Barw Medical Journal | 2024;2(1):2-6
Journal home page: https://barw.krd/index.php/BMJ

Original Article

Comparative Analysis of ChatGPT and Human Decision-Making in Thyroid and Neck Swellings: A Case-Based Study

Zardasht Mahmud Ahamed[1], Hardi Mohammed Dhahir[2], Miran Mawlood Mohammed[3], Rebaz Haji Ali[4], Shko H. Hassan[5], Aso S. Muhialdeen[5*], Yadgar Abdulhameed Saeed[6], Mariwan L. Fatah[7], Aras J. Qaradakhy[8], Rawa M. Ali[5], Shaho F. Ahmed[5], Ari M. Abdullah[9], Hawbash M. Rahim[10], Berun A. Abdalla[11], Abdulwahid M. Salih[5,7], Shvan H. Mohammed[12], Fahmi H. Kakamad[5,7]

1. *Hospital of Treatment Victims Chemical Weapon in Halabja, Halabja, Kurdistan, Iraq*
2. *Shahid Peshraw Hospital, Chamchamal, Sulaimani, Kurdistan, Iraq*
3. *Shar Hospital, Sulaimani, Kurdistan, Iraq*
4. *Hiwa Cancer Hospital, Sulaimani Directorate of Health, Kurdistan, Iraq*
5. *Smart Health Tower, Madam Mitterrand Street, Sulaimani, Kurdistan, Iraq*
6. *Baxshin Hospital, Malik Mahmud Street, Sulaimani, Kurdistan, Iraq*
7. *College of Medicine, University of Sulaimani, Sulaimani, Kurdistan, Iraq*
8. *Department of Radiology, Shorsh Teaching Hospital, Sulaimani, Kurdistan, Iraq*
9. *Sulaimani Teaching Hospital, Sulaimani, Kurdistan, Iraq*
10. *Department of Medical Laboratory Science, College of Health Sciences, University of Human Development, Kurdistan, Sulaimani, Iraq*
11. *Kscien Organization for Scientific Research (Middle East office), Kurdistan, Sulaimani, Iraq*
12. *Chara Laboratory, Shahedan Street, Kalar, Kurdistan, Iraq*

[*] ***Corresponding author***: drasosaeed@gmail.com *(A.S. Muhialdeen). Kaziway 1, House number 306/7, Zip code: 46001, Sulaimani, Iraq*

Keywords:
ChatGPT
Nodules
Head
Neck
AI

Received: October 3, 2023
Revised: October 14, 2023
Accepted: October 20, 2023
First Published: October 24, 2023

Abstract

Introduction

This study aimed to evaluate the performance of Chat Generative Pre-Trained Transformer (ChatGPT), an AI-powered chatbot, in providing treatment recommendations for head and neck nodules.

Methods

Ten diverse cases were examined, including individuals with varying ages and conditions such as thyroid nodules and suspicious neck lesions. The decisions made by ChatGPT were compared to those of physicians. Data were collected from the Smart Health Tower on May 2, 2023.

Results

Analysis of the cases revealed that ChatGPT provided recommendations that aligned with physicians' decisions in seven cases. However, disparities were observed in three cases (N5, N8, N10). Discrepancies were influenced by factors such as nodule size, thyroid function, and the presence of associated symptoms.

Conclusion

Citation: Ahmed ZM, Dhahir HM, Mohammed MM, Ali RH, Hassan SH, Muhialdeen AS et al. Comparative Analysis of ChatGPT and Human Decision-Making in Thyroid and Neck Swellings: A Case-Based Study. Barw Medical Journal. 2024;2(1):2-6. https://doi.org/10.58742/bmj.v1i2.43

The findings suggest that ChatGPT can assist healthcare professionals in formulating individualized diagnoses and treatment strategies for head and neck nodules. However, further research is needed to validate the performance of ChatGPT in larger patient populations and assess its long-term impact on patient outcomes.

1. Introduction

Artificial intelligence (AI) is revolutionizing the healthcare industry by providing new ways to improve patient care and outcomes [1]. Chat Generative Pre-Trained Transformer (ChatGPT) is a cutting-edge natural language processing (NLP) model developed by OpenAI and released in November 2022, which can generate nearly human-quality responses for a wide range of tasks [2,3,4]. ChatGPT has gained significant popularity and garnered attention worldwide [4]. Its exceptional capability to simulate human-like expression and reasoning has attracted over 100 million users [5]. Additionally, it has been evaluated for its accuracy and reliability in providing medical responses [2,3].

One of the key advantages of ChatGPT in medical decision-making is its ability to analyze and interpret complex medical information (1,6). It can assist in reviewing patient data, medical literature, and guidelines, enabling healthcare professionals to access relevant information quickly. This can be particularly helpful in situations where time is critical or when dealing with rare conditions where expertise may be limited [6,7].

ChatGPT's unbiased nature is another crucial factor in medical decision-making. By relying on algorithms and data-driven processes, it has the potential to mitigate human bias and ensure more objective recommendations. This can be especially beneficial in sensitive areas where unconscious bias may influence decision-making, such as in the assessment of patient risk factors or treatment options [6–8].

Moreover, ChatGPT can assist in the differential diagnosis process by considering various symptoms, medical history, and risk factors. It can provide a comprehensive analysis of the available information and generate a list of potential diagnoses, aiding healthcare professionals in their decision-making process [6]. This can improve diagnostic accuracy and reduce the likelihood of missed or delayed diagnoses [8,9].

The integration of ChatGPT into medical decision-making processes has garnered attention across various medical specialties. Several specialties have published articles highlighting the role and potential applications of ChatGPT in their respective fields [10–15].

Two key challenges arise from using ChatGPT in healthcare. Firstly, ChatGPT cannot replace healthcare providers, risking potential harm to patients due to inaccurate or misleading information. Secondly, healthcare professionals may overly rely on ChatGPT, impeding their independent ability to diagnose and treat patients without relying solely on technology [16].

The aim of this study is to analyze the similarities and differences between human and AI decision-making, shedding light on the potential benefits and limitations of incorporating AI technology in healthcare decision-making processes in a case-based fashion.

2. Methods

Data were collected from the Smart Health Tower, a hospital equipped with a multidisciplinary team responsible for discussing cases and making appropriate decisions. The team consists of radiologists, oncologists, pathologists, and surgeons. Data collection took place on May 2, 2023.

3. Results

This study included ten cases: five males and five females, ranging from a 5-year-old female to an 80-year-old male. These cases covered a range of conditions, including thyroid nodules and suspicious neck lesions.

The analysis of the ten cases revealed that ChatGPT provided recommendations that aligned with physicians' decisions in seven cases (70%), while there were disparities in three cases (30%) (N5,8,10). Full details of the cases, ChatGPT's decisions, and the physicians' final decisions are provided in (Table 1).

4. Discussion

This study investigated the use of ChatGPT for recommending management of head and neck nodules and found that the AI-powered chatbot was able to generate recommendations that were in line with the decisions made by physicians in seven of the 10 cases. This suggests that ChatGPT has the potential to be a valuable tool for clinicians as it can provide them with an additional source of information and guidance when making treatment decisions.

ChatGPT, a language model developed by OpenAI, is indeed still under development. OpenAI has released various iterations of the model, with improvements and updates being made over time. The most recent version, as of July 2023, is ChatGPT-4 [5]. However, OpenAI continues to work on refining and expanding the capabilities of ChatGPT to address its limitations and improve its performance in the next version (GPT-5), which introduces improved task-specific features such as abstracting text, providing answers to queries, and interpreting natural language. Additionally, GPT5 might utilize methods derived

Table 1. Full description of the cases and corresponding recommendations.

Case description	Chat GPT recommendation	Physician's decision	Agreements
A 5-year-old female presented with left-sided neck swelling from birth. An old ultrasound (US) showed intramuscular (left sternocleidomastoid muscle) hematoma, and a follow-up US showed enlarged sternocleidomastoid muscle with few small, reactionary lymph nodes. Hematological tests were normal.	Magnetic resonance imaging (MRI) and follow-up.	Magnetic resonance imaging (MRI) and follow-up.	100% agreed.
A 42-year-old female presented with incidental thyroid nodules. US showed bilateral suspicious thyroid nodules (TR3), and fine needle aspiration cytology (FNAC) was reported as suspicious for follicular neoplasm. The patient also had elevated antithyroid peroxidase antibody (ATPO).	Total thyroidectomy	Total thyroidectomy	100% agreed.
A-51-year-old male presented with swelling of the left temporal region for one year. Computed tomography (CT) scan of the head showed a heterogenous lobulated mass around the coronoid process of the mandible measuring 32 x 34 mm. FNA cytology was inconclusive.	Revision of the imaging and consideration for re-biopsy.	Revision of the imaging and consideration for re-biopsy and recommended an excisional biopsy.	100% agreed.
An-80-year-old male presented with an incidental finding of a thyroid nodule. US showed a large lobulated 33 x 17 x 34 mm nodule in the right thyroid lobe with micro and macrocalcifications in addition to 3-mm nodules with microcalcification in the left thyroid lobe and right-sided cervical lymphadenopathy (LAP) in group II, the largest one measuring 12 x 20mm. FNA cytology of the right thyroid lobe nodule and the largest lymph node was suspicious for papillary thyroid carcinoma.	Total thyroidectomy with right-side neck dissection.	Total thyroidectomy with right-side neck dissection.	100% agreed.
A-29 female presented with anterior neck swelling for 18 months. US showed a 13-mm nodule in the thyroglossal duct cyst, suspicious for thyroglossal duct carcinoma. The thyroid was normal in the US. No cervical lymphadenopathy was detected. FNA cytology of the thyroglossal duct cyst was suspicious for papillary carcinoma of the thyroglossal duct.	Thyroglossal duct resection with total thyroidectomy.	the nodule was less than 2 cm, and the thyroid was normal. They recommended resection of the thyroglossal cyst with regular follow-up of the thyroid.	50% agreed.
A 39-year-old male patient presented with the incidental finding of a thyroid nodule and a past surgical history of the removal of an interscapular sarcoma. Neck US revealed bilateral thyroid cystic nodules of less than 8 mm. The right thyroid lobe showed a highly suspicious lesion of malignancy measuring 6 x 5 x 4 mm without cervical LAP. FNA cytology of the suspicious nodule showed papillary thyroid carcinoma.	Thyroid lobectomy or total thyroidectomy	Total thyroidectomy	50% agreed.
A-41-year-old female patient presented with anterior neck swelling for four years. The first neck US showed multiple thyroid nodules, the largest being 26 x 12 mm and containing macrocalcification. FNA cytology was reported as a benign follicular nodule. The last US revealed multiple bilateral TR3 nodules, largest 31 x 30 mm. Thyroid function tests were within normal limits.	FNA	Re-FNA +- Surgery	100% agreed.
A-42-year-old male patient presented with left side neck swelling for two months. Neck US showed bilateral TR3 nodules, largest 27 x 24 mm in the mid-third of left thyroid lobe. FNA cytology was reported as suspicious for follicular neoplasm. The patient was euthyroid.	No specific recommendation (general information regarding consulting an endocrinologist, reviewing the investigations, and assessing the risk factors	Lt thyroid lobectomy +/-	No agreement

Table 1. Continued…

A 66-year-old female patient who had undergone thyroid surgery 35 years ago now had a new neck US which showed a large remnant of thyroid tissue with bilateral TR3 nodules and a 39 x 26-mm TR4 nodule. FNA cytology of the TR4 nodule was reported as a benign follicular nodule. The patient was hypothyroid and on a daily thyroxine dose. Ear, nose, and throat check was normal.	Follow up	Follow up	100% agreed.
A-50-year-old male presented with left lumber pain for a one year. Abdominal ultrasound showed a left-sided renal stone of 6 x 6 mm. Blood investigations revealed a parathyroid hormone (PTH) level of 551 pg/mL and serum calcium level of 11.7 mg/dL. Neck US showed a large multinodular goiter but no parathyroid nodule or lesion. There was no family history of renal stones	Sestamibi scan + FNA from thyroid nodule if suspicious nodule was found; if not, only follow up for thyroid.	Sestamibi scan + total thyroidectomy.	50% agreed.

from image, audio, and video recognition, enhancing its versatility significantly. ChatGPT, or any AI-based system, is not currently capable of fully replacing physicians. While AI and machine learning technologies have made significant advancements in recent years, they still have limitations and cannot replace the complex decision-making, empathy, and experience of human physicians[17] . AI tools like ChatGPT can assist physicians in tasks such as diagnosing diseases, analyzing medical images, and providing treatment recommendations, but they are meant to complement, not replace, human expertise [18]. Moreover, ethical and legal concerns surrounding AI in healthcare also need to be addressed before any potential replacement can be considered [19].

The comparison between ChatGPT recommendations and physicians' decisions revealed a substantial level of agreement in most cases. However, there were instances where disagreements occurred, primarily driven by factors such as nodule size, thyroid functionality, and the presence of associated symptoms. It is important to note that these recommendations were generated based on the available information provided, and there may be additional clinical considerations that influence the final decision-making process.

ChatGPT has been evaluated for its performance on the United States Medical Licensing Exam (USMLE), which showed an accuracy of nearly 60% (20). ChatGPT also showed good performance with many other medical exams [15,21]. For thyroid nodules, AI integration with ultrasound modalities in dynamic AI has demonstrated high diagnostic value for both benign and malignant thyroid nodules. Its accuracy, sensitivity, and specificity were consistently high, surpassing preoperative ultrasound American College of Radiology (ACR) Thyroid Imaging Reporting and Data System (TI-RADS) in several aspects [20]. Dynamic AI was approaching fine-needle aspiration cytology (FNAC) in terms of sensitivity and accuracy. These findings highlight the potential of dynamic AI as an effective tool in improving patient care for thyroid nodules [22,23]. Our study only included 10 cases, which is a relatively small sample size. This means that the study's results may not be generalizable to a larger population of patients. Moreover, it did not include any long-term follow-up of the patients. This means that it is not possible to know patients' long-term outcomes. Real-world validation studies are essential to determine the impact of AI-driven recommendations on patient outcomes and healthcare provider decision-making. Research efforts should be directed towards designing user-friendly interfaces, facilitating seamless integration into existing healthcare workflows, and conducting longitudinal studies to assess long-term benefits and challenges associated with AI integration.

5. Conclusion

There was a significant level of agreement between the ChatGPT recommendations and the decisions made by physicians. This highlights the potential of AI systems to assist healthcare professionals in formulating individualized diagnoses and treatment strategies for patients with thyroid nodules.

Declarations

Conflicts of interest: The author(s) have no conflicts of interest to disclose.

Ethical approval: Not applicable.

Patient consent (participation and publication)**:** Not applicable.

Funding: The present study received no financial support.

Acknowledgements: None to be declared.

Authors' contributions: FHK and AMS major contribution of the idea. ZMA, HMD, MMM, RHA, HRA, SHH, ASM and YAS participated in data collection and interpretation. HMR and BAA designed the study. FHK and HMR participated in preparing the manuscript; MLF, AJQ, RMA, AMA and SHM critically revised the manuscript and preparation of the table; all authors approved the final version of the manuscript.

Data availability statement: Note applicable.

References

1. Amisha, Malik P, Pathania M, Rathaur VK. Overview of artificial intelligence in medicine. J Fam Med Prim care. 2019 Jul;8(7):2328–31. doi:10.4103/jfmpc.jfmpc_440_19
2. Gilson A, Safranek CW, Huang T, Socrates V, Chi L, Taylor RA, et al. 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 Feb;9: e45312. doi:10.2196/45312
3. Ayers JW, Zhu Z, Poliak A, Leas EC, Dredze M, Hogarth M, et al. Evaluating Artificial Intelligence Responses to Public Health Questions. JAMA Netw Open [Internet]. 2023 Jun 7;6(6):e2317517–e2317517. Available from: doi:10.1001/jamanetworkopen.2023.17517
4. ChatGPT sets record for fastest-growing user base-analyst note. Available from: https://www.marketscreener.com/news/latest/ChatGPT-sets-record-for-fastestgrowing-user-base-analyst-note--42873811/.
5. ChatGPT — Release Notes | OpenAI Help Center [Internet]. [cited 2023 Jun 25]. Available from: https://help.openai.com/en/articles/6825453-chatgpt-release-notes.
6. Dave T, Athaluri SA, Singh S. ChatGPT in medicine: an overview of its applications, advantages, limitations, future prospects, and ethical considerations [Internet]. Vol. 6, Frontiers in Artificial Intelligence 2023. doi:10.3389/frai.2023.1169595
7. Khan RA, Jawaid M, Khan AR, Sajjad M. ChatGPT - Reshaping medical education and clinical management. Pakistan J Med Sci. 2023;39(2):605–7. doi:10.12669/pjms.39.2.7653
8. Hirosawa T, Harada Y, Yokose M, Sakamoto T, Kawamura R, Shimizu T. Diagnostic Accuracy of Differential-Diagnosis Lists Generated by Generative Pretrained Transformer 3 Chatbot for Clinical Vignettes with Common Chief Complaints: A Pilot Study. Int J Environ Res Public Health. 2023 Feb;20(4). doi:10.3390/ijerph20043378
9. Liao Z, Wang J, Shi Z, Lu L, Tabata H. Revolutionary Potential of ChatGPT in Constructing Intelligent Clinical Decision Support Systems. Annals of biomedical engineering. United States; 2023. doi:10.1007/s10439-023-03288-w
10. 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(6):269–74. doi:10.1016/j.diii.2023.02.003
11. Balas M, Ing EB. Conversational AI Models for ophthalmic diagnosis: Comparison of ChatGPT and the Isabel Pro Differential Diagnosis Generator. JFO Open Ophthalmol. 2023; 1:100005. doi:10.1016/j.jfop.2023.100005
12. Li W, Zhang Y, Chen F. ChatGPT in Colorectal Surgery: A Promising Tool or a Passing Fad? Ann Biomed Eng. 2023; Available from: doi:10.1007/s10439-023-03232-y.
13. Ma Y. The potential application of ChatGPT in gastrointestinal pathology. Gastroenterol Endosc. 2023;1(3):130–1. doi:10.1016/j.gande.2023.05.002
14. Sinha RK, Deb Roy A, Kumar N, Mondal H. Applicability of ChatGPT in Assisting to Solve Higher Order Problems in Pathology. Cureus. 2023; 15(2): e35237. doi:10.7759/cureus.35237
15. Giannos P. Evaluating the limits of AI in medical specialisation: ChatGPT's performance on the UK Neurology Specialty Certificate Examination. BMJ Neurol Open. 2023; 5(1): e000451. doi:10.1136/bmjno-2023-000451
16. Baumgartner C. The potential impact of ChatGPT in clinical and translational medicine. Clin Transl Med. 2023; 13(3): e1206. doi:10.1002/ctm2.1206
17. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. Nat Med. 2019;25(1):44–56. doi:10.1038/s41591-018-0300-7
18. Jiang F, Jiang Y, Zhi H, Dong Y, Li H, Ma S, et al. Artificial intelligence in healthcare: past, present and future. Stroke Vasc Neurol. 2017; 2(4):230 LP – 243. doi:10.1136/svn-2017-000101
19. Mittelstadt BD, Allo P, Taddeo M, Wachter S, Floridi L. The ethics of algorithms: Mapping the debate. Big Data Soc. 2016; 3(2):2053951716679679. doi:10.1177/2053951716679679
20. Kung TH, Cheatham M, Medenilla A, Sillos C, De Leon L, Elepaño C, et al. Performance of ChatGPT on USMLE: Potential for AI-assisted medical education using large language models. PLOS Digit Heal. 2023; 2(2): e0000198. doi:10.1371/journal.pdig.0000198
21. Hoch CC, Wollenberg B, Lüers J-C, Knoedler S, Knoedler L, Frank K, et al. ChatGPT's quiz skills in different otolaryngology subspecialties: an analysis of 2576 single-choice and multiple-choice board certification preparation questions. Eur Arch oto-rhino-laryngology Off J Eur Fed Oto-Rhino-Laryngological Soc Affil with Ger Soc Oto-Rhino-Laryngology - Head Neck Surg. 2023. doi:10.1007/s00405-023-08051-4
22. Wang B, Wan Z, Li C, Zhang M, Shi Y, Miao X, et al. Identification of benign and malignant thyroid nodules based on dynamic AI ultrasound intelligent auxiliary diagnosis system. Front Endocrinol (Lausanne). 2022; 13:1018321. doi:10.3389/fendo.2022.1018321
23. Nguyen DT, TD, Batchuluun G, Yoon HS, Park KR. Artificial Intelligence-Based Thyroid Nodule Classification Pham Using Information from Spatial and Frequency Domains. J Clin Med. 2019; 8(11). doi:10.3390/jcm8111976