Original Article

Artificial Intelligence Development for Detecting Microcalcification within Mammography

Chalida Aphinives MD[1], Potchavit Aphinives MD[2], Supajit Nawapan MD[1]

[1] Department of Radiology, Faculty of Medicine, Khon Kaen University, Khon Kaen, Thailand

[2] Department of Surgery, Faculty of Medicine, Khon Kaen University, Khon Kaen, Thailand

Background: Artificial Intelligence (AI) is the recently advanced technology in machine learning that is increasingly used to help radiologists, especially when working in arduous conditions. Microsoft Corporation offered a free-trial service called Custom Vision to develop AI for images.

Objective: To study the possibility of AI development from free-trial service for detecting microcalcification within mammography.

Materials and Methods: Radiological images of breast cancer-proven patients who underwent mammography between 2018 and 2019 were used to train AI to detect microcalcification. The training processes were divided into five iterations of 30, 60, 100, 130, and 160 lesion datasets. After each training, the AI was tested as "Performance Per Tag" and clinical performance. There were three types of training, quick, 1-hour, and 2-hour trainings.

Results: The present study included 116 microcalcification images with 206 lesions from 56 breast cancer patients. The 160-tag iteration presented the best performance with a precision of 80.0%, a recall of 12.5%, a mean average precision of 30.5%, and a prediction rate of 32.14%. The performance of the 1-hour training was better than the quick training but was not different from the 2-hour training.

Conclusion: Health personnel can easily develop AI for the detection of microcalcification in mammography. However, the AI development is further required, and the result should be interpreted along with radiologist.

Keywords: AI, Microcalcification, Mammography, Machine learning

Received 25 December 2020 | Revised 8 January 2021 | Accepted 11 January 2021

J Med Assoc Thai 2021;104(4):560-4

Website: http://www.jmatonline.com

Breast cancer is the most common cancer in Thai women with 19,510 new cases in 2018 or 22.8% of all new cancer cases in Thai women(1). In western countries, breast cancer screening programs using mammography reduced the mortality from breast cancer(2-4). Breast cancer is of increasing concern by Thai women year after year. Screening mammography is included in some private hospitals in the annual health check-up but is also requested by many women with breast discomfort.

Several companies have developed computer-aided detection and diagnosis (CAD) for mammography since the late 1960s(5). After the long duration of development, CAD may help radiologists to detect and diagnose microcalcifications. The use of CAD leads to a slightly higher detection rate and longer evaluation times(6) but does not seems to improve diagnosis accuracy(7) in clinical practice.

Artificial intelligence (AI), which is the recently advanced technology in machine learning (ML), may improve CAD for mammography in clinical practice(8-10).

Microsoft Corporation introduced its cloud platform called Azure, supplying over 100 services, some are free trial and some are always free. ML is a feature-based algorithm of the AI before the advent of deep learning (DL), which is the main algorithm for developing AI for medical imaging. Under the budget-constrained situation in the authors' hospital, an attempt was made to develop AI for detecting microcalcification within mammography under "custom vision", which is one of the free trial services from Azure. It was aimed to test the possibility of this service, therefore, the present study was a pilot study conducted solely by clinicians with some guidance from one computer scientist.

Correspondence to:

Aphinives P.

Department of Surgery, Faculty of Medicine, Khon Kaen University, 123 Group 16, Mittraparp Road, Muang, Khon Kaen 40002, Thailand.

Phone: +66-43-363252, +66-89-7115475, **Fax**: +66-43-348393

Email: potchavit@gmail.com

How to cite this article:

Aphinives C, Aphinives P, Nawapan S. Artificial Intelligence Development for Detecting Microcalcification within Mammography. J Med Assoc Thai 2021;104:560-4.

doi.org/10.35755/jmedassocthai.2021.04.11635

Materials and Methods

The present study was approved by the Khon Kaen University Ethics Committee for Human Research based on the Declaration of Helsinki and the ICH Good Clinical Practice Guidelines with reference number HE621498. No informed consents were needed because this was a retrospective study of the stored images in the hospital PACS database.

Radiological images of breast cancer-proven patients who underwent conventional mammography, digital mammography, and tomosynthesis in 2018 and 2019 were retrieved from the hospital PACS database. Only 116 microcalcification-presented images with 206 lesions or groups of microcalcifications from 56 patients were included in the present study.

The training processes were divided into five iterations of 30, 60, 100, 130, and 160 lesion datasets. The images were uploaded, and every lesion was manually tagged to help training the object detector. If an image had three groups of microcalcifications, it would add three tags (lesions) in this dataset. After each training session for one hour, this AI was evaluated with a testing dataset from 10 different images that were not included in the training dataset. The testing dataset was composed of 28 lesions.

The system was presented as the "Performance Per Tag" after the training process into three values:

1. Precision indicates the fraction of identified images that were correct. For example, if the model recognized lesions in 100 images, and 99 of them were actual lesions, then the precision would be 99%.

2. Recall indicates the fraction of actual images that were correctly recognized. For example, if there were 100 images containing lesions, and the model recognized 80 of them, the recall would be 80%.

3. Mean average precision tells the overall precision of the object detector at finding lesions.

The clinical performance of this AI is presented with the amount and percentage of correct detections among the five iterations of training.

Another factor that affects the AI performance should be the duration of training. One-hour training was used as a standard training process as previously mentioned. Then "quick training" and "2-hour training" iterations were performed with the 160 lesions dataset and these performances in both "Performance Per Tag" and clinical performance were compared.

Table 1. The "Performance Per Tag" of 5 iterations with 1-hour training

Dataset (tags)	Precision (%)	Recall (%)	MAP (%)
1. 30	0.0	0.0	2.1
2. 60	0.0	0.0	1.3
3. 100	33.3	5.0	7.6
4. 130	0.0	0.0	4.7
5. 160	80.0	12.5	30.5

MAP=mean average precision

Table 2. The clinical performance of 5 training datasets

Test	Train				
	30 tags	60 tags	100 tags	130 tags	160 tags
1 (4 lesions)	0	0	2	2	2
2 (1 lesion)	0	0	1	1	1
3 (1 lesion)	0	0	0	0	0
4 (1 lesion)	0	0	0	0	0
5 (6 lesions)	1	1	1	1	1
6 (5 lesions)	0	0	0	0	0
7 (3 lesions)	1	0	0	1	0
8 (2 lesions)	0	0	0	2	1
9 (3 lesions)	0	0	0	0	2
10 (2 lesions)	0	0	0	1	2
Total (28 lesions)	2	1	4	8	9
Percent	7.14	3.57	14.29	28.57	32.14

Results

The present study included 116 microcalcification-presented images with 206 lesions from 56 breast cancer patients. The "Performance Per Tag" of five iterations of 30, 60, 100, 130, and 160 tags are presented in Table 1. Tenimages with 28 lesions were tested in each iteration. The false-positive predictions from 30-tag iteration are shown in Figure 1. The 100-tag iteration shows the true-positive predictions in Figure 2. The 130-tag iterations shows the true-positive prediction in Figure 3, and the false-positive prediction in Figure 4. The clinical performance of each training with the same testing dataset of 10 images with 28 lesions, are presented in Table 2.

The "Performance Per Tag" was improved from quick training iteration to 1-hour training iteration, but the 2-hour training iteration showed the same values as 1-hour training iteration (Table 3). The clinical performance showed the same results as the "Performance Per Tag" (Table 4).

Discussion

AI is developed from computer algorithms to simulate intelligent behavior that is capable of

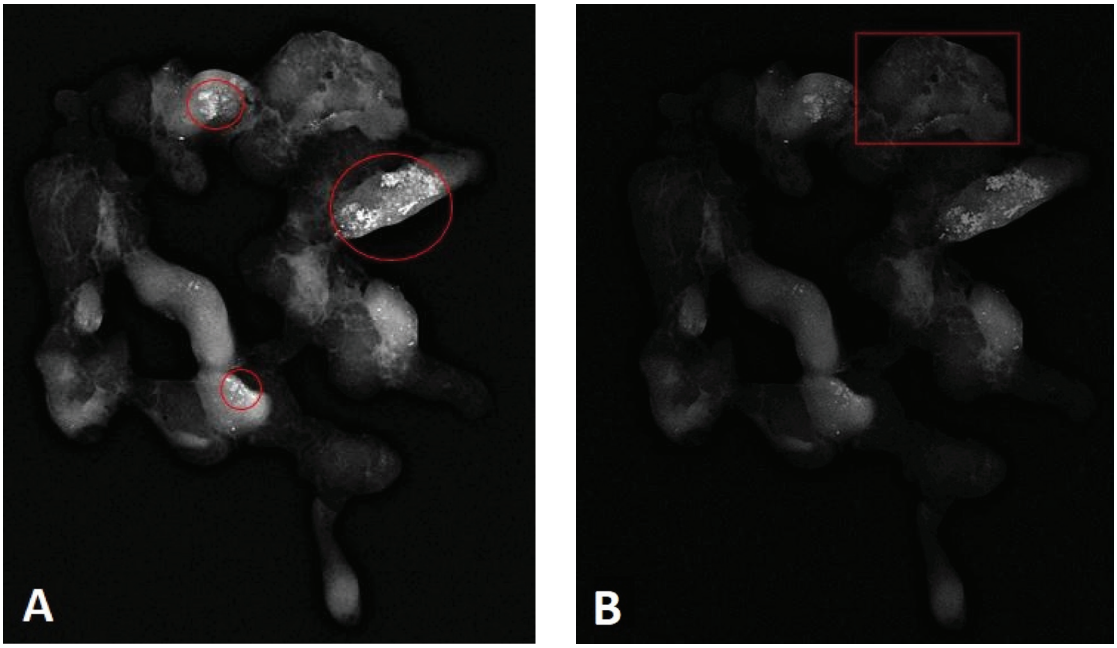

Figure 1. The false-positive prediction from the 30-tag iteration: (A) A radiologist identified 3 lesions (circle) in a testing image, (B) AI predicted 1 lesion (rectangle) at a false position.

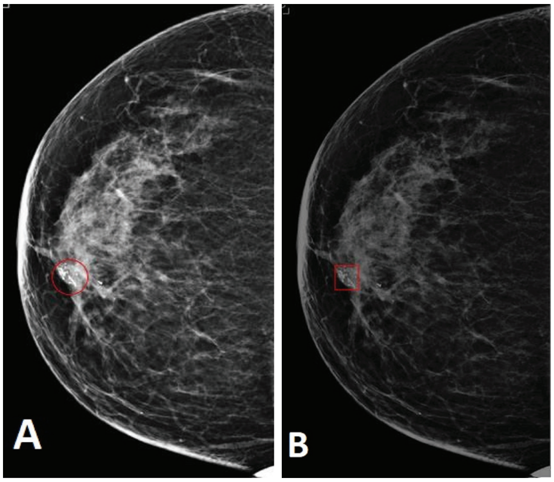

Figure 2. The true prediction from 100-tag iteration: (A) A radiologist identified 1 lesion (circle) in a testing image, (B) AI predicted 1 lesion (rectangle) at the true position.

Table 3. The "Performance Per Tag" of 3 different durations of training with 160 lesions dataset

Duration	Precision (%)	Recall (%)	MAP (%)
Quick	50.0	3.1	10.4
1 hour	80.0	12.5	30.5
2 hours	80.0	12.5	30.5

MAP=mean average precision

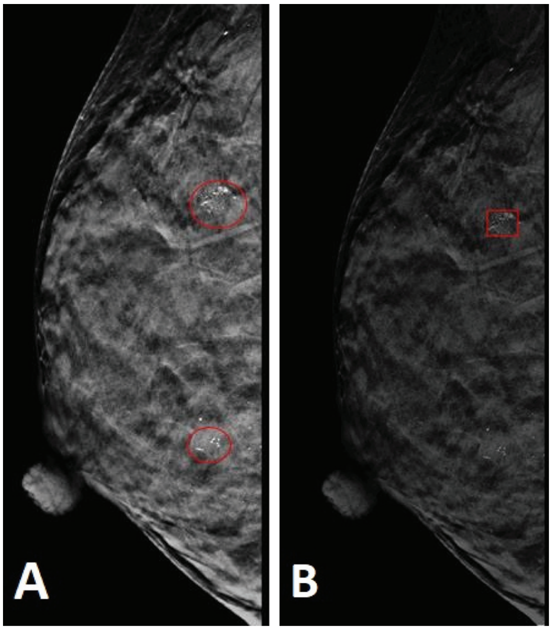

Figure 3. The partial true prediction from 130-tag iteration: (A) A radiologist identified 2 lesions (circle) in a testing image, (B) AI predicted only 1 lesion (rectangle) at the true position.

learning, reasoning, problem-solving, and self-developing. One of the more sophisticated sets of algorithms is often referred to as DL, which is developed from the ML. The ML is the ability of an

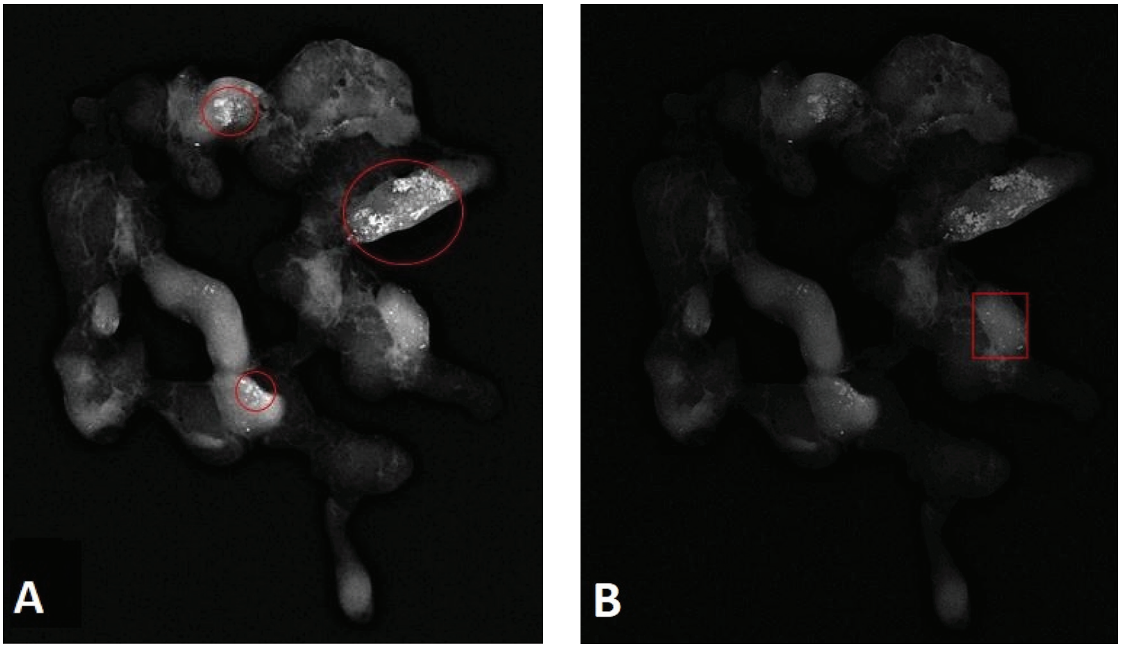

Figure 4. The false-positive prediction from 130-tag iteration: (A) A radiologist identified 3 lesions (circle) in a testing image, (B) AI predicted 1 lesion (rectangle) at a false position that differed from 30-tag iteration in Figure 1.

Table 4. The clinical performance of 3 different durations of training with a 160 lesions dataset

Test	Duration		
	Quick	1 hour	2 hours
1 (4 lesions)	1	2	2
2 (1 lesion)	1	1	1
3 (1 lesion)	0	0	0
4 (1 lesion)	0	0	0
5 (6 lesions)	0	1	1
6 (5 lesions)	0	0	0
7 (3 lesions)	0	0	0
8 (2 lesions)	0	1	1
9 (3 lesions)	0	2	2
10 (2 lesions)	1	2	2
Total (28 lesions)	3	9	9
Percent	10.71	32.14	32.14

AI to extract information from raw data and to learn from experience[11-13]. Microsoft Corporation provides the free-trial service called "custom vision", which health care personnel can use to develop the AI in their daily practices, especially in radiology. This free-trial service, however, can be regarded as an ML level, while DL needs some additional programming. Therefore, DL was not included in the present study.

In theory, more learning makes better AI performance, so the "Performance Per Tag" should improve gradually from 30, 60, 100, 130, and 160-tag iterations. The 160-tag iteration showed the best performance values while other iterations showed inconsistent values. The clinical performances improved gradually from 30, 100, 130, and 160-tag iterations, except for the 60-tag iteration, which showed the results worse than the 30-tag iteration. Many discrete tag varieties, for which each variety had few tag patterns, may have confused the AI on 60-tag iteration. With more tag patterns, the AI made better clinical performances with the best prediction rate at 32.14%.

Rodiguez-Ruiz et al showed a stand-alone AI system achieved a cancer detection accuracy comparable to an average breast radiologist[9]. However, the present study was limited due to retrospective setting, so the AI system needed further investigation in screening setting. Comparing to this sophisticated high-cost AI system, the authors' free-service AI system development is more suitable to alleviate the radiologist's workload in the smaller hospital.

Although the present study showed high precision rate (80%), the recall rate was limited to only 12.5%, which meant only 12.5 cases would be detected correctly from 100 positive cases. There is a need to perform additional studies with large data sets to improve the performance and impact of the present system. Besides radiologist, other clinicians should use the AI system with utmost consideration.

The duration of training should affect the performance as more sophisticated learning needed more time. The 1-hour training model made a better performance than a quick training model. The 2-hour training model, however, was no different in performance from the 1-hour training model. With only 160 tags, the AI needed one hour to experience every pattern thoroughly. One more hour did not help the AI to learn more. If more images were uploaded, 2-hour training may improve AI performance.

Conclusion

Health personnel can easily develop AI for the detection of microcalcification in mammography. AI can predict one-third of microcalcifications correctly after training with only 160 images and the free-trial service.

What is already known on this topic?

AI is currently used to assist the radiologists for arduous works in several fields.

What this study adds?

To show the possibility of self-developed AI for detecting microcalcification with free-trial service.

Acknowledgement

The authors would like to thank Dr. Thanapong Intharah from the Department of Statistics, Faculty of Science, Khon Kaen University for AI system consultation. The authors also would like to acknowledge Emeritus Professor James A. Will, University of Wisconsin-Madison, for editing the manuscript via Publication clinic KKU, Thailand.

Conflicts of interest

The authors declare no conflict of interest.

References

1. Globocan 2020 - Global Cancer Observatory. 764-Thailand-fact-sheets [Internet]. 2020 [cited 2020 Jun 6]. Available from: https://gco.iarc.fr/today/data/factsheets/populations/764-thailand-fact-sheets.pdf.
2. Lauby-Secretan B, Scoccianti C, Loomis D, Benbrahim-Tallaa L, Bouvard V, Bianchini F, et al. Breast-cancer screening--viewpoint of the IARC Working Group. N Engl J Med 2015;372:2353-8.
3. Smith RA, Andrews KS, Brooks D, Fedewa SA, Manassaram-Baptiste D, Saslow D, et al. Cancer screening in the United States, 2017: A review of current American Cancer Society guidelines and current issues in cancer screening. CA Cancer J Clin 2017;67:100-21.
4. Broeders M, Moss S, Nyström L, Njor S, Jonsson H, Paap E, et al. The impact of mammographic screening on breast cancer mortality in Europe: a review of observational studies. J Med Screen 2012;19 Suppl 1:14-25.
5. Winsberg F, Elkin M, Macy J Jr, Bordaz V, Weymouth W. Detection of radiographic abnormalities in mammograms by means of optical scanning and computer analysis. Radiology 1967;89:211-5.
6. Ikeda DM, Birdwell RL, O'Shaughnessy KF, Sickles EA, Brenner RJ. Computer-aided detection output on 172 subtle findings on normal mammograms previously obtained in women with breast cancer detected at follow-up screening mammography. Radiology 2004;230:811-9.
7. James JJ, Giannotti E, Chen Y. Evaluation of a computer-aided detection (CAD)-enhanced 2D synthetic mammogram: comparison with standard synthetic 2D mammograms and conventional 2D digital mammography. Clin Radiol 2018;73:886-92.
8. Geras KJ, Mann RM, Moy L. Artificial intelligence for mammography and digital breast tomosynthesis: Current concepts and future perspectives. Radiology 2019;293:246-59.
9. Rodriguez-Ruiz A, Lång K, Gubern-Merida A, Broeders M, Gennaro G, Clauser P, 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. Sechopoulos I, Mann RM. Stand-alone artificial intelligence - The future of breast cancer screening? Breast 2020;49:254-60.
11. McBee MP, Awan OA, Colucci AT, Ghobadi CW, Kadom N, Kansagra AP, et al. Deep learning in radiology. Acad Radiol 2018;25:1472-80.
12. Suzuki K. Overview of deep learning in medical imaging. Radiol Phys Technol 2017;10:257-73.
13. Saba L, Biswas M, Kuppili V, Cuadrado Godia E, Suri HS, Edla DR, et al. The present and future of deep learning in radiology. Eur J Radiol 2019;114:14-24.