Evaluation of an Artificial Intelligence System for Detection of Invasive Lobular Carcinoma on Digital Mammography

Sylvia Arce [1] , Arunima Vijay [1] , Eunice Yim [1] , Lisa R. Spiguel [2] , Mariam Hanna [1]

1. Department of Radiology, University of Florida College of Medicine, Gainesville, USA 2. Department of Surgery, University of Florida College of Medicine, Gainesville, USA

Corresponding author: Sylvia Arce, stsangs@ufl.edu

Abstract

Introduction

Early breast cancer detection with screening mammography has been shown to reduce mortality and improve breast cancer survival. This study aims to evaluate the ability of an artificial intelligence computer-aided detection (AI CAD) system to detect biopsy-proven invasive lobular carcinoma (ILC) on digital mammography.

Methods

This retrospective study reviewed mammograms of patients who were diagnosed with biopsy-proved ILC between January 1, 2017, and January 1, 2022. All mammograms were analyzed using cmAssist® (CureMetrix, San Diego, California, United States), which is an AI CAD for mammography. The AI CAD sensitivity for detecting ILC on mammography was calculated and further subdivided by lesion type, mass shape, and mass margins. To account for the within-subject correlation, generalized linear mixed models were implemented to investigate the association between age, family history, and breast density and whether the AI detected a false positive or true positive. Odds ratios, 95% confidence intervals, and p-values were also calculated.

Results

A total of 124 patients with 153 biopsy-proven ILC lesions were included. The AI CAD detected ILC on mammography with a sensitivity of 80%. The AI CAD had the highest sensitivity for detecting calcifications (100%), masses with irregular shape (82%), and masses with spiculated margins (86%). However, 88% of mammograms had at least one false positive mark with an average number of 3.9 false positive marks per mammogram.

Conclusion

The AI CAD system evaluated was successful in marking the malignancy in digital mammography. However, the numerous annotations confounded the ability to determine its overall accuracy and this reduces its potential use in real-life practice.

Introduction

Breast cancer is the second leading cause of cancer death in women worldwide with over two million annual diagnoses [1]. Early breast cancer detection with screening mammography can reduce mortality by 20-35% and significantly improve breast cancer survival [2-4]. In the United States, almost 40 million mammographic exams are performed annually, most of which are done as part of screening programs [5]. This generates a large volume of mammograms that are manually interpreted by radiologists. However, manual analysis is both time and labor-intensive.

To assist radiologists and alleviate the problems associated with manual interpretation of mammograms, commercial computer-aided detection (CAD) systems for screening mammography were approved by the Food and Drug Administration in the 1990s [6]. Despite early promise, subsequent studies found that CAD did not improve the diagnostic accuracy of mammography [7-9]. A recent solution to improve diagnostic accuracy has been to integrate artificial intelligence into CAD software (AI CAD) [10]. The development of novel algorithms based on convolutional neural networks, commonly used for image recognition and classification, has allowed for significant advancements in AI CAD performance [10-11]. Research suggests

that this new AI CAD may be used as a "pre-reader" or "second reader" for mammograms to ease the burden of manual interpretation [10,12].

Several studies have shown that AI CAD systems evaluating digital mammograms can achieve a breast cancer detection rate comparable to that of a radiologist [13-15]. However, these studies had some limitations. Most studies were based on cancer-enriched datasets rather than actual screening mammogram data [14]. Furthermore, the radiologists that the AI CAD system was compared with showed a poor performance [15]. Finally, AI CAD algorithms were often not publicly available, making it difficult to understand how the AI CAD arrived at the result [16]. Several AI CAD programs exist, but most have not been validated by an independent third-party [16].

To the best of our knowledge, there is currently no research regarding the application of AI CAD for detection of invasive lobular carcinoma (ILC) on mammography. The lack of work in this area is likely due to the limited number of ILC cases, with ILC representing 5-15% of new breast cancer diagnoses annually [17]. Mammography can detect ILC with a sensitivity of 57-81% [17]. This relatively low sensitivity is partly due to the histological pattern of ILC, which presents diffusely with a single-cell pattern of growth throughout the stroma, and without significant desmoplastic reaction [18]. If AI CAD could improve the detection of ILC on mammography, this may prevent the need for more costly and time-consuming imaging studies such as MRI.

This study evaluates the ability of a commercially available AI CAD to identify biopsy-proven ILC on digital mammography. This study also assesses what characteristics are associated with false positive rates when using the AI CAD.

Materials And Methods

The University of Florida (UF) Institutional Review Boards issued approval (IRB# 202200981) for this retrospective study to review mammograms of UF patients diagnosed with biopsy-proven ILC between January 1, 2017, and January 1, 2022. Patients aged 30 to 84 years old, with mammograms consisting of four standard views (left cranial-caudal, left mediolateral oblique, right cranial-caudal, right mediolateral oblique), and a pathology report confirming ILC were included. Patients were excluded if they lacked a proper mammogram, if there was an inability to access CAD data for their imaging, or if their pathology report was negative for ILC.

The mammograms were analyzed using cmAssist® (CureMetrix, San Diego, California, United States), which is an AI CAD for mammography. cmAssist was trained using mammograms with biopsy-proven benign and malignant lesions and normal mammograms from multiple institutions [19]. cmAssist places markings on a mammogram, labels each marking as a "density" or a "calcification", and assigns each marking a quantitative score (neuScore™) on a scale of 0-100, with higher numbers corresponding to a greater likelihood of malignancy. Figure *1* is an example of how the cmAssist AI CAD marks a spiculated mass in a patient with heterogeneously dense breasts. No mammograms analyzed in this study were used in the development of cmAssist.
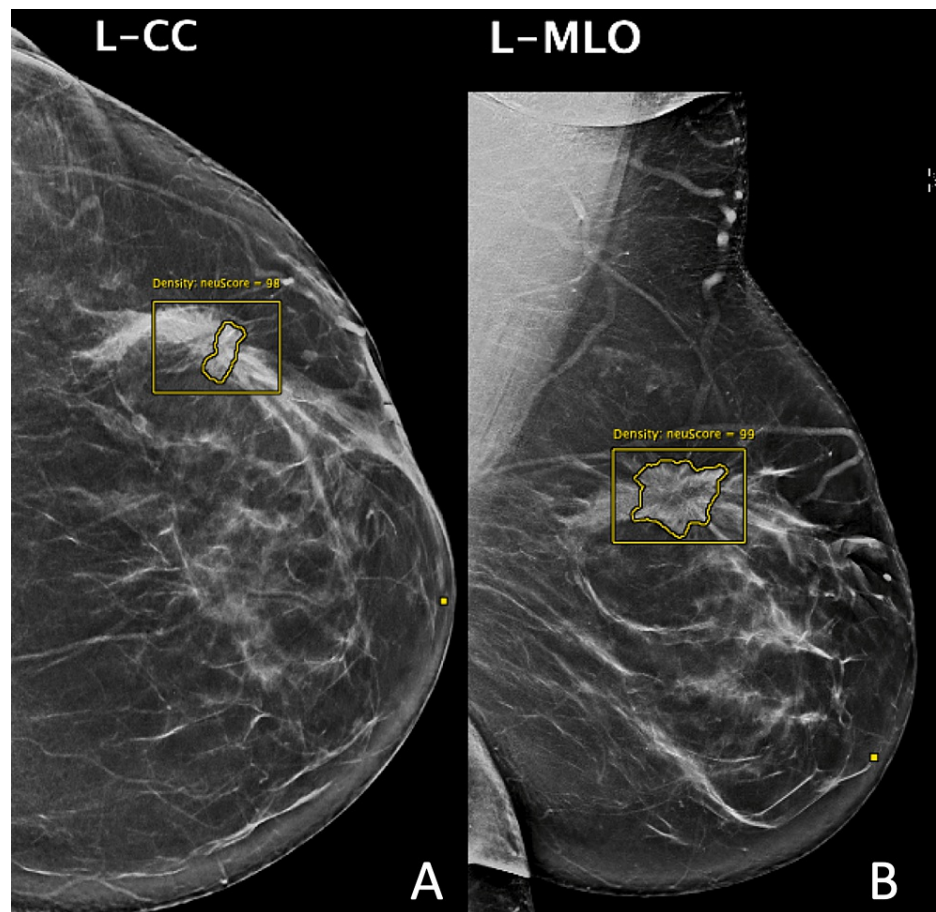
FIGURE 1: Example of cmAssist® AI CAD marking

Mammogram of a 55-year-old woman with heterogeneously dense breasts. (A) Left cranial-caudal (L-CC) view and (B) left mediolateral oblique (L-MLO) view of invasive lobular carcinoma (ILC) presenting as a mass with spiculated margins extending over a 5 x 2 cm region in the left breast. The AI CAD detected the mass as a density and gave it a neuScore™ of 99. Core needle biopsy revealed grade 2 ILC.

cmAssist®: CureMetrix, San Diego, California, United States

Variables included in the dataset for analysis were age, family history of breast cancer (positive if the patient had two or more first-degree relatives with breast cancer, one first-degree relative with a *BRCA* mutation, or a personal history of a *BRCA* mutation), radiologic findings of biopsy-proven cancer (mass, calcification, asymmetry, or distortion), mass size (less than 1 cm or not), mass margins, breast density (fatty, scattered fibroglandular, heterogeneously dense, and extremely dense), mammogram findings and impression, Breast Imaging Reporting & Data System (BI-RADS®) category, ultrasound findings and impression, pathology findings and impression from the breast biopsy, and surgical pathology results. The number and type of marking made by cmAssist were recorded for each mammogram. cmAssist was considered to have correctly marked the lesion location if the AI CAD marking overlapped an area of concern identified in the radiology report. Additionally, cmAssist was considered to have correctly marked the lesion type if it labeled a finding as a "calcification" and the radiologist called it a calcification, or if it labeled a finding as a "density" and the radiologist called it a mass, asymmetry, or distortion. The ground truth was defined in terms of whether ILC was present or absent, and confirmed by histopathology from the breast biopsy.

The AI CAD sensitivity for detecting ILC on mammography was calculated and further subdivided by lesion type, mass shape, and mass margins. To account for the within-subject correlation, generalized linear mixed models were implemented to investigate the association between age, family history, and breast density and whether the AI detected a false positive or true positive. Odds ratios (ORs), 95% confidence intervals (CIs), and p-values were calculated and statistical significance was set at $p < 0.05$.

Results

A total of 124 patients with biopsy-proven ILC were included in this study. Some patients had multiple ILC lesions in one or both breasts, so the total number of ILC lesions was 153. The AI CAD detected ILC on

2023 Arce et al. Cureus 15(5): e38770. DOI 10.7759/cureus.38770

3 of 7

mammography with a sensitivity of 80% (95%CI: 0.73-0.86). The AI CAD had the highest sensitivity for detecting calcifications (100%; 95%CI: 0.87-1.00), and the lowest sensitivity for detecting distortions (55%; 95%CI: 0.33-0.75) (Table 1). Analysis of imaging characteristics showed that the AI CAD had the highest sensitivity for detecting masses with irregular shapes (82%; 95%CI: 0.60-0.95) (Table 2). It also had a sensitivity of 86% (95%CI: 0.71-0.95) for detecting masses with spiculated margins (Table 3). Eighty-eight percent of mammograms had at least one false positive mark with an average number of 3.9 false positive marks per mammogram. Patients who were older (OR=1.02; 95%CI: 1.00-1.04), had a positive family history (OR=1.8; 95%CI: 1.02-3.31), or had extremely dense breast density (OR=7.4; 95%CI: 2.00-28.56) were more likely to have a false positive mark detected by the AI CAD (all p < 0.05) (Table 4).

Lesion Type	n = 153	Sensitivity (%)	95% Confidence Interval
Mass	74	80	0.68 – 0.89
Distortion	22	55	0.33 – 0.75
Asymmetry	23	78	0.56 – 0.93
Calcification	34	100	0.87 – 1.00

TABLE 1: Sensitivity of AI CAD detection on mammography by lesion type

AI CAD: artificial intelligence computer-aided detection

Mass Shape	n = 24	Sensitivity (%)	95% Confidence Interval
Oval	1	0	0.00 – 0.97
Round	1	100	0.03 – 1.00
Irregular	22	82	0.60 – 0.95

TABLE 2: Sensitivity of AI CAD detection on mammography by mass shape

AI CAD: artificial intelligence computer-aided detection

Mass Margins	n = 53	Sensitivity (%)	95% Confidence Interval
Spiculated	37	86	0.71 – 0.95
Obscured	8	88	0.47 – 1.00
Indistinct	5	60	0.15 – 0.95
Circumscribed	3	67	0.09 – 0.99

TABLE 3: Sensitivity of AI CAD detection on mammography by mass margins

AI CAD: artificial intelligence computer-aided detection

	Odds Ratio	95% Confidence Interval	p-value
Intercept	0.37	0.07 - 1.78	0.2089
Age (years)	1.02	1.00 - 1.04	0.0306
Family History of Breast Cancer	1.80	1.02 - 3.31	0.0479
Scattered vs Fatty	1.90	0.74 - 4.72	0.1692
Heterogeneously dense vs Fatty	1.71	0.66 - 4.24	0.2485
Extremely dense vs Fatty	7.38	2.00 - 28.56	0.0027

TABLE 4: Generalized linear mixed model for false positives of AI CAD detection

AI CAD: artificial intelligence computer-aided detection

Discussion

This study demonstrates that recent advances in AI algorithms have narrowed the gap between computer systems and human experts for detecting breast cancer in digital mammography. The sensitivity of the AI CAD for detecting ILC on mammography was 80%, which is comparable to the sensitivity of radiologists (between 57% and 81%) [17]. ILC most commonly presents as a spiculated, ill-defined mass or as an architectural distortion on mammography [17]. The AI CAD was successful at detecting masses with spiculated margins, suggesting that it is well suited to detect the subtle features of ILC in mammography. Many countries engage in double reading of mammograms to improve the sensitivity of detecting cancer [20]. Given the AI CAD's sensitivity for ILC detection, it is possible that this technology could be used as a second reader or as part of a radiologist's toolkit.

Despite its promise, AI CAD still has room for improvement. With a false positive rate of 3.9 marks per mammogram, radiologists using cmAssist must be able to identify and dismiss false positives to prevent an increase in screening recall rates. Also, this could increase the reading time of radiologists significantly. False positives are associated with patient anxiety, benign biopsies, unnecessary intervention or treatment, and increased healthcare spending [21]. The annual cost of false-positive mammograms and overdiagnosis of breast cancer is $4 billion among women aged 40-59 years in the United States [22]. Given that two-dimensional (2D) mammographic screening is the most commonly used screening method worldwide, the accuracy of mammographic readings must be preserved [23].

The limitations of this study include its retrospective design, single institution dataset, and the fact that the dataset only contains biopsy-proven cancer. Because all patients in this dataset had cancer, the specificity of the AI CAD could not be calculated, and a receiver operating characteristic analysis could not be performed. Since AI CAD is intended for use in a screening setting, the performance of AI CAD systems should also be tested on a screening population that includes normal mammograms.

Future work should focus on external validation studies that demonstrate AI CAD's performance in a screening setting using a population with diverse racial and ethnic backgrounds, breast cancer risk factors, imaging vendors, and imaging modalities (e.g. digital breast tomosynthesis vs digital mammography). This data could then be compared to the radiologists' performance in that setting. A prospective, randomized controlled trial involving multiple institutions is needed to further validate the AI CAD system.

Conclusions

The AI CAD system evaluated was successful in marking malignancies although numerous annotations confounded the ability to determine its overall accuracy and decreases its use in real-life practice. However, with improvement in technology, AI CAD can have important implications for screening mammography programs that employ double reading, which is more costly and labor-intensive than single readings. Adoption of the AI CAD as a second reader could significantly decrease the workload of radiologists. Although promising, the AI CAD system can be improved to minimize the number of false positive markings and requires further validation in a screening setting.

Additional Information

Disclosures

Human subjects: Consent was obtained or waived by all participants in this study. The University of Florida Institutional Review Boards issued approval IRB202200981 dated June 29, 2022. This study is approved as

2023 Arce et al. Cureus 15(5): e38770. DOI 10.7759/cureus.38770

5 of 7

exempt because it poses minimal risk and is approved under the following exempt category/categories: (4) (iii) Secondary research for which consent is not required: secondary research uses of identifiable private information or identifiable biospecimens, if at least one of the following criteria is met: (iii) The research involves only information collection and analysis involving the investigator's use of identifiable health information when that use is regulated under 45 CFR parts 160 and 164, subparts A and E, for the purposes of "health care operations" or "research" as those terms are defined at 45 CFR 164.501 or for "public health activities and purposes" as described under 45 CFR 164.512(b). Approval Includes, but is not limited to, Health Insurance Portability and Accountability Act (HIPAA) waiver to enroll approved in accordance with 45CFR § 164.512. **Animal subjects:** All authors have confirmed that this study did not involve animal subjects or tissue. **Conflicts of interest:** In compliance with the ICMJE uniform disclosure form, all authors declare the following: **Payment/services info:** All authors have declared that no financial support was received from any organization for the submitted work. **Financial relationships:** All authors have declared that they have no financial relationships at present or within the previous three years with any organizations that might have an interest in the submitted work. **Other relationships:** All authors have declared that there are no other relationships or activities that could appear to have influenced the submitted work.

References

1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A: Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin. 2018, 68:394-424. 10.3322/caac.21492
2. Marmot MG, Altman DG, Cameron DA, Dewar JA, Thompson SG, Wilcox M: The benefits and harms of breast cancer screening: an independent review. Br J Cancer. 2013, 108:2205-40. 10.1038/bjc.2013.177
3. Hakama M, Coleman MP, Alexe DM, Auvinen A: Cancer screening: evidence and practice in Europe 2008. Eur J Cancer. 2008, 44:1404-13. 10.1016/j.ejca.2008.02.013
4. Saadatmand S, Bretveld R, Siesling S, Tilanus-Linthorst MM: Influence of tumour stage at breast cancer detection on survival in modern times: population based study in 173,797 patients. BMJ. 2015, 351:h4901. 10.1136/bmj.h4901
5. Kooi T, Litjens G, van Ginneken B, et al.: Large scale deep learning for computer aided detection of mammographic lesions. Med Image Anal. 2017, 35:303-12. 10.1016/j.media.2016.07.007
6. Giger ML, Chan HP, Boone J: Anniversary paper: history and status of CAD and quantitative image analysis: the role of medical physics and AAPM. Med Phys. 2008, 35:5799-820. 10.1118/1.3013555
7. Lehman CD, Wellman RD, Buist DS, Kerlikowske K, Tosteson AN, Miglioretti DL: Diagnostic accuracy of digital screening mammography with and without computer-aided detection. JAMA Intern Med. 2015, 175:1828-37. 10.1001/jamainternmed.2015.5231
8. Fenton JJ, Taplin SH, Carney PA, et al.: Influence of computer-aided detection on performance of screening mammography. N Engl J Med. 2007, 356:1399-409. 10.1056/NEJMoa066099
9. Kohli A, Jha S: Why CAD failed in mammography. J Am Coll Radiol. 2018, 15:535-7. 10.1016/j.jacr.2017.12.029
10. Abdelrahman L, Al Ghamdi M, Collado-Mesa F, Abdel-Mottaleb M: Convolutional neural networks for breast cancer detection in mammography: a survey. Comput Biol Med. 2021, 131:104248. 10.1016/j.compbiomed.2021.104248
11. Oakden-Rayner L: The rebirth of CAD: how is modern AI different from the CAD we know?. Radiol Artif Intell. 2019, 1:e180089. 10.1148/ryai.2019180089
12. Tiedeu A, Daul C, Kentsop A, Graebling P, Wolf D: Texture-based analysis of clustered microcalcifications detected on mammograms. Digit Signal Process. 2012, 22:124-32. 10.1016/j.dsp.2011.09.004
13. Wu N, Phang J, Park J, et al.: Deep neural networks improve radiologists' performance in breast cancer screening. IEEE Trans Med Imaging. 2020, 39:1184-94. 10.1109/TMI.2019.2945514
14. Rodriguez-Ruiz A, Lång K, Gubern-Merida A, et al.: Stand-alone artificial intelligence for breast cancer detection in mammography: Comparison with 101 radiologists. J Natl Cancer Inst. 2019, 111:916-22. 10.1093/jnci/djy222
15. McKinney SM, Sieniek M, Godbole V, et al.: International evaluation of an AI system for breast cancer screening. Nature. 2020, 577:89-94. 10.1038/s41586-019-1799-6
16. Salim M, Wåhlin E, Dembrower K, et al.: External evaluation of 3 commercial artificial intelligence algorithms for independent assessment of screening mammograms. JAMA Oncol. 2020, 6:1581-8. 10.1001/jamaoncol.2020.3321
17. Johnson K, Sarma D, Hwang ES: Lobular breast cancer series: imaging. Breast Cancer Res. 2015, 17:94. 10.1186/s13058-015-0605-0
18. Lee JH, Park S, Park HS, Park BW: Clinicopathological features of infiltrating lobular carcinomas comparing with infiltrating ductal carcinomas: a case control study. World J Surg Oncol. 2010, 8:34. 10.1186/1477-7819-8-34
19. Watanabe AT, Lim V, Vu HX, et al.: Improved cancer detection using artificial intelligence: a retrospective evaluation of missed cancers on mammography. J Digit Imaging. 2019, 32:625-37. 10.1007/s10278-019-00192-5
20. Jairam MP, Ha R: A review of artificial intelligence in mammography. Clin Imaging. 2022, 88:36-44. 10.1016/j.clinimag.2022.05.005
21. Schaffter T, Buist DS, Lee CI, et al.: Evaluation of combined artificial intelligence and radiologist assessment to interpret screening mammograms. JAMA Netw Open. 2020, 3:e200265. 10.1001/jamanetworkopen.2020.0265
22. Ong MS, Mandl KD: National expenditure for false-positive mammograms and breast cancer overdiagnoses estimated at $4 billion a year. Health Aff (Millwood). 2015, 34:576-83. 10.1377/hlthaff.2014.1087
23. Kim HE, Kim HH, Han BK, et al.: Changes in cancer detection and false-positive recall in mammography

using Artificial Intelligence: A retrospective, multireader study. Lancet Digit Health. 2020, 2:e138-48.
10.1016/S2589-7500(20)30003-0

2023 Arce et al. Cureus 15(5): e38770. DOI 10.7759/cureus.38770

7 of 7