Poster No. 158.05

International Surgical Week
The World's Congress of Surgery
isw2024.org
Kuala Lumpur, Malaysia
25-29 August 2024

Development of a deep-learning system for clinical diagnosis of BI-RADS4A and higher classifications in breast ultrasound imaging

Authors: Takamichi Yokoe; Tetsu Hayashida; Erina Odani; Masayuki Kikuchi; Aiko Nagayama; Tomoko Seki; Maiko Takahashi; Yuko Kitagawa

Institution: Department of Surgery, Keio University School of Medicine

Introduction

- Background: Breast ultrasound has significantly advanced over the past decade, with notable improvements in resolution and rapid image processing.
- Challenges: The diagnostic accuracy of breast ultrasound is heavily dependent on the observer's skill and experience. The BI-RADS classification was introduced to standardize reporting, but inter-observer variability remains a challenge.
- Purpose: This study aims to develop an AI system capable of distinguishing between BI-RADS 3 or lower and BI-RADS 4a or higher in breast ultrasound images and to verify its accuracy.

Materials and Methods

1. Study design

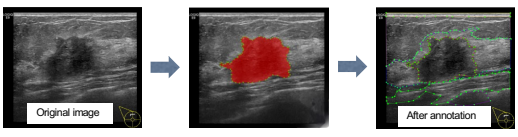
- This was a multicenter exploratory study aimed to establish an AI system for breast ultrasound diagnosis using a deep-learning technology and verify its accuracy.
- The AI diagnostic system determined whether the test image was BI-RADS 3 or lower and BI-RADS 4a or higher.
- These results were compared with the predetermined diagnoses made by human experts, and the sensitivity, specificity, the area under the curve (AUC) were calculated and used for evaluation.

2. Collection of ultrasound images

- Breast ultrasound images were collected using opt-out recruitment methods from eight facilities.
- The images included those from women with histologically confirmed benign or malignant breast tumors or those clinically diagnosed with benign tumors after a follow-up of six months or more.
- Images were selected by breast cancer specialists certified by the Japanese Breast Cancer Society, with each image assigned information about the institution, diagnosis, histological type, and ultrasound machine manufacturer.
- Images with Doppler or elastography or those technically inappropriate for evaluation were excluded.

3. Image evaluation and annotation

- Ultrasound images were evaluated by two independent, certified evaluators who marked all observed lesions and provided assessments based on the 5th edition of BI-RADS. Lesion-by-lesion assessments were collected and analyzed.
- The annotation process used Labelme software. Statistical calculations were performed using Python 3.6 with NumPy and scikit-learn libraries.



Conclusion

This is the first attempt to establish an AI system to classify BI-RADS3 or lower and BI-RADS4 or higher successfully, providing important implications for clinical actions.

Results

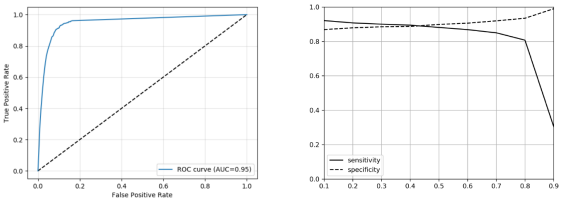
1. Establishment of the AI diagnosis system

- A total of 8,670 lesions were targeted from 7,194 images (training data: 4,028 images with 5,014 lesions, test data: 3,166 images with 3,656 lesions).

	Training data n=3279			Test data n=2730		
BI-RADS	malignant	benign	% of malignant	malignant	benign	% of malignant
1	0	0	0%	0	1470	0%
2	0	437	0%	0	176	0%
3	0	579	0%	0	278	0%
4a	44	701	6%	16	317	5%
4b	291	653	31%	148	251	37%
4c	978	127	89%	420	48	90%
5	1189	15	99%	524	8	98%

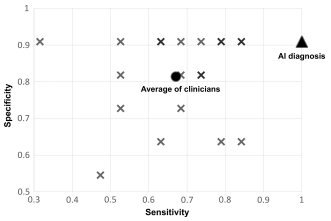
2. Validation of the diagnostic accuracy by AI

- At the optimal balance between sensitivity and specificity, the AUC is 0.95, with a sensitivity of 91.2% and a specificity of 90.7%.



3. Comparison of diagnostic performance between clinicians and AI

- The mean sensitivity and specificity of the diagnosis made by the clinicians were 67.1% (31.6%-84.2%) and 81.4% (47.4%-90.9%), respectively.



Discussion

- Although many reports exist on AI-based diagnosis of breast ultrasound, most focus on technical aspects such as deep learning algorithms, with few addressing clinical applications.
- While there are many reports on distinguishing benign from malignant lesions in static images, the critical clinical issue is determining appropriate medical management for patients with abnormalities in breast ultrasound.
- Despite some biases, the results are promising for clinical application.

This study has been published in a paper.

Tetsu Hayashida, et al. Cancer Sci. 2022 Oct;113(10):3528-3534.