Ultrasound in Med. & Biol., Vol. 48, No. 11, pp. 2267–2275, 2022

0301-5629/$ - see front matter

https://doi.org/10.1016/j.ultrasmedbio.2022.06.019

● *Original Contribution*

APPLICATION OF DEEP LEARNING TO REDUCE THE RATE OF MALIGNANCY AMONG BI-RADS 4A BREAST LESIONS BASED ON ULTRASONOGRAPHY

Zhijin Zhao,[*,†,1] Size Hou,[‡,1] Shuang Li,[§] Danli Sheng,[*,†] Qi Liu,[*,†] Cai Chang,[*,†] Jiangang Chen,[¶,‖] and Jiawei Li[*,†]

* Department of Medical Ultrasound, Fudan University Shanghai Cancer Center, Shanghai, China; † Department of Oncology, Shanghai Medical College, Fudan University, Shanghai, China; ‡ Department of Applied Mathematics, School of Science, Xi'an Jiaotong-Liverpool University, Suzhou, China; § International Business School Suzhou, Xi'an Jiaotong-Liverpool University, Suzhou, China; ¶ Shanghai Key Laboratory of Multidimensional Information Processing, School of Communication & Electronic Engineering, East China Normal University, Shanghai, China; and ‖ Engineering Research Center of Traditional Chinese Medicine Intelligent Rehabilitation, Ministry of Education, Shanghai, China

(*Received* 31 *March* 2022; *revised* 31 *May* 2022; *in final form* 24 *June* 2022)

Abstract—The aim of the work described here was to develop an ultrasound (US) image–based deep learning model to reduce the rate of malignancy among breast lesions diagnosed as category 4A of the Breast Imaging-Reporting and Data System (BI-RADS) during the pre-operative US examination. A total of 479 breast lesions diagnosed as BI-RADS 4A in pre-operative US examination were enrolled. There were 362 benign lesions and 117 malignant lesions confirmed by postoperative pathology with a malignancy rate of 24.4%. US images were collected from the database server. They were then randomly divided into training and testing cohorts at a ratio of 4:1. To correctly classify malignant and benign tumors diagnosed as BI-RADS 4A in US, four deep learning models, including MobileNet, DenseNet121, Xception and Inception V3, were developed. The performance of deep learning models was compared using the area under the receiver operating characteristic curve (AUROC), accuracy, sensitivity, specificity, positive predictive value (PPV) and negative predictive value (NPV). Meanwhile, the robustness of the models was evaluated by five-fold cross-validation. Among the four models, the MobileNet model turned to be the optimal model with the best performance in classifying benign and malignant lesions among BI-RADS 4A breast lesions. The AUROC, accuracy, sensitivity, specificity, PPV and NPV of the optimal model in the testing cohort were 0.897, 0.913, 0.926, 0.899, 0.958 and 0.784, respectively. About 14.4% of patients were expected to be upgraded to BI-RADS 4B in US with the assistance of the MobileNet model. The deep learning model MobileNet can help to reduce the rate of malignancy among BI-RADS 4A breast lesions in pre-operative US examinations, which is valuable to clinicians in tailoring treatment for suspicious breast lesions identified on US. (E-mail addresses: jgchen@cee.ecnu.edu.cn jiaweili2006@163.com)

Key Words: **Deeping learning, Ultrasonography, Breast Imaging Reporting and Data System, Breast neoplasm.**

INTRODUCTION

According to the report in 2020 from the World Health Organization International Agency for Research on Cancer (IARC), breast cancer has replaced the top-ranked lung cancer as the most common malignant tumor worldwide and the leading cause of cancer-related death in women (Ferlay et al. 2021). The early diagnosis of breast cancer plays an essential part in improving prognostic outcomes. Ultrasound (US) is an important screening and diagnostic tool complementary to mammography (MG) for the early diagnosis of breast cancers. Compared with MG, US is more sensitive in identifying intraductal lesions and nodular lesions. It is even superior to MG in young and pregnant women. In China, US is equivalent to MG for the screening of breast cancers, considering the low sensitivity of MG for breast tissues with high density, which are common among Chinese women (Lian and Li 2020). In addition, clinicians prefer to combine the findings of both US and MG to make treatment plans for suspicious breast lesions.

In 2013, the American College of Radiology (ACR) updated the Breast Imaging Reporting and Data System (BI-RADS) for US imaging (Mendelson et al. 2013).

Address correspondence to: Jiawei Li, 270 Dongan Road, Shanghai, China.Address correspondence to: Jiangang Chen, 500 Dongchuan Road, Shanghai, China.E-mail addresses: jgchen@cee.ecnu.edu.cn, jiaweili2006@163.com

[1] These authors contributed equally.

The BI-RADS lexicon helps US physicians and breast surgeons in standardizing the probability of malignancy for breast lesions. It also facilitates breast surgeons in planning proper treatments according to the BI-RADS score. Therefore, the BI-RADS lexicon is widely accepted by US physicians and breast surgeons (Mercado 2014). However, because of the high heterogeneity and variety of breast lesions, the cutoff point for adjacent BI-RADS scores is ambiguous but important, especially for BI-RADS 3/4A and 4A/4B (Stavros et al. 2017; Shang et al. 2019; Mei et al. 2020).

BI-RADS 4A is the crucial cut point in determining treatment strategy at our breast cancer center. For BI-RADS 3 lesions, surgeons usually recommend follow-up or minimally invasive surgery through the Mammotome System in the outpatient unit, whereas for BI-RADS 4A breast lesions, surgical resection at the day surgery ward is usually performed. Patients with BI-RADS 4B lesions are accepted as inpatients to prepare for malignant breast tumors. The treatment for BI-RADS 4A lesions may differ from that at other breast cancer centers where core needle biopsies are recommended. Because there is no pre-operative biopsy for BI-RADS 4A lesions at our center, US physicians are very cautious with the report as it determines the clinical decisions.

On the basis of the statistics gathered at our cancer center, the malignancy rate among BI-RADS 4A breast lesions on US was about 20%−30%, which is much higher than the 2%−10% recommended by ACR. This means approximately 20% of breast lesions with BI-RADS 4A were underestimated at our center, which may delay the treatment for those malignant tumors as the paraffin pathology results are not available until 14 d after surgery. Thus, in contrast to previous studies in which auxiliary methods were used to downgrade a portion of BI-RADS 4A breast lesions to BI-RADS 3 (Koh et al. 2019; Zheng et al. 2019), we intended to upgrade a portion of BI-RADS 4A breast lesions to BI-RADS 4B to reduce the malignancy rate among BI-RADS 4A breast lesions.

The evaluation of US images is generally subjective and is probably affected by US physicians' experience. Computer-aided techniques, especially deep learning, overcome these shortcomings in subjectivity (Zheng et al. 2020; Zhou et al. 2021). In this study, we aimed to evaluate the performance of the deep learning model in reducing the malignancy rate among BI-RADS 4A lesions to achieve more accurate risk stratification.

METHODS

Patients

The ethics committee of Fudan University Shanghai Cancer Center approved this retrospective study with written informed consent waived. From August 2013 to December 2020, a total of 479 lesions in 477 patients diagnosed as BI-RADS 4A on US before core needle biopsy or surgery were enrolled. For each breast lesion, at least two US images were selected to obtain the optimal diagnostic performance for deep learning. Finally, 1748 images were included in this study. The inclusion criteria were as follows: (i) classification of the lesion as BI-RADS 4A by two US physicians; (ii) clarity of lesions in gray-scale images without measurement labels or sample window of color Doppler; (iii) lesion size <5 cm; (iv) pathological confirmation. Patients were excluded from this study if only one image was available in the database or the tumor could not be confirmed pathologically.

Image acquisition and processing

All US images in our study were from the Fudan University Shanghai Cancer Center and acquired with different equipment, including the IU-22 and EPIQ7 (Philips Medical Systems, Andover, MA, USA), LOGIQ-E9 (GE Healthcare, Hatfield, UK), Toshiba-Aplio500 (Canon Medical Systems, Tokyo, Japan), Mindray-Resona7 (Mindray Medical, Shenzhen, China) and MylabTwice (Esaote-Biomedica, Genoa, Italy). Two radiologists experienced in breast US performed all examinations. All lesions were evaluated and scored as BI-RADS 4A in accordance with the US BI-RADS lexicon, and were confirmed by post-operative pathology (Figs. 1 and 2).

We randomly split all US images into a training data set and an independent testing data set with a ratio of 4:1. The training data set was used to train all deep learning models, and the model with the best performance was selected as the final model. The independent testing data set was used to verify the performance of all deep learning models that had been trained.

Deep learning model

The deep learning model was employed as a computer-aided tool in the present study. Other than extracting features manually in traditional machine learning, the “end-to-end” deep learning models contribute to mining features automatically with a sophisticated network structure. In detail, the relevant features were automatically extracted from the US images. In addition, complex function mapping between the US imaging data and the pathological classification of breast cancer was established. The deep learning model contains multiple hidden layers. It combines low-level breast image features together automatically and forms a more abstract high-level representation to discover the actual category of US data.

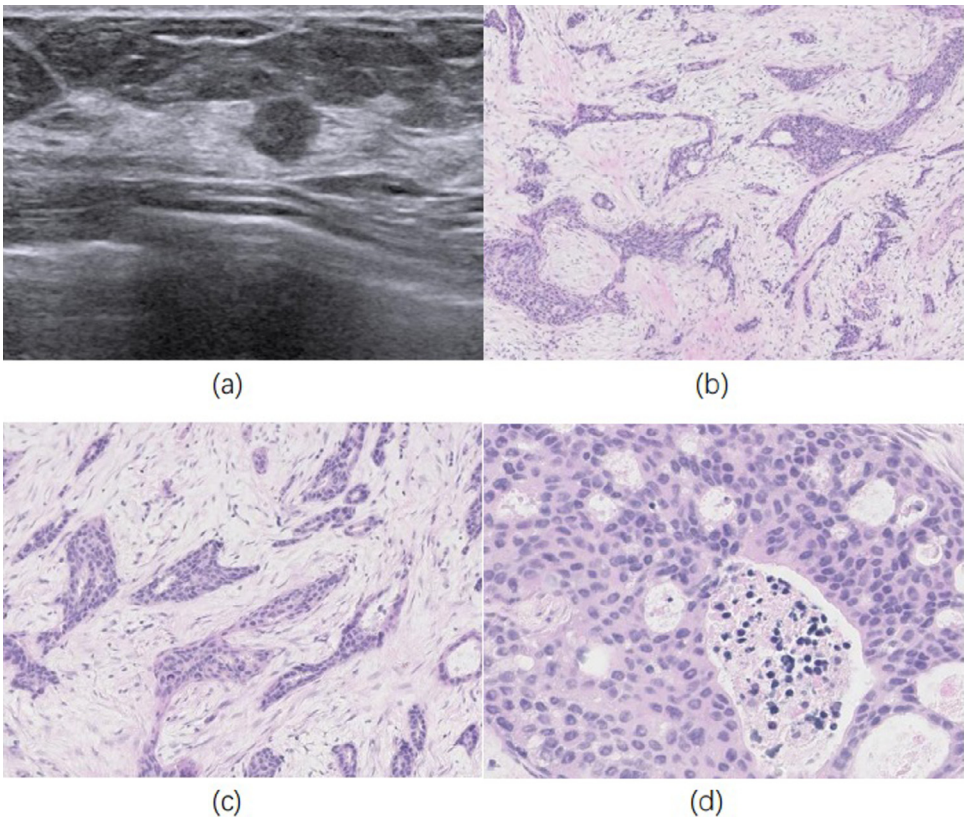

Fig. 1. Thirty-seven-year-old woman with malignant breast lesion (invasive ductal carcinoma, grade II) 0.8 × 0.6 × 0.5 cm. (a) Gray-scale US image. (b) Hematoxylin and eosin (HE) staining. Original magnification: × 100. (c). HE staining. Original magnification: × 200. (d) HE staining. Original magnification: × 400.

To develop the end-to-end mapping of breast US images of BI-RADS 4A, we used five-fold cross-validation to compare the performance of four models including the mature lightweight convolutional neural network MobileNet, the well-known complex deep learning model with fewer parameters DenseNet121, the SOTA multi-scale convolutional neural network InceptionV3 and the most widely used image classification model Xception. In five-fold cross-validation, the ratio of training data (1398 images) to validation data (350 images) was 4:1. In Each data set was equally split into five folds. One of the folds was selected as the validation set and the other folds as the training set to develop the model. This process was repeated five times. The MobileNet model with the best performance (see Results) was selected as the model used in the present study. We used the Tensorflow (Google, Mountain View City, CA, USA) framework to code all the deep learning models. The workflow of our research is illustrated in Figure 3.

The MobileNet model is based on a depthwise separable convolution (Fig. 4) and shortcut connection (Sandler et al. 2018). Depthwise separatable convolution combines different feature channels. It also significantly reduces the number of model parameters and the computational complexity. The shortcut connection structure significantly eliminates the difficulty of training deep neural networks and allows MobileNet to capture more abstract features.

The theoretical basis of MobileNet is that the features of each channel can be mapped to a manifold region in a low-dimensional subspace. After completing the convolution operation, a layer of activation function was usually added to increase the non-linearity of the features. As long as the high-dimensional input data can be restored through the feature map, the computational complexity is vastly reduced. The MobileNet model relies on the fundamental assumption in data science that high-dimensional data must have a low-dimensional structure. Therefore, the high-dimensional breast US image data can be reduced to a low-dimensional computable subspace through MobileNet to accurately classify benign and malignant tumors.

Statistical analysis

The performance of deep learning models was evaluated with respect to sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV)

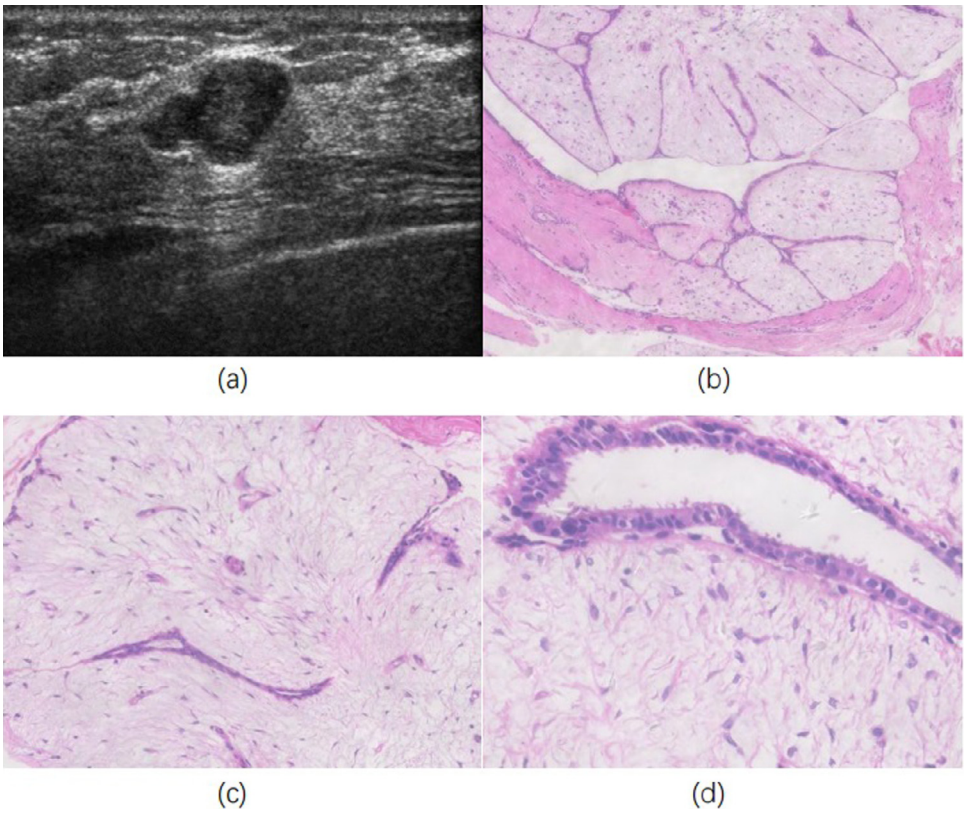

Fig. 2. Forty-four-year-old woman with begin breast lesion (fibroadenoma) 1.5 × 1.4 × 1.4cm. (a) Gray-scale US image; (b) Hematoxylin and eosin (HE) staining. Original magnification: × 100. (c). HE staining. Original magnification: × 200. (d) HE staining. Original magnification: × 400.

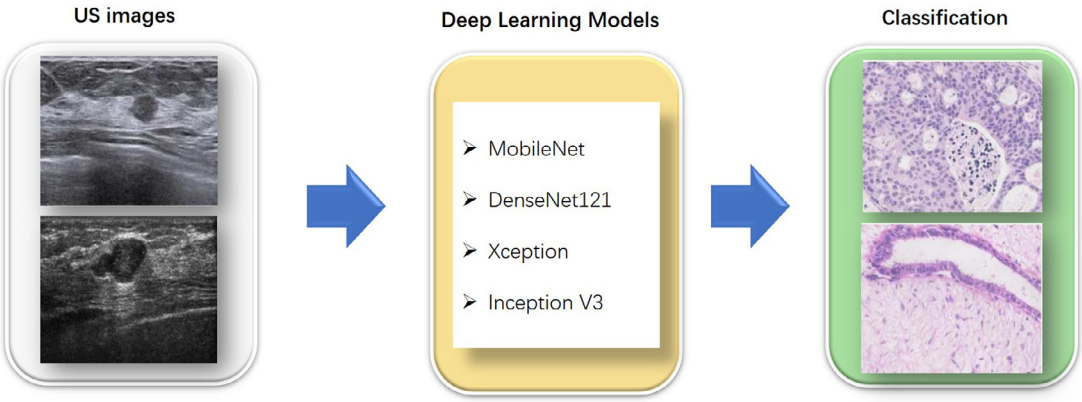

Fig. 3. Workflow of use of deep learning model to classify malignant and benign breast tumors among Breast Imaging Reporting and Data System 4A nodules.

and area under the receiver operating characteristic curve (AUROC).

RESULTS

Four hundred seventy-nine breast lesions from 477 female patients aged between 16 and 90 y were scored as BI-RADS 4A (mean ± standard deviation: 44.3 ± 13.1). Table 1 outlines the pathological subtypes of all breast lesions. There were 362 cases of benign tumors (75.6%) and 117 cases of malignant tumors (24.4%). Most benign tumors were fibroadenoma (153 out of 362, 42.3%), and most malignant tumors were invasive ductal carcinoma (71 out of 117, 60.7%).

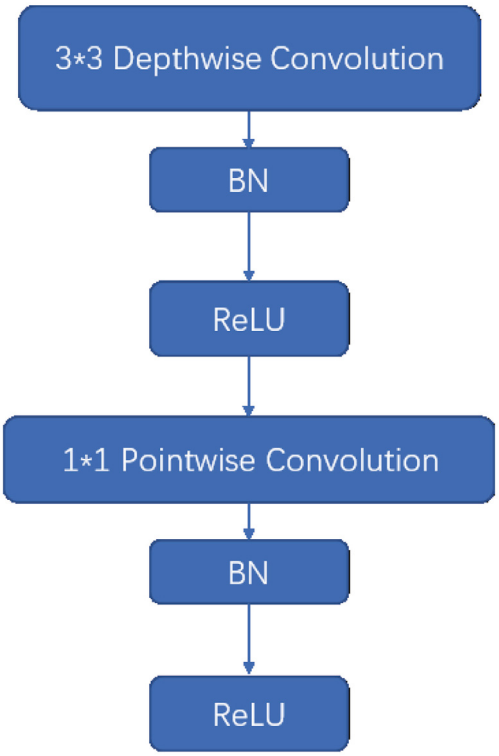

Fig. 4. Depthwise separable convolution.

Performance in the five-fold cross-validation is outlined in Tables 2 and 3. MobileNet's training prediction accuracy is 94.3%, 98.3%, 91.7%, 94.6% and 95.7%, respectively; its validation prediction accuracy is 90.5%, 90.2%, 90.6%, 90.6% and 90.1%, respectively. Figure 5 illustrates the ROC curves in the cross-validation. MobileNet's AUROC in the cross-validation is 89.2%, 86.6%, 82.2%, 89.7% and 88.9%, respectively.

Depicted in Table 4 are the results of the testing data set in the four models. Among the four types of models, MobileNet had the best diagnostic performance with an AUROC of 89.7% and an accuracy of 91.3% in the testing data set. The AUROCs of the other three models ranged from 75.2%−78.7%. Figure 6 illustrates the confusion matrix, which evaluated the model in the testing data set by comparing the predictions with the facts. In the matrix, the columns represent the real labels of the malignant and benign classes, and the rows represent the classes predicted by the MobileNet model. In Figure 7 is the ROC curve in the testing data set. The AUROC of the MobileNet model was 89.7%, which was higher than those of the other three deep learning models.

Table 1. Pathological types of benign and malignant breast lesions

	Pathological types	Number	Percentage
Benign lesions (362 cases)	Fibroadenoma	153	31.9%
	Adenosis	131	27.3%
	Inflammation	18	3.8%
	Intraductal papilloma	54	11.3%
	Others	6	1.3%
Malignant lesions (117 cases)	Invasive ductal carcinoma	71	14.8%
	Ductal carcinoma *in situ*	30	6.3%
	Ductal papillary carcinoma	4	0.8%
	Others	12	2.5%

DISCUSSION

In the present study, 24.4% of breast lesions were malignant carcinomas. This rate is much higher than the malignancy rate of 2%−10% for BI-RADS 4A breast lesions defined by the ACR BI-RADS lexicon. This study was designed to establish and validate an US imaging−based deep learning model that differentiates benign and malignant tumors among BI-RADS 4A breast lesions. Among the four models, the MobileNet model had the optimal performance with an AUROC of 0.897, indicating that the MobileNet model could help US physicians in controlling the malignancy rate to be less than 10% among BI-RADS 4A breast lesions by upgrading some BI-RADS 4A lesions to BI-RADS 4B.

US is recommended primarily for screening and pre-operative examinations of breast lesions. The BI-RADS score provides a valuable reference for breast surgeons in determining the treatment strategy for breast lesions. Among BI-RADS 3, 4A and 4B, BI-RADS 4A is a vital cut point at our cancer center in tailoring treatment for breast lesions to be followed up (BI-RADS 3) or surgical resection (BI-RADS 4A), and surgery as an outpatient (BI-RADS 4A) or inpatient (BI-RADS 4B). Most previous research paid much attention to studying the necessity for downgrading BI-RADS 4A lesions to BI-RADS 3 with the assistance of multiple new technologies, such as elastography, contrast-enhanced ultrasonography (CEUS) and computer-aided techniques to avoid unnecessary biopsies for breast lesions (Li et al. 2016; Koh et al. 2019; Zheng et al. 2019; Weismann 2021). For example, Koh et al. (2019) used strain elastography to downgrade category 4A breast lesions with respect to personal risk factors. This kind of research work was quite necessary at some breast disease centers as most of breast surgeries were performed for benign lesions. US physicians at these centers are usually very cautious with suspicious breast lesions, so that the BI-RADS 4A is quite common rather than BI-RADS 3 for benign breast tumors. As a result, the proportion of benign breast tumors among BI-RADS 4A is usually quite high, requiring the downgrading of BI-RADS 4A to BI-RADS 3 for some breast lesions to avoid unnecessary surgeries.

Table 2. Five-fold cross-validation of the training data set

Model	Accuracy 1	Accuracy 2	Accuracy 3	Accuracy 4	Accuracy 5
MobileNet	94.3%	98.3%	91.7%	94.6%	95.7%
DenseNet121	80.0%	82.6%	81.3%	80.2%	85.1%
Xception	79.7%	87.7%	89.4%	89.8%	90.5%
Inception V3	85.7%	90.9%	87.4%	90.0%	94.6%

Table 3. Five-fold cross-validation of the validation data set

Model	Accuracy 1	Accuracy 2	Accuracy 3	Accuracy 4	Accuracy 5
MobileNet	90.5%	90.2%	90.6%	90.6%	90.1%
DenseNet121	79.2%	80.6%	83.1%	83.6%	83.1%
Xception	86.3%	87.2%	90.5%	91.2%	92.7%
Inception V3	85.6%	86.1%	87.3%	89.4%	89.3%

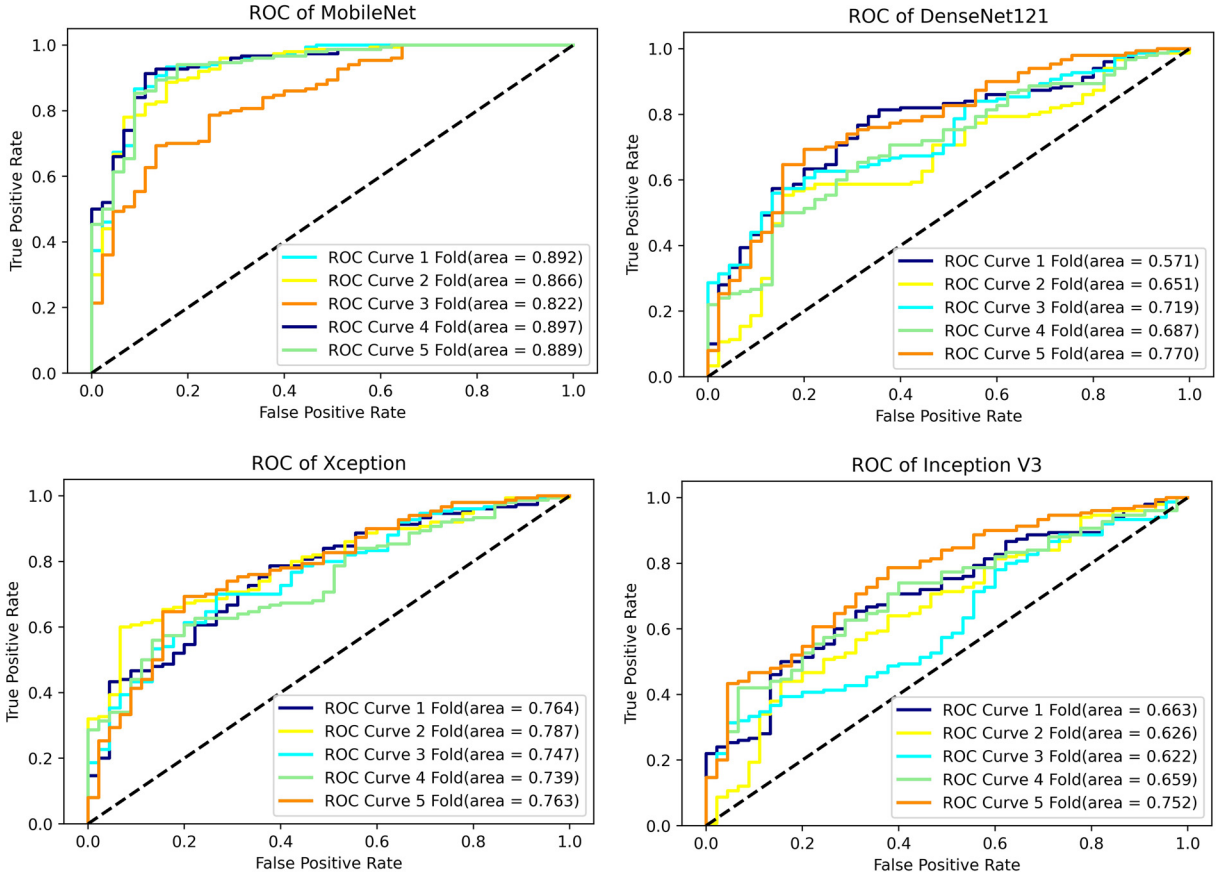

Fig. 5. Receiver operating curves (ROCs) for the four models in the five-fold cross-validation.

Table 4. Diagnostic performance of the deep learning models for the testing data set

Model	Accuracy	Sensitivity	Specificity	PPV	NPV	AUROC
MobileNet	91.3%	92.6%	89.9%	95.8%	78.4%	89.7%
DenseNet121	86.7%	95.3%	59.7%	88.2%	79.4%	77.0%
Xception	83.6%	87.9%	69.5%	90.3%	64.0%	78.7%
Inception V3	85.1%	93.9%	56.5%	87.5%	74.3%	75.2%

AUROC = area under the receiver operating characteristic curve; NPV = negative predictive value; PPV = positive predictive value.

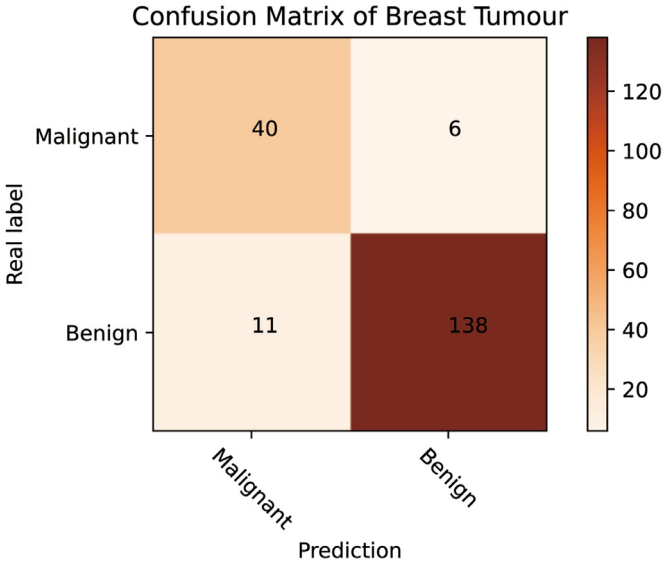

Fig. 6. Confusion matrix of MobileNet in the testing data set.

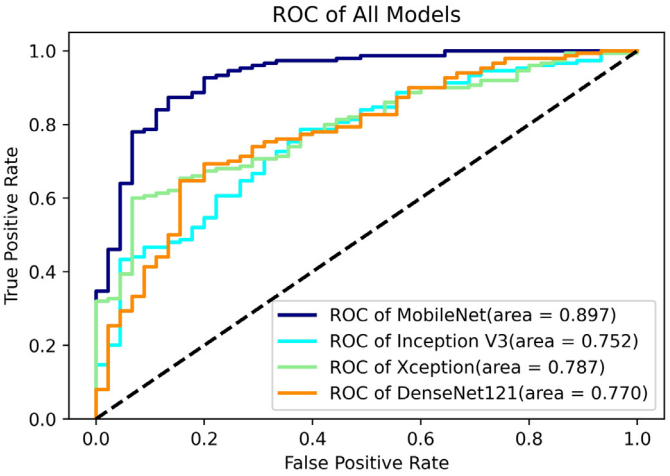

Fig. 7. Receiver operating curves (ROCs) for the four models in the testing data set.

In Fudan University Shanghai Cancer Center, one of the largest breast cancer centers in China, about half of breast surgeries were conducted for malignant breast tumors. The challenge for US physicians is to recognize malignant breast tumors with untypical sonographic features (BI-RADS 4A or BI-RADS 3) to avoid the second surgery after the local resection. On the basis of our statistical data, we found that the problem of underestimation for BI-RADS 4A breast lesions is common among US physicians at our center. The malignancy rate among BI-RADS 4A breast lesions was about 20%−30% per US physician (unpublished data). Most of these patients with malignant breast tumors accepted surgical resection as outpatients and then mastectomy for standard breast cancer treatment. This may increase the chance of hematogenous metastasis (Hu et al. 2003; Ismail et al. 2004; Li et al. 2019). Even for very experienced US physicians, differentiating with the naked eye the subtle difference between benign and malignant breast tumors with similar sonographic appearance is very challenging. Therefore, the precise differentiation of benign and malignant breast tumors among BI-RADS 4A breast lesions reported by US physicians is crucial in optimizing the treatment strategy and improving the clinical outcome.

In this study, we applied deep learning, an artificial intelligence technique, to solve this clinical issue for the first time. Artificial intelligence, including radiomics, machine learning and deep learning, has played an important role in computer-aided diagnosis of breast lesions. The computer is superior to US physicians as it has lower subjectivity and more robust operational capability (Munir et al. 2019; Tagliafico et al. 2020). Numerous

studies have proved that artificial intelligence based on MG, magnetic resonance imaging (MRI) or US imaging can differentiate malignant from benign breast lesions (Dogan et al. 2010; Shia et al. 2021; Zhang et al. 2021). Lee et al. (2018) tried to use US radiomics to distinguish fibroadenomas and triple-negative breast cancers. Fleury and Marcomini (2019) compared the performance of five machine learning methods in quantifying the five BI-RADS radiomic sonographic features with the specified region of interest (ROI) selected. The outlined ROI enables the exact localization of breast tumors; however, it demands great human effort. Meanwhile, in radiomics, the margin of breast lesions needs to be manually outlined, which may bring out inter- and intra-observer variations, especially for those lesions with indistinct margins (Valdora et al. 2018; Isik et al. 2020; Conti et al. 2021). In contrast, the deep learning model can automatically learn and extract features from US images. Deep learning has been used widely to differentiate benign and malignant tumors in breast and thyroid lesions (Niu et al. 2020; Zhou et al. 2020a; Ha and Baek 2021), to predict underestimation in ductal carcinoma *in situ* (Qian et al. 2021) and to predict axillary lymph node metastasis in breast cancers (Zhou et al., 2020b).

In the present study, we built a lightweight deep neural network based on MobileNet to differentiate benign and malignant breast tumors among BI-RADS 4A lesions. Compared with the other three deep learning models, MobileNet had the following advantages: (i) Depthwise separable convolutions were used to build a lightweight deep neural network, which could be embedded in handheld US equipment to be used freely. (ii) MobileNet requires fewer computing resources and can be applied in the hospital environment. To the best of our knowledge, this is the first trial using the MobileNet model to help risk stratification of BI-RADS 4A breast lesions with promising results. The robustness and accuracy of MobileNet are excellent. Exploration of clinical applications is warranted, with collaboration between US physicians and biomedical engineers. Still, some limitations of this study need to be considered. First, this was a retrospective study, which may weaken the confidence level of our results. Second, the number of breast lesions was relatively small for the deep learning model. Finally, data were from a single center with no external verification set. These limitations are expected to be overcome in future studies.

CONCLUSIONS

The deep learning MobileNet model based on US images had stable performance in differentiating benign and malignant tumors among BI-RADS 4A breast lesions. This approach may serve as a complementary tool to assist clinical decision-making by US physicians when there is a need to upgrade BI-RADS 4A lesions to 4B to reduce underestimations.

DECLARATION OF COMPETING INTEREST

The authors declare no competing financial or non-financial interests.

Acknowledgments—This study was funded by National Natural Science Foundation of China (Nos. 81627804, 81830058, 81901760 and 82102070).

REFERENCES

Conti A, Duggento A, Indovina I, Guerrisi M, Toschi N. Radiomics in breast cancer classification and prediction. Semin Cancer Biol 2021;72:238–250.

Dogan BE, Gonzalez-Angulo AM, Gilcrease M, Dryden MJ, Yang WT. Multimodality imaging of triple receptor-negative tumors with mammography, ultrasound, and MRI. AJR Am J Roentgenol 2010;194:1160–1166.

Ferlay J, Colombet M, Soerjomataram I, Parkin DM, Pineros M, Znaor A, Bray F. Cancer statistics for the year 2020: An overview. Int J Cancer 2021;. doi: 10.1002/ijc.33588 Published online April 5.

Fleury E, Marcomini K. Performance of machine learning software to classify breast lesions using BI-RADS radiomic features on ultrasound images. Eur Radiol Exp 2019;3:34.

Ha EJ, Baek JH. Applications of machine learning and deep learning to thyroid imaging: Where do we stand?. Ultrasonography 2021;40:23–29.

Hu XC, Loo WT, Chow LW. Surgery-related shedding of breast cancer cells as determined by RT-PCR assay. J Surg Oncol 2003;82:228–232 discussion 233.

Isik A, Isik N, Kurnaz E. Complete breast autoamputation: Clinical image. Breast J 2020;26:2265–2266.

Ismail MS, Wynendaele W, Aerts JL, Paridaens R, Gaafar R, Shakankiry N, Khaled HM, Christiaens MR, Wildiers H, Omar S, Vandekerckhove P, Van Oosterom AT. Detection of micrometastatic disease and monitoring of perioperative tumor cell dissemination in primary operable breast cancer patients using real-time quantitative reverse transcription-PCR. Clinical Cancer Res 2004;10:196–201.

Koh J, Kim EK, Kim MJ, Yoon JH, Park VY, Moon HJ. Role of elastography for downgrading BI-RADS category 4a breast lesions according to risk factors. Acta Radiol 2019;60:278–285.

Lee SE, Han K, Kwak JY, Lee E, Kim EK. Radiomics of US texture features in differential diagnosis between triple-negative breast cancer and fibroadenoma. Sci Rep 2018;8:13546.

Li DD, Xu HX, Guo LH, Bo XW, Li XL, Wu R, Xu JM, Zhang YF, Zhang K. Combination of two-dimensional shear wave elastography with ultrasound breast imaging reporting and data system in the diagnosis of breast lesions: A new method to increase the diagnostic performance. Eur Radiol 2016;26:3290–3300.

Li S, Yan W, Yang X, Chen L, Fan L, Liu H, Liu K, Zhang Y, Jiang J. Less micrometastatic risk related to circulating tumor cells after endoscopic breast cancer surgery compared to open surgery. BMC Cancer 2019;19:1070.

Lian J, Li K. A review of breast density implications and breast cancer screening. Clin Breast Cancer 2020;20:283–290.

Mei H, Xu J, Yao G, Wang Y. The diagnostic value of MRI for architectural distortion categorized as BI-RADS category 3-4 by mammography. Gland Surg 2020;9:1008–1018.

Mendelson EB, Böhm-Vélez M, Berg WA, Whitman GJ, Feldman MI, Madjar H, Rizzatto G, Baker JA, Zuley M, Stavros AT, Comstock C, Van Duyn Wear V. ACR BI-RADS Ultrasound. ACR BI-RADS Atlas, Breast Imaging Reporting and Data System. 5th edition Reston, VA: American College of Radiology; 2013.

Mercado CL. BI-RADS update. Radiol Clin North Am 2014;52:481–487.

Munir K, Elahi H, Ayub A, Frezza F, Rizzi A. Cancer diagnosis using deep learning: A bibliographic review. Cancers (Basel) 2019;11:1235.

Niu S, Huang J, Li J, Liu X, Wang D, Zhang R, Wang Y, Shen H, Qi M, Xiao Y, Guan M, Liu H, Li D, Liu F, Wang X, Xiong Y, Gao S, Wang X, Zhu J. Application of ultrasound artificial intelligence in the differential diagnosis between benign and malignant breast lesions of BI-RADS 4A. BMC Cancer 2020;20:959.

Qian L, Lv Z, Zhang K, Wang K, Zhu Q, Zhou S, Chang C, Tian J. Application of deep learning to predict underestimation in ductal carcinoma in situ of the breast with ultrasound. Ann Transl Med 2021;9:295.

Sandler M, Howard A, Zhu M, Zhmoginov A, Chen LC. MobileNetV2: Inverted residuals and linear bottlenecks. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE; 2018. p. 4510–4520.

Shang J, Ruan LT, Wang YY, Zhang XJ, Dang Y, Liu B, Wang WL, Song Y, Chang SJ. Utilizing size-based thresholds of stiffness gradient to reclassify BI-RADS category 3-4b lesions increases diagnostic performance. Clin Radiol 2019;74:306–313.

Shia WC, Lin LS, Chen DR. Classification of malignant tumours in breast ultrasound using unsupervised machine learning approaches. Sci Rep 2021;11:1418.

Stavros AT, Freitas AG, deMello GGN, Barke L, McDonald D, Kaske T, Wolverton D, Honick A, Stanzani D, Padovan AH, Moura APC, de Campos MCV. Ultrasound positive predictive values by BI-RADS categories 3-5 for solid masses: An independent reader study. Eur Radiol 2017;27:4307–4315.

Tagliafico AS, Piana M, Schenone D, Lai R, Massone AM, Houssami N. Overview of radiomics in breast cancer diagnosis and prognostication. Breast 2020;49:74–80.

Valdora F, Houssami N, Rossi F, Calabrese M, Tagliafico AS. Rapid review: Radiomics and breast cancer. Breast Cancer Res Treat 2018;169:217–229.

Weismann C. Ultrasound elastography techniques in breast cancer. Radiologe 2021;61:170–176.

Zhang X, Li H, Wang C, Cheng W, Zhu Y, Li D, Jing H, Li S, Hou J, Li J, Li Y, Zhao Y, Mo H, Pang D. Evaluating the accuracy of breast cancer and molecular subtype diagnosis by ultrasound image deep learning model. Front Oncol 2021;11 623506.

Zheng X, Huang Y, Wang Y, Liu Y, Li F, Han J, Wang J, Cao L, Zhou J. Combination of different types of elastography in downgrading ultrasound Breast Imaging-Reporting and Data System category 4a breast lesions. Breast Cancer Res Treat 2019;174:423–432.

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.

Zhou H, Jin Y, Dai L, Zhang M, Qiu Y, Wang K, Tian J, Zheng J. Differential diagnosis of benign and malignant thyroid nodules using deep learning radiomics of thyroid ultrasound images. Eur J Radiol 2020a;127 108992.

Zhou LQ, Wu XL, Huang SY, Wu GG, Ye HR, Wei Q, Bao LY, Deng YB, Li XR, Cui XW, Dietrich CF. Lymph node metastasis prediction from primary breast cancer US images using deep learning. Radiology 2020b;294:19–28.

Zhou BY, Wang LF, Yin HH, Wu TF, Ren TT, Peng C, Li DX, Shi H, Sun LP, Zhao CK, Xu HX. Decoding the molecular subtypes of breast cancer seen on multimodal ultrasound images using an assembled convolutional neural network model: A prospective and multicentre study. eBioMedicine 2021;74 103684.