

AI chatbots in emergency medicine: analyzing agreement with expert physician triage decisions

Ahmad Aalam¹ 

1. Department of Emergency Medicine, King Abdulaziz University, Jeddah, Saudi Arabia

Correspondence to: Ahmad Aalam

*Department of Emergency Medicine, King Abdulaziz University, Jeddah, Saudi Arabia.

Email: DrAalam@DrAalam.com

DOI: 10.24911/SJEMed.72-1740484871



Introduction:

In emergency medicine, accurate triage is vital for patient outcomes and resource management. The Canadian Triage and Acuity Scale (CTAS) has been essential in training healthcare providers to make prompt and precise triage decisions. With the rise of artificial intelligence (AI), chatbots are being considered for their potential to support or even replace human decision-making in various medical situations. This study aims to assess the agreement between AI chatbot triage decisions and those made by experienced emergency physicians using CTAS.

Methods:

This study involved a comparative analysis between an AI chatbot and two expert emergency physicians, each with over 10 years of experience. We used a dataset of 60 emergency case scenarios, which have been utilized for over 8–10 years to train medical personnel at the start of their careers. The AI chatbot received training materials on CTAS and triage before being tasked with assigning appropriate triage levels for each scenario. Meanwhile, the expert physicians independently triaged the same cases. Scenarios where the two experts disagreed on the triage level were excluded, leaving 35 case scenarios for the final analysis.

To evaluate the agreement between the AI chatbot and the expert physicians, we used the Cohen's kappa coefficient. This included determining the Cohen's kappa coefficient value, the *p*-value, and the 95% confidence interval (CI) to assess the statistical significance and reliability of the agreement.

Results:

The Cohen's kappa coefficient value between the AI chatbot and the expert physicians was 0.721, indicating a substantial level of agreement. The *p*-value was < 0.001 , showing a statistically significant correlation. The 95% confidence interval ranged from 0.539 to 0.903, further supporting the reliability of the observed agreement.

Discussion:

The substantial agreement suggests that the AI chatbot can make triage decisions that closely match those of experienced emergency physicians. This finding is important as it demonstrates the potential of AI chatbots to assist in emergency medicine, particularly in settings with limited resources or during peak times when quick decision-making is crucial. However, it is noteworthy that the study excluded scenarios where the experts disagreed, possibly indicating more complex or ambiguous cases. Further research should investigate the chatbot's performance in such scenarios and its ability to manage a broader range of cases.

Conclusion:

This study shows that AI chatbots have the potential to achieve a high level of agreement with expert physicians in triaging emergency cases using CTAS. Integrating AI in emergency departments could improve triage efficiency and support medical staff, ultimately enhancing patient care. Future studies should examine the chatbot's performance in more complex cases and its integration into real-world clinical settings.