A pilot study to evaluate breast cancer screening in low resource areas of the Pacific using portable ultrasound and artificial intelligence

Dustin Valdez¹, Arianna Bunnel¹, Thomas Wolfgruber¹, Jami Fukui¹, Peter Sadowski¹, John A. Shepherd¹  
University of Hawaii Cancer Center, Honolulu, HI, USA ¹

Background

- Advanced stage breast cancer (stages III and IV) rates in the Pacific are much higher than in the USA mainland, especially where mammography services do not exist: Marshall Islands (61%), Palau (94%), and Samoa (79%). Little is known about the performance of portable ultrasound when compared to clinical ultrasound for use in breast cancer screening.

- Introducing a portable handheld ultrasound (US) system coupled with an artificial intelligence (AI) detection algorithm and operated by a trained healthcare worker may be able to reduce advanced stage cancer rates in areas where mammography is not available.

- We developed an AI model using breast ultrasound images from the Hawaii Pacific Island Mammography Registry as well as a reader study tool.

- In this preliminary study, we report the preliminary reader study results from two graduate students (surrogate health worker).

Methods

- A case-control study design was used to minimize variability and data selection was done at both the patient and scan level. The matching was 1:3 by birth year (113 cases to 339 controls). Figure 1 shows a CONSORT style flowchart of our data selection process.

- The ultrasound images were then stripped of PHI and had to be preprocessed (scan area cropping, removal of burnt-in annotations, and splitting of dual-view scans) to be compatible for AI model training. The model selected for our transfer learning was DenseNet121 with a 70%-20%-10% split for training-validation-testing.

- The reader study interface used a VGG Image Annotator software [1] that we modified for our reader study. An example case can be seen in Figure 2.

- The preliminary reader study was run with a random sample of 400 breast ultrasound images comprised of 301 benign and 99 malignant scans from a publicly available breast ultrasound dataset [2].

- Readers were asked to provide a BI-RADS score for each image using the reader tool.

- Sensitivity and specificity analysis was done based on a BI-RADS 4a threshold (2% likelihood of malignancy).

Results

Figure 1. CONSORT style flowchart showing the data selection process at both the patient and scan levels. The red section presents patient-level selection, and the blue section represents the 4812 breast ultrasound scans collected from the 452 case-control patients. After all data selection criteria were met, 3871 total breast ultrasound scans were used for model training and evaluation.

Figure 2. Reader User Interface. Screen capture of the reader user interface from the VGG VIA image annotator tool. Highlighted in blue is the AI output section presented as a percentage likelihood of malignancy and a translated BI-RADS score. Highlighted in purple is where the reader indicates the mass-specific characteristics from the ultrasound lexicon.

Figure 3. Reader Study Preliminary Results. ROC for the AI model is shown as the blue line. The readers Cancer/No Cancer classifications were determined from a BI-RADS score of at least 4a for cancer (corresponding to a 2% likelihood of malignancy).

Results (continued)

- The AI model had an AUC=0.826 on the validation set and AUC=0.776 on the internal held-out test data set. On the external public breast ultrasound data set used for the preliminary reader study the model had an AUC=0.8411.

- The average sensitivity saw an improvement of 0.165 with AI, but a small decrease in specificity of 0.04. The full sensitivity and specificity results with and without AI can be seen in Table 1.

Table 1. Reader Study Sensitivity and Specificity Results.

<table>
<thead>
<tr>
<th>Reader</th>
<th>Sensitivity (Without AI)</th>
<th>Sensitivity (With AI)</th>
<th>Sensitivity Difference</th>
<th>Specificity (Without AI)</th>
<th>Specificity (With AI)</th>
<th>Specificity Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reader 1</td>
<td>0.79</td>
<td>0.83</td>
<td>0.04</td>
<td>0.638</td>
<td>0.601</td>
<td>-0.037</td>
</tr>
<tr>
<td>Reader 2</td>
<td>0.667</td>
<td>0.867</td>
<td>0.2</td>
<td>0.664</td>
<td>0.621</td>
<td>-0.043</td>
</tr>
<tr>
<td>Averages</td>
<td>0.684</td>
<td>0.849</td>
<td>0.165</td>
<td>0.651</td>
<td>0.611</td>
<td>-0.04</td>
</tr>
</tbody>
</table>

Conclusion

- The preliminary reader results are very promising as both readers saw noticeable gains in sensitivity while maintaining only minor losses in specificity. There is potential for AI to be used with ultrasound for non-radiologist healthcare workers, but a full reader study still needs to be done to show if AI can improve performance of a healthcare worker to that of a radiologist.

Future Work

- We are working on improving our AI model before performing a full reader study where different types of readers (radiologist, MDs, and general healthcare workers) are asked to assign a BI-RADS score to breast ultrasound images with and without the aid of an AI system.

- Currently working with a breast radiologist to improve our training materials for non-radiologist readers.

References