Article Type: Research Letter

Title: Automated Detection of Lymphadenopathy in Breast MRI Using a Large-Language Model

Co-authors: Nitin Chetla, BA[1], Tamer Hage, BS[2], Swapna Vaja, BS[3], Shivam Patel[4], Mihir Tandon, BA[5], Sai Samayamanthula, BA[1], Harshita Kacham MBBS[6], Luis Rodriguez BS[7], Kunal Sukhija, MD[8]

[1]University of Virginia School of Medicine, Charlottesville, Virginia, USA
[2]Virginia Tech, Blacksburg, Virginia, USA
[3]Rush Medical College, Chicago, IL, USA
[4]University of Virginia, Charlottesville, Virginia, USA
[5]Albany Medical College, Albany, NY, USA
[6]Osmania University, Hyderabad, India
[7]Johns Hopkins University School of Medicine, Maryland, USA
[8]University of California Los Angeles, California, USA

Corresponding Author:
Tamer Hage, BS
Email: Tamerwh@gmail.com
Phone: 703-894-8362
Virginia Tech, Virginia, USA
ORCID ID: 0009-0002-6407-954X

Funding: None

Acknowledgments: None.

Introduction:

Lymphadenopathy and suspicious lymph nodes in breast cancer patients are critical indicators for staging, prognosis, and treatment planning. Traditionally, radiologists assess these findings through manual interpretation of breast MRI. However, artificial intelligence (AI) models, particularly large multimodal models (LMMs), have increasingly demonstrated potential to enhance medical imaging analysis by providing consistent, objective assessments. AI-driven analyses could significantly reduce variability among readers and improve overall diagnostic efficiency. For instance, AI tools have already shown efficacy in improving breast cancer detection accuracy during mammography screening [1]. Furthermore, convolutional neural networks (CNNs) have achieved promising outcomes in identifying nodal metastasis through multiparametric MRI analysis, indicating AI's evolving role in advanced imaging diagnostics [2]. Despite these advancements, the capability of LMMs, which leverage broader contextual understanding beyond conventional CNN architectures, has yet to be thoroughly evaluated in lymphadenopathy detection specifically. This study assesses the performance of Gemini 2.0, an LMM, in automatically classifying lymphadenopathy or suspicious nodes in breast MRI for patients with biopsy-confirmed invasive breast cancer.

Methods:

Breast MRI images were sourced from the Duke-Breast-Cancer-MRI dataset, encompassing 200 MRI series, each representing an individual patient. Among these, 100 cases were labeled with confirmed lymphadenopathy or suspicious nodes, and 100 were negative. Each MRI series consisted of three to four post-contrast sequences originally in DICOM format and converted to PNG images for analysis.

To evaluate Gemini 2.0's diagnostic accuracy, each MRI series was individually assessed through Gemini's API, prompted with the query:

"This is a series of Breast MRI images from a patient with biopsy-confirmed invasive breast cancer. Based on the images, is there evidence of Lymphadenopathy or Suspicious Nodes? A) Lymphadenopathy or Suspicious Nodes are present B) Lymphadenopathy or Suspicious Nodes are not present. Answer this question with a single letter ONLY."

An automated Python-based recursive script systematically processed all images and documented the model's responses. Diagnostic accuracy was assessed through standard performance metrics: accuracy, precision, recall, F1-score, and confusion matrix analysis.

Results:

Gemini 2.0 achieved an overall diagnostic accuracy of 55% in classifying lymphadenopathy or suspicious nodes. When examining class-specific performance, the model performed better at identifying positive cases. For patients with lymphadenopathy or suspicious nodes (Class A), the model achieved a precision of 0.54, recall of 0.76, and F1-score of 0.63. In contrast, for patients without lymphadenopathy (Class B), precision was slightly higher at 0.59, but recall dropped significantly to 0.34, resulting in a lower F1-score of 0.43. These disparities illustrate that while the model is sensitive to detecting positive findings, it struggles to correctly identify true negative cases. The macro average and weighted average scores were consistent, both yielding precision, recall, and F1-scores around 0.55, reinforcing the overall limited diagnostic value of the model. As visualized in the confusion matrix, Gemini 2.0 correctly identified 76 of the 100 positive cases but misclassified 66 of the 100 negative cases, further highlighting the imbalance in performance across classes. (**Figure 1**)

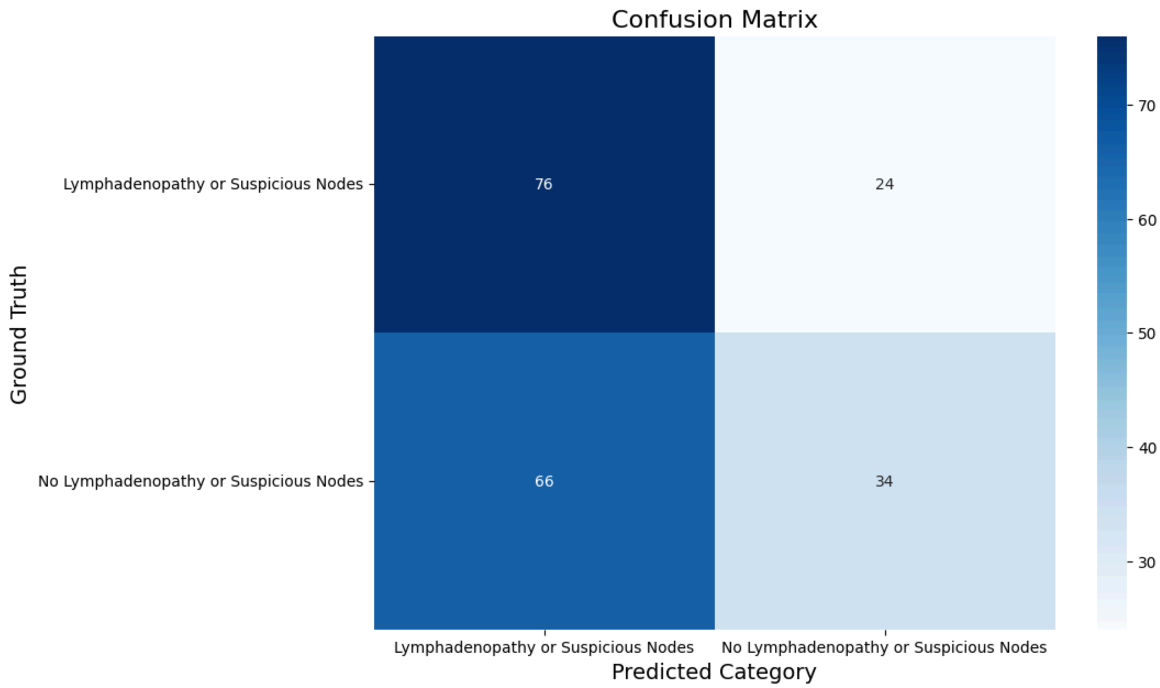

Figure 1. Confusion matrix illustrating Gemini 2.0's performance in classifying lymphadenopathy or suspicious nodes from breast MRI images.

Discussion:

The findings indicate that Gemini 2.0 while promising in sensitivity (76%) for detecting lymphadenopathy, demonstrates significant limitations with an overall accuracy of only 55%. This sensitivity suggests the model can effectively capture positive lymphadenopathy cases but simultaneously tends toward false positives, limiting its reliability as a standalone diagnostic tool. Conversely, its notably low specificity, of only 34% for negative cases, highlights substantial room for improvement.

The current results underscore that general LMMs like Gemini 2.0, though advantageous for multimodal context understanding, require further optimization to achieve clinical applicability in breast imaging. Potential enhancements include utilizing multi-stage prompting techniques, rigorous fine-tuning with extensive, annotated medical imaging datasets, or constructing hybrid diagnostic workflows that integrate both LMMs and specialized CNN-based models.

Notably, studies leveraging specialized CNNs for detecting axillary lymph node metastasis in breast cancer report substantially better diagnostic metrics, suggesting specialized models currently hold advantages over broader context LMMs in precise classification tasks [3].

Despite its limitations, this study illustrates the feasibility and value of employing interactive LMM-based AI as potential adjunctive tools in clinical breast MRI evaluation. For radiologists and oncologists, AI-driven systems could serve as supportive decision-making aids, potentially improving diagnostic efficiency and reducing interpretation variability. Further investigation is necessary to clearly delineate the roles and limitations of LMMs compared to conventional image-based AI approaches, guiding their integration into clinical practice.

References:

1. Díaz O, Rodríguez-Ruíz A, Sechopoulos I. Artificial Intelligence for breast cancer detection: Technology, challenges, and prospects. Eur J Radiol. 2024;175:111457. doi:10.1016/j.ejrad.2024.111457
2. Ren T, Lin S, Huang P, Duong TQ. Convolutional Neural Network of Multiparametric MRI Accurately Detects Axillary Lymph Node Metastasis in Breast Cancer Patients With Pre Neoadjuvant Chemotherapy. Clin Breast Cancer. 2022;22(2):170-177. doi:10.1016/j.clbc.2021.07.002
3. Ren T, Cattell R, Duanmu H, et al. Convolutional neural network detection of axillary lymph node metastasis using standard clinical breast MRI. Clin Breast Cancer. 2020;20(3):E301–E308. doi:10.1016/j.clbc.2019.11.009