Medicine®

Detection and classification the breast tumors using mask R-CNN on sonograms

Jui-Ying Chiao, BS[a], Kuan-Yung Chen, MD[b], Ken Ying-Kai Liao, MEng[c], Po-Hsin Hsieh, BS[a], Geoffrey Zhang, PhD[d], Tzung-Chi Huang, PhD[a,c,e,*]

Abstract

Breast cancer is one of the most harmful diseases for women with the highest morbidity. An efficient way to decrease its mortality is to diagnose cancer earlier by screening. Clinically, the best approach of screening for Asian women is ultrasound images combined with biopsies. However, biopsy is invasive and it gets incomprehensive information of the lesion. The aim of this study is to build a model for automatic detection, segmentation, and classification of breast lesions with ultrasound images. Based on deep learning, a technique using Mask regions with convolutional neural network was developed for lesion detection and differentiation between benign and malignant. The mean average precision was 0.75 for the detection and segmentation. The overall accuracy of benign/malignant classification was 85%. The proposed method provides a comprehensive and noninvasive way to detect and classify breast lesions.

Abbreviations: BI-RADS = breast imaging reporting and data system, CNN = convolutional neural network, mAP = mean average precision, MRI = magnetic resonance imaging, NN = neural network, R-CNN = regions with convolutional neural network, RoI = region of interest, RoIAlign = region of interest alignment, RoIPool = region of interest pooling, RPN = region proposal network, SVM = support vector machine.

Keywords: breast cancer, mask R-CNN, ultrasound

1. Introduction

Breast cancer is a malignant tumor formed by the abnormal division of ducts or lobules. If the breast structure changes, it might produce tumors. Tumors can be classified into benign and malignant tumors according to the histopathology (eg, differentiation ability, cell pleomorphic, nuclear to cytoplasm ratio), or clinical biological indicators (eg, invasion and metastasis). And it is one of the most harmful diseases for women with the highest morbidity. In addition, the course of breast cancer develops rapidly. Thus delayed diagnosis may have a significant impact on patients.[1] If breast cancer diagnosis can be done earlier, its mortality can be decreased. Breast cancer screening is an efficient method to detect indeterminate breast lesions early.

The common way of breast screening is imaging diagnosis, which includes breast magnetic resonance imaging (MRI), mammography, and breast ultrasound. Different indications are associated with different imaging approaches. MRI for breast screening is highly sensitive to soft tissue lesions. However, it is costly, with a relatively long scan time and with a higher rate of false positives. Consequently, breast MRI is mainly recommended for women at high risk of breast cancer.[2] Mammography is highly sensitive to the detection of calcifications but with limitations on people with dense breast tissues.

Breast ultrasound uses the transducer to convert electrical signals into ultrasound signals. Based on the different magnitude of reflected ultrasound waves and echoes time, the reflected sound waves can create an image through computer processing. As a result, ultrasound has the advantage of no ionizing radiation and real-time examination. Clinically, ultrasound is used for echo-guided biopsy examinations. Currently, mammography and breast ultrasound are the most common screening approaches.

Breast imaging reporting and data system (BI-RADS) proposed by the American College of Radiology suggests mammography as the standard imaging approach for breast screening. However, the breast density of Asian women is denser than Western women.[3] Women with dense breast are at greater risk of breast cancer,[4] and the sensitivity of mammography decreases 30% in dense-breasted women.[2,5] Because of this, breast ultrasound plays a vital role for Asian women in comparison to mammography.

Clinically, ultrasound is generally combined with biopsies to aid in the diagnosis of breast lesions. However, biopsy is an invasive procedure at risk of infection. Besides, on account of tumor heterogeneity, biopsy only gets incomprehensive information of

the tumor. To overcome the shortages of the ultrasound/biopsy combination screening, the purpose of this study was to distinguish breast lesions between benign and malignant comprehensively and prevent unnecessary biopsy by objectively analyzing noninvasive breast ultrasound images.

In a previous study of breast cancer classification, local texture features are important characteristics. They applied computer-aided diagnosis in breast ultrasound to quantify lesions by BI-RADS features including shape, orientation, margin, lesion boundary, echo pattern, and posterior acoustic feature classes to find the correlation between the extracted image features and the lesion. However, each feature has significant differences in the correlation of pathological section results.[6] Another research used an artificial neural network based on the 5 characteristics of spiculation, ellipsoid shape, branch pattern, brightness of nodule, number of lobulations to effectively distinguish between benign, and malignant breast lesions.[7] Besides, Li et al used deep learning and feature-based statistical learning to evaluate breast density and compare the effectiveness of the above 2 methods.[8] The results showed that techniques using deep learning are better than feature-based statistical learning. Therefore, this study used deep learning technique in breast lesions classification.

Neural network (NN) is a mathematical model to simulate the structure and function of biological NN. Convolutional neural networks (CNN) has a strong ability in image recognition and has been proven a good tool for judging the characteristics of borders and colors. Regions with CNN (R-CNN) applies CNN in object detection. However, R-CNN is slow to generate region proposal. To increase efficiency, Fast R-CNN combines feature extraction, classifier, and bounding box prediction of R-CNN into 1, and proposes a method called region of interest pooling (RoIPool).[9] The above approaches reduce the number of convolutions and the detection time but still uses the selective search method, which is still time-consuming, to generate region proposal. Consequently, Faster R-CNN proposes extracting region proposal by CNN that shares convolutional layers for getting region proposal, class, bounding box simultaneously to speed up the system. In this study, Mask R-CNN approach was taken, which is based on Faster R-CNN and has the advantage of automatic image segmentation – defining the tumor bounding box, drawing a contour of the tumor area, before lesion classification between benign and malignant.

The aim of this work was to build a model for automatic detection, segmentation, and classification of breast lesions with ultrasound images. And the results of this study were compared with biopsy results, which are the gold standard for breast cancer diagnosis. To establish a benign and malignant classification model of breast cancers, Mask R-CNN was applied to achieve automatic tumor contouring and classification. It also can provide more quantitative information in breast ultrasound images and improve the consistency and accuracy of benign and malignant classification of breast cancers.

2. Material and methods

2.1. Establishment of the imaging database – case collection and tracking

This study retrospectively collected the primary ultrasound images with biopsy histological and diagnostic report from China Medical University Hospital. This study protocol was reviewed and approved by Institutional/Independent Review Board (IRB: CMUH106-REC1-087). Patients who underwent breast ultrasound examination accompanied by biopsy in China Medical University Hospital were included in the study group. The breast ultrasound images, histological confirmation, and clinical information, including the category of BI-RADS and the biopsy report of patients were collected. In this study, a total of 80 cases were recruited and the image datasets were composed of 307 images of ultrasound images obtained during echo guide biopsy.

Ultrasound was performed by radiologists using GE ultrasound machine (LOGIQ S8, GE Medical Systems, Milwaukee, WI) with a 9 to 12-MHz transducer. The original image format was Digital Imaging and Communications in Medicine and the image size was 960×720 pixels, where 1-pixel size corresponded to $0.08 \, mm \times 0.08 \, mm$. Images with artifact and incomplete tumor were excluded. Figure 1 shows ultrasound images of a pair of typical benign and malignant breast lesions.

2.2. Contouring and classification of tumor

After collecting the ultrasound image, the radiologist with 7 years of work experience using image J delineated the contour of the tumor area, and the physician classified the lesions into 6 BI-RADS categories. The categories associated with the clinical
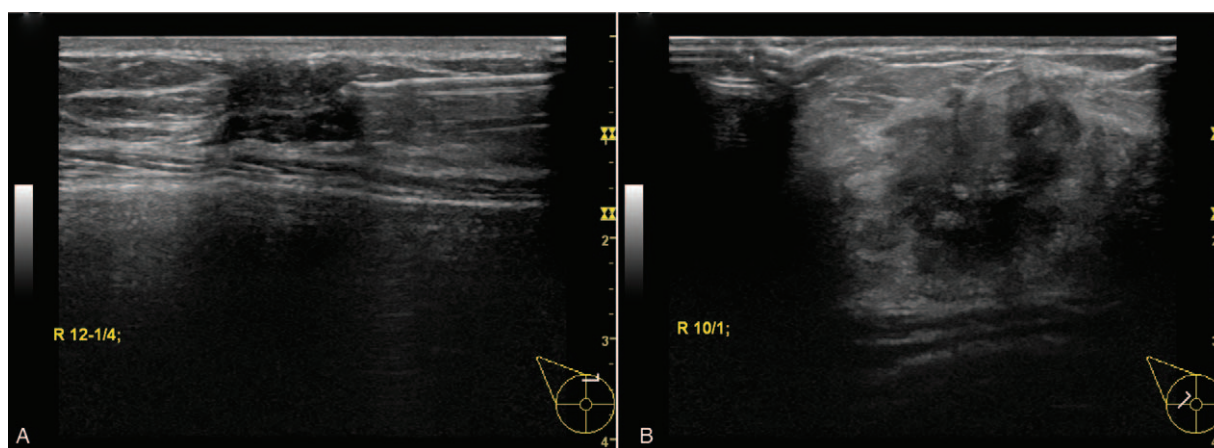

Figure 1. Breast ultrasound images of (a) benign lesion, (b) malignant tumor.

BI-RADS categories associated with the clinical assessment.

Category	Assessment
1	Negative
2	Benign
3	Probably benign
4	Suspicious malignancy
5	Highly suspicious malignancy
6	Proven malignancy

BI-RADS = breast imaging reporting and data system.

assessment are listed in Table 1. Clinically, if the lesion was sorted into category 3, the clinician assessed and determined whether to proceed with biopsy. If the BI-RADS category was 4 or higher, the clinician mostly suggested proceeding with biopsy to aid in the discrimination of lesions' types and benign-malignant classification. In this study, the results of tumor contour and biopsy were used as the ground truth for Mask R-CNN network training.

2.3. Mask R-CNN techniques

Object detection and segmentation are to distinguish different objects in an image and draw the bounding box on a specific object. Mask R-CNN is one of the methods of object detection and segmentation. It can not only draw a bounding box for the target object, but also further mark and classify whether the pixels in the bounding box belong to the object or not, which can be used to identify the object, mark the boundary of the object, and detect key points.

Mask R-CNN is based on Faster R-CNN and extends its application to the field of image segmentation. Its network architecture is illustrated in Figure 2. The process of Mask R-CNN is similar to Faster R-CNN, both using region proposal network (RPN) to extract features, and to classify and tighten bounding boxes. Faster R-CNN uses RoIPool as a feature extraction method for quantifying each RoI region, and solving the problem of sizes of RoI features at different scales by max pooling.[9] However, the process causes the loss of spatial information, making the original image RoI and extraction features misplaced. To solve this problem, Mask R-CNN replaces RoI pooling of Faster R-CNN with ROI alignment (RoIAlign),

and consecutively uses the mask branch to mark the result of RoIAlign for the object area.

After the network architecture was completed, Mask R-CNN was trained using the ultrasound images and the corresponding biopsy data, tumor contours, drawn by a radiologist, as ground truth. The training process randomly split the collected cases into a training set and a validation set, and the model established by the training set data was tested against the validation set in order to ensure the accuracy and stability of the model. The value of the loss function L, $L_{class} + L_{box} + L_{mask}$, in Mask R-CNN was minimized, and the most suitable model through the minimization of the loss function on the training data was used as the NN model. The trained model was applied to predict and analyze with new data, such as the validation set.

The loss function of Mask R-CNN is defined as:

$$L = L_{class} + L_{box} + L_{mask}$$

where $L_{class} + L_{box}$ are identified the same as in Faster R-CNN, $L_{class} + L_{box}$ are defined as:

$$L_{class} + L_{box} = \frac{1}{N_{cls}} \sum_i L_{cls}(p_i, p_i^*) + \frac{1}{N_{box}} \sum_i p_i^* L_1^{smooth}(t_i - t_i^*)$$

$$L_{cls}(\{p_i, p_i^*\}) = -p_i^* log p_i^* - (1 - p_i^*) log(1 - p_i^*)$$

And the L_{mask} is the average binary cross-entropy loss:

$$L_{mask} = -\frac{1}{m^2} \sum_{1 \le i,j \le m} \left[y_{ij} log \circ y_{ij}^k + (1 - y_{ij}) log(1 - \circ y_{ij}^k) \right]$$

| Symbol	Explanation
...p_i	Predicted probability of anchor i being an object
p_i^*	Ground truth label (binary) of whether anchor i is an object
t_i	Predicted 4 parameterized coordinates
t_i^*	Ground truth coordinates
N_{cls}	Normalization term, set to be mini-batch size (~2) in the paper
N_{box}	Normalization term, set to the number of anchor locations (~256) in the paper
λ	A balancing parameter, set to be ~10 in the paper (so that both L_{class} and L_{box} terms are roughly equally weighted).
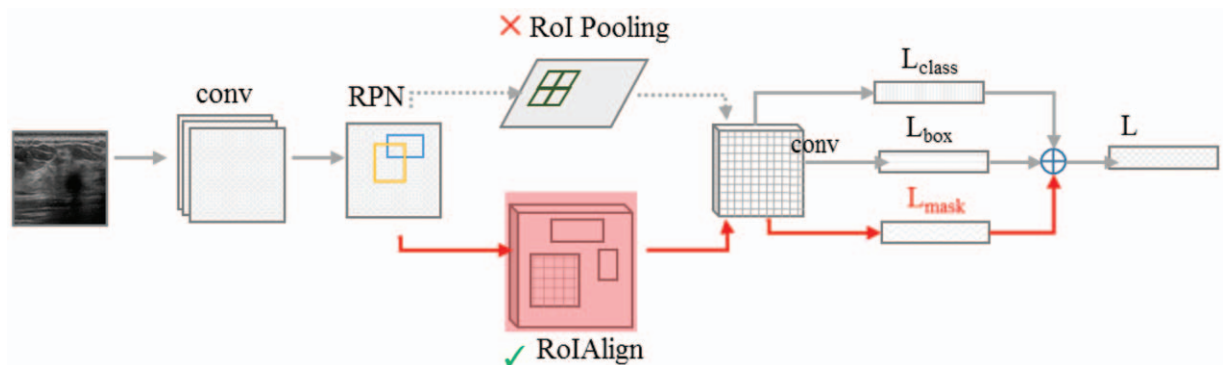
Figure 2. The network architecture of Mask R-CNN. RoIAlign replaces RoI Pooling in Mask R-CNN, and the mask branch is consecutively used to mark the result of RoIAlign. Gray flow chart is the original Faster R-CNN, and the red one is differences and amendments between Mask R-CNN and Faster R-CNN. R-CNN = regions with convolutional neural network, RoI = region of interest, RoIAlign = region of interest alignment, RoIPool = region of interest pooling.
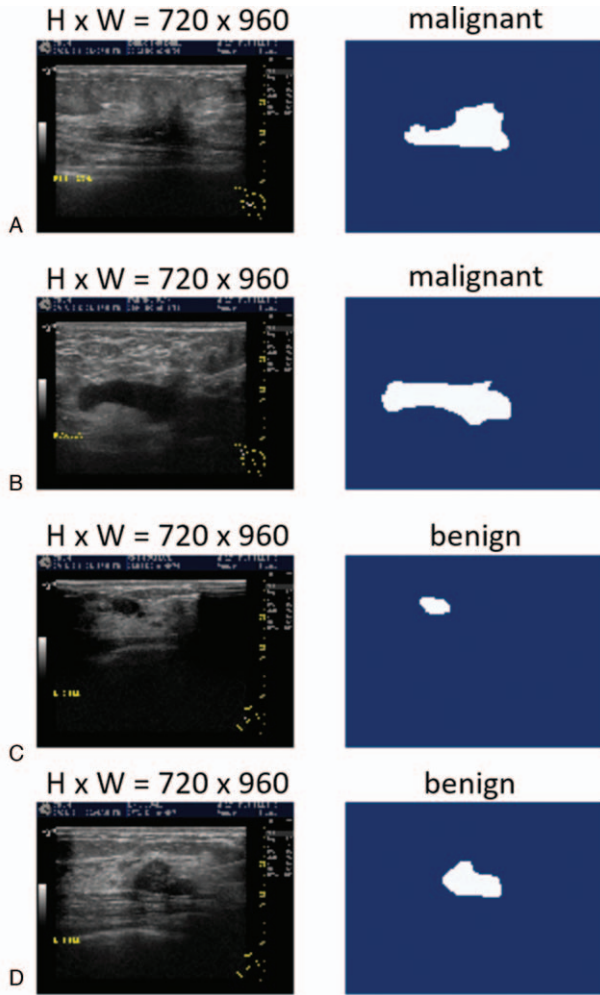
Figure 3. Example of tumor contour. (a, b) An original image of malignant tumor and contour mask (white area); (c, d) an original image of benign tumor and contour mask (white area).

The performance of the trained Mask R-CNN model was quantitatively evaluated by mean average precision (mAP) as the accuracy of lesion detection/segmentation on the validation set:

$$mAP = \frac{A \cap B}{A \cup B} = \frac{1}{N_T} \sum_{i=1}^{N_T} \left(\frac{N_i^{DR}}{N_i^D} \right) \qquad (1)$$

where A is the model segmentation result and B is the corresponding tumor contour delineated by the experienced radiologist, true clinical lesion, as the ground truth. N_T is the number of images; N_i^{DR} is the overlapped area between the model detected lesion and the true clinical lesion regions; and N_i^D is the size of the true clinical lesion.

The overall lesion classification performance of the proposed method was validated by accuracy. The measures of accuracy is evaluated by the following equations:

$$Accuracy = (TP + TN)/(TP + TN + FP + FN) \qquad (2)$$

where TP = true positive, TN = true negative, FP = false positive and FN = false negative.

3. Results

In this study, the 307 cases in the image database (178 benign and 129 malignant) were splitted into 80% as the training set and 20% as the validation set.

Figure 3 shows the results of tumor contour by professional radiologists. Figure 3(a) and (b) are breast ultrasound images of 2 different malignant tumors, (c) and (d) are benign tumors. The left side of the image is the original reference image, and the right side shows the actual mask produced by a professional radiologist referenced from the original image.

Figure 4 shows an example of lesion segmentation evaluation with the contour delineated by a radiologist and the corresponding result of the model segmentation. The mAP was 0.75 for the automatic lesion delineation in validation.

The accuracy of benign-malignant classification of breast cancers compared with histological results was 85% in validation.

The loss is 0.9648; RPN class loss is 0.0159; RPN bounding box loss is 0.1581; Mask R-CNN class loss is 0.0659; Mask R-CNN bounding box loss is 0.2583; Mask R-CNN mask loss is 0.4666; validation loss is 1.5698; validation RPN class loss is 0.0147; validation RPN bounding box loss is 0.5478; validation Mask R-CNN class loss is 0.0829; validation Mask R-CNN bounding box loss is 0.4343; validation Mask R-CNN mask loss is 0.4901.

4. Discussion

The aim of this work is to build a model to automatic detection, segmentation and classification of breast lesions with ultrasound images. The traditional generation of RoI region shape is usually
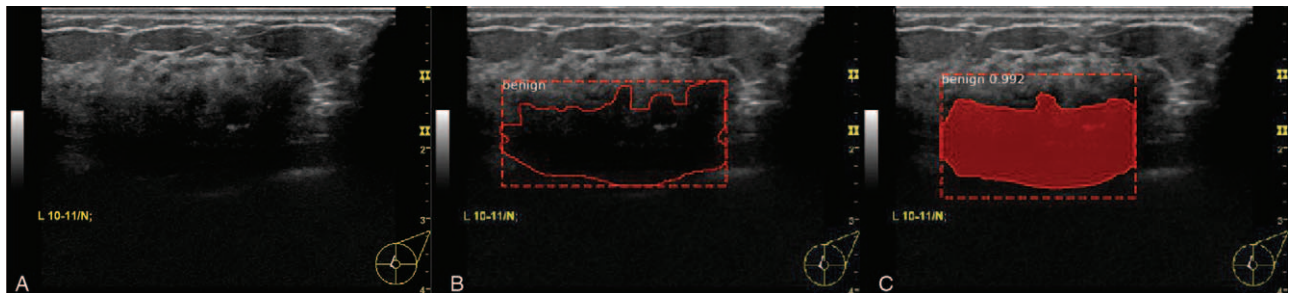

Figure 4. Example of lesion segmentation evaluation. (a) A benign lesion; (b) the radiologist delineated the red contour (solid line), and the rectangular box was calculated according to the manual contour (dashed line); (c) the automatic lesion delineation by the proposed method. The confident score for this case was 0.992.

rectangular which only can delineate lesion contour roughly. And it is difficult to auto-segmentation in ultrasound images due to its low image quality.[10] If more normal tissues in RoI can be excluded, the differentiation between tumor and normal tissues would be more accurate.[11]

A few other recent studies used support vector machine (SVM),[12,13] a method of machine learning, in detection and classification. Those methods needed to extract features form RoI and then the features were given to SVM classifier through SVM detection. Besides, those studies used active contour method in lesion detection, for which statistical features were applied to find seed points and then delineate the lesion.

In this study, RoI regions were automatically delineated and features were extracted from images by CNN layer by layer without previously giving the features. As a result, the proposed method has the advantage of observation lesions comprehensively, not only by analyzing single features.

Ultrasound images is an effective diagnostic tool for breast cancer detection. In order to visualize lesions clearly, the radiologists must change the depth of images along with lesion depth. The way of changing depth is important for identifying deep lesions in breast ultrasound images.[14] But the thickness of the breast in each case is different and each lesion is in different depth. As a result, the change of depth might lead to misinterpretation which in consequence may decrease the accuracy.

Some studies need to preprocess images before extracting features.[13,15] But it was not required in this study. In those studies, preprocessing images was supposed to reduce the noise in the images and thus to improve the accuracy. However, another study concluded that the reduction of speckle noise does not improve the diagnostic performance.[16] And the other study even used the speckle noise as the feature in computer-aided classification of breast masses.[17] As a result, preprocessing images could influence the result of classification, although how it could influence overall performance is uncertain at this point.

5. Conclusions

In this study, a method of automatic detection, segmentation and classification of breast lesions with ultrasound images is proposed. It can accurately delineate the lesion regions and classify the regions into benign or malignant.

By the combination of breast ultrasound images and deep learning, it can provide the information that was not available in traditional diagnostic software in the past. The proposed method can improve the consistency and accuracy of benign–malignant classification of breast lesions and it can serve as a new tool for clinical diagnosis. In the future, the number of cases in the image database is expected to increase and the hyperparameters in deep learning are expected to be more optimized, which will increase the model's accuracy further.

Author contributions

Conceptualization, W.C. Chiang, G. Zhang and T.C. Huang; Methodology, T.C. Huang and Y.K. Liao; Software, Y.K. Liao; Validation, Y.K. Liao; Formal Analysis, Y.K Liao; Investigation, Y.K. Liao and J.Y. Chiao; Resources, T.C. Huang and Y.K. Liao; Data Curation, T.C. Huang, Y.K. Liao, and J.Y. Chiao; Writing-Original Draft Preparation, J.Y. Chiao; Writing-Review and Editing, Y.K. Liao, G. Zhang and T.C. Huang; Visualization, Y. K. Liao; Supervision, G. Zhang and T.C. Huang; Project Administration, T.C. Huang.

Conceptualization: Kuan-Yung Chen, Tzung-Chi Huang.
Data curation: Ying-Kai Ken Liao.
Formal analysis: Jui-Ying Chiao, Ying-Kai Ken Liao.
Methodology: Jui-Ying Chiao, Kuan-Yung Chen, Ying-Kai Ken Liao.
Resources: Kuan-Yung Chen, Tzung-Chi Huang.
Software: Tzung-Chi Huang.
Supervision: Tzung-Chi Huang.
Validation: Jui-Ying Chiao, Ying-Kai Ken Liao, Po-Hsin Hsieh.
Visualization: Ying-Kai Ken Liao.
Writing – original draft: Jui-Ying Chiao, Ying-Kai Ken Liao, Po-Hsin Hsieh.
Writing – review and editing: Kuan-Yung Chen, Po-Hsin Hsieh, Geoffrey Zhang, Tzung-Chi Huang.

References

[1] Shieh S-H, Hsieh VC-R, Liu S-H, et al. Delayed time from first medical visit to diagnosis for breast cancer patients in Taiwan. J Formosan Med Assoc 2014;113:696–703.

[2] Kelly KM, Dean J, Comulada WS, et al. Breast cancer detection using automated whole breast ultrasound and mammography in radiographically dense breasts. Eur Radiol 2010;20:734–42.

[3] Maskarinec G, Meng L, Ursin G. Ethnic differences in mammographic densities. Int J Epidemiol 2001;30:959–65.

[4] Boyd NF, Martin LJ, Yaffe MJ, et al. Mammographic densities and breast cancer risk. Cancer Epidemiol Prevent Biomark 1998;7:1133–44.

[5] Mandelson MT, Oestreicher N, Porter PL, et al. Breast density as a predictor of mammographic detection: comparison of interval-and screen-detected cancers. J Natl Cancer Inst 2000;92:1081–7.

[6] Shen W-C, Chang R-F, Moon WK, et al. Breast ultrasound computer-aided diagnosis using BI-RADS features. Acad Radiol 2007;14:928–39.

[7] Joo S, Yang YS, Moon WK, et al. Computer-aided diagnosis of solid breast nodules: use of an artificial neural network based on multiple sonographic features. IEEE Trans Med Imaging 2004;23:1292–300.

[8] Li SF, Wei J, Chan HP, et al. Computer-aided assessment of breast density: comparison of supervised deep learning and feature-based statistical learning. Phys Med Biol 2018;63:2.

[9] Ren S, He K, Girshick R, Sun J. Faster r-cnn: Towards real-time object detection with region proposal networks. Advances in Neural Information Processing Systems 2015;39:91–9.

[10] Noble JA, Boukerroui D. Ultrasound image segmentation: a survey. IEEE Trans Med Imaging 2006;25:987–1010.

[11] Xian M, Zhang Y, Cheng H. Fully automatic segmentation of breast ultrasound images based on breast characteristics in space and frequency domains. Pattern Recognit 2015;48:485–97.

[12] Prabhakar T and Poonguzhali S. Automatic detection and classification of benign and malignant lesions in breast ultrasound images using texture morphological and fractal features. in Biomedical Engineering International Conference (BMEiCON), 2017 10th. 2017. IEEE.

[13] Menon RV, et al. Automated detection and classification of mass from breast ultrasound images. in Computer Vision, Pattern Recognition, Image Processing and Graphics (NCVPRIPG), 2015 Fifth National Conference on. 2015. IEEE.

[14] Park JM, Yang L, Laroia A, et al. Missed and/or misinterpreted lesions in breast ultrasound: reasons and solutions. Can Assoc Radiol J 2011;62:41–9.

[15] Marcomini KD, Carneiro AA, Schiabel H. Application of artificial neural network models in segmentation and classification of nodules in breast ultrasound digital images. Int J Biomed Imaging 2016;2:1–3.

[16] Tseng HS, Wu HK, Chen ST, et al. Speckle reduction imaging of breast ultrasound does not improve the diagnostic performance of morphology-based CAD System. J Clin Ultrasound 2012;40:1–6.

[17] Moon WK, Lo CM, Chang JM, et al. Computer-aided classification of breast masses using speckle features of automated breast ultrasound images. Med Phys 2012;39:6465–73.