Diagn Interv Radiol 2025; DOI: 10.4274/dir.2025.253680

ARTIFICIAL INTELLIGENCE AND INFORMATICS

LETTER TO THE EDITOR

©Copyright 2026 Author(s) - Available online at dirjournal.org.
Content of this journal is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

Reply: challenges of applying large language models to image-based interpretation in abdominal radiology

Alperen Elek[1]
Duygu Doğa Ekizalioğlu[2]
Ezgi Güler[2]

[1]Ege University Faculty of Medicine, İzmir, Türkiye

[2]Ege University Faculty of Medicine, Department of Radiology, İzmir, Türkiye

Dear Editor,

We thank our colleagues for their valuable comments[1] on our study.[2] Below, we address the points raised in the context of our study's aim and design choices, and we outline potential improvements for future research.

The primary aim of our work was to provide an objective baseline assessment of a general-purpose large language model (LLM) in the most straightforward and realistic "out-of-the-box," browser-based scenario. This approach was intentionally chosen to highlight the current limitations of model architectures and training data, particularly the absence of radiology-specific pretraining. Recent reviews[3-5] have highlighted that LLMs continue to face challenges in data scarcity, coarse visual embeddings, and limited explainability, all of which contribute to difficulties in capturing subtle signal or texture patterns.

We agree that radiologists base their decisions on volumetric, multiplanar, and multiphase image series. Our study, however, was designed as a standardized and ethically safe browser-based scenario to establish a "minimum requirement" baseline. Future studies will incorporate sequential and volumetric inputs, as well as multiphase evaluation. Of course, this will require LLMs to become technically capable of ingesting larger and more complex inputs.

To isolate image-based signal interpretation, patient history was excluded, allowing us to measure the model's pure image-based performance. In clinical reality, the integration of imaging and history is essential. Yet, as Bulut et al.[6] recently demonstrated, even when clinical findings were provided, overall accuracy remained low. Although their work focused on pneumothorax detection, these results still indicate that the performance of current models remains questionable, even in a clinical context.

Moving forward, we believe improvements should include: (i) the ingestion of sequential/volumetric and multiphase data; (ii) radiology-specific pretraining and/or adapter fine-tuning; (iii) structured prompt libraries and chain-of-thought reasoning; (iv) the integration of clinical metadata; and (v) blinded multicenter studies comparing different levels of radiologist expertise. Industry collaboration will be crucial to achieving these goals.

In conclusion, although many of the limitations highlighted by our colleagues were already acknowledged in our original manuscript, we view these comments as an opportunity to expand the scope of our subsequent studies and to establish a concrete roadmap for future research.

Handling editor: Mehmet Ruhi Onur

Corresponding author: Ezgi Güler

E-mail: gulerezgi@yahoo.com

Received 02 October 2025; accepted 13 October 2025.

Epub: 04.11.2025

Publication date: 02.11.2026

DOI: 10.4274/dir.2025.253680

You may cite this article as: Elek A, Ekizalioğlu DD, Güler E. Reply: challenges of applying large language models to image-based interpretation in abdominal radiology. *Diagn Interv Radiol*. 2026;32(6):767-768.

Conflict of interest disclosure

The authors declared no conflicts of interest.

References

1. Letter to the editor: challenges of applying large language models to image-based interpretation in abdominal radiology. Diagn Interv Radiol. Ahead of Print. [Crossref]
2. Elek A, Ekizalioğlu DD, Güler E. Evaluating Microsoft Bing with ChatGPT-4 for the assessment of abdominal computed tomography and magnetic resonance images. *Diagn Interv Radiol*. 2025;31(3):196-205. [Crossref]
3. Nam Y, Kim DY, Kyung S, et al. Multimodal large language models in medical imaging: current state and future directions. *Korean J Radiol*. 2025;26(10):900-923. [Crossref]
4. Zhang A, Zhao E, Wang R, Zhang X, Wang J, Chen E. Multimodal large language models for medical image diagnosis: challenges and opportunities. *J Biomed Inform*. 2025;169:104895. [Crossref]
5. Lanzafame LRM, Gulli C, Mazziotti S, et al. Chatbots in radiology: current applications, limitations and future directions of ChatGPT in medical imaging. *Diagnostics (Basel)*. 2025;15(13):1635. [Crossref]
6. Bulut B, Öz M, Genç M, et al. New frontiers in radiologic interpretation: evaluating the effectiveness of large language models in pneumothorax diagnosis. *PLoS One*. 2025;20(9):e0331962. [Crossref]