Clinical Application of Preliminary Breast Cancer Screening for Dense Breasts Using Real-Time AI-Powered Ultrasound with Deep-Learning Computer Vision

Zhenzhong Zhou1*, Xueqin Xie2, Zongjin Yang3, Zhongxiong Feng3, Xiaoling Zheng3, Qian Huang1

1School of Business, Macau University of Science and Technology, Macau 999078, China
2Ultrasound Department, Duanzhou District Women and Children’s Hospital, Zhaoqing 526040, China
3iCare365 Technologies, Inc., Guangzhou 510515, China

*Corresponding author: Zhenzhong Zhou, joe.zhou@icare365.cn

Abstract: Objective: We propose a solution that is backed by cloud computing, combines a series of AI neural networks of computer vision; is capable of detecting, highlighting, and locating breast lesions from a live ultrasound video feed, provides BI-RADS categorizations; and has reliable sensitivity and specificity. Multiple deep-learning models were trained on more than 300,000 breast ultrasound images to achieve object detection and regions of interest classification. The main objective of this study was to determine whether the performance of our AI-powered solution was comparable to that of ultrasound radiologists. Methods: The noninferiority evaluation was conducted by comparing the examination results of the same screening women between our AI-powered solution and ultrasound radiologists with over 10 years of experience. The study lasted for one and a half years and was carried out in the Duanzhou District Women and Children’s Hospital, Zhaoqing, China. 1,133 females between 20 and 70 years old were selected through convenience sampling. Results: The accuracy, sensitivity, specificity, positive predictive value, and negative predictive value were 93.03 %, 94.90 %, 90.71 %, 92.68 %, and 93.48 %, respectively. The area under the curve (AUC) for all positives was 0.91569 and the AUC for all negatives was 0.90461. The comparison indicated that the overall performance of the AI system was comparable to that of ultrasound radiologists. Conclusion: This innovative AI-powered ultrasound solution is cost-effective and user-friendly, and could be applied to massive breast cancer screening.

Keywords: Breast cancer screening; Ultrasound; Lesion detection; BI-RADS; Deep learning; Computer vision; Cloud computing

Online publication: July 18, 2024

1. Introduction

Female breast cancer has surpassed lung cancer as the leading cause of global cancer incidence in 2020, with an estimated 2.3 million new cases representing 11.7% of all cancer cases. Among women, breast cancer accounts
for 1 in 4 cancer cases and 1 in 6 cancer deaths, ranking first for incidence in the vast majority of countries (159 of 185 countries) and mortality in 110 countries [1].

Currently, China is facing a similar situation. In 2015, breast cancer was the most frequently diagnosed cancer in Chinese women, with an estimated 304,000 new cases. This remained the case for 2018 (19.2% of total cancer cases) [2]. The mean age of diagnosis for Chinese female breast cancer patients is 45–55 years old, and its peak is estimated to be in the 50–54-year-old range [3].

In the absence of a vaccine for breast cancer, population-wide screening programs are critical for reducing breast cancer mortality through early detection and effective treatment [1]. Achieving high population coverage and adherence to the target population with appropriate screening methods are vital for effective screening. Owing to preliminary screening and improved treatments in the United States, the breast cancer death count decreased by 41% from 1989 to 2018. Unfortunately, the early breast cancer detection rate in China is still low; only 5.2% of breast cancer cases are detected through regular screening as opposed to 60% in the United States. Of all the patients examined, 82.1% showed apparent symptoms.

Ultrasound offers several advantages over mammography for preliminary breast cancer screening.

1. It is more effective at examining small and dense breasts, which is a common feature of Asian women [4,5]. Because of the relatively small breast sizes among Asian women, it is difficult to maintain precise positioning during mammography examination, which can cause misdiagnosis or missed diagnosis. Breast cancer can be easily obscured and missed in mammographically dense breast tissue due to overlapping surrounding fibroglandular tissue.

2. Ultrasound is cheaper than mammography in terms of procurement, operation, and maintenance.

3. Ultrasound is radioactive-free and can therefore be used repetitively, even on pregnant women, while mammography is radioactive.

4. Ultrasound machines are already installed in many frontline clinical institutions around the world, even in remote areas, and are portable.

Previous studies have also suggested that ultrasound achieves better screening results than mammography [6]. Considering the above comparison between mammography and ultrasound, ultrasound is the most logical clinical choice for preliminary breast cancer screening for women with small and dense breasts and has been designated as the examination method for screening according to China’s national standard. However, due to the enormous shortage of qualified ultrasound radiologists in China, the coverage of breast cancer screening has been very low for years.

In recent years, several new techniques have been developed to detect breast cancer. For example, computer-aided diagnosis (CADx) systems have been developed to help radiologists analyze mammograms, particularly breast cancer mammograms [7]. Since the invention of convolutional neural networks (CNNs), deep CNNs have performed well and have been used to revolutionize medical image analysis. Deep-learning-based CADx systems have achieved human-level performance in detecting breast cancer lesions with mammography and have contributed to higher diagnostic accuracy [8]. The outstanding results of deep learning in breast cancer screening for mammography have led to its frequent application in the UK to tackle the crisis of radiology workforce shortages [9].

Other papers have discussed the application of deep learning to analyze static ultrasound breast images manually captured by radiologists [10], which could be used for academic research purposes only but not for breast cancer screening in the real world, since during breast examination or screening using ultrasound, the analysis of ultrasound video by AI must be performed in real-time. Further, for breast cancer screening the entire breast region must be scanned and analyzed, and the lesion must be able to be pinned point whenever detected. To address this real-world challenge, we have proposed an AI-based breast cancer ultrasound screening solution that can detect, locate, diagnose, and classify breast cancer lesions with CNNs in the real-time processing of ultrasound video. We
used deep-learning computer vision techniques to ensure complete scan coverage of the breasts for quality control. Our system can automatically recommend BI-RADS \cite{11} classification based on detected lesions \cite{12}. This study is designed to evaluate the performance of our solution for the preliminary breast cancer screening on small and dense breasts and determine whether the performance of our AI-powered solution is comparable to that of ultrasound radiologists with over 10 years of experience. We believe that this study might be one of the first attempts to combine AI with ultrasound for preliminary breast cancer screening on small and dense breasts.

2. Materials, methods, and algorithms

We applied deep CNNs and computer vision techniques to develop a solution for real-time breast lesion detection and classification while enforcing a complete breast scan and locating lesions. Our training dataset contained over 300,000 breast ultrasound images generated by various ultrasound machines of diverse brands, including healthy images and images with malignant and benign lesions, cysts, and lymph nodes. All images were labeled by ultrasound radiologists with over 10 years of experience, and malignant lesions were confirmed by biopsies. The dataset was then split into three subsets: 70% for training, 20% for validation, and 10% for testing. This 7:2:1 ratio provides better training performance. PyTorch was used as a deep-learning framework to train our deep-learning models. We preprocessed the dataset using computer vision denoising techniques. For experimental purposes, we performed data augmentation for training. A stochastic gradient descent approximation algorithm with an initial learning rate of 10^{-2} was used. We trained our deep learning models on a server with eight NVIDIA GeForce RTX 3090 Ti 24GB GPU. During the training of all the neural network models, we selected the epoch that achieved the highest accuracy during testing.

While performing screening, our multi-model system analyzed the breast ultrasound video signals at a frequency of 24 frames/sec. The AI system analyzed each frame and highlighted detected lesions, including malignant and benign masses, cysts, and lymph nodes. When these types of lesions were detected, the ultrasound images were captured and the lesion was automatically framed with a red rectangle; the edge of the lesion was denoted with green polylines, and corresponding clock-based position diagrams were simultaneously generated by the ultrasound probe tracing system, which demonstrated the positions of lesions; all images were compressed with a lossless algorithm and uploaded to the data repository on the cloud. The size of each detected lesion was also recorded. Multiple lesions could be detected on one image if they exist.

This real-time process repeated until breast scanning was completed, and all images with potential lesions were captured and stored on the cloud, which allows ultrasound radiologists and physicians to download stored images for quality control and/or referral for medical treatment. This facilitates follow-up examinations and diagnostic processes.

2.1. Neural network architecture

Our solution’s automatic detection of lesions, BI-RADS classification, and complete scan enforcement have the following stages.

(1) Data preprocessing: The quality and resolution of breast ultrasound images vary widely depending on the manufacturer and model of the ultrasound equipment. To ensure the generalization of the network models and enhance the quality of breast ultrasound images, images were preprocessed with denoising algorithms, such as non-local means (NLM) and histogram equalization, in addition to the fact that images from various ultrasound machines manufactured by different manufacturers were used for training. Ultrasound images typically include Gaussian noise, and NLM works well for this type of noise. Furthermore, we specifically focused on improving the contrast between suspicious lesion areas and surrounding normal breast tissues to distinguish them.
(2) Object detection, lesion classification, and model evaluation: Deep-learning neural networks for image recognition and object detection are commonly used in medical imaging analysis \[^{13}\]. In our proposed method, we concatenated different types of unique models as a pipeline for the detection of breast lesions on a live ultrasound video feed during screening. The first model is a valid/invalid model that receives preprocessed breast ultrasound images and filters out low-quality images due to inappropriate operations during ultrasound scanning. The second model uses a YOLOX \[^{14}\] to identify potential lesion regions as regions of interest (ROIs). For ROIs detected by the YOLOX model, we used SE-ResNeXt-50 \[^{15,16}\] for the Positive/Negative (P/N) CNN models to filter out false positive ROIs from true positive ROIs to achieve solid specificity. We structured our P/N models such that the output of one model is the input of another, making the filtering process more effective. The last model was used to classify true-positive lesions that successfully passed all the previous models into one of four types of lesions: malignant masses, benign masses, cysts, and lymph nodes. Figure 1 shows the workflow of the multi-model structure. We further evaluated the overall performance of our models by using benchmark metrics including accuracy, sensitivity, specificity, positive predictive value, and negative predictive value, and drew ROC curves for our solution. In Section 3, we compare our results with those obtained by ultrasound radiologists with over 10 years of experience.

![Figure 1. Workflow of our deep-learning-based lesion detection models](image)

(3) We used YOLOX to detect the ultrasound probe, trace the movement of the ultrasound probe, and calculate the area of the breast scanned thus far while the ultrasound probe moves and scans the breast. Our system reminds the operator to scan missed regions if occurred. This function ensures a complete scan of the entire breast region as a quality control measure to prevent missed diagnoses because of miss-scanned areas. Furthermore, whenever a lesion is detected, its location is displayed on a clock-based diagram after calculating the position of the ultrasound probe.

Our solution offers multiple advantages:
(1) Fast: capable of analyzing a real-time breast ultrasound video during scanning, which means that the algorithms are highly efficient.
(2) Accurate: The accuracy was significantly improved by using our updated deep CNNs, and false positive ROIs were eliminated by constructing a pipeline of P/N models.
(3) Cost-effective: Our ultrasound-based solution is more affordable in terms of procurement, operation, and maintenance than mammography and MRIs.

### 2.2. BI-RADS classification

We applied the BI-RADS categories to define the breast health status of screened women. After passing through the pipeline of P/N models, the ROIs identified as true positive lesions were categorized as (1) malignant, (2) benign, (3) cyst, or (4) lymph node, as shown in Figure 2.

![Figure 2. Lesion classification flowchart of our solution](image)

Then, depending on whether lesions are detected and the type of lesions if detected, our AI system classifies women’s breast health status using BI-RADS categories from 1 to 5. If no lesion is detected, the screened breast is considered healthy or without breast cancer, and a BI-RADS category of 1 is issued. When a malignant mass is detected, the BI-RADS category is 4 or 5 depending on the features of the malignant mass, such as mass shapes, echogenicity, margins, etc. A BI-RADS category of 3, 4, or 5 signals that further clinical verification is urgently needed for the screened woman.

### 3. Results

Figure 3 shows statistics of the training results; the visualization indicates that small ROIs of both dimensions within 150 pixels presented the maximum number of false positives and impacted the overall performance negatively. Since the performance of P/N models varied significantly for ROIs with different sizes, to ensure solid specificity, we trained different P/N models to analyze ROIs for potential lesions of different sizes as an effective measure to filter out small-size false positive ROIs.

![Figure 3. Relationship of easy and difficult false positives; width – width of ROI; height – height of ROI; easy false positive – the false positive ROI which is easy to be classified as false positive by Positive/Negative classification models; difficult false positive – the false positive ROI which is difficult to be classified as false positive by Positive/Negative classification models](image)
Figures 4–7 illustrate the four types of lesions that our solution can detect (malignant masses, benign masses, cysts, and lymph nodes). Lesion detection includes information on the lesion type, size, and location in the breast. The test results show that the AI solution, based on our efficient multi-model neural networks, can effectively locate, identify, and categorize lesions.

**Figure 4.** A 43-year-old female was found to have a benign mass in her right breast during the breast cancer ultrasound preliminary screening. The following images were stored in our cloud platform: (A) original ultrasound image captured by the AI system; (B) highlighted lesion in a red rectangle and its exact shape with green polyline; (C) recorded location of the lesion on a clock-based diagram when the lesion was detected. This can greatly facilitate relocating the lesions in follow-up treatment; (D) An AI Breast Cancer Preliminary Screening Report was automatically generated by our system, containing the BI-RADS category, most serious lesions, and other essential information upon successful completion of the preliminary screening.

**Figure 5.** Screening report of a 48-year-old female, where a malignant tumor was identified in her left breast.
Figure 6. Screening report of a 21-year-old female, where a cyst was identified in her right breast.

Figure 7. Screening report of a 63-year-old female, where a lymph node was identified in her left breast.

Our AI system, named “Dr. J,” went through a clinical validation of breast cancer screening on a total of 1,133 individuals, including healthy individuals without breast lesions and individuals with breast lesions, in the Duanzhou District Women and Children’s Hospital located in Zhaoqing, China, one of the third-grade hospitals ranked at the top of China’s three-tier hospital grading system. Furthermore, malignant lesions were confirmed...
by biopsy. Ethical approval was obtained from the Ethics Committee of the Duanzhou District Women and Children’s Hospital.

Figure 8 presents the distribution of women undergoing breast cancer screening with different BI-RADS categories by age group from AI screening.

![Figure 8](image)

**Figure 8.** Distribution of breast cancer screening results with different BI-RADS categories by age group, based on results from the screening performed by our AI solution

Screening performance was compared between ultrasound radiologists with over 10 years of experience and our AI system.

To assess the accuracy of our solution, we calculated the percentages of our diagnoses with respect to those of ultrasound radiologists. Table 1 presents a more detailed comparison between the results of the AI system and those of radiologists. As can be seen from Table 1, our solution is more accurate in diagnosing individuals with lesions than in diagnosing healthy individuals. In other words, our solution has a slightly higher sensitivity than specificity. We calculated the accuracy, sensitivity, specificity, positive predictive value, and negative predictive value, as shown in Table 2.

**Table 1.** Comparison of breast cancer screening results between radiologists and AI system

<table>
<thead>
<tr>
<th></th>
<th>Number of healthy individuals diagnosed by ultrasound radiologists (X1)</th>
<th>506</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Number of healthy individuals diagnosed by our solution within X1 (Y1)</td>
<td>459</td>
</tr>
<tr>
<td>% Of healthy individuals correctly diagnosed by our solution ($\frac{Y1}{X1} \times 100%$)</td>
<td>90.71 %</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Number of individuals with lesions diagnosed by ultrasound radiologists (X2)</td>
<td>627</td>
</tr>
<tr>
<td></td>
<td>Number of individuals with lesions diagnosed by our solution within X2 (Y2)</td>
<td>595</td>
</tr>
<tr>
<td>% of individuals with lesions correctly diagnosed by our solution ($\frac{Y2}{X2} \times 100%$)</td>
<td>94.90 %</td>
<td></td>
</tr>
</tbody>
</table>

*True negative: TN = Y1 = 459; True positive: TP = Y2 = 595; False negative: FN = X2 - TP = 32; False positive: FP = X1 - TN = 47
Table 2. Evaluations of our AI solution

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Formula</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuracy</td>
<td>(\frac{TP+TN}{TP+FP+TN+FN})</td>
<td>93.03 %</td>
</tr>
<tr>
<td>Sensitivity (Se)</td>
<td>(\frac{TP}{TP+FN})</td>
<td>94.90 %</td>
</tr>
<tr>
<td>Specificity (Sp)</td>
<td>(\frac{TN}{FP+TN})</td>
<td>90.71 %</td>
</tr>
<tr>
<td>Positive Predictive Value (PPV)</td>
<td>(\frac{TP}{TP+FP})</td>
<td>92.68 %</td>
</tr>
<tr>
<td>Negative Predictive Value (NPV)</td>
<td>(\frac{TN}{TN+FN})</td>
<td>93.48 %</td>
</tr>
</tbody>
</table>

*True negative: TN = Y1 = 459; True positive: TP = Y2 = 595; False negative: FN = X2 - TP = 32; False positive: FP = X1 - TN = 47

As shown in Figure 9, the validation results indicated that the overall performance of the AI system was comparable to that of the average radiologist with over 10 years of experience (95% confidence interval, CI) in terms of reliability and accuracy.

![Non-Inferiority Evaluation](image)

Figure 9. Noninferiority evaluation

In addition to performing the statistical noninferiority evaluation based on screening individuals between the AI system and ultrasound radiologists, we chose ROC curves and the AUC to evaluate the accuracy of our solution on static ultrasound images. By examining the area under the ROC curves, we can determine the effectiveness of our solution. Figure 10 shows that the AUC for all positives is 0.91569 and the AUC for all negatives is 0.90461. Our solution is more suitable for women with small and dense breasts than some of the latest AI breast cancer screening methods, such as AI solutions with mammography. [7]

![Figure 10](image)

Figure 10. (Left) Positive ROC; (Right) Negative ROC
4. Discussion

Comprehensive work has been conducted to detect breast cancer using neural networks, such as deep learning (DL). The multi-U-net algorithm was developed to segment suspicious breast masses by using 433 clinical breast ultrasound images from 258 patients as training, validation, and testing data, which successfully segmented the breast masses, achieving a mean Dice coefficient of 0.82, a true positive fraction of 0.84, and a false positive fraction (FPF) of 0.01 [18]. A DL model was developed for automated feature learning and classification of malignant/benign lesions using ultrasound shear-wave elastography (SWE). The proposed two-layer DL-based SWE architecture performed more efficiently by using a point-wise gated Boltzmann machine and restricted Boltzmann machine, with an accuracy, sensitivity, specificity, and receiver operating characteristic (ROC) curve of 93.4%, 88.6%, 97.1%, and 0.947, respectively [19]. A deep neural network was able to predict breast cancer during screening using mammography, with an area under curve (AUC) of 0.895. The network was trained and tested on over 200,000 examinations (over 1,000,000 mammography images). A custom ResNet network was used as the backbone of the model. A reader study with 14 readers, each reading 720 screening mammogram examinations, indicated that this model was as accurate as experienced radiologists. Prediction using the average probability of malignancy predicted by radiologists combined with the probability of that by the model was more accurate than that predicted individually by radiologists or the model [20]. However, these studies did not mention how to enforce the complete scan of breasts or locate lesions once detected during ultrasound breast cancer screening, and both are critical issues to be overcome.

Our study suggested that breast cancer lesion detection on live ultrasound video feed is much more challenging than lesion detection and classification tasks on static medical images, including CT, mammography, and MRI. A noteworthy challenge was to achieve a solid specificity since typical breast cancer screening produces over 7,200 ultrasound images. If a false positive occurs on a single image, this screening would have a false positive result; thus, to achieve good specificity, we developed a series of P/N models to filter out false positive ROIs, and only true positive ROIs can pass through the P/N model pipeline successfully. Meanwhile, other challenges included how to ensure the complete scanning of the entire breast for quality control and how to locate the lesions upon detection. To overcome these challenges, we utilized YOLOX with a camera to trace the movement of the ultrasound probe and calculate the coverage of scanning. When lesions were detected by AI, the location of the ultrasound probe was marked on the clock-based diagram to locate the lesion. Unlike previous studies that focused mainly on the detection of lesions in stored static breast ultrasound images, our work defined an AI-powered breast cancer preliminary screening solution for real-world applications. In addition, our AI system can detect lesions as small as 3 × 3 mm, which might be considered a minor breast problem and ignored by ultrasound radiologists. Since we adopted a more cautious and sensitive method for preliminary screening of breast cancer to avoid missed diagnosis, a few screened women initially classified as healthy subjects by ultrasound radiologists were later verified to have lesions as alerted by our AI system.

Our AI system demonstrated solid performance, satisfying the requirement of breast cancer preliminary screening, which received an enthusiastic welcome from frontline medical staff during field application for the value-added support of the comprehensive breast ultrasound video analysis provided by our system. Since most women are free from breast cancer trouble while the supply of ultrasound radiologists is scarce in most areas around the world, our proposed solution can optimize the utilization of radiologist resources by focusing their expert services on only those suspected cases filtered out by our AI system. Our system has been undergoing evaluation for real-world applications with over a hundred thousand screening subjects, more massive breast cancer screening results will be revealed in the near future. Furthermore, we are researching new deep learning algorithms including transformers and multimodal deep learning to optimize our system network architecture.
5. Conclusion
In this study, we utilized deep-learning computer vision to develop a breast cancer preliminary screening system that is highly cost-effective and suggests the possibility of mass distribution for breast cancer screening of Asian women with small and dense breasts. Our system was able to successfully identify breast lesions with comparable performance to experienced ultrasound radiologists. By combining AI power with a cost-effective, non-invasive, and non-radioactive ultrasound device to assist breast cancer diagnosis, we have managed to overcome the bottleneck setback of expert service shortages in promoting preliminary screening for breast cancer. In the future, the inflow of growing data from field applications will expand our database and improve the accuracy and reliability of our developed system. We will apply this AI-powered ultrasound solution in the screening of more types of high-risk diseases, such as carotid plaque, for stroke screening in the future.

Funding
This research project was fully funded by iCare365 Technologies, Inc.

Disclosure statement
The authors declare no conflict of interest.

Author contributions
Conceptualization: Zhenzhong Zhou
Data curation: Xueqin Xie, Xiaoling Zheng
Methodology: Zhenzhong Zhou
Funding acquisition: Zhenzhong Zhou, Zhongxiong Feng
Resources: Zhenzhong Zhou, Xueqin Xie, Xiaoling Zheng
Supervision: Zhenzhong Zhou
Software: Zhenzhong Zhou, Zongjin Yang
Investigation: Zhenzhong Zhou, Zongjin Yang
Formal analysis: Zhenzhong Zhou, Zongjin Yang, Qian Huang
Validation: Xueqin Xie
Visualization: Zongjin Yang
Writing – original draft: Zhenzhong Zhou
Writing – reviewing & editing: Zhenzhong Zhou, Zhongxiong Feng

References


Publisher’s note
Bio-Byword Scientific Publishing remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.