Title: Evaluation of Gemini 2.0 AI in Classifying Breast Lesion Status from Dynamic Contrast-Enhanced MRI: A Preliminary Study

Coauthors: Nitin Chetla, Shivam Patel, Rahul Reddy, Kunal Sukhjia

Introduction

Breast cancer represents one of the most prevalent cancers affecting women worldwide, with early detection significantly improving prognosis and survival outcomes. Breast magnetic resonance imaging (MRI), particularly dynamic contrast-enhanced (DCE) MRI, is increasingly employed for its superior sensitivity in identifying breast lesions compared to conventional imaging modalities. However, accurately distinguishing between benign and malignant lesions remains clinically challenging and heavily reliant on radiologist expertise. Artificial intelligence (AI)-driven tools, including Gemini 2.0, offer the potential to augment diagnostic accuracy, reduce variability, and expedite clinical workflow. This study evaluates Gemini 2.0's capability in classifying breast lesion status from MRI images using prompts directed through its Application Programming Interface (API).

Methods

The study utilized breast MRI images sourced from the publicly available fastMRI Breast dataset developed by NYU Langone Health. Images were originally acquired as axial dynamic contrast-enhanced MRI sequences using a 3D Golden-angle Radial Sparse Parallel (GRASP) acquisition protocol. Each patient's MRI was initially provided in Digital Imaging and Communications in Medicine (DICOM) format, subsequently converted into Portable Network Graphics (PNG) files suitable for API processing.

Two distinct classification prompts were employed to evaluate Gemini 2.0's performance:

Prompt 1 classified lesion status as either "A) Benign or Malignant Lesion Status" or "B) Negative Lesion Status." This set included 100 MRI scans (50 with benign or malignant lesions, 50 negative).

Prompt 2 distinguished exclusively between "A) Benign Lesion Status" and "B) Malignant Lesion Status." This second assessment involved 180 distinct patient scans, split evenly with 90 benign and 90 malignant cases.

Each patient's image series was evaluated individually through Gemini 2.0's API using a recursive Python script, automatically prompting the system to provide a classification decision based solely on the specified prompts. API outputs were systematically recorded, and subsequent analyses included calculations of standard key performance metrics: accuracy, precision, recall, F1-score, and support.

Results

For Prompt 1, Gemini 2.0 demonstrated an overall accuracy of 50%. Although it achieved high recall (100%) in identifying benign or malignant lesions, the model incorrectly classified all negative cases as positive, resulting in zero recall for negative lesion status. Precision was 50%, with an F1-score of 0.67 for benign/malignant lesions and 0 for negative lesions.

For Prompt 2, Gemini 2.0 displayed an overall accuracy of 52%. It notably showed high recall (97%) for malignant lesions but very low recall (7%) for benign lesions, indicating a strong bias toward classifying lesions as malignant. Precision was moderate at 67% for benign and 51% for malignant lesions, with an overall weighted average F1-score of 0.39. These findings reflect Gemini 2.0's ability to identify malignant lesions with high sensitivity but also indicate a high false-positive rate when differentiating malignant from benign lesions.

Discussion

Gemini 2.0 exhibits potential as an adjunct diagnostic tool for breast lesion classification but demonstrates significant limitations in accurately distinguishing negative cases and differentiating benign from malignant lesions. The high false-positive rate observed, particularly in negative lesion classification, suggests the need for algorithmic refinements and extensive training on larger, more varied datasets incorporating broader clinical contexts.

Future research should involve prospective clinical trials comparing AI diagnostic outcomes to radiologist interpretations, which are essential to determining Gemini 2.0's real-world clinical utility.

In conclusion, Gemini 2.0 shows promise in MRI-based breast lesion classification, particularly in identifying lesion presence but requires additional refinement for clinical application in accurately classifying lesion type.

References:

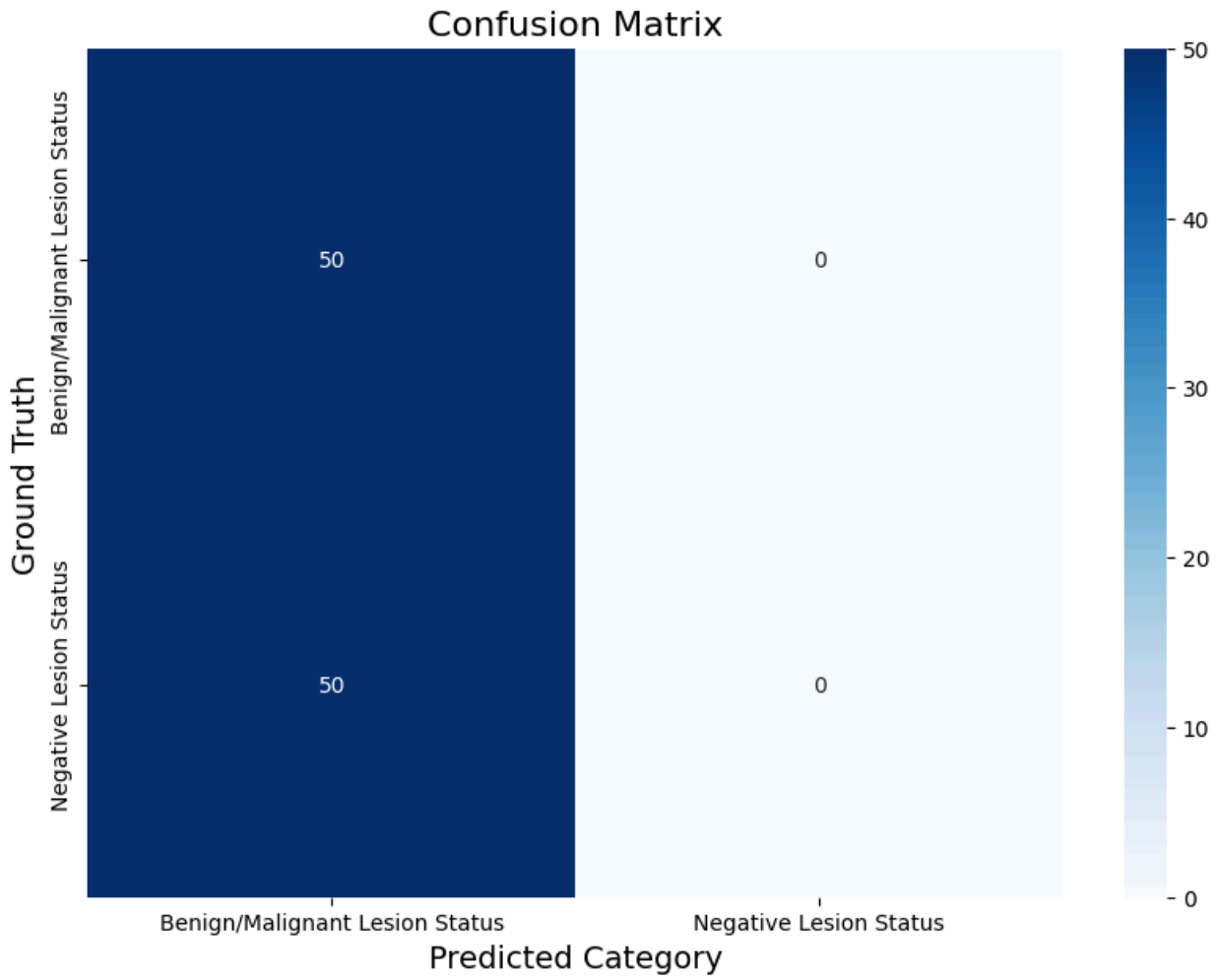
Dataset: https://fastmri.med.nyu.edu/

Prompt:
"This is a series of Breast MRI images from a patient. Based on the images, what would the lesion status be? A) Benign or Malignant Lesion Status B) Negative Lesion Status. Answer this question with a single letter ONLY. For example, if you believe there is evidence of Benign or Malignant Lesions, answer 'A' and NOTHING ELSE."

Methods:
Our study aimed to assess the abilities of Gemini 2.0 in correctly identifying the lesion status of a patient through the use of Gemini's Application Programming Interface (API). Breast MRI

images were obtained from the fastMRI Breast dataset. The images used for each patient included axial DCE-MR using a 3D GRASP sequence. The images were provided in a Dicom format and each image was converted into PNG format. In total, this study employed 100 MRI scans each from an individual patient where 50 were identified as having Negative Lesion Status, and the other 50 were identified as having Benign or Malignant Lesion Status.

To achieve reliable results, each series of MRI images for a patient was next processed through a recursive Python loop, which queried the API with the prompt "This is a series of Breast MRI images from a patient. Based on the images, what would the lesion status be? A) Benign or Malignant Lesion Status B) Negative Lesion Status. Answer this question with a single letter ONLY. For example, if you believe there is evidence of Benign or Malignant Lesions, answer 'A' and NOTHING ELSE.”. Output was recorded and key performance metrics including accuracy, precision, recall, F1-score, and support values were calculated for Gemini 2.0’s test outcomes.



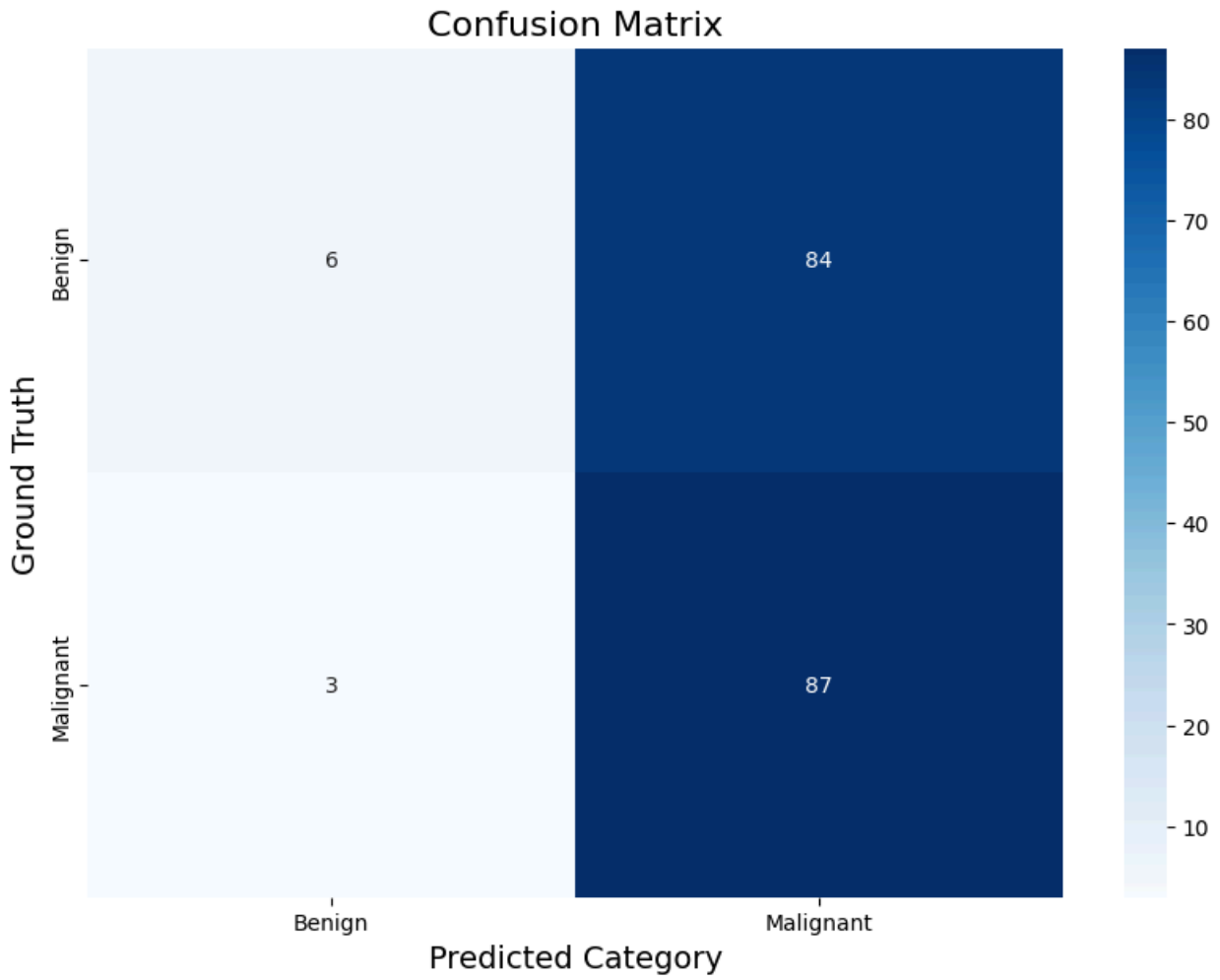
Metric	Precision	Recall	F1-score	Support
Benign/Malignant Lesion Status (A)	0.5	1.0	0.67	50
Negative Lesion Status (B)	0.0	0.0	0.0	50
Accuracy			0.5	100
Macro Avg	0.25	0.5	0.33	100
Weighted Avg	0.25	0.5	0.33	100

Second Prompt:
"This is a series of Breast MRI images from a patient. Based on the images, what would the lesion status be? A) Benign Lesion Status B) Malignant Lesion Status. Answer this question with a single letter ONLY. For example, if you believe there is evidence of Benign Lesions, answer 'A' and NOTHING ELSE."

Methods:

Our study aimed to assess the abilities of Gemini 2.0 in correctly identifying the lesion status of a patient through the use of Gemini's Application Programming Interface (API). Breast MRI images were obtained from the fastMRI Breast dataset. The images used for each patient included axial DCE-MR using a 3D GRASP sequence. The images were provided in a Dicom format and each image was converted into PNG format. In total, this study employed 180 MRI scans each from an individual patient where 90 were identified as having Benign Lesion Status, and the other 90 were identified as having Malignant Lesion Status.

To achieve reliable results, each series of MRI images for a patient was next processed through a recursive Python loop, which queried the API with the prompt "This is a series of Breast MRI images from a patient. Based on the images, what would the lesion status be? A) Benign Lesion Status B) Malignant Lesion Status. Answer this question with a single letter ONLY. For example, if you believe there is evidence of Benign Lesions, answer 'A' and NOTHING ELSE.". Output was recorded and key performance metrics including accuracy, precision, recall, F1-score, and support values were calculated for Gemini 2.0's test outcomes.



Metric	Precision	Recall	F1-score	Support
Benign (A)	0.67	0.07	0.12	90
Malignant (B)	0.51	0.97	0.67	90
Accuracy			0.52	180
Macro Avg	0.59	0.52	0.39	180
Weighted Avg	0.59	0.52	0.39	180