($p=0.022$). These findings further indicate that AI models, including ChatGPT, may adhere to general treatment patterns but often miss the clinical nuance needed for accurate management decisions.^[15,16,19] In addition, ChatGPT-4.0 demonstrated a significant improvement in the interpretations of radiological images. Despite the improvements, as the latest version of OpenAI, ChatGPT-4.0 is still incompetent in preliminary diagnosis of dentomaxillofacial lesions and detection of anatomical structures, as this study revealed.^[20]

Several limitations can be considered for this study. First, this study used 52 radiographic images from a single institution, limiting the generalizability of the results. The dataset may not represent the full range of anatomical and pathological conditions, especially rare cases. Image selection based on suboptimal quality could introduce bias, affecting AI performance with higher-quality or more diverse images. Diagnostic accuracy depended on including specific prompts, which may affect reproducibility in other clinical scenarios. Additionally, the study was conducted under controlled conditions, lacking real-world clinical validation where factors like patient history and clinician judgment influence diagnostics. Finally, ChatGPT-4.0's ability to generate a comprehensive list of differential diagnoses, especially for complex cases, was not fully assessed.

CONCLUSIONS

The findings of this study indicate that ChatGPT-4.0 in its current iteration lacks sufficient efficiency in accurately detecting oral and maxillofacial pathologies. Although incorporating specific clues into the prompt did enhance the ChatGPT-4.0's ability to generate a more comprehensive and accurate preliminary diagnosis list, the results revealed that its diagnostic capabilities remain limited.

Consequently, further development and refinement are necessary to improve the ChatGPT-4.0's diagnostic accuracy and usability before it can be considered a viable option for patient use. Future research should also explore integrating more advanced algorithms and expanding the dataset to encompass a broader range of pathologies, which could ultimately contribute to the ChatGPT-4.0's evolution into a more reliable diagnostic aid in the field of oral and maxillofacial radiology.

REFERENCES

1. Jeong H, Han SS, Yu Y, et al. How well do large language model-based chatbots perform in oral and maxillofacial radiology? *Dentomaxillofacial Radiology*. 2024; 53(6):390–5.
2. Srivastav S, Chandrakar R, Gupta S, et al. ChatGPT in radiology: the

advantages and limitations of artificial intelligence for medical imaging diagnosis. *Cureus* 2023; 15(7):e41435

3. Bhayana R, Bleakney RR, Krishna S. GPT-4 in radiology: improvements in advanced reasoning. *Radiology* 2023; 307(5):e230987.
4. Lecler A, Duron L, Soyer P. Revolutionizing radiology with GPT-based models: current applications, future possibilities, and limitations of ChatGPT. *Diagn Interv Imaging* 2023; 104:269–74.
5. Ueda D, Mitsuyama Y, Takita H, et al. ChatGPT's diagnostic performance from patient history and imaging findings on the Diagnosis Please quizzes. *Radiology* 2023; 308(1). doi: 10.1148/radiol.231040
6. Horiuchi D, Tatekawa H, Shimono T, et al. Accuracy of ChatGPT-generated diagnosis from patient's medical history and imaging findings in neuroradiology cases. *Neuroradiology* 2024; 66(1):73–9.
7. Kottlors J, Bratke G, Rauen P, et al. Feasibility of differential diagnosis based on imaging patterns using a large language model. *Radiology* 2023; 308(1):e231167.
8. Adams LC, Truhn D, Busch F, et al. Leveraging GPT-4 for post hoc transformation of free-text radiology reports into structured reporting: a multilingual feasibility study. *Radiology* 2023; 307(4):e230725.
9. Gertz RJ, Bunck AC, Lennartz S, et al. GPT-4 for automated determination of radiological study and protocol based on radiology request forms: a feasibility study. *Radiology* 2023; 307(5):e230877.
10. Lyu Q, Tan J, Zapadka ME, et al. Translating radiology reports into plain language using ChatGPT and GPT-4 with prompt learning: results, limitations, and potential. *Vis Comput Ind Biomed Art* 2023; 6(1):9.
11. Mago J, Sharma M. The potential usefulness of ChatGPT in oral and maxillofacial radiology. *Cureus* 2023; 15(7):e42133. doi: 10.7759/cureus.42133
12. Jeblick K, Schachtner B, Dexl J, et al. ChatGPT makes medicine easy to swallow: an exploratory case study on simplified radiology reports. *Eur Radiol* 2024; 34(5):2817–25.
13. Silva TP, Andrade-Bortoletto MF, Ocampo TS, et al. Performance of a commercially available Generative Pre-trained Transformer (GPT) in describing radiolucent lesions in panoramic radiographs and establishing differential diagnoses. *Clin Oral Investig* 2024; 28(3):204.
14. Hu Y, Hu Z, Liu W, et al. Exploring the potential of ChatGPT as an adjunct for generating diagnosis based on chief complaint and cone beam CT radiologic findings. *BMC Medical Informatics and Decision Making* 2024; 24(1):55.
15. Rao A, Kim J, Kamineni M, et al. Evaluating GPT as an adjunct for radiologic decision making: GPT-4 versus GPT-3.5 in a breast imaging pilot. *J Am College Radiol* 2023; 20(10):990–7.
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. Shen Y, Heacock L, Elias J, et al. ChatGPT and other large language models are double-edged swords. *Radiology* 2023; 307(2):e230163.
18. Vaishya R, Iyengar KP, Patralekh MK, et al. Effectiveness of AI-powered Chatbots in responding to orthopaedic postgraduate exam questions - an observational study. *International Orthopaedics* 2024; 15:1–7.
19. Albagieh H, Alzeer ZO, Alasmari ON, et al. Comparing artificial intelligence and senior residents in oral lesion diagnosis: a comparative study. *Cureus* 2024; 16(1):e51584. doi:10.7759/cureus.51584
20. OpenAI. What is ChatGPT? OpenAI. Available from: <https://help.openai.com/en/articles/6783457-chatgpt-general-faq>. Accessed February 1, 2023.

ChatGPT-4.0 в рентгенологии полости рта и челюстно-лицевой области: прогнозирование анатомических и патологических состояний по рентгенографическим изображениям

Шила Кахалиан¹, Марие Раджабзадех¹, Мелиса Йочбе², Махмут Сабри Медисоглу³

¹ Факультет дентальной медицины, Университет здоровья и технологий Коджаэли, Коджаэли, Турция

² Кафедра оральной и челюстно-лицевой рентгенографии, Факультет дентальной медицины, Университет здоровья и технологий Коджаэли, Коджаэли, Турция

³ Кафедра анатомии, Факультет дентальной медицины, Университет здоровья и технологий Коджаэли, Коджаэли, Турция

Адрес для корреспонденции: Мелиса Йочбе, Кафедра оральной и челюстно-лицевой рентгенографии, Факультет дентальной медицины, Университет здоровья и технологий Коджаэли, 29 Yeniköy Mahallesi Ilca Caddesi квартал Йеникёй, ул. „Кадеси“ № 26, 21275 Башпискеле, Коджаэли, Турция; Email: melisabozkurt@windowslive.com, тел.: +905332946233

Дата получения: 26 августа 2024 г. ♦ **Дата приемки:** 26 ноября 2024 г. ♦ **Дата публикации:** 31 декабря 2024 г.

Образец цитирования: Kahalian S, Rajabzadeh M, Öçbe M, Medisoglu MS. ChatGPT-4.0 in oral and maxillofacial radiology: prediction of anatomical and pathological conditions from radiographic images. Folia Med (Plovdiv) 2024;66(6):863-868. doi: 10.3897/folmed.66.e135584.

Резюме

Введение: ChatGPT может генерировать текст, похожий на человеческий, анализировать и понимать медицинские изображения с помощью алгоритмов обработки естественного языка (NLP). Он может генерировать диагностику в реальном времени и распознавать закономерности и учиться на предыдущих случаях для повышения точности путём объединения истории болезни пациента, симптомов и характеристик изображений. В последнее время он использовался нами для изучения челюстно-лицевых заболеваний, написания и перевода отчётов по рентгенологии и определения анатомических ориентиров, среди прочего.

Материалы и методы: В этом исследовании 52 рентгенологических изображения были запрошены в приложении OpenAI ChatGPT-4.0. Ответы оценивались с использованием и без использования подсказок для определённых рентгенограмм, чтобы увидеть, улучшает ли добавление подсказок во время поиска точность диагностики.

Результаты: Истинный показатель предварительной диагностики без какой-либо подсказки составил 30.7%. При добавлении одной подсказки этот показатель значительно увеличился до 56.9%. Не было значительной разницы в точности диагностики анатомических ориентиров, кист и опухолей ($p > 0.05$). Однако включение информации о внутренней структуре повысило точность диагностики ($p < 0.05$)

Вывод: GPT-4.0 продемонстрировал тенденцию к неправильной диагностике близко расположенных анатомических структур, а при добавлении дополнительных подсказок его производительность улучшилась, в то время как его способность распознавать различные дифференциальные диагнозы остается ограниченной.

Ключевые слова

анатомические ориентиры, искусственный интеллект, ChatGPT, рентгенография полости рта