Short Communication

Evaluation of The Diagnostic Capabilities of Artficial Intelligence (CHATGPT-4) in The Cardiology Department of Bogodogo University Hospital, Burkina Faso

Wendlassida Martin NACANABO[1]*, Taryètba André Arthur SEGHDA[1], Yannick Laurent Tchenadoyo BAYALA[3], Georges MILLOGO[2], Anna THIAM[2], Nobila Valentin YAMEOGO[2], Koudnoaga André SAMADOULOUGOU[1], and Patrice ZABSONRE[2]

[1]*Department of Cardiology, Bogodogo University Hospital Center, Burkina Faso*
[2]*Department of Cardiology, Yalgado Ouedraogo University Hospital Center, Burkina Faso*
[3]*Department of Rheumatology, Bogodogo University Hospital Center, Burkina Faso*

Summary

Introduction: ChatGPT is an artificial intelligence developed by OpenAI. It can be used to generate positive and differential diagnoses. However, its effectiveness in a cardiology department in Africa has not been studied.

Objectives: To evaluate the diagnostic accuracy of ChatGPT4 in the cardiology department of the Bogodogo University Hospital Center.

Patients and Methods: This was a retrospective descriptive study conducted from 1st April to 30 May 2024 in the cardiology department of the Bogodogo University Hospital. Our primary endpoint was whether the main diagnosis of ChatGPT corresponded to the final diagnosis made by the cardiologists.

Results: Out of 50 patients collected, ChatGPT found the diagnosis on the basis of clinical data in 35.19%. In 81.48% of cases, ChatGPT's diagnosis was one of the cardiologist's three hypotheses and in 64.81% of cases the diagnosis was found with certainty by ChatGPT. The differential diagnosis listed by ChatGPT was score 5 in 46 patients. All the diagnoses of the aetiological groups were found by ChatGPT in 100% of cases except in the hypertensive and ischaemic cardiomyopathy groups.

Conclusion: ChatGPT demonstrated a variable ability to generate accurate diagnoses, with a significant improvement when paraclinical data was included.

INTRODUCTION

ChatGPT is an artificial intelligence developed by OpenAI [1]. It is a large language model based on automatic natural language processing, also known as "generative pretrained transform" (GPT) [2]. ChatGPT is capable of generating textual responses that sound human in response to queries written by users [2]. Previous studies have reported that the diagnostic accuracy of differential diagnosis lists generated by ChatGPT for clinical vignettes ranged from 64% to 83% [3]. Extensive research is currently being conducted in a variety of areas, including cardiovascular disease, using ChatGPT [4]. In our African context, the findings of these studies could be beneficial due to the inadequacy of medical equipment and the obvious shortage of medical specialists. However, there is a lack of studies addressing the competence of ChatGPT in the diagnosis of cardiovascular disease, based on clinical and paraclinical data in a black African population. The aim of this study was to assess the diagnostic accuracy of ChatGPT when provided with clinical and paraclinical data and to compare its performance with that of cardiologists in a cardiology department in Burkina Faso.

MATERIALS AND METHODS

Clinical information from 50 consecutive patients admitted to the cardiology department of Bogodogo University Hospital Center between 1er April and 30 May 2024 was reviewed. Patients without a clear and precise diagnosis were excluded. We used the Chat GPT 4 template application (June 14 version; ChatGPT4, OpenAI, LLC). Clinical and paraclinical information was anonymised, transcribed and entered into Chat GPT 4, followed by the question "what is the most likely diagnosis?" and then "what are the possible diagnoses?". Our predefined primary endpoint was whether the main diagnosis in ChatGPT matched the final diagnosis made by cardiologists. The secondary endpoints were whether the final diagnosis matched by providing clinical data only. Then the presence of the final diagnosis in the possible differential and the differential quality score of the model using a 5-point ordinal scoring system previously published by Bond *et al.*, [5]. This score is based on accuracy and utility (in which a score of 5 is assigned for a differential including the correct diagnosis and a score of 0 is assigned when no diagnosis is close) [5].

The source documents were hospital registers, medical records and reports of paraclinical results.

The data were entered into an Excel database and all the analysis was carried out using SPSS software version 20.0. Missing data were treated as missing data during analysis.

We calculated the frequencies of the exact diagnoses found by ChatGPT (clinical data only, then clinical + paraclinical data). We also calculated the frequencies of the final diagnosis as a function of the differential diagnosis scores.

Identifiers were assigned to each patient during the collection process, so that no names are on our database, thus preserving anonymity and confidentiality.

RESULTS

In this study, ChatGPT found the diagnosis solely on the basis of clinical data in 35.19% of cases. In 81.48% of cases, ChatGPT's diagnosis was one of the three diagnostic hypotheses put forward by the clinicians. After inclusion of the paraclinical data, he reported 64.81% certainty with the physicians' diagnoses. The main diagnoses such as hypertensive heart disease, ischaemic heart disease, toxic heart disease and valvular heart disease were found in 19.23%, 28.84%, 15.38% and 15.38% respectively. All the diagnoses of the aetiological groups were found by ChatGPT in 100% of cases except in the hypertensive and ischaemic cardiomyopathy groups.

DISCUSSION

Artificial intelligence (AI) occupies a prominent place in contemporary medical practice [3]. The aim of our study was to evaluate the performance of ChatGPT in the context of African cardiology practice. To our knowledge, this is the first study to evaluate the diagnostic capabilities of ChatGPT in a cardiology department in sub-Saharan Africa.

Our study in cardiology revealed that ChatGPT correctly identified patients' diagnoses in 64.81% of cases, a significantly better result than that observed by Stoneham et al. in dermatology, where ChatGPT correctly identified the diagnosis in 56% of cases. [6]. In terms of differential diagnoses, ChatGPT listed the cardiologist's diagnosis among its hypotheses with a score equal to 5 in 85.18% of cases, compared with 100% in the dermatology study [6]. This disparity can be attributed to several factors. Firstly, our study used the latest version of ChatGPT, version 4, which benefits from significant improvements in terms of accuracy and analytical capabilities. In contrast, the study by Stoneham *et al*. used an earlier version of the AI, which could explain the lower performance [6]. In addition, the inherent complexity of dermatological diagnoses often requires a very precise clinical semiological description, a task that may be more difficult for ChatGPT to accomplish without high-quality input data. In cardiology, ChatGPT has the advantage of being able to draw on a multitude of paraclinical examinations, such as electrocardiograms, echocardiograms and laboratory analyses, to refine its diagnoses. This wealth of clinical data enables ChatGPT to produce more accurate and reliable diagnoses. Our superior results in cardiology illustrate not only the technological evolution of ChatGPT, but also the importance of available data in improving its diagnostic capabilities.

ChatGPT's capabilities are proving to be an invaluable asset for the future of medical practice. Indeed, the integration of generative AI such as ChatGPT can offer immediate support to doctors in complex cases, reducing diagnostic errors and improving patient outcomes [4]. In an educational context, ChatGPT could play a crucial role in the training of future cardiologists, in particular by refining clinical reasoning and the acquisition of medical knowledge [4]. Interaction with generative AI exposes learners to a variety of diagnoses, preparing them for complex clinical situations.

ChatGPT, is becoming increasingly important in cardiology, as demonstrated by several Western studies. For example, the study by Gunay *et al*., concluded that ChatGPT

outperformed cardiologists on common questions, while its performance closely aligned with that of cardiologists as the complexity of the questions increased [7]. Similarly, the study by Guo *et* al., found that the application of AI in the diagnosis and treatment of cardiac arrhythmias was superior to that of medical specialists [8].

The use of ChatGPT in the diagnosis of cardiovascular disease presents both risks and challenges. Risks include concerns about confidentiality, ethics, bias and discrimination. ChatGPT can be used intentionally or unintentionally to create false evidence and material, thanks to its impressive ability to produce information with a high degree of plausibility [9]. This includes "hallucinations", where the content generated is not based on reality, creating entirely fabricated facts [9]. Another major risk is the reproduction of biases present in training data. In the field of health, where precision is crucial, errors or inaccuracies can be catastrophic [9]. To minimise these risks, rigorous human evaluation is essential, as is compliance with standards of accuracy, reliability and interpretability. In addition, security measures must be put in place to protect patient information, including encryption, access control, secure data storage and compliance with confidentiality regulations.

Challenges include the need for ChatGPT to have the medical expertise to understand the complex relationships between conditions and treatments. Being limited to data, it lacks recent medical advances, which may affect its clinical utility [10]. Our study showed that in 81.48% of cases, the diagnosis of ChatGPT matched the cardiologists' assumptions, highlighting its potential usefulness despite the challenges mentioned. Another challenge is the potential help of ChatGPT for non-cardiologists in the triage and referral of patients, especially in countries such as ours where there is a critical shortage of specialist physicians, but also with the security challenge of not being able to evacuate certain patients.

Our study of the diagnostic capabilities of ChatGPT in cardiology has several important limitations. Firstly, as with any retrospective study, it suffers from missing data in some patients, which may affect the representativeness of the results. Secondly, although the cases studied provide valuable insight into diagnostic scenarios, they may not reflect the full range of clinical presentations, including atypical cases or diagnostic challenges encountered in the cardiology department. In addition, some diagnoses could have been refined if ChatGPT had access to the clinical course of patients, as a clinician would. These limitations must be taken into account when interpreting the results, and underline the need for further studies to fully assess the potential and limitations of ChatGPT in the diagnosis of cardiovascular disease.

CONCLUSION

In conclusion, our study of the diagnostic capabilities of ChatGPT in a cardiology department in Africa reveals promising results. In 64.81% of cases, ChatGPT established a diagnosis concordant with that of the cardiologist, particularly in the nosological groups associated with valvular and hypertensive cardiomyopathies. These results demonstrate the potential of ChatGPT as a diagnostic aid, while underlining the importance of continuing research to refine these results, particularly in developing countries such as ours. The challenges encountered, particularly in relation to the variability of clinical presentations and the lack of longitudinal follow-up of patients, need to be overcome by future studies.

DECLARATIONS

Ethics approval and consent to participate

We have obtained the 'informed consent' of the subjects concerned. All measures are taken to preserve the confidentiality of information concerning them. Approval was obtained from the ethics committee of Bogodogo University Hospital. The procedures used in this study adhere to the tenets of the Declaration of Helsinki.

Consent for publication

We have obtained the 'informed consent' of the subjects concerned for publication

Availability of data and materials

The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

Competing interests

The authors state that they have no conflicts of interest that might have influenced the outcome of this research.

Author contributions

Wendlassida Martin NACANABO: Conceptualization;

methodology; data curation; writing–review and editing; writing–original draft; project administration.

Taryètba André Arthur SEGHDA: project administration; investigation; methodology.

Yannick Laurent Tchenadoyo BAYALA: Conceptualization investigation; data curation, project administration.

Georges MILLOGO: Supervision; investigation;

Anna THIAM: Supervision; investigation;

Nobila Valentin YAMEOGO: Supervision; investigation;

André Kounoaga SAMADOULOUGOU: supervision; validation

Patrice ZABSONRE: Supervision

Author details (optional)

The lead author Wendlassida Martin NACANABO is a resident in cardiology department of Bogodogo university hospital

REFERENCES

1. Curtis N, ChatGPT. To ChatGPT or not to ChatGPT? The Impact of Artificial Intelligence on Academic Publishing. Pediatr Infect Dis J. 2023; 42: 275.

2. 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 Health. 2023; 2: e0000198.

3. Kanjee Z, Crowe B, Rodman A. Accuracy of a Generative Artificial Intelligence Model in a Complex Diagnostic Challenge. JAMA. 2023; 330: 78-80.

4. Hirosawa T, Kawamura R, Harada Y, Mizuta K, Tokumasu K, Kaji Y, et al. ChatGPT-Generated Differential Diagnosis Lists for Complex Case-Derived Clinical Vignettes: Diagnostic Accuracy Evaluation. JMIR Med Inform. 2023; 11: e48808.

5. Bond WF, Schwartz LM, Weaver KR, Levick D, Giuliano M, Graber ML. Differential Diagnosis Generators: an Evaluation of Currently Available Computer Programs. J Gen Intern Med. 2012; 27: 213-219.

6. Stoneham S, Livesey A, Cooper H, Mitchell C. Chat GPT vs Clinician: challenging the diagnostic capabilities of A.I. in dermatology. Clin Exp Dermatol. 2024; 49: 707-710.

7. Günay S, Öztürk A, Özerol H, Yiğit Y, Erenler AK. Comparison of emergency medicine specialist, cardiologist, and chat-GPT in electrocardiography assessment. Am J Emerg Med. 2024; 80: 51-60.

8. Guo R-X, Tian X, Bazoukis G, Tse G, Hong S, Chen KY, et al. Application of artificial intelligence in the diagnosis and treatment of cardiac arrhythmia. PACE. 2024; 47: 789-801.

9. Cascella M, Montomoli J, Bellini V, Bignami E. Evaluating the Feasibility of ChatGPT in Healthcare: An Analysis of Multiple Clinical and Research Scenarios. J Med Syst. 2023; 47: 33.

10. Vaishya R, Misra A, Vaish A. ChatGPT: Is this version good for healthcare and research? Diabetes Metab Syndr. 202317: 102744.