



Letter to the Editor: diagnostic sensitivity of ChatGPT for detecting hemorrhages in cranial computed tomography scans

Ali Salbas¹
 Raşit Eren Büyüktoka²
 Ali Murat Koc¹

¹İzmir Katip Çelebi University, Atatürk Training and Research Hospital, Department of Radiology, İzmir, Türkiye

²İzmir City Hospital, Department of Radiology, İzmir, Türkiye

Dear Editor,

We read the article by Bayar-Kapıcı et al.¹ regarding the diagnostic sensitivity of ChatGPT for detecting hemorrhages in cranial computed tomography (CT) scans with great interest. This study provides valuable insights into how large language models (LLMs) might support radiologists in acute clinical settings. As researchers focused on the intersection of artificial intelligence and radiology, we appreciate this contribution to the literature. Studies exploring the diagnostic boundaries of these systems are essential for understanding their potential and limitations. Recent studies indicate that multimodal LLMs demonstrate more limited performance in image interpretation than text-based reasoning.²⁻⁴ The findings of Bayar-Kapıcı et al.¹ reflect these diagnostic limitations within the context of cranial CT scans.

To facilitate a more nuanced understanding of the results presented and to support the reproducibility of this research, we believe that certain technical and procedural aspects warrant further clarification. Given that the figures in the article appear to reflect the web-based version of ChatGPT, we were curious if a new chat session was started for each of the 216 images. If multiple cases were analyzed within a single session, the model may have been influenced by prior contextual information, which could, in turn, affect diagnostic accuracy in subsequent questions.

Furthermore, just as specific window settings and slice thickness are essential for clinical interpretation by radiologists, these parameters may substantially influence how LLMs process and identify findings in radiological images. Information regarding these technical standards would be highly beneficial for future comparative studies.

Regarding the model version, although the study refers to ChatGPT-4V, the research timeline from March to May 2025 suggests the potential use of more recent iterations. Considering that ChatGPT-4o was introduced in May 2024 and GPT-4.5 in February 2025, it is possible that one of these later versions was the model actually utilized during the data collection period.⁵ Clarifying the specific version used would provide essential context for future benchmarking and would be particularly helpful for meta-analyses evaluating the evolution of model performance in neuroimaging.

Finally, knowing the image format and whether the files were obtained via direct screenshots or through software-based conversion from DICOM to JPG or PNG would be valuable for ensuring the reproducibility of future studies. Different image formats and compression methods can substantially impact image quality and data integrity.⁶ These technical variations may influence the diagnostic performance of LLMs, particularly when identifying subtle radiological findings.

In conclusion, the finding that diagnostic accuracy improves significantly with guided prompts is particularly encouraging. This underscores the potential of LLMs as supportive tools when embedded within supervised diagnostic systems. We believe these details will further enhance the impact of this successful study. We thank the authors for their inspiring work.

Corresponding author: Ali Salbas

E-mail: dralisalbas@gmail.com

Received 28 December 2025; revision requested 09 January 2026; accepted 16 January 2026.



Epub: 04.02.2026

Publication date:

DOI: 10.4274/dir.2026.263839

Conflict of interest disclosure

The authors declared no conflicts of interest.

References

1. Bayar-Kapıcı O, Altunışık E, Musabeyoğlu F, Dev Ş, Kaya Ö. Artificial intelligence in radiology: diagnostic sensitivity of ChatGPT for detecting hemorrhages in cranial computed tomography scans. *Diagn Interv Radiol.* 2026;32(1):27-32. [\[Crossref\]](#)
2. Nguyen D, Kim GHJ, Bedayat A. Evaluating ChatGPT's performance across radiology subspecialties: A meta-analysis of board-style examination accuracy and variability. *Clin Imaging.* 2025;125:110551. [\[Crossref\]](#)
3. Güneş YC, Cesur T, Çamur E, Günbey Karabekmez L. Evaluating text and visual diagnostic capabilities of large language models on questions related to the Breast Imaging Reporting and Data System Atlas 5th edition. *Diagn Interv Radiol.* 2025;31(2):111-129. [\[Crossref\]](#)
4. Payne DL, Purohit K, Borrero WM, et al. Performance of GPT-4 on the American College of Radiology in-training examination: evaluating accuracy, model drift, and fine-tuning. *Acad Radiol.* 2024;31(7):3046-3054. [\[Crossref\]](#)
5. OpenAI. OpenAI Help Center ChatGPT Release Notes. 2025. [\[Crossref\]](#)
6. Urbaniak, I. A. (2024). Using compressed JPEG and JPEG2000 medical images in deep learning: a review. *Applied Sciences.* 2024;14(22):10524. [\[Crossref\]](#)