Submitted: 24 January, 2025 Accepted: 15 May, 2025 Published: 08 January, 2026 DOI:10.22514/sv.2025.138

Open Access
Signa Vitae

ORIGINAL RESEARCH

Is ChatGPT able to interpret arterial blood gas analysis? A comparative cross-sectional study

Uğur Serkan Çitilcioğlu[1,*], Barış Arslan[1], Hatice Kaya Özdoğan[1], Hakan Yalım[1], Ümit Kara[1]

[1]Department of Anesthesiology and Reanimation, SBU Adana City Training and Research Hospital, 010650 Adana, Türkiye

***Correspondence**
ugur.citilcioglu@saglik.gov.tr
(Uğur Serkan Çitilcioğlu)

Abstract

Background: This study aimed to evaluate the performance of ChatGPT in interpreting arterial blood gas results and compare it to the performance of physicians with varying levels of experience. **Methods**: In this comparative and cross-sectional study, 30 selected clinical cases encompassing simple and mixed acid-base disorders, respiratory abnormalities, and electrolyte disturbances, were analyzed by ChatGPT and 45 anesthesiology physicians who were divided into three groups: specialists, experienced residents, and inexperienced residents. Participants assessed arterial blood gases across five domains, including acid-base diagnosis, respiratory evaluation, fluid and electrolyte disturbance, other abnormalities, and treatment planning. Both ChatGPT and participant responses were scored by two experienced anesthesiology academicians based on a 5-point Likert scale. The overall score was calculated as the sum of the scores from the five individual subdomains. **Results**: The overall scores of ChatGPT and the physicians were similar (21.75, 21.97). ChatGPT demonstrated high accuracy in diagnosing primary acid-base disorders and providing treatment recommendations, though slight variability was observed in mixed disorders. A strong correlation was observed between ChatGPT's scores and those of physicians (r = 0.912, $p <$ 0.001). **Conclusions**: ChatGPT demonstrated a performance comparable to that of physicians, suggesting that it could assist in the decision-making processes of healthcare professionals and contribute to workload reduction.

Keywords

Artificial intelligence; Arterial blood gas; Critical care; Anesthesia; ChatGPT; Acid-base disorders; Decision support

1. Introduction

Arterial blood gas (ABG) analysis is a cornerstone in critical care and anesthesia practice, offering vital information about a patient's respiratory and metabolic status. ABG interpretation is crucial for monitoring ventilation, oxygenation, and acid-base balance during surgeries, particularly in high-risk patients or complex procedures [1]. However, interpreting ABG results can be challenging, even for experienced clinicians, due to the sheer volume of data and the need for rapid and accurate decision-making in a dynamic clinical environment [2]. One of the significant challenges in ABG interpretation lies in synthesizing multiple parameters, such as pH, pCO_2 (partial pressure of carbon dioxide), pO_2 (partial pressure of oxygen), HCO_3^- (bicarbonate), electrolytes, glucose, and lactate, while simultaneously considering the patient's clinical context [3]. Errors or oversights can occur, particularly under time pressure, leading to delayed or suboptimal treatment decisions [4]. Moreover, less experienced clinicians may struggle to identify subtle abnormalities or formulate appropriate treatment recommendations, further complicating patient management [5].

Artificial intelligence (AI) has emerged as a transformative tool across various medical specialties, offering new opportunities to address such challenges. Among AI systems, ChatGPT, developed by OpenAI, has generated considerable excitement due to its ability to process and synthesize complex information efficiently. ChatGPT is a large language model trained using reinforcement learning from human feedback (RLHF) (OpenAI, https://openai.com/blog/chatgpt), allowing it to interact with users and provide outputs based on prompts. Since its launch in November 2022, ChatGPT has been explored for its potential applications in medicine, including clinical and laboratory diagnostics, patient management, and medical education [6–9].

AI has demonstrated success in specialties such as radiology and pathology, where large datasets and structured data drive decision-making [10, 11]. However, early studies have highlighted the limitations of AI systems, particularly concerns regarding performance variability and bias in clinical diagnostics [12–16]. According to the available literature, its potential

in dynamic, context-heavy tasks such as ABG interpretation remains unexplored. Due to the complexity of ABG analysis, especially in cases involving mixed acid-base disorders or coexisting respiratory and metabolic abnormalities, AI systems may play a significant role in reducing cognitive workload and improving accuracy [17]. By systematically interpreting ABG results and providing evidence-based treatment recommendations, ChatGPT could assist clinicians in rapid decision-making, enhancing both patient safety and clinical efficiency [4]. Furthermore, ChatGPT offers potential as an educational tool, helping residents and junior clinicians strengthen their understanding of ABG interpretation through case-based learning and interactive feedback [8, 9].

This study aims to evaluate ChatGPT's performance in interpreting ABG results by comparing it with a group of physicians. Additionally, it explores ChatGPT's potential as a valuable decision support tool for enhancing patient safety and clinical efficiency in the fields of anesthesia and intensive care.

2. Materials and methods

2.1 Study design and case collection

This comparative and cross-sectional study was conducted at SBU Adana City Education and Research Hospital between 02 January and 20 April 2025, after obtaining ethical approval. The study utilized blood gas analysis results and the clinical medical histories of patients who underwent surgical procedures or were admitted to the intensive care units of the hospital during this period. A total of 160 ABG analyses and patient histories were reviewed, from which 30 representative cases, illustrating both primary and mixed acid-base disorders, were selected during a session organized by the authors.

Inclusion criteria for case selection were as follows:

- Adult patients ($\geq$18 years) who underwent arterial blood gas (ABG) analysis during the study period,
- Availability of complete clinical documentation, including demographic data, clinical history, and ABG values,
- Representation of distinct acid-base disorders, including both primary and mixed disturbances.

Exclusion criteria included:

- Pediatric patients ($<$18 years of age),
- Incomplete or missing ABG data,
- Lack of sufficient clinical information (*e.g.*, undocumented comorbidities, medications, or presenting complaints).

2.2 ChatGPT for diagnosis and treatment recommendations

For each case, ABG results were presented to ChatGPT (GPT-4.0, OpenAI, San Francisco, CA, USA) in the form of photographic images taken directly from printed outputs of the hospital's blood gas analyzer. These images reflected real-world ABG printouts, preserving original formatting and values. Additionally, each case included a separate photograph of a handwritten note summarizing the patient's demographic data and brief clinical history. The model was tasked with diagnosing abnormalities and providing treatment recommendations based on five predefined domains (Table 1). To ensure unbiased outcomes, all responses were documented following the model's initial inquiry, as ChatGPT retains contextual memory within a single session.

2.3 Physicians

The cases were presented to physicians in the same format. Participants were instructed to analyze the cases and provide answers under the same five domains. The study included 45 physicians with varying levels of experience in intensive care and anesthesia (Table 2).

2.4 Assessment

A standardized 5-point Likert scale (Table 3) was used to assess each response, as commonly applied in prior studies evaluating AI performance in clinical decision-making [18–20].

The assessment of acid-base disorders was performed using a standardized three-step approach: (1) primary disorder identification (pH, PCO_2, HCO_3^- thresholds), (2) compensation analysis (Winter's formula, respiratory compensation rules), and (3) anion gap evaluation (calculation and delta ratio interpretation), utilizing established physiological ranges and widely accepted clinical equations [21].

The following diagnostic thresholds were applied for other parameters based on established clinical guidelines: hyponatremia was defined as serum sodium $<$135 mmol/L and hypernatremia as $>$145 mmol/L; hypokalemia as potassium $<$3.5 mmol/L and hyperkalemia as $>$5.0 mmol/L; hypoglycemia as glucose $<$70 mg/dL and hyperglycemia as $>$140 mg/dL; hypoxia as arterial partial oxygen pressure (PaO_2) $<$80 mmHg; hyperlactatemia as lactate $>$2.0 mmol/L; and clinically significant methemoglobinemia as $>$1.5% of total hemoglobin [22–25].

The treatment of acid-base disorders was evaluated according to four fundamental principles: (1) targeted treatment of the underlying etiology, (2) correction of coexisting blood sugar, fluid and electrolyte abnormalities, and (3) implementation of appropriate respiratory interventions when necessary (including mechanical ventilation for severe respiratory acidosis or supplemental oxygen for hypoxemia), (4) recognition of hemoglobinopathies (methemoglobinemia) [26–28].

Two experienced anesthesiology academicians (BA, with 15 years of experience, and HKO, with 25 years of experience) evaluated the responses provided by ChatGPT and the physicians using the 5-point Likert scale. To ensure objectivity, all evaluations were conducted independently and blinded to participant identity. Also, responses generated by ChatGPT were transcribed into standardized evaluation forms without any indication of their origin, ensuring that evaluators were unaware of whether a response came from a human participant or the AI model. Each category was scored by the two raters based on the clinical accuracy and appropriateness of the response. To assess the inter-rater reliability between the two independent raters, Cohen's kappa (κ) statistic was calculated. The overall kappa value was found to be 0.82, indicating almost perfect agreement between the raters.

In total, 1380 individual assessments were performed, covering the interpretation of 30 distinct ABG cases by 45 physician participants and ChatGPT. Each case was evaluated across

TABLE 1. Five domains of ABG evaluation scale.

Diagnosis/Assessment	Description
Acid-Base Disorder Diagnosis	Determining the type of disorder (respiratory, metabolic, mixed), whether it is compensated or not, and whether it involves acidosis or alkalosis.
Fluid and Electrolyte Disorder	Identifying abnormalities in electrolytes or fluid balance.
Respiratory Parameters Evaluation	Assessing for hypoxia or hypercapnia.
Other Anomalies	Highlighting any abnormalities such as elevated lactate, carboxyhemoglobin (CO), methemoglobin, or glucose levels.
Treatment Recommendations	Proposing treatment strategies based on the identified abnormalities.

TABLE 2. Demographics of the physicians.

Group	Participants	Description
Specialist Group	15	Physicians with at least 5 years of experience in anesthesia and intensive care
Experienced Resident Group	15	Residents with 2 or more years of training in anesthesia and intensive care
Inexperienced Resident Group	15	Residents with less than 2 years of training

TABLE 3. 5-point Likert scale.

Score	Description
1	Completely incorrect/Irrelevant response
2	Mostly incorrect with minimal relevant information
3	Partially correct but incomplete or vague
4	Mostly correct with minor omissions or inaccuracies
5	Completely correct and clinically appropriate

five predefined clinical domains. In cases of scoring discrepancies, the average score was considered final.

2.5 Statistical analysis

Continuous variables are presented as mean ± standard deviation (SD) or median (interquartile range (IQR)), while categorical variables are expressed as numbers (proportions). Normality of continuous variables was assessed using the Shapiro-Wilk test. Comparisons of continuous variables between groups were performed by Kruskal-Wallis and Mann-Whitney U-tests. For categorical data, comparisons between groups were made using the chi-square test or Fisher's exact test. Correlation analysis between scores was conducted using Spearman's rank correlation coefficient. All statistical analyses were performed using the Statistical Package for Social Sciences (SPSS; version 18.0; IBM Corp., Chicago, IL, USA), and a p-value of < 0.05 was considered statistically significant.

3. Results

3.1 Overall performance

The median overall score for all participants was 21.97 (IQR: 20.9–23), indicating a generally high performance across groups. No statistically significant differences were observed among the three participant groups regarding total score or individual parameter scores ($p > 0.05$). ChatGPT's total score was 21.75 (IQR: 20.7–22.2), closely aligning with the median score of participants (Fig. 1).

3.2 Parameter-specific analysis

3.2.1 Diagnosis of acid-base disorders

The median scores for acid-base disorder diagnosis were 4.3 for specialists, 4.5 for experienced residents, and 4.3 for inexperienced residents. ChatGPT received a median score of 4.3, indicating a performance that was comparable to that of the human participants.

Physicians and ChatGPT both demonstrated high accuracy in diagnosing primary acid-base disorders. For mixed acid-base disorders, participants achieved a mean score of 3.9 ± 0.8 out of 5, while ChatGPT scored 4 ± 1.2. Despite slightly higher variability, ChatGPT's performance was consistent with the participant groups (Fig. 2).

3.2.2 Other parameters

The accuracy of respiratory issue diagnosis, electrolyte disturbance identification, treatment planning, and detection of other anomalies was consistent across all participant groups ($p > 0.05$). Detailed scores for these categories are summarized in Table 4. ChatGPT demonstrated similar performance to the participants across all parameters (Table 5).

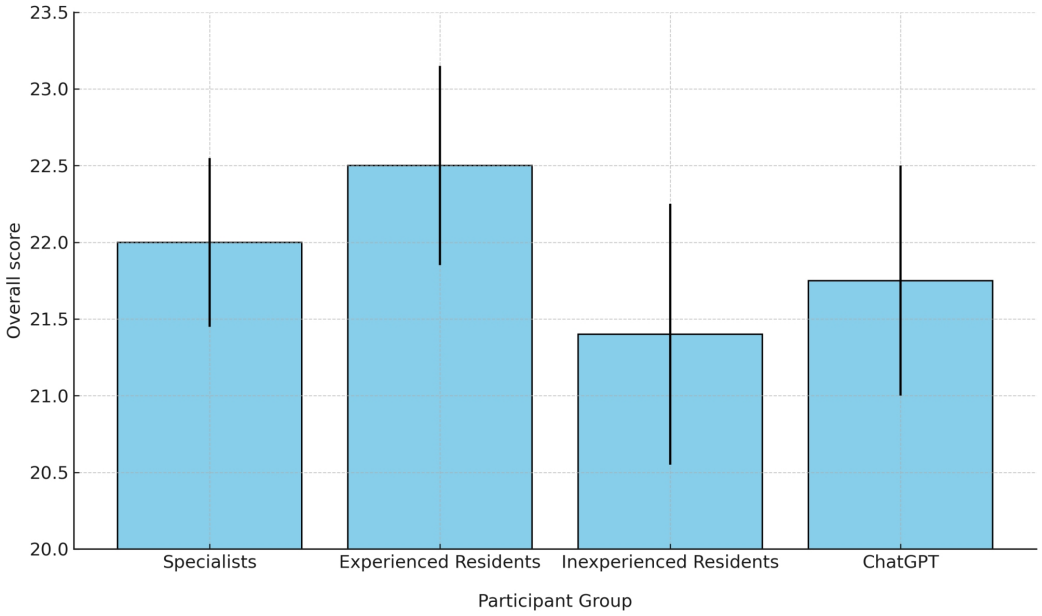

FIGURE 1. Overall performance of the participant groups and ChatGPT.

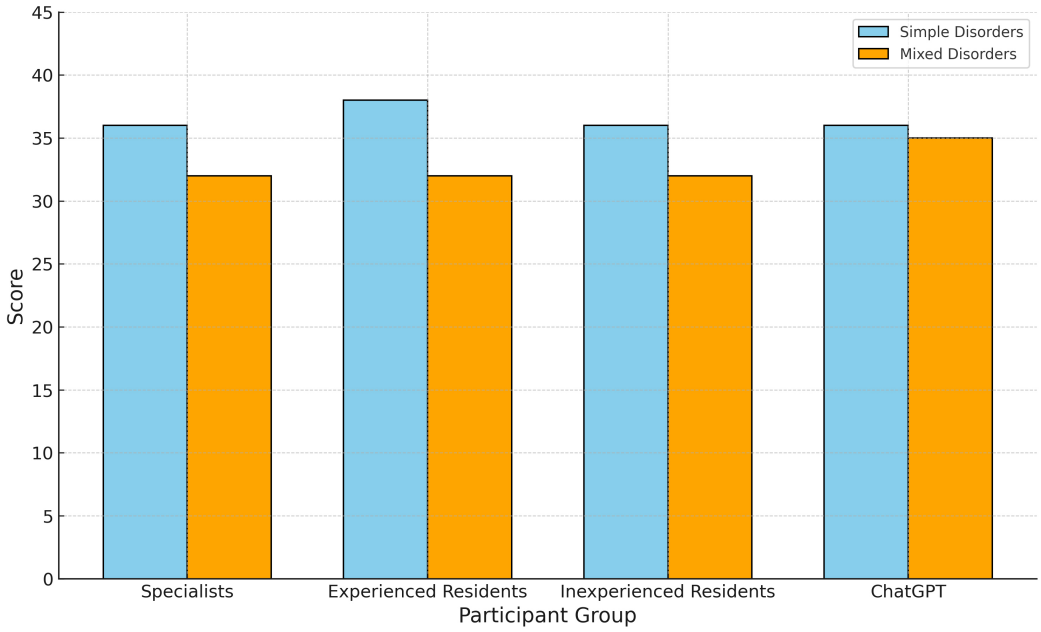

FIGURE 2. Performance in simple and mixed acid-base disorders.

TABLE 4. Comparison of diagnostic and treatment performance of participants.

Variable	Specialist Group (Median (IQR))	Experienced Resident Group (Median (IQR))	Inexperienced Resident Group (Median (IQR))	*p*-value
Diagnosis of acid-base disorder	4.3 (4.1–4.7)	4.5 (4.4–4.6)	4.3 (4.1–4.4)	0.056
Diagnosis of respiratory issue	4.5 (4.2–4.7)	4.6 (4.4–4.9)	4.3 (3.9–4.5)	0.062
Diagnosis of electrolyte disorder	4.4 (4.2–4.6)	4.5 (4.4–4.7)	4.3 (4.0–4.8)	0.265
Treatment plan	4.3 (4.0–4.7)	4.4 (4.3–4.6)	4.2 (3.7–4.3)	0.053
Other issue	4.5 (4.2–4.6)	4.5 (4.4–4.8)	4.3 (4.0–4.6)	0.054
Overall score	22.0 (20.9–23.0)	22.5 (21.7–23.0)	21.4 (20.4–22.1)	0.067

- *Data are presented as medians with interquartile ranges (IQR).*
- *p-values were calculated to assess statistical differences between groups.*

TABLE 5. Comparison of ChatGPT and participants.

Variable	Participants	ChatGPT	*p*-value
Diagnosis of acid-base disorder	4.37 (4.10–4.70)	4.30 (4.10–4.70)	0.243
Diagnosis of respiratory issue	4.47 (4.10–4.70)	4.34 (4.10–4.70)	0.194
Diagnosis of electrolyte disorder	4.40 (4.10–4.70)	4.45 (4.10–4.70)	0.354
Treatment plan	4.30 (4.10–4.70)	4.21 (4.10–4.70)	0.390
Other issue	4.43 (4.10–4.70)	4.45 (4.10–4.70)	0.574
Overall score	21.97 (20.90–23.00)	21.75 (20.70–22.20)	0.632

- *Data are presented as medians with interquartile ranges (IQR).*
- *p-values were calculated to assess statistical differences between groups.*

3.3 Group comparisons

No statistically significant differences were observed among the specialist, experienced residents, and inexperienced resident groups across any subcategories ($p > 0.05$) (Table 4). There were no statistically significant differences between ChatGPT and the participants across the five subdomains (Table 5). A strong correlation was observed between ChatGPT's scores and those of physicians ($r = 0.912$, $p < 0.001$).

4. Discussion

In this comparative cross-sectional study, we found that ChatGPT demonstrated performance comparable to that of physicians with varying levels of experience in interpreting ABG results.

ABG analysis is a challenging skill in anesthesia and intensive care medicine, where timely and accurate interpretation of acid-base, respiratory, and metabolic abnormalities is essential. This challenge becomes even more apparent, particularly in scenarios where quick decision-making is crucial [29]. Therefore, artificial intelligence, with its ability to perceive details and perform rapid analysis, can significantly ease the workload of clinicians. In this study, both participants and ChatGPT demonstrated high accuracy in diagnosing simple acid-base disorders. For mixed acid-base disorders, which are inherently more complex, participants achieved slightly lower scores compared to primary disorders. ChatGPT's performance in these cases was also variable, but within the range of human participants. For instance, one representative case involving a patient who had undergone resuscitation following an anaphylactic shock. The arterial blood gas revealed a mixed acid-base disorder characterized by respiratory acidosis (pH: 7.230, pCO_2: 59.6 mmHg) and borderline metabolic acidosis (HCO_3^-: 20.1 mmol/L, Base Excess: −2.9), along with borderline lactate elevation and hypoxemia (Lactate: 4 mmol/L, pO_2: 69 mmHg). Clinically, this profile was compatible with the patient's post-resuscitation status and tissue hypoperfusion. However, despite receiving the full clinical context, ChatGPT categorized the case solely as a primary respiratory acidosis and failed to recognize the coexisting metabolic component and the clinical implications of rising lactate levels. This may indicate a limitation in the model's ability to integrate complex physiological patterns, particularly in the absence of explicit prompting. In a real-world scenario, such a misclassification could obscure the diagnosis of an ongoing shock state, potentially leading to delays in hemodynamic support or further diagnostic work-up. This example underscores the importance of human oversight in interpreting AI outputs, particularly in cases involving multiple, interrelated pathophysiological processes. Our findings are consistent with previous research. Earlier studies have demonstrated that ChatGPT performs quite well in solving problems that are well-structured and less detailed, but its success rate can fluctuate in scenarios involving rare conditions and complex details [30, 31].

In our study we find strong correlation between ChatGPT's scores and participant performance, which highlights its potential as a decision support tool for clinicians. Our study also highlights that, in its current state, ChatGPT exhibits slightly lower overall performance compared to human participants, although this difference is not statistically significant. This finding underscores the importance of having its recommendations reviewed by a clinician in clinical decision-making processes to ensure patient safety. A previous study highlighted the potential of using a hybrid system where clinicians leverage AI support in clinical judgment processes. When this application is integrated into clinician's workflows, it has been observed to significantly enhance their ability to manage patients and make balanced and accurate decisions [32].

Recent advancements in artificial intelligence have demonstrated significant potential in automating repetitive and data-intensive tasks, enabling clinicians to focus on more complex and critical decision-making processes [33]. For example, Wong *et al.* [31] (2021) validated an AI-based sepsis prediction model that performed comparably to expert clinicians, showcasing the potential of AI in critical care decision-making. Opportunities presented by AI, such as improving medical imaging diagnostics and reducing human errors, are particularly noteworthy [34–36]. As healthcare continues to evolve, integrating AI frameworks will be essential for addressing the challenges of modern medicine while maintaining the critical human connection in patient care [37]. When integrated with electronic healthcare systems (EHS), ChatGPT can facilitate rapid data analysis and provide instant feedback in point-of-care (POC) applications [38]. In fields such as anesthesia and critical care, it has the potential to reduce cognitive load and minimize errors by offering systematic arterial blood gas (ABG) interpretations and evidence-based treatment recommendations.

Artificial intelligence has certain limitations, ChatGPT's performance heavily depends on the accuracy and completeness of the input data [39]. In real-world applications, errors in entering ABG values or missing critical context, such as comorbidities, medications, or clinical history, can lead to inappropriate or incomplete recommendations. Furthermore, ChatGPT cannot independently verify or question the validity of the provided data and lacks the ability to integrate this information into a broader clinical framework [40]. For instance, ABG interpretation often relies on understanding the patient's clinical trajectory, underlying conditions, and current treatment plans [5]. These nuances, which human clinicians use to refine their decisions, are beyond ChatGPT's capabilities. Over-reliance on AI systems without adequate human supervision can lead to errors and serious consequences [41]. Additionally, concerns about data privacy and security in AI-based systems must be addressed before widespread implementation [42].

While this study offers encouraging results regarding the use of ChatGPT in ABG interpretation, further research is warranted to explore its applicability across different clinical scenarios and healthcare settings. Future investigations may also provide a deeper understanding of its role in clinical decision-making and integration into healthcare workflow processes.

5. Conclusions

This study demonstrates that ChatGPT performs comparably to clinicians in ABG interpretation and treatment planning, offering significant potential as a decision-support tool in anesthesia and critical care. By enhancing clinical accuracy and reducing cognitive workload, ChatGPT could play an increasingly valuable role in modern healthcare. However, its use should remain complementary to clinical expertise, emphasizing the importance of validation and ethical considerations in its application.

AVAILABILITY OF DATA AND MATERIALS

The data presented in this paper are available upon reasonable request from the corresponding author.

AUTHOR CONTRIBUTIONS

USÇ—development of methodology, conception, and study design; data collection, analysis, interpretation, drafting of the manuscript, critical revision, and final approval. BA—development of methodology, conception and study design, analysis, statistics critical revision and final approval. HKÖ—development of methodology, analysis, and final approval. HY—development of methodology, analysis, and final approval. ÜK—conception and study design, critical revision, and final approval. All the authors have read and approved the final version of this manuscript.

ETHICS APPROVAL AND CONSENT TO PARTICIPATE

This study complied with the principles of the 1964 Declaration of Helsinki and its amendments. Patient data were anonymized to ensure confidentiality. Ethical approval was obtained from the Institutional Review Board on 02 January 2025 (Approval Number: 02.01.2025-9-285) SBU Adana City Education and Research Hospital. Additionally, all participating physicians provided written informed consent prior to their inclusion in the study.

ACKNOWLEDGMENT

Not applicable.

FUNDING

This research received no external funding.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

REFERENCES

[1] Weimar Z, Smallwood N, Shao J, Chen XE, Moran TP, Khor YH. Arterial blood gas analysis or venous blood gas analysis for adult hospitalised patients with respiratory presentations: a systematic review. Internal Medicine Journal. 2024; 54: 1531–1540.

[2] Korpi-Steiner N, Horowitz G, Tesfazghi M, Suh-Lailam BB. Current issues in blood gas analysis. The Journal of Applied Laboratory Medicine. 2023; 8: 372–381.

[3] Mohammed HM, Abdelatief DA. Easy blood gas analysis: implications for nursing. Egyptian Journal of Chest Diseases and Tuberculosis. 2016; 65: 369–376.

[4] Rodríguez-Villar S, Poza-Hernández P, Freigang S, Zubizarreta-Ormazabal I, Paz-Martín D, Holl E, *et al*. Automatic real-time analysis and interpretation of arterial blood gas sample for point-of-care testing: clinical validation. PLOS ONE. 2021; 16: e0248264.

[5] Musa Hussain EY, Sidahmed Abdullah AM, Mahgoub Idris RM, Hashim Gabir ZT, Mohammed Diab RA, Mustafa Ahmed RN, *et al*. Evaluation and improving the quality of arterial blood gas interpretation among junior doctors in Aswan University Hospital: a clinical audit. Cureus. 2024; 16: e74906.

[6] Rizwan A, Sadiq T. The use of AI in diagnosing diseases and providing management plans: a consultation on cardiovascular disorders with ChatGPT. Cureus. 2023; 15: e43106.

[7] Mu Y, He D. The potential applications and challenges of ChatGPT in the medical field. International Journal of General Medicine. 2024; 17: 817–826.

[8] Scherr R, Halaseh FF, Spina A, Andalib S, Rivera R. ChatGPT interactive medical simulations for early clinical education: case study. JMIR Medical Education. 2023; 9: e49877.

[9] Puleio F, Lo Giudice G, Bellocchio AM, Boschetti CE, Lo Giudice R. Clinical, research, and educational applications of ChatGPT in dentistry: a narrative review. Applied Sciences. 2024; 14: 10802.

[10] Najjar R. Redefining radiology: a review of artificial intelligence integration in medical imaging. Diagnostics. 2023; 13: 2760.

[11] Försch S, Klauschen F, Hufnagl P, Roth W. Artificial intelligence in pathology. Deutsches Ärzteblatt International. 2021; 118: 194–204.

[12] Irfan B, Yaqoob A. ChatGPT's epoch in rheumatological diagnostics: a critical assessment in the context of Sjögren's Syndrome. Cureus. 2023; 15: e47754.

[13] Köroğlu EY, Fakı S, Beştepe N, Tam AA, Çuhacı Seyrek N, Topaloglu O,

et al. A novel approach: evaluating ChatGPT's utility for the management of thyroid nodules. Cureus. 2023; 15: e47576.

[14] Oca MC, Meller L, Wilson K, Parikh AO, McCoy A, Chang J, *et al*. Bias and Inaccuracy in AI chatbot ophthalmologist recommendations. Cureus. 2023; 15: e45911.

[15] Puladi B, Gsaxner C, Kleesiek J, Hölzle F, Röhrig R, Egger J. The impact and opportunities of large language models like ChatGPT in oral and maxillofacial surgery: a narrative review. International Journal of Oral and Maxillofacial Surgery. 2024; 53: 78–88.

[16] Sallam M, Al-Salahat K. Below average ChatGPT performance in medical microbiology exam compared to university students. Frontiers in Education. 2023; 8: 1333415.

[17] Hadi A, Tran E, Nagarajan B, Kirpalani A. Evaluation of ChatGPT as a diagnostic tool for medical learners and clinicians. PLOS ONE. 2024; 19: e0307383.

[18] Scheschenja M, Viniol S, Bastian MB, Wessendorf J, König AM, Mahnken AH. Feasibility of GPT-3 and GPT-4 for in-depth patient education prior to interventional radiological procedures: a comparative analysis. Cardiovascular and Interventional Radiology. 2024; 47: 245–250.

[19] Tangadulrat P, Sono S, Tangtrakulwanich B. Using ChatGPT for clinical practice and medical education: cross-sectional survey of medical students' and physicians' perceptions. JMIR Medical Education. 2023; 9: e50658.

[20] Cocci A, Pezzoli M, Lo Re M, Russo GI, Asmundo MG, Fode M, *et al*. Quality of information and appropriateness of ChatGPT outputs for urology patients. Prostate Cancer and Prostatic Diseases. 2024; 27: 103–108.

[21] Marino PL. Chapter 33: Acid-Base Disorders. Marino's The ICU Book (pp. 544–557). 5th edn. Philadelphia: Lippincott Williams & Wilkins; 2024.

[22] Kidney Disease: Improving Global Outcomes (KDIGO) CKD Work Group. KDIGO 2024 clinical practice guideline for the evaluation and management of chronic kidney disease. Kidney International. 2024; 105: S117–S314.

[23] Cederholm T, Barazzoni R, Austin P, Gonzalez MC, Fukushima R, Correia MITD, *et al*. GLIM criteria for the diagnosis of malnutrition – A consensus report from the global clinical nutrition community*. Journal of Cachexia, Sarcopenia and Muscle. 2019; 10: 207–217.

[24] McDonagh TA, Metra M, Adamo M, Gardner RS, Baumbach A, Böhm M, *et al*.; ESC Scientific Document Group. 2021 ESC Guidelines for the diagnosis and treatment of acute and chronic heart failure. European Heart Journal. 2021; 42: 3599–3726.

[25] ElSayed NA, Aleppo G, Aroda VR, Bannuru RR, Brown FM, Bruemmer D, *et al*. 6. Glycemic targets: standards of care in diabetes—2023. Diabetes Care. 2023; 46: S97–S110.

[26] Sood P, Paul G, Puri S. Interpretation of arterial blood gas. Indian Journal of Critical Care Medicine. 2010; 14: 57–64.

[27] Evans L, Rhodes A, Alhazzani W, Antonelli M, Coopersmith CM, French C, *et al*. Surviving sepsis campaign: international guidelines for management of sepsis and septic shock 2021. Intensive Care Medicine. 2021; 47: 1181–1247.

[28] Iolascon A, Bianchi P, Andolfo I, Russo R, Barcellini W, Fermo E, *et al*.; SWG of red cell and iron of EHA and EuroBloodNet. Recommendations for diagnosis and treatment of methemoglobinemia. American Journal of Hematology. 2021; 96: 1666–1678.

[29] Yee J, Frinak S, Mohiuddin N, Uduman J. Fundamentals of arterial blood gas interpretation. Kidney360. 2022; 3: 1458–1466.

[30] Cross JL, Choma MA, Onofrey JA. Bias in medical AI: implications for clinical decision-making. PLOS Digital Health. 2024; 3: e0000651.

[31] Wong A, Otles E, Donnelly JP, Krumm A, McCullough J, DeTroyer-Cooley O, *et al*. External validation of a widely implemented proprietary sepsis prediction model in hospitalized patients. JAMA Internal Medicine. 2021; 181: 1065–1070.

[32] Hirosawa T, Suzuki T, Shiraishi T, Hayashi A, Fujii Y, Harada T, *et al*. Adapting artificial intelligence concepts to enhance clinical decision-making: a hybrid intelligence framework. International Journal of General Medicine. 2024; 17: 5417–5422.

[33] Patel BN, Rosenberg L, Willcox G, Baltaxe D, Lyons M, Irvin J, *et al*. Erratum: author correction: human-machine partnership with artificial intelligence for chest radiograph diagnosis. NPJ Digital Medicine. 2019; 2: 129.

[34] Wang S, Zhao Z, Ouyang X, Liu T, Wang Q, Shen D. Interactive computer-aided diagnosis on medical image using large language models. Communications Engineering. 2024; 3: 133.

[35] Lee TC, Staller K, Botoman V, Pathipati MP, Varma S, Kuo B. ChatGPT answers common patient questions about colonoscopy. Gastroenterology. 2023; 165: 509–511.e7.

[36] Xue VW, Lei P, Cho WC. The potential impact of ChatGPT in clinical and translational medicine. Clinical and Translational Medicine. 2023; 13: e1216.

[37] Wang L, Zhang Z, Wang D, Cao W, Zhou X, Zhang P, *et al*. Human-centered design and evaluation of AI-empowered clinical decision support systems: a systematic review. Frontiers in Computer Science. 2023; 5: 1187299.

[38] Coskun AB, Elmaoglu E, Buran C, Alsaç SY. Integration of ChatGPT and e-health literacy: opportunities, challenges, and a look towards the future. Journal of Health Reports and Technology. 2024; 10: e139748.

[39] Koubaa A, Boulila W, Ghouti L, Alzahem A, Latif S. Exploring ChatGPT capabilities and limitations: a survey. IEEE Access. 2023; 11: 118698–118721.

[40] Sallam M, Salim NA, Barakat M, Al-Tammemi AB. ChatGPT applications in medical, dental, pharmacy, and public health education: a descriptive study highlighting the advantages and limitations. Narra J. 2023; 3: e103.

[41] Jones C, Thornton J, Wyatt JC. Artificial intelligence and clinical decision support: clinicians' perspectives on trust, trustworthiness, and liability. Medical Law Review. 2023; 31: 501–520.

[42] Sallam M. The utility of ChatGPT as an example of large language models in healthcare education, research and practice: systematic review on the future perspectives and potential limitations. Healthcare. 2023; 11: 887.

How to cite this article:
Uğur Serkan Çitilcioğlu, Barış Arslan, Hatice Kaya Özdoğan, Hakan Yalim, Ümit Kara. Is ChatGPT able to interpret arterial blood gas analysis? A comparative cross-sectional study. Signa Vitae. 2026; 22(1): 58-64. doi: 10.22514/sv.2025.138.