Original Article

Deep learning model for the diagnosis of breast cancers smaller than 1 cm with ultrasonography: integration of ultrasonography and clinical factors

Jae Hwan Bong[1#], **Tae Hee Kim**[2#], **Seongkyun Jeong**[1]

[1]Department of Human Intelligence Robot Engineering, Sangmyung University, Cheonan, South Korea; [2]Department of Radiology, Ajou University School of Medicine, Suwon, South Korea

Contributions: (I) Conception and design: All authors; (II) Administrative support: S Jeong, TH Kim; (III) Provision of study materials or patients: JH Bong, TH Kim; (IV) Collection and assembly of data: TH Kim, JH Bong; (V) Data analysis and interpretation: All authors; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

[#]These authors contributed equally to this work.

Correspondence to: Seongkyun Jeong, PhD. Department of Human Intelligence Robot Engineering, Sangmyung University, 31 Sangmyungdae-gil, Dongnam-gu, Cheonan 31066, South Korea. Email: skjeong@smu.ac.kr.

Background: The aim of this study was to evaluate the diagnostic performance of a deep learning (DL) algorithm for breast masses smaller than 1 cm on ultrasonography (US). We also evaluated a hybrid model that combines the predictions of the DL algorithm from US images and a patient's clinical factors including age, family history of breast cancer, *BRCA* mutation, and mammographic breast density.

Methods: A total of 1,041 US images (including 633 benign and 408 malignant masses) were obtained from 1,041 patients who underwent US between January 2014 and June 2021. All US images were randomly divided into training (513 benign and 288 malignant lesions), validation (60 benign and 60 malignant lesions), and test (60 benign and 60 malignant lesions) data sets. A mask region-based convolutional neural network (R-CNN) was used to generate a feature map of the input image with a CNN and a pre-trained ResNet101 structure. For the clinical model, the multilayer perceptron (MLP) structure was used to calculate the likelihood that the tumor was benign or malignant from the clinical risk factors. We compared the diagnostic performance of an image-based DL algorithm, a combined model with regression, and a combined model with the decision tree method.

Results: Using the US images, the area under the receiver operating characteristics curve (AUROC) of the DL algorithm was 0.85 [95% confidence interval (CI), 0.78–0.92]. With the combined model using a regression model, the sensitivity was 78.3% (95% CI, 67.9–88.8%) and the specificity was 85% (95% CI, 76–94%). The sensitivity of the combined model using a regression model was significantly higher than that of the imaging model (P=0.003). The specificity values of the two models were not significantly different (P=0.083). The sensitivity and specificity of the combined model using a decision tree model were 75% (95% CI, 62.1–85.3%) and 91.7% (95% CI, 81.6–97.2%), respectively. The sensitivity of the combined model using the decision tree model was higher than that of the image model but the difference was not statistically significant (P=0.081). The specificity values of the two models were not significantly different (P=0.748).

Conclusions: The DL model could feasibly be used to predict breast cancers smaller than 1 cm. The combined model using clinical factors outperformed the standalone US-based DL model.

Keywords: Breast cancer; deep learning (DL); ultrasonography (US)

Submitted Aug 25, 2022. Accepted for publication Feb 11, 2023. Published online Mar 09, 2023.
doi: 10.21037/qims-22-880
View this article at: https://dx.doi.org/10.21037/qims-22-880

Introduction

Breast ultrasonography (US) is the most common supplemental screening modality for women with dense breasts and can serve as an adjunct to mammography. Breast US is also the primary imaging modality for the differential diagnosis of benign and malignant breast lesions in a diagnostic setting. Despite its wide applicability, the interpretation of breast US is highly dependent on the operators' experience, leading to intra- and interobserver variability in diagnostic performance (1). The high rate of false positives is a major limitation of breast US, resulting in unnecessary biopsies or short-interval follow-ups (2,3). Previous studies have reported that only 7–8% of US-guided biopsies could identify breast cancers (4,5).

Several studies have reported the utility of commercial computer-aided diagnosis (CAD) systems to overcome these limitations. In a study by Cho *et al.* (6), CAD systems showed higher specificity compared with two radiologists (90.8% compared to 49.2% and 55.4%), but lower sensitivity (72.2% compared to 94.4% and 94.4%). Meanwhile, Park *et al.* (7) reported the diagnostic performance was improved with the aid of CAD, especially for less-experienced radiologists.

Machine learning techniques allow for the detection and classification of breast cancer in mammography, hand-held US, and automated breast US (ABUS) (8-10). Numerous studies of artificial intelligence (AI) systems in breast US have reported high diagnostic performances, with area under the receiver operating characteristic (ROC) curve (AUROC) values of 0.84–0.98 (11-18).

Recent studies have investigated more advanced multitask learning approaches for the diagnosis of breast cancer on mammography and US (19-22). Zhou *et al.* proposed a multitask learning framework for the joint segmentation and classification of breast tumors on automated breast US and demonstrated better results in both tumor segmentation and classification compared to the individually trained single-task models (20). Zhang *et al.* proposed Breast Imaging Reporting and Data System (BI-RADS)-Net, a multitask learning approach incorporating tasks for explaining and classifying breast tumors in US (22). Explanations of the predictions (benign or malignant) are provided in terms of the morphological features of the BI-RADS lexicon that are used by radiologists in clinical practice. These multitask learning methods could improve the diagnostic accuracy of breast imaging modalities via the simultaneous use of segmentation, classification, and BI-RADS lexicons.

However, previous studies have not evaluated the diagnostic performance of AI in relation to lesion size. In particular, when the size is smaller than 1 cm, there may be considerable overlap in US images between benign and malignant lesions, which could make an accurate diagnosis difficult for a radiologist to perform. A previous study reported a sensitivity of 50% and specificity of 66.7% for lesions smaller than 1 cm when handheld US was used (23). In a study by Chen *et al.* (24), the sensitivity of breast US was 85.1% for breast cancers smaller than 1 cm and 92.9% for breast cancers in the range of 1.1–2.0 cm. Although the diagnostic accuracy of the deep learning (DL) algorithm could be affected by the tumor size, there have been limited studies on breast cancers smaller than 1 cm.

In this study, we evaluated the diagnostic performance of the AI system for masses smaller than 1 cm using US. Furthermore, we propose a hybrid model that combines the predictions of AI from US images and patients' clinical factors, including age, family history of breast cancer, *BRCA* mutation, and mammographic breast composition. We present the following article in accordance with the TRIPOD reporting checklist (available at https://qims.amegroups.com/article/view/10.21037/qims-22-880/rc).

Methods

Patients and datasets

This retrospective study was conducted in accordance with the Declaration of Helsinki (as revised in 2013) and was approved by the Institutional Review Board of Ajou University Medical Center (No. AJOUIRB-MDB-2022-051). The requirement for individual consent for this retrospective analysis was waived. All data were fully anonymized before we accessed them. A total of 1,041 US images (including 633 benign and 408 malignant masses) were obtained from 1,041 patients who underwent US between January 2014 and June 2021. All lesions were smaller than 1 cm in size. All of the malignant lesions were

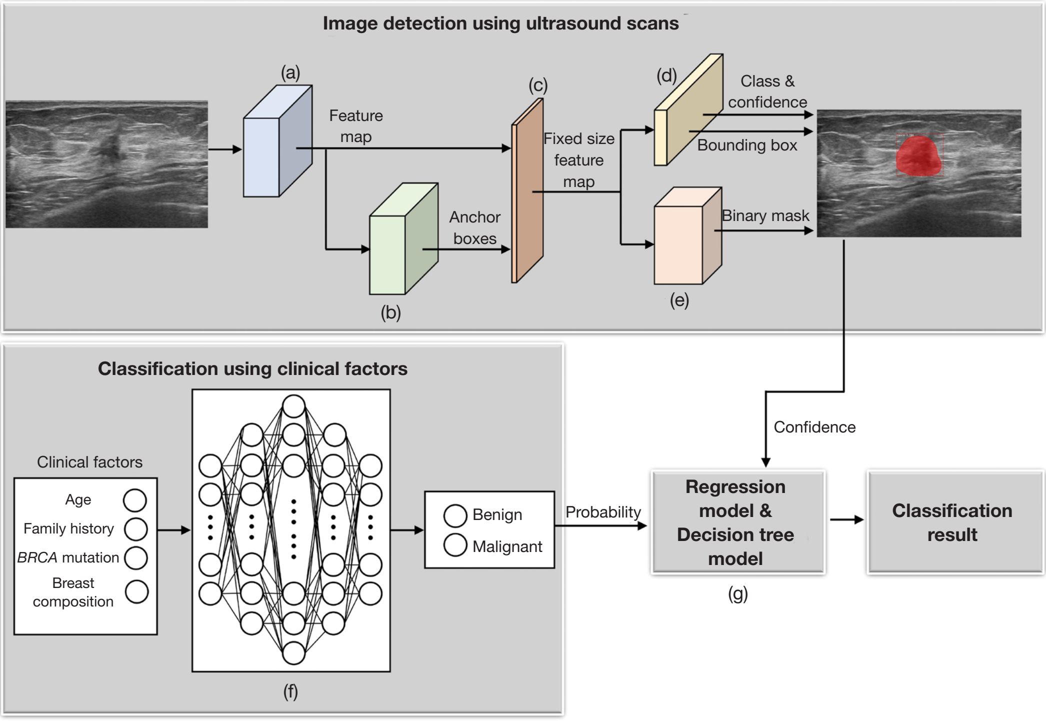

Figure 1 The conceptual architecture of the deep learning model and combined model.

confirmed with core needle biopsy and/or surgery. Benign lesions were confirmed with core needle biopsy or a lack of change in follow-up for more than 2 years.

All 1,041 US images were resized to 1,024×1,024 pixels and randomly split into training (513 benign and 288 malignant lesions), validation (60 benign and 60 malignant lesions), and test (60 benign and 60 malignant lesions) data sets. The US images were obtained using 5 different devices (RS80 A, Samsung Medison Co., Ltd.; ARIETTA 850, Hitachi Healthcare; ACUSON S2000 System, Siemens Healthineers; HI VISION Ascendus, Hitachi Ltd; Philips IU22, Philips Healthcare). Four experienced radiologists specializing in breast imaging performed all US examinations. The lesions in the 801 training-set images and 120 validation-set images were segmented by a radiologist with 15 years of experience in breast imaging to generate the target binary mask images. All radiologists were blinded to the final pathologic results of the target lesions. The target binary mask images were same size as the resized images (i.e., 1,024×1,024 pixels). The target binary mask image was generated by converting all of the pixel values in the US image to 1 or 0. The pixel values of the malignant lesions were converted to 1. The remaining pixel values in the US image were set to 0 in the target binary mask image.

In addition, clinical factors that included the patient's age, family history, *BRCA* mutation, and mammographic breast composition were recorded for the combined model.

Algorithm for analysis

The DL algorithm was developed to localize a tumor and determine whether it was benign or malignant using US images and 4 clinical factors: age, family history, *BRCA* mutation, and mammographic parenchymal density (*Figure 1*). *Figure 1* shows the process of detecting benign and cancerous ultrasound images using a mask region-based convolutional neural network (R-CNN), which was implemented using Matterport's library (25) in the TensorFlow and Keras environments.

Mask R-CNN generates a feature map of the input image using a CNN [(a) in *Figure 1*]. A pretrained ResNet101 structure (26) was used to generate the feature map of the ultrasound image. A region proposal network (RPN) (27) was used to generate anchor boxes from the feature map extracted from the area of the US image that was expected to contain cancerous or benign tumors [(b) in *Figure 1*].

A fixed-size feature map was generated through processing the predicted anchor boxes and feature map information using the region of interest (ROI) pooling method (28), which was then delivered to two types of deep neural network structures [(c) in *Figure 1*]. Anchor boxes were classified with a fully connected layer into 1 of 3 categories: cancer, benign, and background. The coordinates of the bounding boxes for malignant or benign tumors were then predicted through regression [(d) in *Figure 1*]. Using a feature pyramid network (FPN) (29), the AI predicted the categories—cancerous, benign, or background—for each pixel in the US image and generated a "binary mask image" [(e) in *Figure 1*].

For the clinical model, the multilayer perceptron (MLP) structure was used to calculate the likelihood that the tumor was benign or malignant based on the patient's clinical variables, including age, family history, *BRCA* mutation, and mammographic breast density [(f) in *Figure 1*]. The MLP structure was determined through hyperparameter tuning using the Bayesian optimization method. The MLP structure had 3 neurons in the input layer and 5 hidden layers. These hidden layers had 30, 60, 120, 60, and 30 neurons, and rectified linear unit (ReLU) was used as the activation function. Batch normalization was performed between the hidden layers. The output layer had 2 neurons, and Softmax was used as the activation function.

Statistical analysis

We used the AUROC to assess the diagnostic performance of the DL algorithm. US images analyzed using the DL algorithm were considered malignant if the abnormality score produced with the DL algorithm was lower than the operating point. The operating point was chosen using the Youden index from the ROC analysis.

To predict malignancy, multiple logistic regression analysis and a decision tree model were used for the combined model, which included the US score and the clinical risk factor score [(g) in *Figure 1*]. Logistic regression and decision tree models were applied to lesions predicted to be benign in the image model of the DL algorithm.

The exact McNemar test was used to compare the sensitivity and specificity values of the image and combined models. Statistical analyses were performed using SPSS version 21 (IBM Corp., Armonk, NY, USA) and the MedCalc software, version 19.5 (Mariakerke, Belgium). A P value <0.05 was considered to indicate statistical significance.

Results

The baseline characteristics of the training, validation, and test sets are presented in *Table 1*. The mean ages of patients with benign lesions were 46.4, 45.48, and 45.5 years while those of breast cancer patients were 51.73, 50.62, and 52.2 years for the training, validation, and test sets, respectively; for benign lesions, the mean lesion sizes were 0.71, 0.65, and 0.75 cm, respectively, while for malignant lesions, they were 0.79, 0.74, and 0.78 cm, respectively.

In the detection of breast cancers with US images, the AUROC value of the DL algorithm was 0.85 [95% confidence interval (CI), 0.78–0.92]. When the Youden index was used, the sensitivity was 63.3% (95% CI, 51.1–75.5%) and the specificity was 90% (95% CI, 82.4–97.6%). *Figure 2* shows the results of the DL algorithm when the US images were used for the detection of breast cancers smaller than 1 cm.

Table 2 shows the diagnostic performance of the combined model. When the combined model used a regression model, the sensitivity was 78.3% (95% CI, 67.9–88.8%) and the specificity was 85% (95% CI, 76–94%). The sensitivity of the combined model with a regression model was significantly higher than that of the imaging model (P=0.003). The specificities of the two models were not significantly different (P=0.083).

The sensitivity and specificity of the combined model using a decision tree model were 75% (95% CI, 62.1–85.3%) and 91.7% (95% CI, 81.6–97.2%), respectively. The sensitivity of the combined model using the decision tree model tended to be higher than that of the image model, but the difference was not statistically significant (P=0.081). Furthermore, the specificity values of the two models were not significantly different (P=0.748). Representative cases are shown in *Figure 3*.

Discussion

In this study, we evaluated the performance of the DL algorithm for the localization and diagnosis of breast cancers smaller than 1 cm, obtaining an acceptable performance (AUROC 0.85), which was not inferior to those reported in previous reports. When the AI-driven score from the clinical factors was added, the sensitivity increased from 63% to 75–78% without a significant change in specificity.

Two methods could be used to train the DL model: a fully supervised method that requires experts to manually annotate the lesions in each image and a weakly supervised

Table 1 Clinical characteristics of patients in the training, validation, and testing data sets

Characteristics	Training data set		Validation data set		Testing data set	
	Benign (n=513)	Malignant (n=288)	Benign (n=60)	Malignant (n=60)	Benign (n=60)	Malignant (n=60)
Age (years)	46.4±8.6	51.73±9.1	45.48±10.04	50.62±8.5	45.5±9.11	52.2±8.34
Lesion size (cm)	0.71±0.46	0.79±0.31	0.65±0.2	0.74±0.2	0.75±0.42	0.78±0.24
Family history of breast cancer						
Yes	16 (3.1)	34 (11.8)	0 (0.0)	7 (11.7)	1 (1.7)	9 (15.0)
No	497 (96.9)	254 (88.2)	60 (100.0)	53 (88.3)	59 (98.3)	51 (85.0)
BRCA mutation						
Yes	0 (0.0)	7 (2.4)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
No	513 (100.0)	281 (97.6)	60 (100.0)	60 (100.0)	60 (100.0)	60 (100.0)
Breast composition						
Fatty breast	7 (1.4)	14 (4.9)	3 (5.0)	2 (3.3)	0 (0.0)	3 (5.0)
Scattered fibroglandular density	72 (14.0)	95 (33.0)	7 (11.7)	13 (21.7)	6 (10.0)	11 (18.3)
Heterogeneously dense	257 (50.1)	133 (46.1)	33 (55.0)	29 (48.3)	40 (66.7)	27 (45.0)
Extremely dense	177 (34.5)	46 (16.0)	17 (28.3)	16 (26.7)	14 (23.3)	19 (31.7)

Data are expressed as n (%) or mean ± standard deviation.

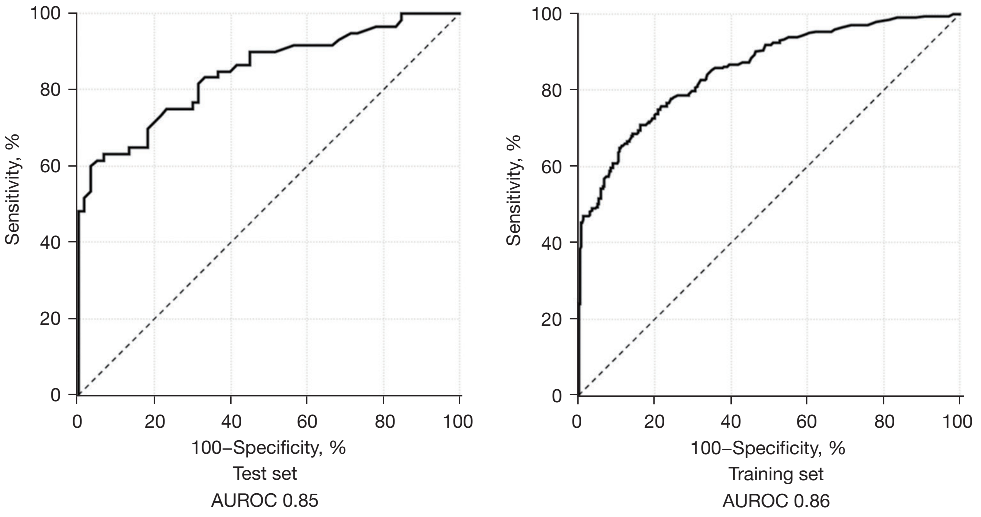

Figure 2 Receiver operating characteristic curves of the deep learning algorithm. AUROC of the DL algorithm using the US image was 0.85 for the test set and 0.86 for the training set. AUROC, the area under the receiver operating characteristics curve; DL, deep learning; US, ultrasonography.

method that uses unannotated images with only image-level labels (i.e., benign or malignant). Classification methods can be classified into region-based and image-based. In region-based classification, the ROI for a lesion should be determined manually or automatically prior to classification, whereas no ROI is needed in image-based classification. Herein, we propose an image-based classification method with a fully supervised DL model.

Numerous studies have used fully supervised algorithms and reported high diagnostic performances, with AUROC

Table 2 Comparison of the results of the predictive models in the validation and testing data sets

Dataset	Models	Sensitivity (%)	Specificity (%)	Accuracy (%)
Test set	Image model	63.3 (51.1–75.5)	90 (82.4–97.6)	76.7 (70.9–83.7)
	Combined model (regression model)	78.3 (67.9–88.8)	85 (76–94)	81.7 (74.4–86.8)
	Combined model (decision tree model)	75 (62.1–85.3)	91.7 (81.6–97.2)	83.3 (75.4–89.5)
Training set	Image model	67.1 (61.8–72.4)	85.3 (82.3–88.3)	78.7 (74.7–82.5)
	Combined model (regression model)	76 (71.2–80.8)	80.4 (77.1–83.8)	78.8 (75.6–81.6)
	Combined model (decision tree model)	72 (66.6–77)	83.1 (79.6–86.1)	79.1 (76.2–81.8)

Data in parentheses are the 95% confidence interval.

values of 0.84–0.94 (11-15). Recent studies using weakly supervised algorithms have reported noninferior or better diagnostic performances, with AUROC values of 0.86–0.98 (16-18). In a previous large study by Shen *et al.* (17), the AI system showed a higher diagnostic performance (AUROC 0.98) than did prior studies of AI systems. In this study, AI achieved a higher AUROC than the average of 10 breast radiologists and reduced radiologists' false-positive rates by 37.3%, while maintaining the same level of sensitivity. Gao *et al.* (18) reported that the semisupervised model can achieve similar performance to the fully supervised model for the detection of breast nodules on US. This semisupervised method could reduce the number of labeled images required for training, thereby alleviating the difficulty in data preparation of medical AI.

In another large study using 7,408 US images, the DL algorithm showed an AUROC of over 0.95, an accuracy of approximately 90%, a sensitivity of 83%, and a specificity of 95% (13). Compared to these previous studies, our results showed a relatively lower AUROC of 0.85, a sensitivity of 63.3%, and a specificity of 90%. However, we only analyzed breast lesions smaller than 1 cm in size. Since the shape and margins of small breast cancers tend to be less irregular and appear less severe, accurate diagnosis is more difficult for small breast lesions. Previous studies also reported relatively lower sensitivities of 50% and 85.1% for small breast cancers (<1 cm) (23,24). We expect that our DL algorithm and combined model will help radiologists to make accurate diagnoses of small breast cancers on US.

Recent studies have investigated the combined use of images and patients' clinical information. A combined model incorporating mammography and clinical information demonstrated favorable AUROC, sensitivity, and specificity values when predicting BI-RADS 4 malignant calcifications, outperforming the clinical and image models (30). In a study by Zheng *et al.* (31), DL-based radiomics combined with clinical parameters yielded the best diagnostic performance for axillary lymph node metastasis, with an AUC of 0.902. Sun *et al.* (32) reported a higher diagnostic performance for the DL algorithm when combining US images and information on molecular subtypes such as human epidermal growth factor receptor 2 (HER2)-positive and triple-negative cancers. In our study, for the combined model, the DL algorithm produced risk scores from patients' clinical data, and we integrated the AI-driven score and US image-based results using a logistic regression model and decision tree model. With the combined model, the sensitivity improved to 78.3% and 75%, respectively, with noninferior specificity.

DL technology is expected to play a role in assisting radiologists in clinical practice. DL models were reported to provide automated detection of ductal carcinoma *in situ* with microinvasion from US images, with an acceptable AUC of 0.803 (33). In a study by Yala *et al.* (34), a DL model was developed to triage mammograms as cancer free. When the DL-triage workflow was simulated, radiologists skipped mammograms triaged as cancer free, which reduced their workload by 19.3%. The specificity improved from 93.5% to 94.2%, with a noninferior sensitivity. In a study by McKinney *et al.* (35), an AI system reduced false positives by 5.7% and 1.2% and false negatives by 9.4% and 2.7% in the United States and United Kingdom, respectively. In addition, in a simulation in which the AI system participated in the double-reading process, the workload of the second reader could be reduced by 88%. Although there have been few studies on the triage of US images, DL technology could effectively be used in clinical practice by selecting a threshold that ensures a very high level of sensitivity.

The present study had several limitations that should be noted. First, our study was a single-center investigation, and

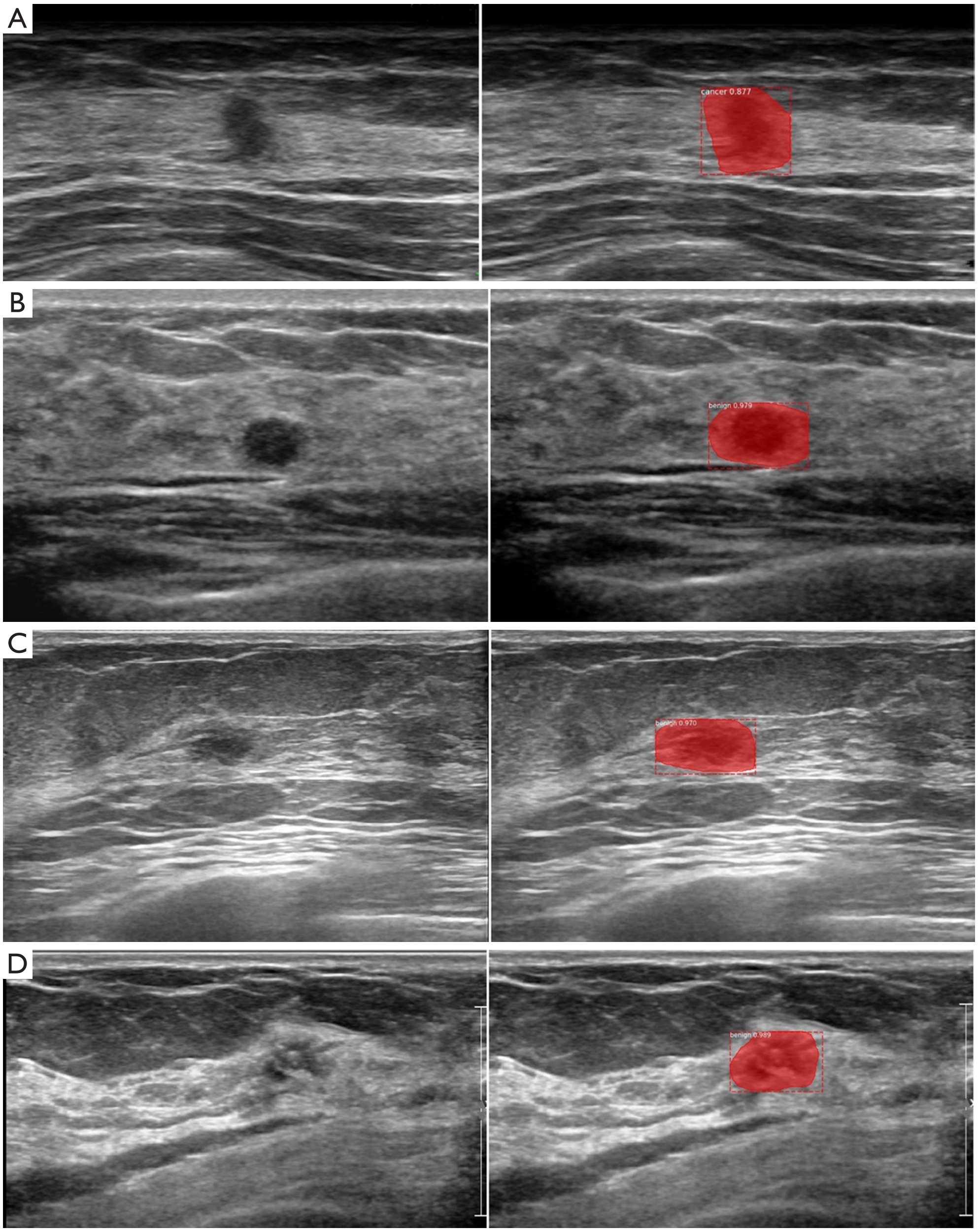

Figure 3 Segmentation and classification results of breast masses on ultrasonography. Examples appeared first as true positive (A), true negative (B), and false negative but later became true positive by combining the ultrasonography image and clinical factors (C,D). (A) US image showing a 0.6-cm irregular isoechoic mass, which was predicted to be malignant. This lesion was diagnosed as invasive ductal carcinoma, histologic grade 2. (B) US image showing a 0.6-cm oval shape, microlobulated mass, which was predicted to be benign. This lesion did not change during the 52-month follow-up period and was considered benign. (C) US image shows a 0.7-cm irregular hypoechoic mass, which was predicted to be benign at first. After results of US image and clinical factors were combined using the deep learning algorithm, this lesion was predicted to be malignant. The pathologic result was ductal carcinoma in situ, histologic grade 3. (D) US image showing a 0.9-cm irregular hypoechoic mass, which was predicted to be benign at first. After results of US image and clinical factors were combined using the deep learning algorithm, this lesion was predicted to be malignant. The pathologic result was ductal carcinoma in situ, histologic grade 1. US, ultrasonography.

the DL algorithm was not trained using a large-scale data set. To improve the reliability of the predictive model, a larger sample size from various institutions is required. Second, we included four clinical factors: patient age, family history, *BRCA* mutation, and mammographic breast density. Other risk factors not included in this study could have improved the diagnostic performance of the combined model. Third, there are primarily two types of deep neural network (DNN) models for image segmentation: transformer-based and CNN-based. In this study, we only used a CNN-based DNN model to extract features from US images. We intend to investigate the performance of feature extraction using the transformer-based DNN model and compare it to the performance of the CNN-based DNN model in future work. Fourth, we did not compare the diagnostic accuracy of the DL algorithm with that of the radiologists. For the evaluation of its usefulness in clinical practice, it is necessary to compare the diagnostic capabilities with that of radiologists with various clinical experiences.

Conclusions

In conclusion, our study demonstrated that the DL model could feasibly be used to predict breast cancers smaller than 1 cm. The combined model outperformed the standalone US-based DL model.

Acknowledgments

Funding: None.

Footnote

Reporting Checklist: The authors have completed the TRIPOD reporting checklist. Available at https://qims.amegroups.com/article/view/10.21037/qims-22-880/rc

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at https://qims.amegroups.com/article/view/10.21037/qims-22-880/coif). The authors have no conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. This retrospective study was conducted in accordance with the Declaration of Helsinki (as revised in 2013) and was approved by the Institutional Review Board of Ajou University Medical Center (No. AJOUIRB-MDB-2022-051). The requirement for individual consent for this retrospective analysis was waived.

References

1. Berg WA, Blume JD, Cormack JB, Mendelson EB. Operator dependence of physician-performed whole-breast US: lesion detection and characterization. Radiology 2006;241:355-65.
2. Sprague BL, Stout NK, Schechter C, van Ravesteyn NT, Cevik M, Alagoz O, Lee CI, van den Broek JJ, Miglioretti DL, Mandelblatt JS, de Koning HJ, Kerlikowske K, Lehman CD, Tosteson AN. Benefits, harms, and cost-effectiveness of supplemental ultrasonography screening for women with dense breasts. Ann Intern Med 2015;162:157-66.
3. Nothacker M, Duda V, Hahn M, Warm M, Degenhardt F, Madjar H, Weinbrenner S, Albert US. Early detection of breast cancer: benefits and risks of supplemental breast ultrasound in asymptomatic women with mammographically dense breast tissue. A systematic review. BMC Cancer 2009;9:335.
4. Yang L, Wang S, Zhang L, Sheng C, Song F, Wang P, Huang Y. Performance of ultrasonography screening for breast cancer: a systematic review and meta-analysis. BMC Cancer 2020;20:499.
5. Corsetti V, Houssami N, Ghirardi M, Ferrari A, Speziani M, Bellarosa S, Remida G, Gasparotti C, Galligioni E, Ciatto S. Evidence of the effect of adjunct ultrasound screening in women with mammography-negative dense breasts: interval breast cancers at 1 year follow-up. Eur J Cancer 2011;47:1021-6.
6. Cho E, Kim EK, Song MK, Yoon JH. Application of Computer-Aided Diagnosis on Breast Ultrasonography: Evaluation of Diagnostic Performances and Agreement of Radiologists According to Different Levels of Experience.

J Ultrasound Med 2018;37:209-16.

7. Park HJ, Kim SM, La Yun B, Jang M, Kim B, Jang JY, Lee JY, Lee SH. A computer-aided diagnosis system using artificial intelligence for the diagnosis and characterization of breast masses on ultrasound: Added value for the inexperienced breast radiologist. Medicine (Baltimore) 2019;98:e14146.
8. Hejduk P, Marcon M, Unkelbach J, Ciritsis A, Rossi C, Borkowski K, Boss A. Fully automatic classification of automated breast ultrasound (ABUS) imaging according to BI-RADS using a deep convolutional neural network. Eur Radiol 2022;32:4868-78.
9. Wang Y, Choi EJ, Choi Y, Zhang H, Jin GY, Ko SB. Breast Cancer Classification in Automated Breast Ultrasound Using Multiview Convolutional Neural Network with Transfer Learning. Ultrasound Med Biol 2020;46:1119-32.
10. Mendelson EB. Artificial Intelligence in Breast Imaging: Potentials and Limitations. AJR Am J Roentgenol 2019;212:293-9.
11. Byra M, Galperin M, Ojeda-Fournier H, Olson L, O'Boyle M, Comstock C, Andre M. Breast mass classification in sonography with transfer learning using a deep convolutional neural network and color conversion. Med Phys 2019;46:746-55.
12. Becker AS, Mueller M, Stoffel E, Marcon M, Ghafoor S, Boss A. Classification of breast cancer in ultrasound imaging using a generic deep learning analysis software: a pilot study. Br J Radiol 2018;91:20170576.
13. Han S, Kang HK, Jeong JY, Park MH, Kim W, Bang WC, Seong YK. A deep learning framework for supporting the classification of breast lesions in ultrasound images. Phys Med Biol 2017;62:7714-28.
14. Mango VL, Sun M, Wynn RT, Ha R. Should We Ignore, Follow, or Biopsy? Impact of Artificial Intelligence Decision Support on Breast Ultrasound Lesion Assessment. AJR Am J Roentgenol 2020;214:1445-52.
15. Dong F, She R, Cui C, Shi S, Hu X, Zeng J, Wu H, Xu J, Zhang Y. One step further into the blackbox: a pilot study of how to build more confidence around an AI-based decision system of breast nodule assessment in 2D ultrasound. Eur Radiol 2021;31:4991-5000.
16. Kim J, Kim HJ, Kim C, Lee JH, Kim KW, Park YM, Kim HW, Ki SY, Kim YM, Kim WH. Weakly-supervised deep learning for ultrasound diagnosis of breast cancer. Sci Rep 2021;11:24382.
17. Shen Y, Shamout FE, Oliver JR, Witowski J, Kannan K, Park J, et al. Artificial intelligence system reduces false-positive findings in the interpretation of breast ultrasound exams. Nat Commun 2021;12:5645.
18. Gao Y, Liu B, Zhu Y, Chen L, Tan M, Xiao X, Yu G, Guo Y. Detection and recognition of ultrasound breast nodules based on semi-supervised deep learning: a powerful alternative strategy. Quant Imaging Med Surg 2021;11:2265-78.
19. Sainz de Cea MV, Diedrich K, Bakalo R, Ness L, Richmond D. Multi-task Learning for Detection and Classification of Cancer in Screening Mammography. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, Cham, 2020:241-50.
20. Zhou Y, Chen H, Li Y, Liu Q, Xu X, Wang S, Yap PT, Shen D. Multi-task learning for segmentation and classification of tumors in 3D automated breast ultrasound images. Med Image Anal 2021;70:101918.
21. Yang K, Suzuki A, Ye J, Nosato H, Izumori A, Sakanashi H. CTG-Net: Cross-task guided network for breast ultrasound diagnosis. PLoS One 2022;17:e0271106.
22. Zhang B, Vakanski A, Xian M. BI-RADS-NET: an explainable multitask learning approach for cancer diagnosis in breast ultrasound images. IEEE Int Workshop Mach Learn Signal Process 2021.
23. Wang HY, Jiang YX, Zhu QL, Zhang J, Dai Q, Liu H, Lai XJ, Sun Q. Differentiation of benign and malignant breast lesions: a comparison between automatically generated breast volume scans and handheld ultrasound examinations. Eur J Radiol 2012;81:3190-200.
24. Chen HL, Zhou JQ, Chen Q, Deng YC. Comparison of the sensitivity of mammography, ultrasound, magnetic resonance imaging and combinations of these imaging modalities for the detection of small (≤2cm) breast cancer. Medicine (Baltimore) 2021;100:e26531.
25. Matterport Inc., Mask r-cnn for object detection and instance segmentation on keras and tensorflow. 2018, (Online). Available online: https://github.com/matterport/Mask_RCNN.git
26. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. Proc IEEE Comput Soc Conf Comput Vis Pattern Recognit 2016;12:770-8.
27. Ren S, He K, Girshick R, Sun J. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. IEEE Trans Pattern Anal Mach Intell 2017;39:1137-49.
28. Girshick R. Fast R-CNN. Proc IEEE Int Conf Comput Vis 2015:1440-8.
29. Li X, Lai T, Wang S, Chen Q, Yang C, Chen R. Weighted

feature pyramid networks for object detection. Proc. - 2019 IEEE Intl Conf Parallel Distrib. Process. With Appl. Big Data Cloud Comput. Sustain. Comput. Commun. Soc. Comput. Networking, ISPA/BDCloud/SustainCom/SocialCom. 2019:1500-4.
30. Liu H, Chen Y, Zhang Y, Wang L, Luo R, Wu H, Wu C, Zhang H, Tan W, Yin H, Wang D. A deep learning model integrating mammography and clinical factors facilitates the malignancy prediction of BI-RADS 4 microcalcifications in breast cancer screening. Eur Radiol 2021;31:5902-12.
31. Zheng X, Yao Z, Huang Y, Yu Y, Wang Y, Liu Y, Mao R, Li F, Xiao Y, Wang Y, Hu Y, Yu J, Zhou J. Deep learning radiomics can predict axillary lymph node status in early-stage breast cancer. Nat Commun 2020;11:1236.
32. Sun Q, Lin X, Zhao Y, Li L, Yan K, Liang D, Sun D, Li ZC. Deep Learning vs. Radiomics for Predicting Axillary Lymph Node Metastasis of Breast Cancer Using Ultrasound Images: Don't Forget the Peritumoral Region. Front Oncol 2020;10:53.
33. Zhu M, Pi Y, Jiang Z, Wu Y, Bu H, Bao J, Chen Y, Zhao L, Peng Y. Application of deep learning to identify ductal carcinoma in situ and microinvasion of the breast using ultrasound imaging. Quant Imaging Med Surg 2022;12:4633-46.
34. Yala A, Schuster T, Miles R, Barzilay R, Lehman C. A Deep Learning Model to Triage Screening Mammograms: A Simulation Study. Radiology 2019;293:38-46.
35. McKinney SM, Sieniek M, Godbole V, Godwin J, Antropova N, Ashrafian H, et al. International evaluation of an AI system for breast cancer screening. Nature 2020;577:89-94.

Cite this article as: Bong JH, Kim TH, Jeong S. Deep learning model for the diagnosis of breast cancers smaller than 1 cm with ultrasonography: integration of ultrasonography and clinical factors. Quant Imaging Med Surg 2023;13(4):2486-2495. doi: 10.21037/qims-22-880