

Multidisciplinary AI-assisted Diagnosis and Management of Acute Abdominal Pain Syndrome: A Pilot Study

✉ Mete Üçdal¹, ✉ Evren Ekingen²

¹Etimesgut Şehit Sait Ertürk State Hospital, Clinic of Internal Medicine, Ankara, Türkiye

²Etimesgut Şehit Sait Ertürk State Hospital, Clinic of Emergency Medicine, Ankara, Türkiye

Dear Editor,

Acute abdominal pain syndrome represents one of the most common and diagnostically challenging presentations in emergency medicine, accounting for approximately 5-10% of all emergency department visits (1). The differential diagnosis spans diverse surgical emergencies including acute appendicitis, acute cholecystitis, intestinal obstruction, and volvulus, each requiring distinct management strategies ranging from conservative therapy to emergent surgical intervention (2). Diagnostic accuracy is paramount, as delayed recognition of surgical emergencies significantly increases morbidity and mortality. Recent advances in multimodal artificial intelligence (AI) systems have demonstrated promise in clinical decision support; however, their application across the heterogeneous spectrum of acute abdominal emergencies remains limited (3). We present a pilot evaluation comparing a multidisciplinary AI system with a single-model AI for diagnosing and managing acute abdominal pain syndrome.

We retrospectively analyzed 36 consecutive patients who presented to Etimesgut Şehit Sait Ertürk State Hospital (Ankara, Türkiye) with acute abdominal pain between September 2024 and January 2025. A multidisciplinary AI system integrating GPT-4V for radiological interpretation (ultrasonography, computed tomography findings), Med-PaLM 2 for emergency medicine assessment, BioGPT for intensive care unit (ICU) triage, and LLaMA 4 for surgical decision-making was compared against ChatGPT-5.0 and decisions made by the clinical team. GPT-4V was prompted using standardized radiological descriptor templates to identify pathognomonic imaging features: appendiceal diameter ≥ 6 mm with periappendiceal fat stranding (appendicitis); gallbladder

wall thickening ≥ 3 mm, pericholecystic fluid, and sonographic Murphy sign (cholecystitis); and whirl sign, coffee-bean sign, and transition point (volvulus). Med-PaLM 2 processed structured clinical parameters including vital signs, symptom onset, physical examination findings (rebound tenderness, guarding, McBurney's tenderness), and laboratory values (white blood cell, C-reactive protein, bilirubin, lipase, lactate) using PICO-formatted prompts. BioGPT assessed the necessity for ICU admission based on Sequential Organ Failure Assessment scores, the systemic inflammatory response syndrome criteria, hemodynamic parameters, and lactate levels according to the Surviving Sepsis Campaign 2023 guidelines. LLaMA 4 integrated American Society of Anesthesiologists physical status classification, comorbidity profile, and current guideline recommendations [World Society of Emergency Surgery (WSES), Tokyo Guidelines 2018, American Society for Gastrointestinal Endoscopy (ASGE)] to generate urgency-stratified surgical management decisions. All AI models, which did not have access to clinical team decisions, received standardized, pre-specified prompts that were developed a priori and applied uniformly across all 36 cases. Final diagnoses, established by consensus of the attending clinical team (emergency physician, surgeon, and radiologist), were confirmed by operative findings, pathological reports, or clinical course as appropriate, and served as the reference standard for all AI performance evaluations. AI model assessments were performed independently and were blinded to the clinical team's decisions, using standardized, pre-specified prompts applied uniformly across all cases. The cohort (mean age 52.4 ± 17.6 years; 55.6% male) comprised acute appendicitis (n=12, 33.3%), acute cholecystitis (n=10, 27.8%), paralytic ileus



Corresponding Author: Mete Üçdal MD, Etimesgut Şehit Sait Ertürk State Hospital, Clinic of Internal Medicine, Ankara, Türkiye

E-mail: meteuçdal@hacettepe.edu.tr **ORCID ID:** orcid.org/0000-0002-2321-2642

Cite this article as: Üçdal M, Ekingen E. Multidisciplinary AI-assisted diagnosis and management of acute abdominal pain syndrome: a pilot study. Eurasian J Emerg Med. 2026;25: 272-4.

Received: 09.01.2026

Accepted: 30.03.2026

Published: 01.04.2026



©Copyright 2026 The Author(s). The Emergency Physicians Association of Turkey / Eurasian Journal of Emergency Medicine published by Galenos Publishing House. Licensed by Creative Commons Attribution-NonCommercial-NoDerivatives (CC BY-NC-ND) 4.0 International License

(n=8, 22.2%), sigmoid volvulus (n=4, 11.1%), and cecal volvulus (n=2, 5.6%).

The multidisciplinary AI achieved 94.4% overall diagnostic accuracy (34/36), significantly outperforming ChatGPT-5.0 (80.6%, 29/36; p=0.012). For acute appendicitis, the multidisciplinary AI system correctly identified 11/12 cases (91.7%), including 3 cases of complicated appendicitis with perforation, compared with 9/12 (75.0%) identified by ChatGPT-5.0; integration of the Alvarado score further enhanced diagnostic precision. Detection of acute cholecystitis achieved 90.0% sensitivity (9/10) by accurately identifying gallbladder wall thickening and pericholecystic fluid on ultrasonography. The critical differentiation between surgical emergencies (appendicitis, cholecystitis, volvulus) and candidates for conservative management (ileus) showed 96.4% accuracy (27/28) for the multidisciplinary system, versus 82.1% (23/28) for ChatGPT-5.0 (p=0.024). Volvulus cases demonstrated a detection sensitivity of 100% with recognition of the coffee-bean and whirl signs (Table 1).

Analysis of guideline adherence revealed concordance rates of 94.4% (34/36) for multidisciplinary AI compared with 77.8% (28/36) for ChatGPT-5.0 (p=0.018). For appendicitis, the AI correctly recommended early laparoscopic appendectomy per WSES Jerusalem guidelines in all uncomplicated cases

(4). Acute cholecystitis management was aligned with Tokyo Guidelines 2018 criteria, with appropriate early cholecystectomy recommended in 9/10 cases (90.0%) versus 7/10 (70.0%) for ChatGPT-5.0 (5). Sigmoid volvulus cases received guideline-concordant endoscopic detorsion recommendations (100% vs 50.0%) per ASGE criteria (6). ICU triage accuracy was 91.7% versus 77.8% (p=0.032), with BioGPT demonstrating superior sepsis recognition per Surviving Sepsis Campaign criteria (7). Clinical outcomes were: surgical intervention in 22 patients (61.1%), endoscopic intervention in 3 patients (8.3%), and conservative management in 11 patients (30.6%); there was one in-hospital death (perforated appendicitis, 2.8%) (Table 2).

Our findings support the emerging paradigm of task-distributed AI in acute care. Schouten et al. (8) reported in a comprehensive scoping review analyzing 432 studies that multimodal AI models provide an average 6.2-point area under the curve improvement over unimodal systems, consistent with our 13.8 percentage point accuracy difference. Notably, the 94.4% diagnostic accuracy achieved by the multidisciplinary AI system is contextually comparable to published benchmarks: Kaczmarczyk et al. (9) reported 95.8% accuracy for collective human decision-making in complex diagnostic tasks; however, given the retrospective single-center design of the current pilot study, direct equivalence with human specialist performance cannot be inferred and

Table 1. Diagnostic performance of multidisciplinary ai across acute abdominal pain subtypes

Condition	n (%)	Multidisc. AI	ChatGPT-5.0	Sens. (%)	Key finding
Acute appendicitis	12 (33.3)	11/12	9/12	91.7 vs 75.0	Alvarado score
Acute cholecystitis	10 (27.8)	9/10	7/10	90.0 vs 70.0	Wall thickening
Paralytic ileus	8 (22.2)	8/8	7/8	100 vs 87.5	Diffuse dilation
Sigmoid volvulus	4 (11.1)	4/4	3/4	100 vs 75.0	Coffee-bean sign
Cecal volvulus	2 (5.6)	2/2	2/2	100 vs 100	Whirl sign
Overall	36 (100)	34/36	29/36	94.4 vs 80.6*	-

*p=0.012, Multidisc: Multidisciplinary, Sens: Sensitivity, AI: Artificial intelligence (multidisc. AI vs ChatGPT-5.0)

Table 2. Guideline adherence and clinical outcomes

Parameter	Multidisciplinary AI	ChatGPT-5.0
Overall guideline adherence, n/N (%)	34/36 (94.4)	28/36 (77.8)*
Appendicitis: early surgery (WSES)	11/12 (91.7)	9/12 (75.0)
Cholecystitis: early cholecystectomy (Tokyo)	9/10 (90.0)	7/10 (70.0)
Volvulus: endoscopic/surgical (ASGE/WSES)	6/6 (100)	4/6 (66.7)
Ileus: conservative management (WJES)	8/8 (100)	7/8 (87.5)
ICU triage accuracy (SSC criteria)	33/36 (91.7)	28/36 (77.8)
Surgical vs conservative differentiation	27/28 (96.4)	23/28 (82.1)
Decision time, minutes (mean ± SD)	2.2±0.4	2.9±0.5
In-hospital mortality, n (%)	-	1 (2.8)

*p=0.018 vs. multidisciplinary, AI. WSES: World Society of Emergency Surgery, Tokyo: Tokyo Guidelines 2018, ASGE: American Society for Gastrointestinal Endoscopy, WJES: World Journal of Emergency Surgery, SSC: Surviving Sepsis Campaign, AI: Artificial intelligence, ICU: Intensive care unit, SD: Standard deviation

prospective head-to-head comparison with clinical teams is required. Kroner et al. (10) demonstrated in Information Fusion that specialized model integration consistently outperforms single comprehensive systems in complex clinical scenarios requiring cross-domain expertise. Recent work by Wiest et al. (11) in *Nature Reviews Gastroenterology & Hepatology* highlighted that domain-specific large language models achieve optimal performance through integration rather than through general-purpose approaches. The 2025 multimodal AI review in *World Journal of Gastroenterology* emphasized the transformative potential of AI in precision diagnosis of gastrointestinal emergencies, with particular relevance to acute abdominal presentations. Our superior performance in differentiating appendicitis, cholecystitis, and ileus addresses a critical diagnostic challenge in which clinical and radiological overlap frequently delays appropriate management. This study has several limitations, including a small sample size, a single-center design, and retrospective data collection, all of which preclude definitive conclusions regarding generalizability; however, these preliminary results demonstrating high diagnostic accuracy within this limited pilot context support larger multicenter prospective validation trials to establish the clinical utility of multidisciplinary AI integration across the acute abdominal pain spectrum.

Ethics

Ethics Committee Approval: Approved by the Ankara Provincial Health Directorate Non-Interventional Clinical Research Ethics Committee (decision number: 2025-10-3, date: 24.10.2025).

Footnotes

Author Contributions

Concept: M.Ü., E.E., Design: M.Ü., E.E., Data Collection or Processing: M.Ü., Analysis or Interpretation: M.Ü., E.E., Literature Search: M.Ü., E.E., Writing: M.Ü., E.E.

Conflict of Interest: No conflict of interest was declared by the authors.

Financial Disclosure: The authors declared that this study received no financial support.

References

1. Dadeh AA. Factors associated with unfavorable outcomes in patients with acute abdominal pain visiting the emergency department. *BMC Emerg Med.* 2022;22:195.
2. Yew KS, George MK, Allred HB. Acute abdominal pain in adults: evaluation and diagnosis. *Am Fam Physician.* 2023;107:585-96.
3. Wu YM, Tang FY, Qi ZX. Multimodal artificial intelligence technology in the precision diagnosis and treatment of gastroenterology and hepatology: innovative applications and challenges. *World J Gastroenterol.* 2025;31:109802.
4. Di Saverio S, Podda M, De Simone B, Ceresoli M, Augustin G, Gori A, et al. Diagnosis and treatment of acute appendicitis: 2020 update of the WSES Jerusalem guidelines. *World J Emerg Surg.* 2020;15:27.
5. Yokoe M, Hata J, Takada T, Strasberg SM, Asbun HJ, Wakabayashi G, et al. Tokyo guidelines 2018: diagnostic criteria and severity grading of acute cholecystitis (with videos). *J Hepatobiliary Pancreat Sci.* 2018;25:41-54. Epub 2018 Jan 9.
6. Naveed M, Jamil LH, Fujii-Lau LL, Al-Haddad M, Buxbaum JL, Fishman DS, et al. American Society for Gastrointestinal Endoscopy guideline on the role of endoscopy in the management of acute colonic pseudo-obstruction and colonic volvulus. *Gastrointest Endosc.* 2020;91:228-235. Epub 2019 Nov 30. Erratum in: *Gastrointest Endosc.* 2020;91:721.
7. Dellinger RP, Rhodes A, Evans L, Alhazzani W, Beale R, Jaeschke R, et al. Surviving sepsis campaign. *Crit Care Med.* 2023;51:431-44.
8. Schouten D, Nicoletti G, Dille B, Chia C, Vendittelli P, Schuurmans M, et al. Navigating the landscape of multimodal AI in medicine: a scoping review on technical challenges and clinical applications. *Med Image Anal.* 2025;105:103621.
9. Kaczmarczyk R, Wilhelm TI, Martin R, Roos J. Evaluating multimodal AI in medical diagnostics. *NPJ Digital Medicine.* 2024;7:205.
10. Kroner F, Marikkar U, Parsons G, Szmul A, Mahdi A. Review of multimodal machine learning approaches in healthcare. *Information Fusion.* 2025;114:102690.
11. Wiest IC, Bhat M, Clusmann J, Schneider CV, Jiang X, Kather JN. Large language models for clinical decision support in gastroenterology and hepatology. *Nat Rev Gastroenterol Hepatol.* 2025;22:773-87.