

RESEARCH ARTICLE

Editorial Process: Submission:11/13/2022 Acceptance:03/20/2023

An Automatic Breast Tumor Detection and Classification including Automatic Tumor Volume Estimation Using Deep Learning Technique

Prinda Labcharoenwongs*, Suteera Vonganansup, Orawan Chunhapran, Duangjai Noolek, Tongjai Yampaka

Abstract

Objective: This study aims to develop automatic breast tumor detection and classification including automatic tumor volume estimation using deep learning techniques based on computerized analysis of breast ultrasound images. When the skill levels of the radiologists and image quality are important to detect and diagnose the tumor using handheld ultrasound, the ability of this approach tends to assist the radiologist's decision for breast cancer diagnosis. **Material and Methods:** Breast ultrasound images were provided by the Department of Radiology of Thammasat University and Queen Sirikit Center of Breast Cancer of Thailand. The dataset consists of 655 images including 445 benign and 210 malignant. Several data augmentation methods including blur, flip vertical, flip horizontal, and noise have been applied to increase the training and testing dataset. The tumor detection, localization, and classification were performed by drawing the appropriate bounding box around it using YOLO7 architecture based on deep learning techniques. Then, the automatic tumor volume estimation was performed using a simple pixel per metric technique. **Result:** The model demonstrated excellent tumor detection performance with a confidence score of 0.95. In addition, the model yielded satisfactory predictions on the test sets, with a lesion classification accuracy of 95.07%, a sensitivity of 94.97%, a specificity of 95.24%, a PPV of 97.42%, and an NPV of 90.91%. **Conclusion:** An automatic breast tumor detection and classification including automatic tumor volume estimation using deep learning technique yielded satisfactory predictions in distinguishing benign from malignant breast lesions. In addition, automatic tumor volume estimation was performed. Our approach could be integrated into the conventional breast ultrasound machine to assist the radiologist's decision for breast cancer diagnosis.

Keywords: Deep learning- ultrasonography- breast cancer diagnosis- artificial intelligence

Asian Pac J Cancer Prev, 24 (3), 1081-1088

Introduction

The new cases of breast cancer is increasing every years (Siegel et al., 2022). When early detection of breast cancer is an effective method to decrease the mortality rate, ultrasound is used to detect and diagnose breast lesions when abnormalities are identified by other imaging modalities or on palpation (Kornecki, 2011). In addition, the ultrasound (US) had a higher sensitivity and diagnostic accuracy (Shen et al., 2015). When the handheld ultrasound is used, the skill levels of the radiologists and image quality are importance to detect and diagnose the tumor (Komatsu et al., 2021). Therefore, computerized analysis of breast images has been widely introduced to increase the efficiency breast screening using a computer system to help radiologists detect and diagnose abnormalities (Jiang et al., 1999; Giger, 2000).

In recent years, artificial intelligence (AI) using deep learning methods has applied in medical fields. For instance, the detection of tuberculosis on chest radiographs, detection and diagnosis of lung nodules on chest CT, and segmentation of brain tumor on MRI were success using deep learning methods (Budd et al., 2021; Akkus et al., 2017; Lakhani and Sundaram, 2017). Furthermore, deep learning has proven useful in the field of automatic breast ultrasound (ABUS).

According to the state-of-the-art of ABUS for the detection and segmentation of tumors in BUS images was reviews. Kumar et al., (2018) proposed a segmentation method based on an ensemble of ten U-Net networks to reduce the uncertainty of finding the minima associated with the random initialization of each network. This method achieved an F1-score (F1s) of 0.82. You Only Look Once (YOLO) network and Single Shot MultiBox

Detector (SSD) are commonly used for real-time object detection. In the experiments, the SSD model obtained the highest F1s score of 0.79. Chiao et al. (2019) used an extension of the Faster R-CNN for tumor segmentation. Their model called The Mask R-CNN model obtained an Intersection over Union (IoU) of 0.75. Amiri et al. (2020) developed a two-stage segmentation method based on the U-Net architecture. This method attained F1s = 0.86 and F1=0.80 with and without test time augmentation procedure, respectively. The Dense skip U-Net (DsUnet) network that was proposed by Cui et al. (2020) is a segmentation approach that is based on the U-Net model. From the experiments, the results show that the DsUnet model reached F1s = 0.86.

According to our survey, many studies are successful in breast ultrasound detection and segmentation. However, no one study proposed automatic tumor volume estimation. Therefore, this study aims to develop automatic breast tumor detection and classification including automatic tumor volume estimation using deep learning technique. The ability of the model tends to assist the radiologist's decision for breast cancer diagnosis.

Materials and Methods

Materials

Breast ultrasound images were provided by the Department of Radiology of Thammasat University and Queen Sirikit Center of Breast Cancer of Thailand. The dataset consists of 655 images including 445 benign and 210 malignant. The dataset consists of 655 images including 445 benign and 210 malignant. Figure 1 shows some example images from our experiment dataset.

If a dataset is very small, may still not be enough for a given problem. The accuracy of deep learning models largely depends on the quality, quantity, and contextual meaning of training data. However, data scarcity is one of the most common challenges in building deep learning models. In production use cases, collecting such data can be costly and time-consuming. Hence, several data augmentation methods including blur, flip vertical, flip horizontal, and noise have been applied to improve the classification performance. Table 1 Data augmentation techniques were used in this study. In addition, Table 2 shows the original sample size compared with the augmentation sample size.

Methods

Deep learning for object detection and classification

Deep learning for object detection involves not only recognizing and classifying every object in an image, but localizing each one by drawing the appropriate bounding box around it. This technique is extended from traditional computer vision and image classification. In recent years, many model architectures were successful approaches to object detection such as R-CNN, Faster R-CNN, and YOLO (Redmon et al., 2016, Ren et al., 2017, and Virasova et al., 2021). Realtime object detection advances with the release of YOLO. YOLOv7 infers faster and with greater accuracy than its previous versions (i.e. YOLOv5), pushing the state of the art in object detection

to new heights. These features are combined and mixed in the neck, and then they are passed along to the head of the network YOLO predicts the locations and classes of objects around which bounding boxes should be drawn.

Figure 2 shows the training process. First, input images were fed to the convolution layer based on YOLOv7 architecture. This process iteratively trained until archive the best model performance. Second, the model tries to localizing each tumor by drawing the appropriate bounding box around it. Third, the tumor volume was estimated from the size of minimum bounding box. Finally, detected object classify the tumor in benign or malignant. Once a deep learning model has been trained, it can be used to make predictions about new data. To do this, we pass the new data through the network and use the output of the final layer to make our predictions.

The object detector is responsible for identifying which pixels in an image belong to an object, and the regressor is responsible for predicting the coordinates of the bounding box around that object. The output of the object detector will typically be a set of bounding boxes around the detected objects, along with a confidence score for each bounding box.

The regressor is then trained on these bounding boxes to learn how to predict the coordinates of the tightest possible bounding box around an object. After both the object detector and regressor have been trained, they can be combined into a single model that can be used to detect and localize objects in new images.

Performance Evaluation

The model evaluation such as precision, recall, and accuracy were used to evaluate the model performance. The calculation types of the metrics are shown in Equations (1)–(3), where TP, FN, FP, TN represent the number of true positives, false negatives, false positives and true negatives.

$$precision = \frac{TP}{TP + FP} \quad (1)$$

$$recall = \frac{TP}{TP + FN} \quad (2)$$

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \quad (3)$$

Intersection over Union (IoU) was used to quantifies the degree of overlap between two boxes. In the case of object detection and segmentation, IoU evaluates the overlap of Ground Truth and Prediction region that helps the model measure the correctness of a prediction. Fig x shows the example to understand how IoU is calculated.

Figure 3 showed that the predicted box of model (a) has more overlap with the Ground Truth as compared to model (b). However, model (c) has an even higher overlap with the ground truth. But it also has a high overlap with the background. It is clear that model (b) and (c) not just about matching the Ground Truth, but how closely the prediction matches the Ground Truth.

Two bounding boxes over iteration were compared for

all the detected objects using the Intersection over Union (IoU) as follow:

$$IoU = \frac{\text{Area of Intersection}}{\text{Area of Union}} \quad (4)$$

In addition, Recall (RE), and the mAP. An Average Precision (AP) formally presents in:

$$AP = \frac{1}{N} \sum_{k=1}^N P(k) \Delta r(k) \quad (5)$$

Where the $P(k)$ refers to the precision at a specifically given threshold k , and $\Delta r(k)$ as the shift in the Recall. For multiple object detection, the mAP calculates the mean of all AP for each category as follow:

$$mAP = \frac{1}{N} \sum_{i=1}^N AP_i \quad (6)$$

Volume measurement

This section explains how to extract the boxes from the raw image and measures the object size. In yolo, a bounding box is represented by four values [x_center , y_center , $width$, $height$]. The x_center and y_center are the normalized coordinates of the center of the bounding box. To make coordinates normalized, we take pixel values of x and y , which marks the center of the bounding box on the x -axis and y -axis. Then we divide the value of x by the width of the image and value of y by the height of the image. The width and height represent the width and the height of the bounding box. When the object pixel was found, the pixel density was used to estimated the object size. Figure 4 shows the volume measurement process.

Assuming the pixel density is 96 dpi, there are 96

pixels per inch. We know that 1 inch is equal to 2.54 cm. So, there are 96 pixels per 2.54 cm. Than 1 pixel = (2.54 / 96) cm. Finally, there are 0.026458333 centimeters in a pixel. Figure 5 show the tumor volume size.

System integration

Integration of various systems of medicine provides the best available therapeutic care to the patient without undue delay, making way for a better prognosis. In recent years, new approaches for existing diseases or newly emerging diseases are thought out. However, developing new software, new tools, or new systems is time-consuming and costly. Instead of developing new systems, this study proposed an automatic breast detection and classification including automatic tumor volume estimation that could be integrated into the conventional breast ultrasound machine. The proposed system is shown in Figure 6.

Results

Data augmentation result

In this section, we present the comparative classification performance of two approach (original dataset and augmentation dataset) using AI models as summarized in Table 3 The quantitative comparison of precision, recall, and accuracy were estimated over the testing dataset.

The original breast ultrasound dataset achieved a lesion classification a precision of 0.92, a recall of 0.84, and an accuracy of 0.88. The effectiveness of the model was evaluated from the two classes, with a precision of 0.89, a recall of 0.85, and an accuracy of 0.86 from the benign class. Moreover, the malignant class achieved higher than the benign class, with a precision of 0.95, a recall of 0.86, and an accuracy of 0.87.

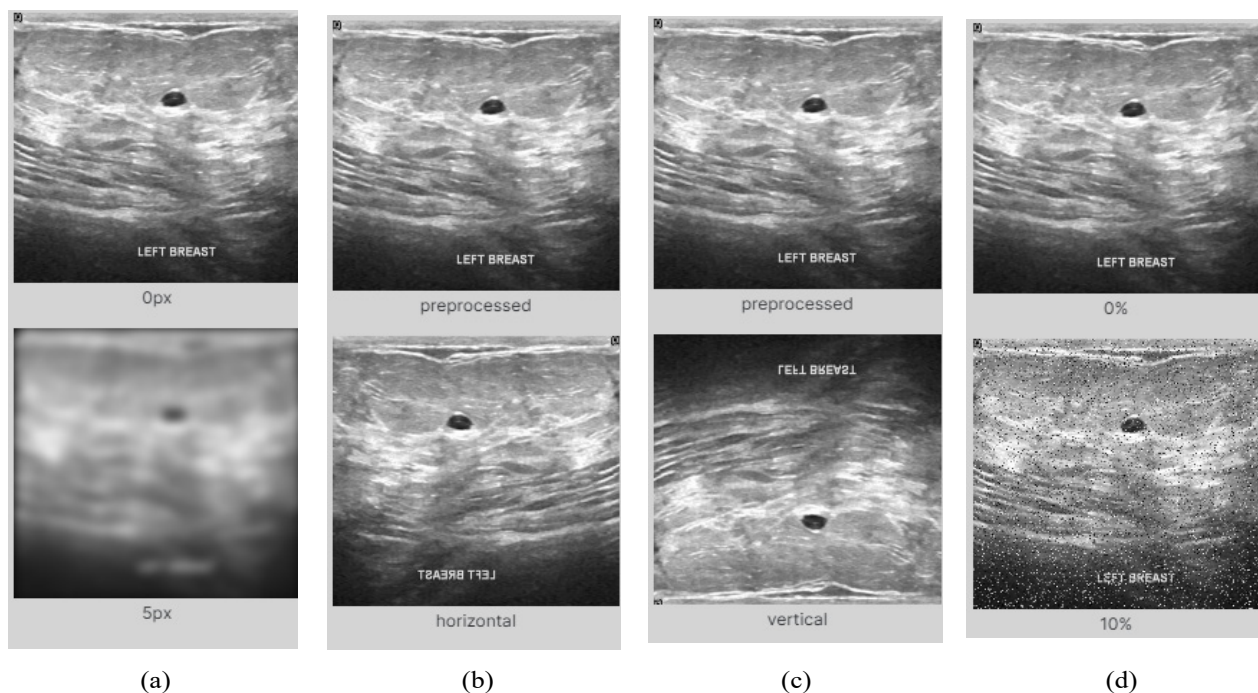


Figure 1. Example Breast Ultrasound Using Augmentation Techniques Including (a) blur 5px, (b) flip horizontal, (c) flip vertical, and noise 10%

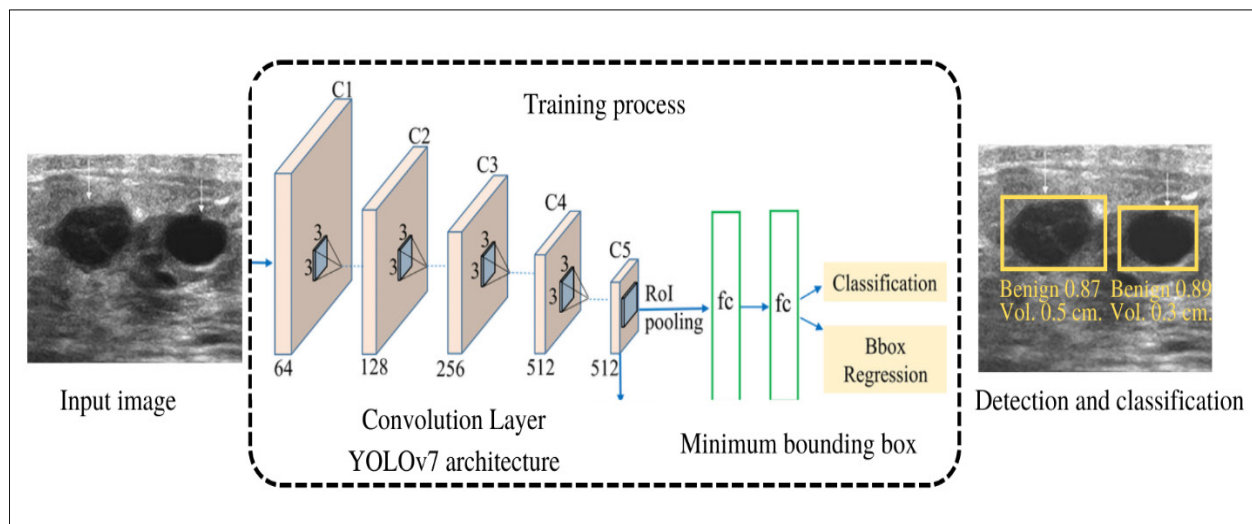


Figure 2. The Process of Iterative Object Detection and Classification.

Table 1. Data Augmentation Techniques have been Used in This Study.

Technique	Facility
Blur	5px
Noise	5%
Flip vertical	180°
Flip horizontal	180°

The results from data augmentation tend to improve the model performance. All classes classification achieved a precision of 0.97, a recall of 0.89, and an accuracy of 0.95. The effectiveness of the model was evaluated from the two classes, with a precision of 0.94, a recall of 0.88, and an accuracy of 0.93 from the benign class. Moreover, the malignant class achieved higher than the benign class, with a precision of 1, a recall of 0.89, and an accuracy of 0.95.

The empirical results showed that data augmentation is useful in improving the performances and outcomes of ABUS models. It could reduce the cost of the data collection process by transforming new synthetic images for image classification.

Table 2. The Sample Size of Each Class that has been Used in This Study

Dataset	Original	Augmentation
Benign	445	1,335
Malignant	210	630
Total	655	1,965

RoI extraction and bounding-box regression results

The lesion detection in ABUS usually uses a bounding box to describe the spatial location of the tumor. The bounding box is rectangular, which is determined by the x and y coordinates of the upper-left corner of the rectangle and the such coordinates of the lower-right corner. Another commonly used bounding box representation is the (x, y)-axis coordinates of the bounding box center, and the width and height of the box. Figure 7 shows the results of RoI extraction using minimum bounding box.

In addition, the mAP compares the ground-truth bounding box to the detected box and returns a score. The higher the score represents the more accurate the model in its detections. Figure 8 shows the mean Average Precision or mAP score by taking the mean AP over all classes and/or overall IoU thresholds. The result show that the

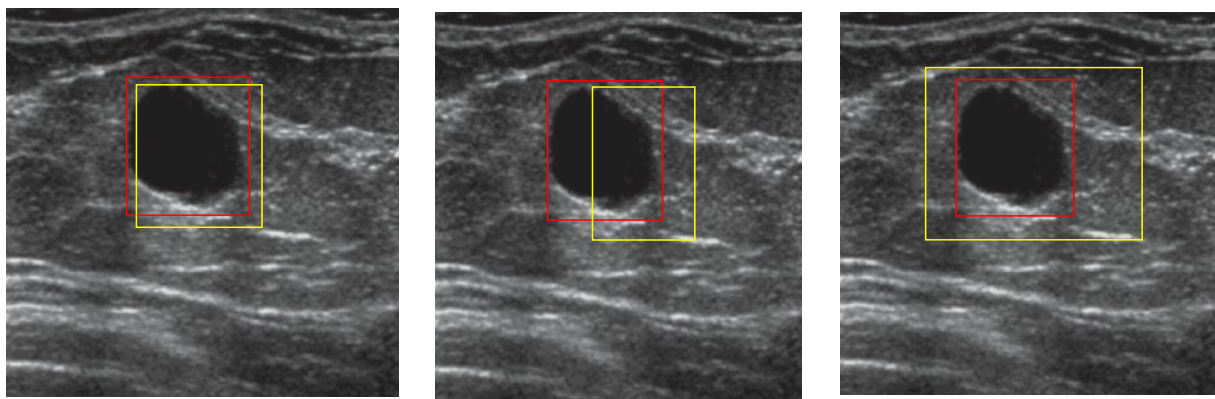


Figure 3. There are Three Models- a, b, and c, Trained to Predict Tumor. An image was passed through the models where trained model already know the Ground Truth (marked in red). The image shows predictions of the models (marked in yellow).

Table 3. The Comparative Classification Performance of Two Approach (Original Dataset and Augmentation Dataset) Using AI Models

Class	Original dataset			Augmentation dataset		
	Precision	Recall	Accuracy	Precision	Recall	Accuracy
Benign	0.89	0.85	0.89	0.94	0.88	0.93
Malignant	0.95	0.86	0.87	1	0.89	0.89
All classes	0.92	0.84	0.88	0.97	0.89	0.95

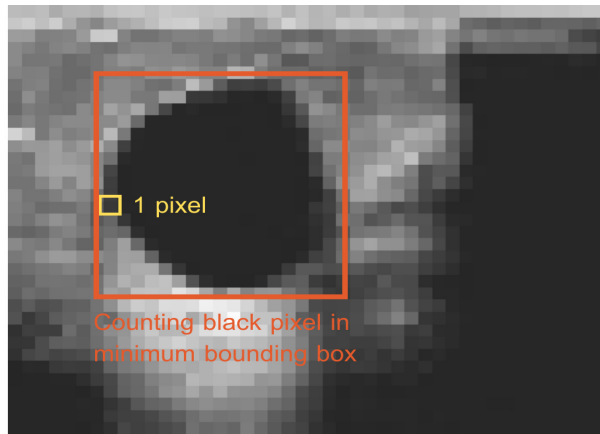


Figure 4. The Tumor Volume was Estimated. First, the minimum bounding box was detected the tumor. Then, the black pixels were automatic measured the tumor volume instead of manually.

mAP@0.5 score achieve 0.95, while the mAP@0.5-0.95 achieve 0.75.

Tumor prediction

The confusion matrices of the model for predicting breast cancer with the test set is shown in Figure 9. The ABUS model achieved a high performance in distinguishing benign from malignant breast lesions when applied to the breast US images of the test set. The model achieved a lesion classification accuracy of 95.07%, a sensitivity of 94.97%, a specificity of 95.24%, a PPV of 97.42%, and an NPV of 90.91% (Table 3).

Table 4. The Model Performance Evaluation in Sensitivity, Specificity, Positive and Negative Predictive Value as well as accuracy are expressed as percentages, confidence intervals for sensitivity, specificity and accuracy are "exact" Clopper-Pearson confidence intervals.

Statistic	Value	95% CI
Sensitivity	94.97%	90.95% to 97.56%
Specificity	95.24%	89.24% to 98.44%
Positive Predictive Value (*)	97.42%	94.14% to 98.89%
Negative Predictive Value (*)	90.91%	84.51% to 94.82%
Accuracy (*)	95.07%	91.99% to 97.21%

Tumor volume estimation

In previous experiment results, this study illustrated the important results for cropping the tumor using bounding box coordinates in a top-left, top-right, bottom-right, and bottom-left arrangement. This section illustrated the computation results in the size of tumor. The ABUS model can measure the size of tumor in an image using a simple "pixels per metric" technique which describes the number of pixels that can "fit" into a given number of inches, millimeters, meters, etc.

Discussion

This paper proposes artificial intelligence model that not only automatically detect the breast tumor lesions but also classify the breast tumor in benign or malignant follow by the tumor volume measurement using effective deep learning technique. The breast ultrasound images were used for model training and testing. When the

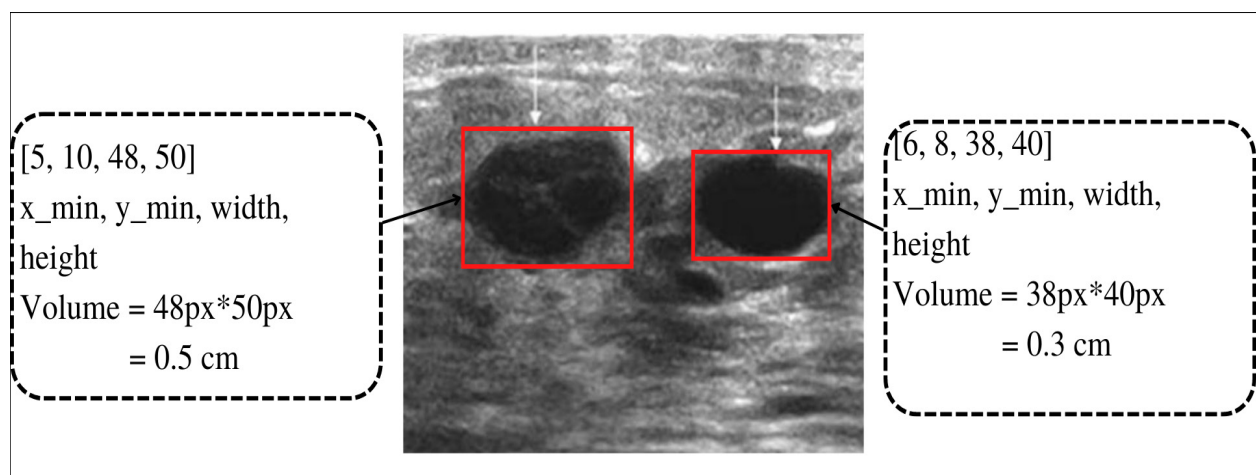


Figure 5. The Tumor Volumes were Estimated from Minimum Bounding Box Coordinator. The coordinators were calculated the object pixel followed by convert the pixel to centimeter.

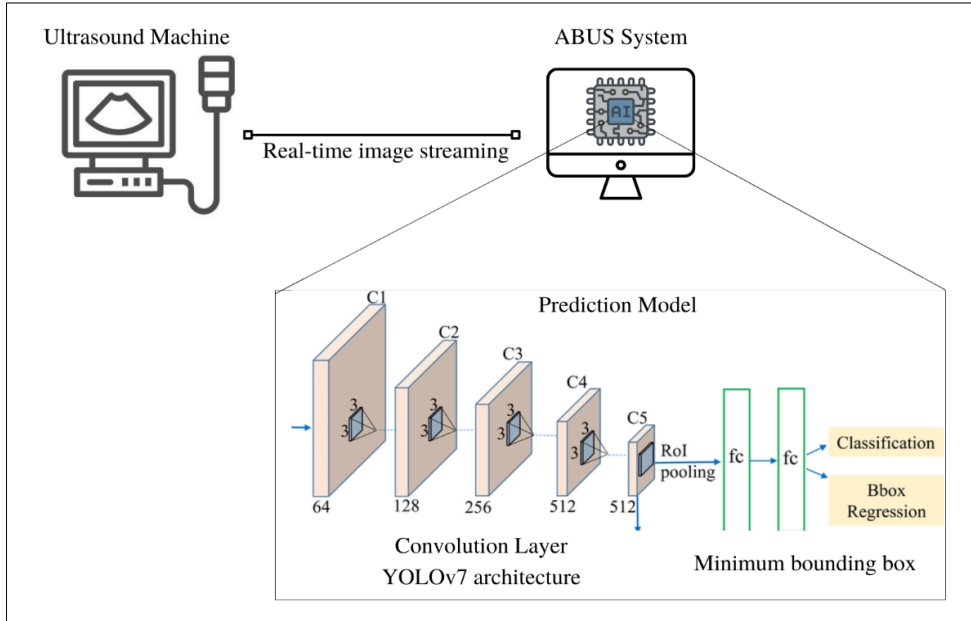


Figure 6. The Proposed Integrated System. First, the physician used the main ultrasound machine. Then, the breast ultrasound image was real-time streamed to the ABUS detection and classification system that was embedded in embedded devices.

quantity and diversity of data are important factors in the effectiveness of most machine learning models (e.g. deep learning neural network models), data augmentation has been used in this study to enhance the amount of

data producing synthetic data from existing data. Our experiment results showed the augmentation dataset tend to improve the model performance. This result is consistent with previous papers (Han et al., 2017, Zheng

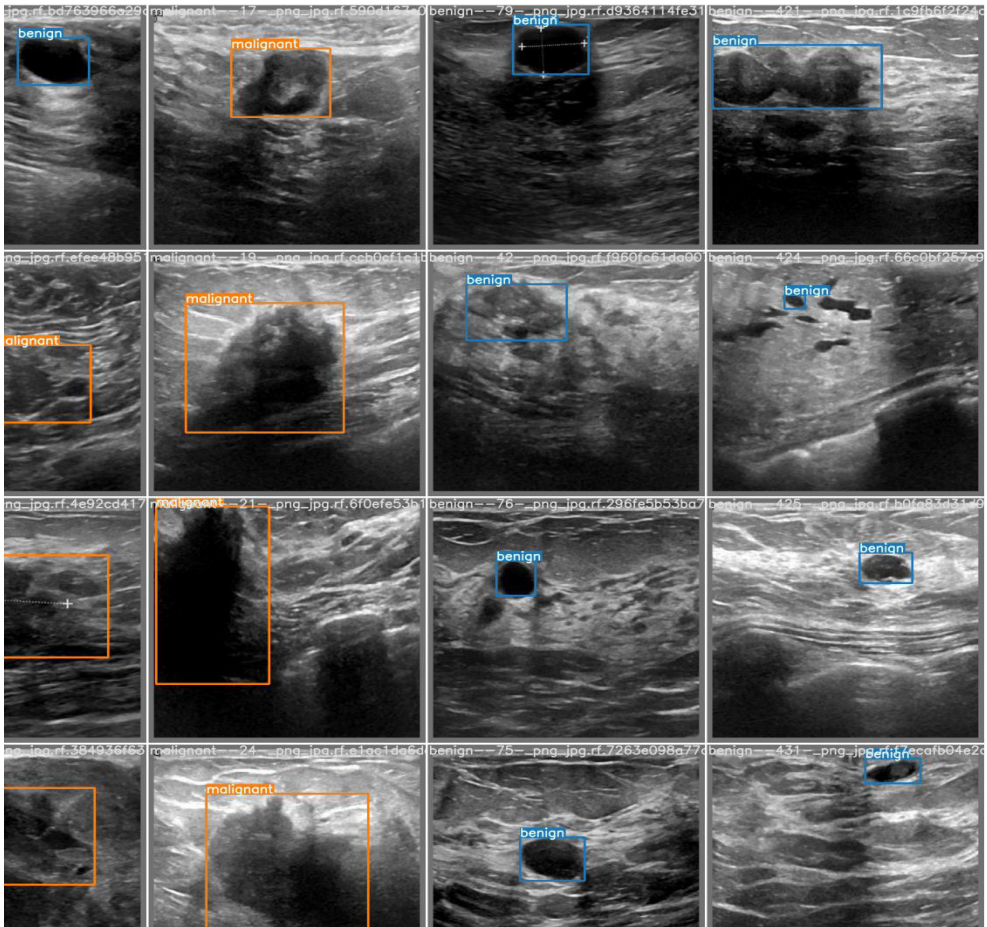


Figure 7. The Results of RoI Extraction Using a Minimum Bounding Box, in which the Orange Boxes are the Rectangle that Detected the Malignant Tumor, and the Blue Boxes are the Rectangle that Detected the Benign Tumor.

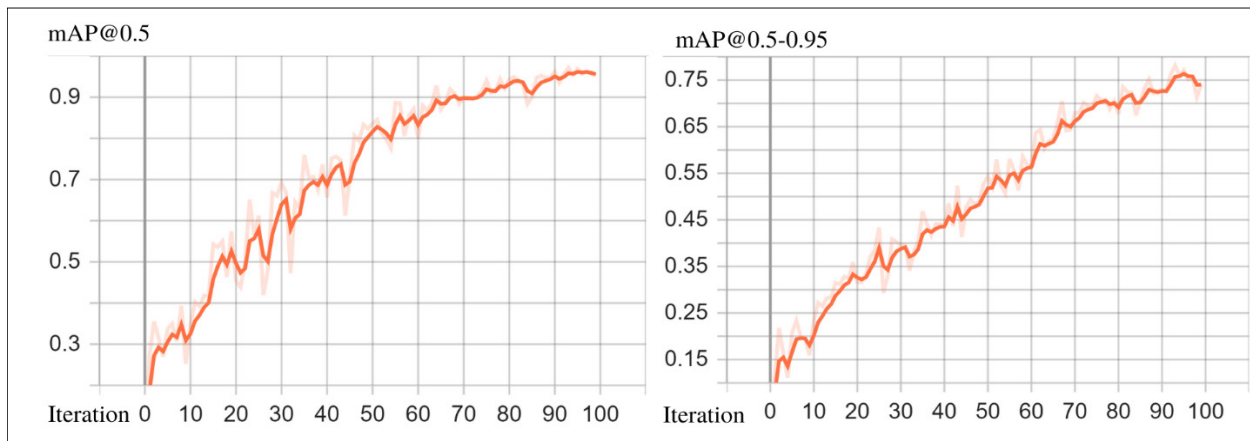


Figure 8. The Mean Average Precision or mAP Score is Calculated by Taking the Mean AP Over All Classes and/or Overall IoU Thresholds, mAP@0.5 is calculated for an IoU threshold of 0.5, while mAP@0.5-0.95 is calculated for an IoU threshold from 0.5 to 0.95.

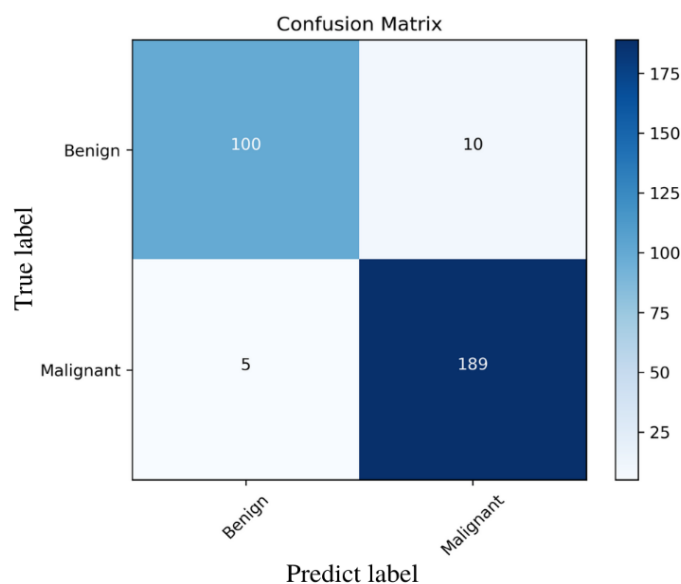


Figure 9. The Confusion Matrices of the Model in Distinguishing benign and Malignant Breast Lesions with the Test Set.

et al., 2020, Jiang M, 2021). They concluded that a large dataset of US breast images showed excellent diagnostic performance in the diagnosis of breast cancer.

The model effectively extracts RoIs in the US images using YOLO7 deep learning architecture. The RoIs obtained by selective model can investigate the features of lesion for the further classification. The model yielded satisfactory detections on the test sets, with the mAP@0.5 score of 0.95, and the mAP@0.5-0.95 of 0.75. These results are consistent with Yujie Li et al, (2022) study. They proposed BUSNet that showed the performance for the breast US images using the backbone network for the classification of RoIs and bounding box regression. As same as the G'omez-Flores et al, (2020) study, the well-established CNN models have been developed by the computer vision community for the automatic segmentation of BUS images using semantic segmentation.

The diagnostic performances using ABUS were

discussed. The model yielded satisfactory predictions on the test sets, with a lesion classification accuracy of 95.07%, a sensitivity of 94.97%, a specificity of 95.24%, a PPV of 97.42%, and an NPV of 90.91% (Table 4). The diagnostic performance of our model was similar to previous papers on AI methods for breast US analysis (Kim KE et al., 2020, Wan KW et al., 2021, and Boumaraf S et al., 2011). In recent years, many new techniques have been developed to compensate for the deficiencies of conventional US (Yampaka and Chongstitvatana, 2020). In particular, automatic breast detection and classification including automatic tumor volume estimation using artificial intelligence can provide a second opinion or supportive decision and significantly improve the efficiency and effectiveness of the radiologists' diagnosis (Chan et al., 2020).

In summary, this study proposed an automatic breast tumor detection and classification including automatic tumor volume estimation for US images. In addition,

our ABUS model can measure the size of tumor in an image using a simple pixels per metric technique. There are several limitations in our study. First, the training dataset consist of only B-mode US images. Therefore, other US mode such as Doppler or Elastography mode can include in the further development. Second, our model classifies the tumor in benign and malignant. Additionally, the BI-RADS assessment is a standard tumor diagnosis. Consequently, the ABUS classification in BI-RADS assessment can be further applied to develop AI-based diagnostic support technologies for breast disease screening.

Author Contribution Statement

All authors contributed equally in this study.

Acknowledgements

General

We are grateful for Biomedical Engineering Unit of Sirindhorn International Institute of Technology that presents a database of ultrasound images of breast cancer provided by the Department of Radiology of Thammasat University and Queen Sirikit Center of Breast Cancer of Thailand.

Conflict of Interest

The authors declare that they have no conflict of interest.

References

- Akkus Z, Galimzianova A, Hoogi A, Rubin DL, Erickson BJ (2017). Deep Learning for Brain MRI segmentation: State of the art and Future Directions. *J Digit Imaging*, **30**, 449–59.
- Al-Dhabyani W, Gomaa M, Khaled H, Fahmy A (2020). Dataset of breast ultrasound images. *Data Brief*, **28**, 104863.
- Amiri M, Brooks R, Behboodi B, Rivaz H (2020). Two-stage ultrasound image segmentation using U-Net and test time augmentation. *Int J Comput Assisted Radiol Surg*, **15**, 981–8.
- Boumaraf S, Liu X, Wan Y, et al (2021). Conventional machine learning versus deep learning for magnification dependent histopathological breast cancer image classification: a comparative study with visual explanation. *Diagnostics*, **11**, 528.
- Budd S, Robinson EC, Kainz B (2021). A survey on active learning and human-in-the-loop deep learning for medical image analysis. *Med Image Anal*, **71**, 102062.
- Chan HP, Samala RK, Hadjiiski LM (2020). Cad and AI for Breast Cancer—recent development and challenges. *Br J Radiol*, **93**, 20190580.
- Chiao JY, Chen KY, Liao KYK, et al (2019). Detection and classification the breast tumors using mask R-CNN on Sonograms. *Medicine*, **98**, e15200.
- Cui S, Chen M, Liu C (2020). DsUnet: A new network structure for detection and segmentation of ultrasound breast lesions. *J Med Imaging Health Infor*, **10**, 661–6.
- Deep learning based Object Detection Methods: A Review (2022). Medicon Engineering Themes.
- Giger ML (2000). Computer-aided diagnosis of breast lesions in medical images. *Comput Sci Eng*, **2**, 39–45.
- Gómez-Flores W, Coelho de Albuquerque Pereira W (2020). A comparative study of pre-trained convolutional neural networks for semantic segmentation of breast tumors in ultrasound. *Comput Biol Med*, **126**, 104036.
- Han S, Kang HK, Jeong JY, et al (2017). A deep learning framework for supporting the classification of breast lesions in ultrasound images. *Phys Med Biol*, **62**, 7714–28.
- Jiang Y, Nishikawa RM, Schmidt RA, et al (1999). Improving breast cancer diagnosis with computer-aided diagnosis. *Acad Radiol*, **6**, 22–33.
- Jiang M, Zhang D, Tang SC, et al (2021). Deep learning with convolutional neural network in the assessment of breast cancer molecular subtypes based on US images: a multicenter retrospective study. *Eur Radiol*, **31**, 3673–82.
- Kim KE, Kim JM, Song JE, et al (2020). Development and validation of a deep learning system for diagnosing glaucoma using optical coherence tomography. *J Clin Med*, **9**, 2167.
- Komatsu M, Sakai A, Dozen A, et al (2021). Towards clinical application of artificial intelligence in ultrasound imaging. *Biomedicine*, **9**, 720.
- Kornecki A (2011). Current status of breast ultrasound. *Can Assoc Radiol J*, **62**, 31–40.
- Kumar V, Webb JM, Gregory A, et al (2018). Automated and real-time segmentation of suspicious breast masses using convolutional neural network. *PLoS One*, **13**, e0195816.
- Lakhani P, Sundaram B (2017). Deep learning at chest radiography: Automated Classification of pulmonary tuberculosis by using convolutional neural networks. *Radiology*, **284**, 574–82.
- Li Y, Gu H, Wang H, Qin P, Wang J (2022). BUSnet: A deep learning model of breast tumor lesion detection for ultrasound images. *Front Oncol*, **12**.
- Redmon J, Divvala S, Girshick R, Farhadi A (2016). You only look once: Unified, real-time object detection. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR).
- Ren S, He K, Girshick R, Sun J (2017). Faster R-CNN: Towards real-time object detection with region proposal networks. *IEEE Trans Pattern Anal Mach Intell*, **39**, 1137–49.
- Shen S (2015). A multi-centre randomised trial comparing ultrasound vs mammography for screening breast cancer in high-risk Chinese women. *Br J Cancer*, **112**, 998–1004.
- Siegel RL (2022). Cancer statistics, 2022. *Cancer J Clin*, **72**, 7–33.
- Virasova AY, Klimov DI, Khromov OE, et al (2021). Rich feature hierarchies for accurate object detection and semantic segmentation. *Radioengineering*, **2021**, 115–26.
- Wan KW, Wong CH, Ip HF, et al (2021). Evaluation of the performance of traditional machine learning algorithms, convolutional neural network and AutoML Vision in ultrasound breast lesions classification: a comparative study. *Quant Imaging Med Surg*, **11**, 1381–93.
- Yampaca T, Chongstitvatana P (2020). Combination of B-mode and color doppler mode using mutual information including canonical correlation analysis for breast cancer diagnosis. *Med Ultrason*, **1**, 49.
- Zheng X, Yao Z, Huang Y, et al (2020). Deep learning radiomics can predict axillary lymph node status in early-stage breast cancer. *Nat Commun*, **11**, 1236.



This work is licensed under a Creative Commons Attribution-Non Commercial 4.0 International License.