Retrieval-augmented generation for answering Breast Imaging Reporting and Data System (BI-RADS)-related questions with large language models

 Esat Kaba

Recep Tayyip Erdoğan University Training and Research Hospital, Department of Radiology, Rize, Türkiye

Dear Editor,

I read with great interest the article 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" published in Diagnostic and Interventional Radiology.[1] The study explores how large language models (LLMs) respond to multiple-choice and some image-based questions based on the Breast Imaging Reporting and Data System (BI-RADS) 5th edition and presents the impressive results achieved by these models. Research of this kind is crucial to understanding the growing potential role of artificial intelligence technologies, particularly LLMs, in radiology decision-making processes. As a contribution to the valuable findings of this study, I believe that considering the retrieval-augmented generation (RAG) approach could be beneficial for more effectively combining information retrieval and text generation in such scenarios.

Retrieval-augmented generation enables language models to address existing knowledge gaps by accessing external information sources, allowing them to generate more accurate, up-to-date, and contextually appropriate text.[2] It consists of two main components: retrieval and generation. In the retrieval phase, queries are converted into vector format (e.g., using OpenAI embeddings) to create text embeddings. These vectors are then compared with pre-indexed documents using similarity search algorithms to retrieve the most relevant content (top-k retrieval). In the generation phase, the retrieved information is added to the input of the LLM, which then generates text based on this context.[3,4] This method holds strong potential, especially in fields that require complex information processing, such as radiology and detailed analyses based on BI-RADS.

In a study highlighting the effectiveness of this method in radiology, Tozuka et al.[5] performed tumor, node, metastasis staging of lung cancer using LLMs with and without RAG. In this study, Google's NotebookLM, a system incorporating RAG, achieved the highest performance in lung cancer staging. GPT-4o was also tested with and without RAG, and the use of RAG resulted in more successful outcomes across all stages of staging.[5] Given the limitations of current static models–such as knowledge gaps and the risk of generating misleading content (hallucinations)–RAG offers a promising approach to mitigate these issues and provide a more practical solution in both radiology education and clinical practice.

I would like to express my gratitude once again for your study's contribution to the field. I believe that incorporating RAG in future research, particularly in studies evaluating the knowledge level of LLMs on radiology guidelines, as in this case, could further enhance model accuracy and reliability.

Corresponding author: Esat Kaba

E-mail: esatkaba04@gmail.com

Footnotes

Conflict of interest disclosure

The author declared no conflicts of interest.

References

1. 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]

2. 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]

3. Toro S, Anagnostopoulos AV, Bello SM, et al. Dynamic retrieval augmented generation of ontologies using artificial intelligence (DRAGON-AI). *J Biomed Semantics*. 2024;15(1):19. [Crossref]

4. Zakka C, Shad R, Chaurasia A, et al. Almanac - Retrieval-Augmented language models for clinical medicine. *NEJM AI*. 2024;1(2):10.1056/aioa2300068. [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]