Comparison of Advanced Cardiac Life Information of Emergency Department Specialist with Artificial Intelligence: Multicenter Study

Resmiye Nur Okudan, Ökkeş Zortuk, Cihan Bedel, Fatih Selvi

Healt Science University Antalya Training and Research Hospital, Antalya, Turkey

DOI: https://doi.org/10.47772/IJRISS.2026.100500005

Received: 25 April 2026; Accepted: 01 May 2026; Published: 21 May 2026

ABSTRACT

Background and aim: The delivery of cardiopulmonary resuscitation (CPR) after sudden cardiac death is a critical training component in emergency medicine. This study aims to investigate the performance of artificial intelligence applications in this area, particularly in relation to evidence-based medical systems and current CPR guidelines.

Patients and methods: This study was conducted as a multi-centre study involving emergency medicine specialists. A 20-question test based on the Advanced Life Support (ALS) guidelines published by the American Heart Association (AHA) was administered to LLM-2 and LLM-1. Correct answers were scored as 1 point each and the data collected was analysed.

Resutls: The study included 22 Emergency Medicine Specialists (EMS) with a mean age of 32.36±3.84 years. It was observed that the performance of EMS and LLM-1 was significantly higher than that of LLM-2, while there was no significant difference between EMS and LLM-1. Correlation analysis among the participants revealed a negative correlation between age, years of professional experience, years of working in the emergency department and the average score.

Conclution: Artificial intelligence systems still have many limiting factors. Although the responses provided by LLM-2 were found to be inadequate, it appears that LLM-1 can be used as a supporting system.

Keywords: Artificial intelligence, Gemini, Chatgpt, Advanced Life Support, Emergency Medicine

INTRODUCTION

Sudden cardiac death is a condition characterised by the cessation of cardiac activity and the cessation of a person's breathing and circulation. Cardiopulmonary resuscitation (CPR) and defibrillation procedures are performed. The way in which these procedures are carried out and their application procedures are guided by guidelines from two different associations [1, 2]. Emergency medical training is the most comprehensive training in this field and the most comprehensive demonstration of the basic intervention. In Turkey, the curriculum, which is of great importance and is determined by the Medical Specialisation Board of the Ministry of Health, provides training in the most up-to-date form of evidence-based medicine [3].

There are different ways of accessing data today. In the past, these were educational mobile applications and web systems created by analysing primary data, but now there are artificial intelligence systems that can analyse and summarise existing data.

A study by Gunay et al. evaluated the electrocardiogram identification function of the Chat-GPT artificial intelligence system [4]. In addition, the Chat-GPT 3.5 program developed by OpenAI, equipped with a learning model, analyses data on the Internet to produce results and answer questions posed to it [5] Another important AI tool in both voice and text-based applications today is LLM-2, developed by Google. This tool analyses and generates results based on the current state of online data with instant Internet access [6]. AI tools provide significant results for diagnostic approaches in current situations. Several studies in this area have questioned their safety. However, it is still expected to produce results that require further development and deeper algorithms [7].

In this study, we aimed to analyse the everyday use of AI tools. Using a comparative group experimental design, we compared the state of knowledge of emergency physicians with current CPR practices using AI.

MATERIALS AND METOT

The study was conducted after ethics committee approval and was based on information from the ALS 2020 guidelines published by the AHA, with 20 questions comparing AI and emergency medicine specialists (EMS). Participants who agreed to participate in the study and completed the demographic data and questionnaire form were included, while physicians with missing data or those trained in another specialty were excluded.

A total of 22 EMS were included in the study. In addition, LLM-1 (ChatGPT-3.5) and LLM-2 (GEMINI) were used. The demographic data of the participants and the questionnaire consisting of 20 multiple-choice questions were distributed to the participants. For the other two comparison groups using AI applications, namely LLM-1 and LLM-2, the questions were randomised and repeated 20 times to collect the data.

Simple Size and Data Collecting

The G-Power programme determined that the study would require a sample size of 22 participants, with 12 participants per group, for a significance level of 0.05, a power of 0.99, and an effect size of 1.75 [8]. A total of 22 EMS participated in the study. For the other two comparison groups using AI applications, namely LLM-1 and LLM-2, the questions were randomised and repeated 20 times to collect the data.

Statistical Anaylsis

The database created from the data obtained in this study was analysed using IBM SPSS version 27, and Graphpad Prism version 10 was used to create the graphs. The data were classified and categorical data were expressed as frequencies and percentages. For numerical data subjected to distribution analysis, those conforming to a normal distribution were presented as mean ± SD, while those not conforming to a normal distribution were defined as median and interquartile range (IQR). For normally distributed binary groups, t-tests were performed, ANOVA analysis was performed for three-group comparisons, and post-hoc Tukey tests were applied. Pearson correlation tests were performed for correlation analysis between participants. Data with p-values below 0.05 were considered statistically significant according to the analyses performed in the study.

RESULT

A total of 22 emergency physicians participated in the study. Female physicians comprised 18.2% of the participants, with a mean age of 32.36±3.84 years. The characteristics of the EMS participants are shown in Table 1. The median correct scores determined after the 20 questions for EMS, LLM-1 and LLM-2 are compared in Table 2. The mean score of the EMS participants was 14.14±1.58, the mean score of the LLM-1 participants was 14.30±2.20 and the mean score of the LLM-2 participants was 6.65±1.84. In the comparison made between them, it was observed that the success demonstrated by EMS and LLM-1 was significantly higher than that of LLM-2, while no significant difference was observed between EMS and LLM-1 ($F=109.3$, $p<0.001$, Figure 1).

When comparing the CPR training status of EMS within the last year, the mean score of the 11 participants who had not received training was 14.18±1.94 and the mean score of the 11 participants who had received training was 14.09±1.22, with no significant difference observed between them ($p=0.897$, Figure 2). Correlation analysis between the participants revealed a negative correlation between age, years of professional experience, years of working in the emergency department and the average score (Table 3).

DISCUSTIONS

As data accessibility has improved and evidence-based medicine has become fundamental to service delivery, artificial intelligence (AI) systems have been introduced to not only access data, but also analyse and interpret it. This raises a new question for us as healthcare professionals regarding access to reliable data. In our study, we focused on cardiopulmonary resuscitation (CPR), which is a critical aspect of emergency medicine training and one of the most basic life-saving interventions. We created a 20-question test based on the guidelines used in the field and compared the responses of emergency medical services (EMS) professionals with those of two widely used AI systems.

Various methods have emerged, particularly in the emergency department, such as triage applications [9, 10], imaging systems [11] and interpretation of the images obtained, as well as electrocardiogram interpretation [4]. The strengths of the systems tested in this area have been demonstrated by various initiatives. For example, a study by Gunay et al. showed that GPT-4.0 outperformed paramedics in electrocardiogram interpretation.

A study by Watari et al. [12] analysed examinations used in emergency medicine training in Japan in comparison with the LLM-1. While emphasising the limited approach of the GPT-4.0, it was found that it scored lower, particularly in attitudes towards patients requiring professional skills. Another study by Meral et al. [13] compared EMS, LLM-2 and LLM-1 on the basis of Emergency Severity Index triage and demonstrated successful results in triage, suggesting its potential use as an adjunct in the emergency department.

In our study, we found that the Google-developed LLM-2 system provided significantly lower responses in ALS compared to both EMS and LLM-1. In our analysis focused on ALS, which will bring a different perspective to any field of medicine, we found that these AI applications did not provide more effective responses than EMS professionals.

AI is a system based on machine learning and represents a significant development in computer science. Although its development dates back to 1956, it has become popular today thanks to open source and free software [14, 15]. Its effective use is widespread in many areas of healthcare [16]. An analysis of AI systems found that their impact on ALS was ineffective in complex scenarios. However, it is noted that there is potential for a potentially beneficial future [17].

Based on the results of our study, it is clear that the current state of effective use of AI at today's stage of development is still not optimal. The way in which data is collected, analysed and interpreted can pose various challenges to the reliability of the results. As seen in the literature, although they are used as auxiliary tools, it is recommended to re-evaluate the security aspect of the data obtained.

Keeping the information databases used by AI programmes under control will contribute to the advancement of analysis. As observed in our study, the success of open source and free systems in providing reliable results is weak.

LİMİTATİONS

Our study has several limitations. Firstly, it was conducted within a single specialty. Multicentre studies involving several specialties could provide more comprehensive findings.

CONCLUTIONS

In the emergency department, the conditions to be identified in cases of cardiac arrest are characterised by individual solutions. Current artificial intelligence systems should be considered as supportive ideas rather than definitive recommendations for this situation.

ACKNOWLEDGE

Conflict of Interest: The authors declare that they have no conflict of interest.

Financial Support: no.

REFERENCES

1. Berg, K.M., et al., Adult Advanced Life Support: 2020 International Consensus on Cardiopulmonary Resuscitation and Emergency Cardiovascular Care Science With Treatment Recommendations. 2020. **142**(16_suppl_1): p. S92-S139.
2. Soar, J., et al., European Resuscitation Council Guidelines 2021: Adult advanced life support. Resuscitation, 2021. **161**: p. 115-151.
3. TUK. Emergency Medicine Specialist Education. 2021 [cited 2024; Available from: https://tuk.saglik.gov.tr/TR,50050/acil-tip.html.
4. Gunay, S., et al., Comparison of emergency medicine specialist, cardiologist, and chat-GPT in electrocardiography assessment. Am J Emerg Med, 2024. **80**: p. 51-60.
5. OpenAI. Chat GPT. 2022; Available from: openai.com/chatgpt.
6. Google. Gemini. 2023; Available from: https://gemini.google.com.
7. Martinez-Selles, M. and M. Marina-Breysse, Current and Future Use of Artificial Intelligence in Electrocardiography. J Cardiovasc Dev Dis, 2023. **10**(4).
8. Chang, K.C., et al., Usefulness of Machine Learning-Based Detection and Classification of Cardiac Arrhythmias With 12-Lead Electrocardiograms. Can J Cardiol, 2021. **37**(1): p. 94-104.
9. Neri, E., et al., Artificial intelligence: Who is responsible for the diagnosis? Radiol Med, 2020. **125**(6): p. 517-521.
10. Ramlakhan, S., et al., Understanding and interpreting artificial intelligence, machine learning and deep learning in Emergency Medicine. Emerg Med J, 2022. **39**(5): p. 380-385.
11. Ehrlich, H., M. McKenney, and A. Elkbuli, The niche of artificial intelligence in trauma and emergency medicine. Am J Emerg Med, 2021. **45**: p. 669-670.
12. Watari, T., et al., Performance Comparison of ChatGPT-4 and Japanese Medical Residents in the General Medicine In-Training Examination: Comparison Study. JMIR Med Educ, 2023. **9**: p. e52202.
13. Meral, G., et al., Comparative analysis of ChatGPT, Gemini and emergency medicine specialist in ESI triage assessment. Am J Emerg Med, 2024. **81**: p. 146-150.
14. Moor, J.J.A.M., The Dartmouth College artificial intelligence conference: The next fifty years. 2006. **27**(4): p. 87-87.
15. Solomonoff, R.J.J.H.S.M., The time scale of artificial intelligence: Reflections on social effects. 1985. **5**(2): p. 149-153.
16. Jumper, J., et al., Highly accurate protein structure prediction with AlphaFold. Nature, 2021. **596**(7873): p. 583-589.
17. Semeraro, F., et al., Clinical questions on advanced life support answered by artificial intelligence. A comparison between ChatGPT, Google Bard and Microsoft Copilot. Resuscitation, 2024. **195**: p. 110114.

ISSN No. 2454-6186 | DOI: 10.47772/IJRISS | Volume X Issue V May 2026

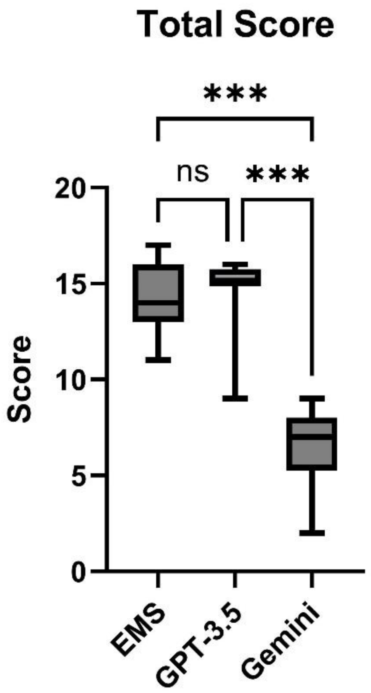

Figure 1. Comparison of AI models

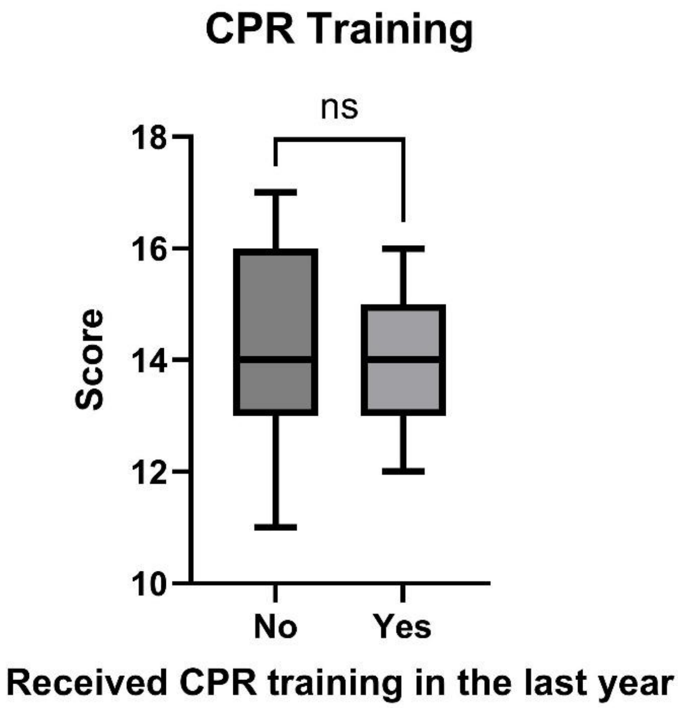

Figure 2. Comparison of CPR training

TABLE LEGEND

Table 1: Descriptive characteristics of EMS participating

Table 2: Comparison between groups according to questions

Table 3: Correlation analysis between age, working time and working time in the emergency department and total score

ISSN No. 2454-6186 | DOI: 10.47772/IJRISS | Volume X Issue V May 2026

Figure Legend

Fig. 1: Comparison of average scores by groups

Fig. 2: Comparison of EMSs by CPR training status

Table 1: Descriptive characteristics of EMS participating

Spesification	
Gender, female (n, %)	4 (18,2)
Age	32,36±3,84
Worktime in Medical Doctor	7,00±3,32
Worktime in ED	6,68±3,21
Received CPR training in the last year (n, %)	11 (50)

Table 2: Comparison between groups according to questions

Question	EMS[a]	LLM-1[b]	LLM-2[c]	F	p-Value	
1 (Median,IQR)	0 (0)	0 (0)	0 (0)		0,255	
2 (Median,IQR)	1 (0)	1 (0)	0 (1)		<0,001	
3 (mean±SD)	0,95±0,21	0,90±0,30	0	125,44	<0,001	b>a,c
4 (Median,IQR)	0 (0)	0 (0)	0 (0)		0,403	
5 (Median,IQR)	1 (1)	1 (0)	0 (0)		<0,001	
6 (mean±SD)	0,95±0,21	0,75±0,44	0,80±0,39	2,56	0,086	-
7 (Median,IQR)	1 (1)	0 (1)	0 (0)		<0,001	
8 (mean±SD)	0,95±0,21	1,00±0,00	0	399,66	<0,001	a,b>c
9 (mean±SD)	0,81±0,21	1,00±0,00	0,90±0,030	2,016	0,142	-
10 (Median,IQR)	0,5 (1)	0 (0,75)	1 (1)		0,075	
11 (mean±SD)	0,91±0,29	1,00±0,00	0,80±0,41	2,358	0,103	-
12 (mean±SD)	0,82±0,39	1,00±0,00	1,00±0,00	4,229	0,019	a<b,c
13 (Median,IQR)	1 (1)	1 (0)	0 (0)		<0,001	
14 (mean±SD)	0,81±0,39	0,80±0,41	0	40,481	<0,001	c<a,b
15 (Median,IQR)	1 (0)	1 (0)	0 (1)		<0,001	
16 (mean±SD)	0,55±0,50	0	0	22,84	<0,001	a>b,c
17 (Median,IQR)	0 (0,25)	0 (0)	0 (0)		<0,001	
18 (mean±SD)	1,00±0,00	1,00±0,00	0,55±0,51	16,35	<0,001	a>c, b>c
19 (mean±SD)	0,73±0,46	0,80±0,41	0,52±0,50	30,89	<0,001	a>c, b>c
20 (mean±SD)	0,96±0,21	1,00±0,00	0,95±0,22	0,479	0,622	-

Table 3: Correlation analysis between age, working time and working time in the emergency department and total score

		Score	Age	Worktime in MD	Worktime in ED
Score	Pearson Correlation	1			
	Sig. (2-tailed)				
	N	62			
Age	Pearson Correlation	-0,368	1		
	Sig. (2-tailed)	0,092			
	N	22	22		
Work time	Pearson Correlation	-,425*	,916**	1	

ISSN No. 2454-6186 | DOI: 10.47772/IJRISS | Volume X Issue V May 2026

	Sig. (2-tailed)	0,048	0,000		
	N	22	22	22	22
Worktime in ED	Pearson Correlation	-,441*	,915**	,919**	1
	Sig. (2-tailed)	0,040	0,000	0,000	
	N	22	22	22	22
*. Correlation is significant at the 0.05 level (2-tailed).					
**. Correlation is significant at the 0.01 level (2-tailed).					