



Reply: evaluating text and visual diagnostic capabilities of large language models on questions related to the Breast Imaging Reporting and Data System (BI-RADS) Atlas 5th edition

Yasin Celal Güneş¹

Turay Cesur²

Eren Çamur³

Leman Günbey Karabekmez⁴

¹Kırıkkale Yüksek İhtisas Hospital, Clinic of Radiology, Kırıkkale, Türkiye

²Mamak State Hospital, Clinic of Radiology, Ankara, Türkiye

³Ankara 29 Mayıs State Hospital, Clinic of Radiology, Ankara, Türkiye

⁴Ankara Yıldırım Beyazıt University, Ankara Bilkent City Hospital, Department of Radiology, Ankara, Türkiye

KEYWORDS

BI-RADS, breast radiology, ChatGPT, large language models, retrieval-augmented generation

Dear Editor,

We sincerely thank the author for their insightful comments¹ and valuable suggestions regarding our manuscript titled "Evaluating text and visual diagnostic capabilities of large language models on questions related to the Breast Imaging Reporting and Data System Atlas 5th edition.² We appreciate the author's interest and the constructive proposal to incorporate retrieval-augmented generation (RAG) methodologies.¹

We fully agree that employing RAG could enhance the accuracy, contextual relevance, and reliability of responses generated by large language models (LLMs), particularly when addressing complex clinical scenarios such as those encountered in breast radiology.² As noted, RAG effectively mitigates limitations inherent in static models, including knowledge gaps and the risk of hallucinations, by dynamically retrieving relevant external information.^{3,4}

The examples provided by the author, including the promising results reported by Tozuka et al.⁵ in lung cancer tumor, node, metastasis staging using Google's NotebookLM with RAG, clearly demonstrate the considerable potential of this approach in radiological contexts.

Given these compelling points, we agree that incorporating RAG methodologies in future research would be highly valuable. Our current study did not include RAG, as its initial scope was limited to evaluating the inherent capabilities of standalone LLMs compared with radiologists based solely on the models' existing training data. Nonetheless, we acknowledge that future investigations involving retrieval-based augmentation could yield further insights into enhancing LLM performance in clinical radiology decision-making and guideline adherence.⁶

We appreciate the constructive input and believe that combining LLMs with RAG techniques in future work could substantially advance radiology education and clinical practice.

Thank you again for your thoughtful recommendations and contribution to this important discussion.

Footnotes

Conflict of interest disclosure

The authors declared no conflicts of interest.

Corresponding author: Yasin Celal Güneş

E-mail: gunesyasin cel@gmail.com

Received 24 March 2025; accepted 20 April 2025.



Epub: 20.05.2025

Publication date: 02.01.2026

DOI: 10.4274/dir.2025.253360

You may cite this article as: Güneş YC, Cesur T, Çamur E, Günbey Karabekmez L. Reply: evaluating text and visual diagnostic capabilities of large language models on questions related to the Breast Imaging Reporting and Data System (BI-RADS) Atlas 5th edition. *Diagn Interv Radiol.* 2026;32(1):25-26.

References

1. Kaba E. Retrieval-augmented generation for answering breast imaging reporting and data system (BI-RADS)-related questions with large language models. *Diagn Interv Radiol.* 2025. [\[Crossref\]](#)
2. 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 (BI-RADS) Atlas 5th edition. *Diagn Interv Radiol.* 2025;31(2):111-129. [\[Crossref\]](#)
3. Zakka C, Shad R, Chaurasia A, et al. Almanac - retrieval-augmented language models for clinical medicine. *NEJM AI.* 2024;1(2):10. [\[Crossref\]](#)
4. Steybe D, Poxleitner P, Aljohani S, et al. Evaluation of a context-aware chatbot using retrieval-augmented generation for answering clinical questions on medication-related osteonecrosis of the jaw. *J Craniomaxillofac Surg.* 2025;53(4):355-360. [\[Crossref\]](#)
5. Tozuka R, Johno H, Amakawa A, et al. Application of NotebookLM, a large language model with retrieval-augmented generation, for lung cancer staging. *Jpn J Radiol.* 2025;43(4):706-712. [\[Crossref\]](#)
6. Liu S, Wright AP, Patterson BL, et al. Using AI-generated suggestions from ChatGPT to optimize clinical decision support. *J Am Med Inform Assoc.* 2023;30(7):1237-1245. [\[Crossref\]](#)