SIIM CMIMI24
Oct 21-22 | Boston, MA
Conference On Machine Intelligence In Medical Imaging

Classifying Common Breast Pain Symptoms for Patients Using a Large Language Model, ChatGPT

Hana Haver, MD, MSc, Mass General Brigham; Manisha Bahl, MD; Maggie Chung, MD

Introduction/Background

Breast pain is a common symptom for which diagnostic imaging evaluation is recommended based on clinical significance according to the American College of Radiology's (ACR) Appropriateness Criteria. Imaging is not recommended for clinically insignificant breast pain, which is defined as nonfocal, diffuse, or cyclical pain. This study aims to use ChatGPT GPT-4 (March 2023 release, OpenAI) to automate the classification of common breast pain symptoms based on clinical significance.

Methods/Intervention

The authors created a library of 150 breast pain symptoms representing breast pain variants described in the ACR Appropriateness Criteria, including clinically insignificant and significant pain, and non-pain-related clinically significant symptoms (e.g., palpable lump, pathologic nipple discharge). A zero-shot prompt for the LLM was developed to characterize breast concerns as clinically insignificant or clinically significant, "Use the ACR appropriateness criteria for breast pain. Respond with only is this 'clinically significant breast symptom' or 'not clinically significant symptom.'" Each breast symptom was submitted with the prompt in three independent tests in June 2024. Clinical significance was determined by the mode of the three tests and compared to the ground truth, established by radiologist consensus based on the ACR Appropriateness Criteria for breast pain.

Results/Outcome

ChatGPT GPT-4 assigned the appropriate clinical significance, in agreement with the breast imaging radiologists, in 74.7% (112/150) of breast pain symptoms (Table 1). ChatGPT GPT-4 correctly identified 89.1% (57/64) of clinically significant breast symptoms. Among instances where the model did not agree with the ground truth, the majority (81.6%; 31/38) were clinically insignificant cases that ChatGPT GPT-4 considered to be clinically significant. All 30 pain symptoms with non-pain-related clinically significant symptoms (e.g., palpable lump, pathologic nipple discharge) were correctly assessed by ChatGPT GPT-4 as clinically significant. Eighty-nine point three percent (134/150) of LLM-generated results were identical across three independent tests.

Conclusion

We demonstrate the first known potential application of an LLM to classify breast pain symptoms as clinically significant or clinically insignificant.

Statement of Impact

To automate ascertaining breast pain clinical significance, prior to patient scheduling, could influence decision-making about imaging evaluation, as only clinically significant symptoms would be indicated for imaging evaluation.

Table 1. Summary of clinical classification of breast symptoms by ChatGPT GPT-4: 150 breast pain vignettes. Ground truth clinical significance was determined by the consensus of two fellowship-trained breast radiologists and one breast radiology fellow.

Features of Breast Pain	**Ground truth clinical significance**	**Correct classification of clinical significance [%]**	**Agreement in assigned clinical significance across three tests [%]**
Non-focal	Not significant	66.7 (10/15)	80.0 (12/15)
Intermittent/Cyclical	Not significant	100.0 (15/15)	100.0 (15/15)
Focal/Intermittent	Not significant	80.0 (12/15)	80.0 (12/15)
Nonfocal/Constant	Not significant	6.7 (1/15)	86.7 (13/15)
Nonfocal/Intermittent	Not significant	100.0 (15/15)	93.3 (14/15)
Focal	Significant	53.3 (8/15)	66.7 (10/15)
Focal/Constant	Significant	100.0 (15/15)	100.0 (15/15)
Significant Pain + Concerning symptom	Significant	100.0 (15/15)	100.0 (15/15)
Nonsignificant Pain + Concerning symptom	Significant	100.0 (15/15)	100.0 (15/15)
Constant	Mixed significance*	40.0 (6/15)	86.7 (13/15)
All		74.7 (112/150)	89.3 (134/150)

*Refers to variation in the clinical characterization of breast symptoms with constant pain.

Keywords

Large language model; Breast Imaging; Clinical decision support