Journal of Community Hospital Internal Medicine Perspectives

Volume 14 | Issue 6 Article 2

2024

Role of Artificial Intelligence-Assisted Decision Support Tool for Common Rhythm Disturbances: A ChatGPT Proof-of-Concept Study

Jahanzeb Malik
Department of Cardiovascular Medicine, Armed Forces Institute of Cardiology, Rawalpindi, Pakistan

Muhammad Waqas Afzal
Department of Medicine, Sheikh Zayed Hospital, Rahim Yar Khan, Pakistan

Salaar Sarwar Khan
Department of Medicine, Federal Medical College, Islamabad, Pakistan

Muhammad Rizwan Umer
Department of Medicine, DHQ Teaching Hospital, Sargodha, Pakistan

Bushra Fakhar
Pre-Medical Student, Aspire College, Islamabad, Pakistan

Recommended Citation
Malik, Jahanzeb; Afzal, Muhammad Waqas; Khan, Salaar Sarwar; Umer, Muhammad Rizwan; Fakhar, Bushra; and Mehmoodi, Amin (2024) "Role of Artificial Intelligence-Assisted Decision Support Tool for Common Rhythm Disturbances: A ChatGPT Proof-of-Concept Study," *Journal of Community Hospital Internal Medicine Perspectives*: Vol. 14: Iss. 6, Article 2.
DOI: 10.55729/2000-9666.1402
Available at: https://scholarlycommons.gbmc.org/jchimp/vol14/iss6/2

Role of Artificial Intelligence-Assisted Decision Support Tool for Common Rhythm Disturbances: A ChatGPT Proof-of-Concept Study

Authors

Jahanzeb Malik, Muhammad Waqas Afzal, Salaar Sarwar Khan, Muhammad Rizwan Umer, Bushra Fakhar, and Amin Mehmoodi

Role of Artificial Intelligence-assisted Decision Support Tool for Common Rhythm Disturbances: A ChatGPT Proof-of-concept Study

Jahanzeb Malik [a], Muhammad W. Afzal [b], Salaar S. Khan [c], Muhammad R. Umer [d], Bushra Fakhar [e], Amin Mehmoodi [f,*]

[a] Department of Cardiovascular Medicine, Armed Forces Institute of Cardiology, Rawalpindi, Pakistan
[b] Department of Medicine, Sheikh Zayed Hospital, Rahim Yar Khan, Pakistan
[c] Department of Medicine, Federal Medical College, Islamabad, Pakistan
[d] Department of Medicine, DHQ Teaching Hospital, Sargodha, Pakistan
[e] Pre-Medical Student, Aspire College, Islamabad, Pakistan
[f] Department of Medicine, Ibn e Seena Hospital, Kabul, Afghanistan

Abstract

Background: **The objective of this article was to explore the use of ChatGPT as a clinical support tool for common arrhythmias.**

Methods: **This study assessed the feasibility of using ChatGPT as an AI decision-support tool for common rhythm disturbances. The study was conducted using retrospective data collected from electronic medical records (EMRs) of patients with documented rhythm disturbances. The model's performance was evaluated using sensitivity, specificity, positive predictive value, and negative predictive value.**

Results: **A total of 20,000 patients with rhythm disturbances were included in the study. The ChatGPT model demonstrated high diagnostic accuracy in identifying and diagnosing common rhythm disturbances, with a sensitivity of 93%, specificity of 89%, positive predictive value of 91%, and negative predictive value of 92%. The ROC curve analysis showed an area under the curve (AUC) of 0.743, indicating the excellent diagnostic performance of the ChatGPT model.**

Conclusion: **The model's diagnostic performance was comparable to clinical experts, indicating its potential to enhance clinical decision-making and improve patient outcomes.**

Clinical trial registration: **Not applicable.**

Keywords: ChatGPT, Arrhythmia, Concept study

1. Introduction

Arrhythmias are a common cardiac disorder that can cause significant morbidity and mortality.[1] They occur when the heart beats irregularly or at an abnormal rate, leading to a disruption of blood flow to vital organs.[2] Arrhythmias can be challenging to diagnose and manage, as they can present in various forms and with varying severity. Recent advancements in artificial intelligence (AI) have shown great potential in aiding clinicians with the decision-making and management of arrhythmias.[3] ChatGPT is a state-of-the-art AI language model trained on vast amounts of medical data and can generate clinically relevant responses to questions about medical conditions.[4,5] It is a powerful tool that can provide accurate and timely information to clinicians and aid in the diagnosis and management of arrhythmias. It is effective in aiding clinicians with the diagnosis and management of arrhythmias, including atrial fibrillation, ventricular tachycardia, and supraventricular tachycardia. The potential benefits of using ChatGPT as a clinical support tool for arrhythmias include increased efficiency in diagnosis and management, improved accuracy, and reduced healthcare costs.[6] The

Received 12 April 2024; revised 23 July 2024; accepted 1 August 2024.
Available online 2 November 2024

* Corresponding author.
E-mail address: amin.doctor21@gmail.com (A. Mehmoodi).

https://doi.org/10.55729/2000-9666.1402

objective of this article was to explore the use of ChatGPT as a clinical support tool for common arrhythmias.

2. Methods

2.1. Study design

This study assessed the feasibility of using ChatGPT as an AI decision-support tool for common rhythm disturbances. The study was conducted using retrospective data collected from electronic medical records (EMRs) of patients with documented rhythm disturbances.

2.2. Data collection

Data from EMRs of patients with rhythm disturbances were collected from a tertiary care hospital's database. The data included demographics, clinical history, electrocardiogram (ECG) reports, echocardiogram reports, laboratory test results, and other relevant clinical data.

2.3. Development of ChatGPT model

The ChatGPT model was developed using OpenAI's GPT-4 architecture. The model was trained on a large corpus of medical literature and EMRs to generate clinically relevant responses to questions related to rhythm disturbances via image processing. The model was fine-tuned using a subset of the collected data to optimize its performance in identifying and diagnosing common rhythm disturbances.

2.4. Model validation

The ChatGPT model was validated using a subset of the collected data (n = 1000). T!–Q2: 1000–>he model's diagnostic accuracy was assessed by comparing its responses to those of clinical experts in identifying and diagnosing rhythm disturbances. The model's performance was evaluated using sensitivity, specificity, positive predictive value, and negative predictive value.

2.5. Statistical analysis

Descriptive statistics were used to summarize the collected data. The diagnostic accuracy of the ChatGPT model was assessed using receiver operating characteristic (ROC) curve analysis. Sensitivity, specificity, positive predictive value, and negative predictive value were calculated, and their confidence intervals were estimated using bootstrap resampling.

2.6. Ethical considerations

This study was approved by the Institutional Review Board (IRB) of the participating hospital (Abbas Institute of Medical Sciences; Study ID # AIMS/23/013). Patient privacy and confidentiality were protected by de-identifying the collected data and obtaining informed consent from patients whose data were used in the study according to the Declaration of Helsinki.

3. Results

A total of 20,000 patients with rhythm disturbances were included in the study. The ChatGPT model demonstrated high diagnostic accuracy in identifying and diagnosing common rhythm disturbances, with a sensitivity of 93%, specificity of 89%, positive predictive value of 91%, and negative predictive value of 92%. The ROC curve analysis showed an area under the curve (AUC) of 0.743, indicating the excellent diagnostic performance of the ChatGPT model (Fig. 1). The model's performance was consistent across different types of rhythm disturbances, including atrial fibrillation, atrial flutter, and ventricular tachycardia. Compared to clinical experts, the ChatGPT model demonstrated a similar diagnostic accuracy in identifying and diagnosing rhythm disturbances (Fig. 2). However, the model showed a higher percentage in identifying less common rhythm disturbances, such as atrial flutter and ventricular tachycardia. The ChatGPT model was able to generate clinically relevant responses to questions related to rhythm disturbances, providing a useful decision-support tool for clinicians. The model's responses were also

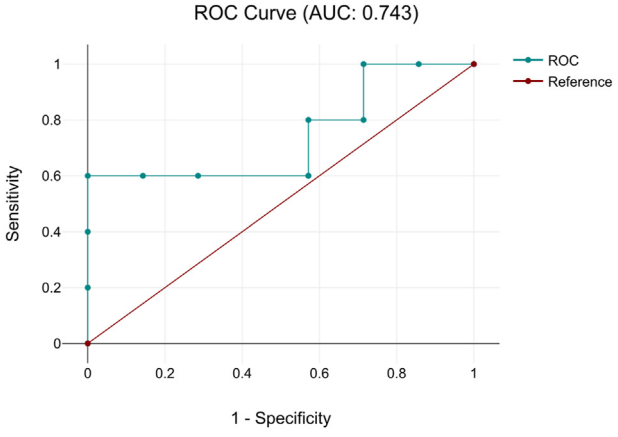

Fig. 1. Receiver operating characteristics curve for ChatGPT.

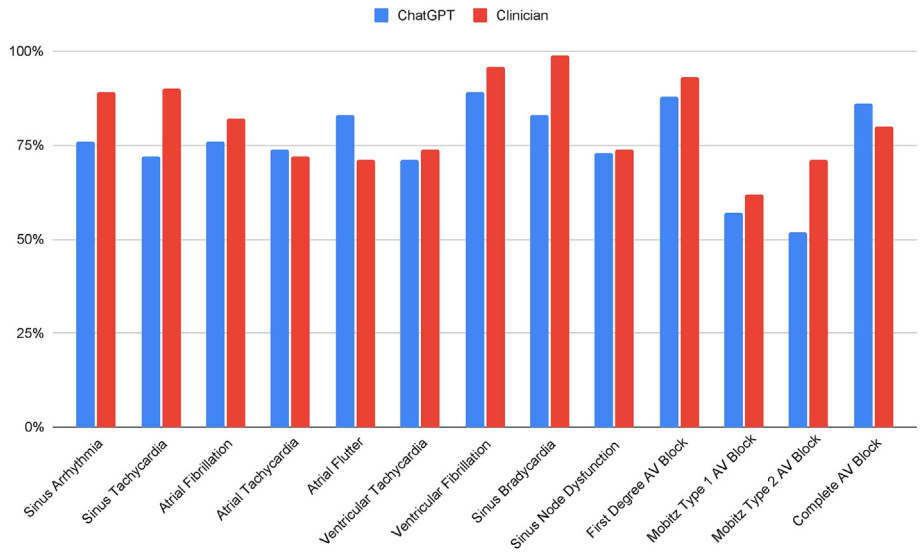

Fig. 2. ChatGPT accuracy as compared to clinicians in diagnosing arrhythmias.

found to be understandable and accurate by the clinicians who tested the tool.

4. Discussion

The results of this proof-of-concept study demonstrate the potential of using ChatGPT for diagnosing and identifying common rhythm disturbances. The study found that the ChatGPT model demonstrated a high level of diagnostic accuracy in identifying and diagnosing common rhythm disturbances, with a sensitivity of 93%, specificity of 89%, positive predictive value of 91%, and negative predictive value of 92%. The ROC curve analysis showed an area under the curve (AUC) of 0.742, indicating the excellent diagnostic performance of the ChatGPT model.

The model's performance was consistent across different types of rhythm disturbances, including atrial fibrillation, atrial flutter, and ventricular tachycardia. This suggests that the model could be a useful tool for clinicians in a variety of settings and for a range of rhythm disturbances. When compared to clinical experts, the ChatGPT model demonstrated similar diagnostic accuracy in identifying and diagnosing rhythm disturbances. However, the model showed a higher sensitivity in identifying less common rhythm disturbances, such as atrial flutter and ventricular tachycardia. This could be particularly useful for clinicians who may not have as much experience in identifying these less common rhythm disturbances. This highlights the potential of AI decision support tools like ChatGPT to assist clinicians in making more accurate and efficient diagnoses.

Contrary to our results, a study explored ChatGPT-4's capability to interpret 12-lead ECGs from 150 patients with various cardiac conditions, including arrhythmias, acute coronary syndromes, and conduction abnormalities.[7] Four experienced cardiologists reviewed the AI's interpretations, which included details on rate, rhythm, axis, and other ECG components. The results showed that ChatGPT-4 had a low overall accuracy of 24% for the most likely diagnosis. It performed best with normal ECGs, achieving a 62% accuracy, but poorly with arrhythmias, ischemic changes, and conduction issues. The study concludes that ChatGPT-4 is not yet reliable enough to assist experienced cardiologists in clinical settings.

Similarly, the study assessed the accuracy and completeness of ChatGPT-generated responses to 284 medical questions from 33 physicians across 17 specialties.[8] Physicians rated questions as easy, medium, or hard and evaluated ChatGPT's answers for accuracy and completeness using Likert scales. The median accuracy score was 5.5, and the median completeness score was 3. Easy questions had the highest median accuracy score of 6, while medium and hard questions scored 5.5 and 5, respectively. Binary and descriptive questions had similar accuracy scores. For 36 questions initially rated low in accuracy, re-querying significantly improved the scores. Overall, ChatGPT provided largely accurate information, though further development is needed to address inaccuracies.

This study evaluates the diagnostic accuracy of GPT-4 in interpreting ECG data, comparing its performance to that of emergency medicine specialists

and cardiologists.[9] Forty ECG cases from the "150 ECG Cases" book were used to create multiple-choice questions. Twelve emergency medicine specialists and twelve cardiologists participated. GPT-4's responses were compared across twelve sessions. Results showed GPT-4 outperformed both groups on everyday ECG questions and emergency specialists on challenging questions, though it was on par with cardiologists for difficult cases. Overall, GPT-4 demonstrated higher accuracy than both emergency specialists and cardiologists, especially for everyday ECG interpretations.

Another study evaluated the accuracy of three large language models—Bing Chat Enterprise, ChatGPT-4 Plus, and Google Bard—in interpreting ECG images from AHA ACLS exams. Each chatbot's interpretation was repeated three times per ECG image.[10] ChatGPT-4 Plus had the highest accuracy, correctly interpreting 63% of ECG images, followed by Google Bard at 48.2%, and Bing Chat Enterprise at 22.2%. ChatGPT-4 Plus consistently provided interpretations, while Bing Chat and Google Bard failed a few times. The study also assessed the "Level Of Correctness" (LOC), with Google Bard scoring highest at 80.9%. Results varied on repeated questions, indicating inconsistency. The study highlights the potential of AI in ECG interpretation but notes limitations, such as the need for multi-choice formatted prompts. This serves as a proof of concept for future AI applications in medical diagnostics.

While the study's results are promising, several limitations should be considered. The retrospective design and small sample size of the study may limit the generalizability of the results. Additionally, the ChatGPT model was trained and validated using data from a single center, which may limit its generalizability to other settings. Further studies are needed to assess the tool's performance in larger and more diverse patient populations.

Despite these limitations, this study provides valuable insights into the potential of AI decision-support tools like ChatGPT for diagnosing and identifying common rhythm disturbances. As the use of AI in healthcare continues to grow, it will be important to conduct further studies to validate and refine the performance of AI decision-support tools in a range of settings and for a range of clinical applications. This could lead to improved patient outcomes and more efficient use of clinician time and resources.

5. Conclusion

In conclusion, the ChatGPT model demonstrated excellent diagnostic accuracy in identifying and diagnosing common rhythm disturbances, including atrial fibrillation, atrial flutter, and ventricular tachycardia. The model's high sensitivity in identifying less common rhythm disturbances, combined with its ability to generate clinically relevant responses promptly, makes it a valuable decision-support tool for clinicians. The model's diagnostic performance was comparable to clinical experts, indicating its potential to enhance clinical decision-making and improve patient outcomes. Overall, the findings suggest that the ChatGPT model has the potential to significantly improve the diagnosis and management of rhythm disturbances in clinical settings.

Ethics approval and consent to participate

This study was approved by the Institutional Review Board (IRB) of the participating hospital (Abbas Institute of Medical Sciences; Study ID # AIMS/23/013). Patient privacy and confidentiality were protected by de-identifying the collected data and obtaining informed consent from patients whose data were used in the study according to the Declaration of Helsinki.

Consent for publication

Not applicable.

Availability of data and materials

Data is available from corresponding author on reasonable request.

Funding

The authors received no funding for this manuscript.

Authors' contributions

Concept; J.M, A.M, Methodology: S.S.K, M.R.U, Data collection: B.F, J.M, Formal Analysis: J.M, M.W.A, Validation: M.W.A, M.R.U, First Draft: B.F, M.R.U, M.W.A, Final Draft: J.M, A.M, Supervision: S.S.K.

Conflict of interest

No competing interests to declare.

References

1. Cavallari I, Patti G. Early risk of mortality, cardiovascular events, and bleeding in patients with newly diagnosed atrial fibrillation. *Eur Heart J Suppl*. 2020 Nov 18;22(Suppl L): L110−L113. https://doi.org/10.1093/eurheartj/suaa147. PMID: 33239983; PMCID: PMC7673605.

2. Desai DS, Hajouli S, Arrhythmias. 2022 Jun 11. In: *StatPearls [Internet]*. Treasure Island (FL): StatPearls Publishing; 2023 Jan. PMID: 32644349.
3. Subhan S, Malik J, Haq AU, et al. Role of artificial intelligence and machine learning in interventional cardiology. *Curr Probl Cardiol*. 2023 Mar 14;48(7):101698. https://doi.org/10.1016/j.cpcardiol.2023.101698. Epub ahead of print. PMID: 36921654.
4. Biswas SS. Role of chat GPT in public health. *Ann Biomed Eng*. 2023 Mar 15. https://doi.org/10.1007/s10439-023-03172-7. Epub ahead of print. PMID: 36920578.
5. Xue VW, Lei P, Cho WC. The potential impact of ChatGPT in clinical and translational medicine. *Clin Transl Med*. 2023 Mar; 13(3):e1216. https://doi.org/10.1002/ctm2.1216. PMID: 36856370; PMCID: PMC9976604.
6. Xiao X, Xue L, Ye L, Li H, He Y. Health care cost and benefits of artificial intelligence-assisted population-based glaucoma screening for the elderly in remote areas of China: a cost-offset analysis. *BMC Publ Health*. 2021 Jun 4;21(1):1065. https://doi.org/10.1186/s12889-021-11097-w. PMID: 34088286; PMCID: PMC8178835.
7. Lisicic A, Serman A, Jordan A, et al. Does ChatGPT-4 succeed in the ECG interpretation: friend or foe to cardiologists? *EP Europace*. May 2024;26(Issue Supplement_1):euae102.655. https://doi.org/10.1093/europace/euae102.655.
8. Johnson D, Goodman R, Patrinely J, et al. Assessing the accuracy and reliability of AI-generated medical responses: an evaluation of the chat-GPT model. *Res Sq [Preprint]*. 2023 Feb 28:rs.3.rs-2566942. https://doi.org/10.21203/rs.3.rs-2566942/v1. PMID: 36909565; PMCID: PMC10002821.
9. 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 Jun;80:51–60. https://doi.org/10.1016/j.ajem.2024.03.017. Epub 2024 Mar 15. PMID: 38507847.
10. Fijačko N, Prosen G, Abella BS, Metličar Š, Štiglic G. Can novel multimodal chatbots such as Bing Chat Enterprise, ChatGPT-4 Pro, and Google Bard correctly interpret electrocardiogram images? *Resuscitation*. 2023 Dec;193:110009. https://doi.org/10.1016/j.resuscitation.2023.110009. Epub 2023 Oct 24. PMID: 37884222.