Artificial intelligence: competitor or best friend of breast cancer radiologists?

Yutonia Tang[1,2], Logan Leak[3]

Abstract

Around the world, over half a million women die of breast cancer every year. Screening mammograms were introduced to aid in early breast cancer diagnosis more than 30 years ago. Traditionally, disease determination from mammograms has been dependent upon radiologist interpretation. However, technologies have been developed to improve the sensitivity and specificity of mammograms beyond the capabilities of humans. Artificial intelligence (AI) using deep learning and convolutional neural networks has been employed in the past few years to improve early detection of breast cancer in digital mammography and breast tomosynthesis. Here, we aggregated data from past studies comparing the performance of experienced breast cancer radiologists, AI, and AI-informed radiologists. Our review showed that the performance of the AI system was greater than that of radiologists alone and that the combination of AI and radiologist input showed the greatest promise in detecting breast cancer. Our review demonstrates the potential of AI while also acknowledging potential pitfalls of this technology. Prospective studies in hospital settings are needed to fully establish the potential benefits of utilizing AI in breast cancer screening.

Abbreviations

AI, Artificial intelligence; AUC, area under the curve; DM, digital mammography; SD, standard deviation

Author Information

1, corresponding author (yutoniatang@fcds.org)
2, Forsyth Country Day School
3, Department of Biology, Stanford University, Stanford, CA 94305

Introduction

Breast cancer is one of the most common life-threatening diseases in the world. Every year, more than half a million women die of breast cancer[1]. Screening and early diagnosis are vital to decrease breast cancer mortality[2].

Computer-assisted detection of mammogram lesions was developed in the late 1990's; however, it was proved to not increase interpretive accuracy, mainly because of a low specificity[4]. In the past decade, the field of AI has been rapidly progressing due to the success of novel algorithms based on deep learning convolutional neural networks. AI is a broad feature that affects many aspects of life. It allows individuals to reevaluate ways to use information and leverage these emerging discoveries to better everyday life across several areas including healthcare. The combination of AI and humans has the potential to surpass modern methods.

There is a popular debate about whether radiologists or AI perform better in cancer early detection. When used in collaboration with radiologists, AI can assist to speed up and improve the accuracy of image analysis. As AI advances, it is conceivable that it will play an increasingly prominent role in breast cancer detection and treatment long into the future. It is important to investigate the use and accuracy of AI in breast cancer early diagnosis, the pitfalls of AI and how to avoid these problems,

and the tradeoff of accuracy and efficiency between AI and traditional radiologists. In this study, we retrospectively compare the breast cancer detection performance of AI to that of radiologists.

Methods

In this review, we conducted a systematic review of PubMed and used data from other published medical journals to perform a meta-analysis to assess whether AI algorithms could meet or beat the performance of radiologists in mammogram interpretation. We used area under the curve (AUC) as our primary metric of evaluating the algorithm accuracy and the performance. AUC is the level of separability, meaning the higher the AUC, the better the ability of the model to distinguish between classes. We looked through six retrospective studies back to 2013, comparing performance of diagnosis accuracy between AI and experienced radiologists. Most studies were done in the U.S and Europe. We summarized each study's sample size as well as the AUC of AI, radiologist alone, and combined AI and radiologist in table 1. Weighted averages (weighted by sample size) of the AUCs of AI, radiologist, and combined AI and radiologist were calculated along with standard deviations. The performance of AI was compared with radiologists with a non-inferiority null hypothesis based on differences in the AUC. T-test was used to assess the average AUC differences. The study was performed with as many data as could be gathered on PubMed search to get the most robust conclusion possible.

Table 1. Seven studies comparing performance of AI, radiologist and combined AI and radiologist

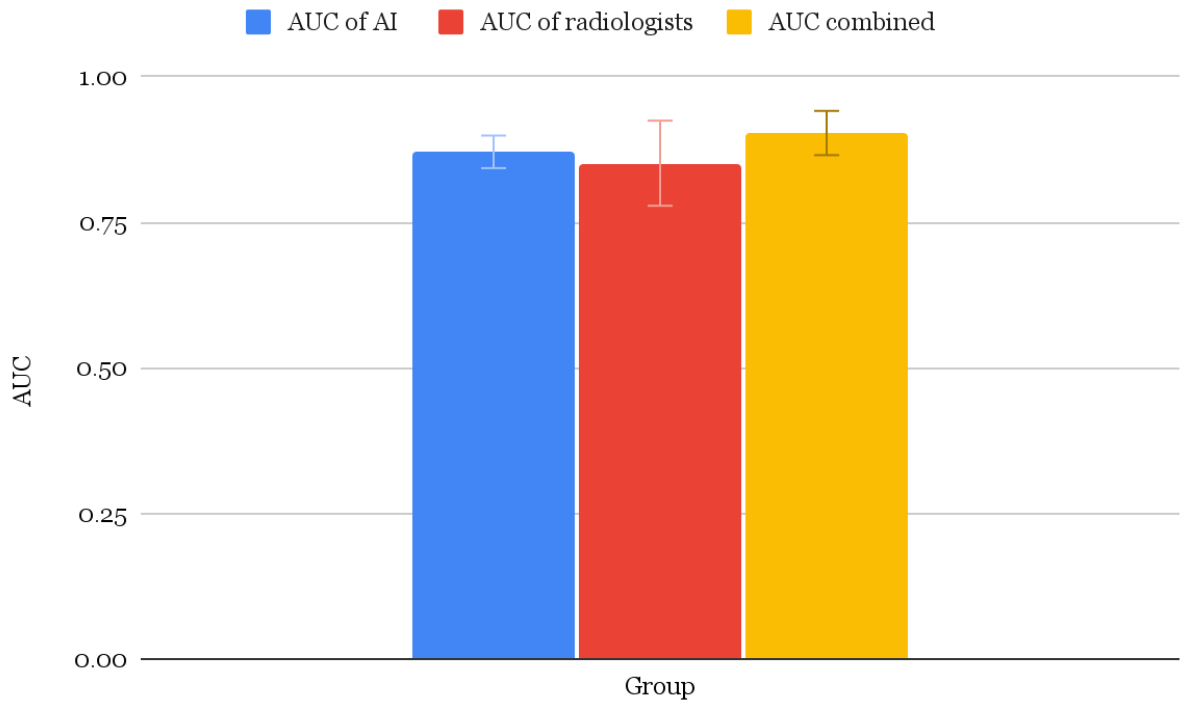
Studies	AUC of AI	AUC of radiologists	AUC combined	sample size
Rodríguez-Ruiz et al. (February 2019)[5]	0.84	0.814	NA	2,652
Rodríguez-Ruiz et al. (September 2019)[6]	0.89	0.87	0.89	240
Conant et al. (July 2019)[7]	NA	0.795	0.852	65
Sechopoulos et al. (February 2020)[8]	0.84	0.814	NA	2,000,000
Schaffer et al. (March 2020 (U.S.))[9]	0.858	0.905	0.942	144,231
Schaffter et al. (March 2020 (Sweden))[9]	0.903	0.985	0.942	166,578
Wu et al. (April 2020)[10]	0.895	0.778	0.891	229,426
Weighted average of AUC of all studies	0.874	0.0735	0.903	
Standard deviation	0.0285	0.0735	0.0386	

Results

On average, the AUC of AI alone was 0.87 (SD 0.028), the AUC of radiologists was 0.85 (SD 0.073), and the AUC of combined AI with radiologists was 0.9034 (SD 0.038). By comparing the AUC among three groups, top-performing AI algorithms combined with radiologist assessments resulted in a higher area under the curve of 0.9034. However, the difference is not statistically significant as determined by aTtest. T-tests of AI v.s. Radiologist, AI v.s. Combined, radiologist v.s. Combined all came back with P value >0.05. The result is summarized in Figure 1.

Our research shows that the tested AI system based on deep learning algorithms has non-inferior performance compared with radiologists alone for detecting breast cancer in mammography. These results were consistently observed across multi-center, multi-countries mammograms. The combination of AI algorithms with radiologists showed increased overall breast cancer diagnosis accuracy, though the T-test in our study was statistically nonsignificant. We acknowledge that our study has its limitations: all studies we looked at were done in the US or Europe, and all studies were retrospective studies. In the future, worldwide studies and prospective studies will help to overcome these pitfalls.

Figure 1. Comparison of AUC among AI, radiologist and combined AI and radiologist



Discussion

AI has the ability to significantly enhance the productivity and functionality of a wide range of sectors and is expected to have an expanding crucial role in the coming years. There has been improvement: with more research being done to avoid algorithmic error in the field of early breast cancer detection, there have been promising results that are comparable to radiologists. Additional advancements in algorithms, paired with information from more specific training sets, may result in AI playing a fundamental role in breast cancer detection. With AI engagement in breast cancer treatment, the future seems optimistic. The collaboration between radiologists and AI will also help reduce the recall rate, decrease financial burden, and avoid unnecessary diagnostic work-up.

We must also keep in mind that there exist pitfalls associated with using AI. A machine learning algorithm designed to assess risk of breast cancer, for instance, may have been built on biased data that fails to correctly reflect the varied population at risk for breast cancer. For example, many of the research

studies use samples from one country or a specific place. These countries, including the United States and various countries across Europe, tend to be more developed. As a consequence, certain groups of people may receive misleading or incorrect results from the program due to overfitting of the model to a specific sub-population. There is also a lack of a gold standard breast cancer detection method while constructing algorithms, which indicates that it has not been extensively tested or proven to be reliable. It is necessary to establish the data utilized to develop these systems is varied and inclusive in order to minimize the chance of bias in AI systems. There is also historical bias that uses data that is no longer accurate to the present time. Additionally, it is critical to assess and monitor AI systems on a regular basis to guarantee that there are no biased results.

All current AI studies in breast cancer diagnosis are in retrospective setting. Therefore, the performance and impact of AI in early breast cancer diagnosis needs further investigation with prospective studies

References

1. Bray, F et al. (2018). Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA: a cancer journal for clinicians, 68(6), 394-424.
2. NelsonHD, et al.Screening for breast cancer: an update for the U.S. Preventive Services Task Force. Ann Intern Med. 2009;151(10):727-737
3. Sadoughi, F., Kazemy, Z., Hamedan, F., Owji, L., Rahmanikatigari, M., & Azadboni, T. T. (2018). Artificial intelligence methods for the diagnosis of breast cancer by image processing: a review. Breast Cancer: Targets and Therapy, 219-230.
4. Lehman,CD, et al. Diagnostic accuracy of digital screening mammography with and without computer-aided detection. JAMA Intern Med. 2015;175(11):1828-1837.
5. Rodriguez-Ruiz et al. Detection of Breast Cancer with Mammography: Effect of an Artificial Intelligence Support System. Radiology. 2019 Feb;290(2):305-31
6. Rodriguez-Ruiz et al. Detection of Breast Cancer with Mammography: Effect of an Artificial Intelligence Support System. Radiology. 2019 Feb;290(2):305-31
7. Conant E et al. Radiol Artif Intell. 2019 Jul; 1(4): e180096
8. Sechopoulos I et al. Stand-alone artificial intelligence - The future of breast cancer screening? Breast 2020 Feb;49:254-260
9. Schaffter T et al. Evaluation of Combined Artificial Intelligence and Radiologist Assessment to Interpret Screening Mammograms. JAMA Network Open.
10. Wu N et al. Deep Neural Networks Improve Radiologists' Performance in Breast Cancer Screening. EE Trans Med Imaging. 2020 Apr;39(4):1184-1194.