Diagn Interv Radiol 2026; DOI: 10.4274/dir.2026.263959

ARTIFICIAL INTELLIGENCE AND INFORMATICS

LETTER TO THE EDITOR

Reply: Response to the letter regarding our study on ChatGPT and cranial CT hemorrhage detection

Olga Bayar Kapıcı[1]
Erman Altunışık[2]
Feyza Musabeyoğlu[2]
Şeyda Dev[2]
Ömer Kaya[3]

[1]Seyhan State Hospital, Clinic of Radiology, Adana, Türkiye

[2]University of Health Sciences Türkiye, Gaziantep City Hospital, Clinic of Neurology, Gaziantep, Türkiye

[3]Çukurova University Faculty of Medicine, Department of Radiology, Adana, Türkiye

Dear Editor,

We sincerely thank the authors[1] for their interest in our article[2] and for their constructive, methodologically focused comments. We welcome the opportunity to clarify the rationale of our study design and to further contextualize our findings regarding the use of a general-purpose vision–language model for intracranial hemorrhage assessment on non-contrast cranial computed tomography (CT).[2]

Regarding inter-session variability, we agree that probabilistic model behavior and output stochasticity are important considerations for any system contemplated for clinical decision support, and that reliability requires rigorous evaluation before real-world use.[3] Our study was conceived as an initial feasibility and characterization analysis rather than a formal reproducibility study. We used a fixed image set and a standardized question framework to describe baseline behavior and prompt sensitivity under controlled conditions.[2] Nevertheless, we concur that repeated measurements across independent sessions and, where possible, controlled inference settings are necessary to quantify repeatability and to estimate the stability of diagnostic metrics, especially if such tools are to be considered for high-stakes clinical pathways.[3]

Concerning the increase in sensitivity under guided prompting, we agree that this condition should not be interpreted as primary detection performance. The guided prompt was intentionally designed to evaluate the model's conditional classification behavior once the presence of hemorrhage is assumed, thereby separating detection from subtype identification under guidance. More broadly, accumulating evidence shows that prompt design and input conditions can materially affect multimodal model outputs and consistency, reinforcing the need to interpret prompt-conditioned improvements cautiously.[4] For clinical safety, we emphasized the unguided condition as the more appropriate reflection of autonomous behavior and highlighted the low baseline sensitivity as a key limitation. The main message remains that, in its current form, a general-purpose system is not suitable for unsupervised hemorrhage detection in acute care settings. This is consistent with emerging reports that demonstrate potential but also clinically relevant limitations when Generative Pre-Trained Transformer 4 with Vision (GPT-4V) class systems are applied to cranial CT hemorrhage detection tasks.[5]

With respect to the use of selected two-dimensional slices, we fully acknowledge that clinical CT interpretation is inherently volumetric and interactive, involving multi-slice review, windowing, and anatomical continuity assessment. We chose representative slices to standardize inputs and enable reproducible querying in a multimodal chat interface, recognizing that this approach does not replicate routine radiology workflows and may influence both sensitivity and specificity. Future investigations should prioritize a scan-level evaluation of full volumetric datasets, include workflow-representative testing where feasible, and analyze performance by hemorrhage subtype, size, and location to better characterize failure modes.[5]

Finally, we agree that GPT-4V was not optimized or trained specifically for radiologic imaging, and that benchmarking against specialized hemorrhage detection models and human

Corresponding author: Olga Bayar Kapıcı

E-mail: olgasahbayar@gmail.com

Received 24 February 2026; accepted 26 February 2026.

Epub: 11.03.2026

Publication date:

DOI: 10.4274/dir.2026.263959

You may cite this article as: Kapıcı OB, Altunışık E, Musabeyoğlu F, Dev Ş, Kaya Ö. Reply: Response to the letter regarding our study on ChatGPT and cranial CT hemorrhage detection. *Diagn Interv Radiol*. 11 March 2026 DOI: 10.4274/dir.2026.263959 [Epub Ahead of Print].

readers would strengthen its interpretability and clinical relevance. We also agree that post-deployment performance can drift or vary across settings, underscoring the importance of continuous monitoring frameworks when AI systems are integrated into clinical workflows.[6] Our intent was to evaluate a widely accessible, general-purpose model to delineate its limitations and prompt dependence—findings we believe to be informative for clinicians who may encounter such tools.

We appreciate the authors' contribution to this dialogue and share the view that future work should emphasize reproducibility, workflow-representative validation, clinically grounded benchmarking, and ongoing performance monitoring to ensure safety and reliability.

Conflict of interest disclosure

The authors declared no conflicts of interest.

References

1. Büyükceran EU, Seyfettin A, Babatürk A, Letter to the Editor: Artificial intelligence in radiology: diagnostic sensitivity of ChatGPT for detecting hemorrhages in cranial computed tomography scans. *Diagn Interv Radiol.* XXX March 2026 DOI: 10.4274/dir.2026.263940 [Crossref]
2. 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]
3. Rajpurkar P, Chen E, Banerjee O, Topol EJ. AI in health and medicine. *Nat Med*. 2022;28(1):31-38. [Crossref]
4. Sng GGR, Xiang Y, Lim DYZ, Tung JYM, Tan JH, Chng CL. A multimodal large language model as an end-to-end classifier of thyroid nodule malignancy risk: usability study. *JMIR Form Res*. 2025;9:e70863. [Crossref]
5. Zhang D, Ma Z, Gong R, et al. Using natural language processing (GPT-4) for computed tomography image analysis of cerebral hemorrhages in radiology: retrospective analysis. *J Med Internet Res*. 2024;26:e58741. [Crossref]
6. Rohren E, Ahmadzade M, Colella S, et al. Post-deployment monitoring of AI performance in intracranial hemorrhage detection by ChatGPT. *Acad Radiol*. 2025;32(10):6104-6113. [Crossref]