Can artificial intelligence derived ultrasound breast density provide comparable breast cancer risk estimates to density derived from mammograms

UNIVERSITY OF HAWAI'I CANCER CENTER

Dustin Valdez[12], Arianna Bunnell[12], Thomas Wolfgruber[1], Aleen V. Altamirano[3], Brandon Quon[1] , Gertraud Maskarinec[1], Peter Sadowski[2] ,John A. Shepherd[1]

University of Hawaii Cancer Center, Honolulu, HI, USA [1], University of Hawaii at Manoa, Honolulu, HI, USA [2], Instituto Radiodiagnostico, Managua, Nicaragua[3]

SHEPHERD RESEARCH LAB

Background

- Breast Cancer is the second leading cause of cancer-related death among women in Hawaii and the Pacific. The lack of access to mammography screening and low screening participation rates across many Pacific countries outside of Hawaii contributes to very high advanced breast cancer rates, in most cases over 50%.
- Portable breast ultrasound is a promising screening technology for low resource areas, however without mammography, mammographic density is not available for risk modeling to determine who should participate in screening programs.
- In this study, we ask if breast ultrasound (BUS) images can be used to derive an equivalent mamographic density for risk modeling. We utilized artificial intelligence (AI) to derive breast density from diagnostic ultrasound images and compared to BI-RADS mammographic density in an established breast cancer risk model[1] and logistic regression model.

Methods

Study Design: A total of 1317 women received screening or diagnostic breast ultrasound imaging from 2009 to 2021 at one clinical site from within the Hawai'i and Pacific Islands Mammography Registry (HIPIMR). Temporally-matched negative mammographic and ultrasound images, cancer outcome status, cancer risk information, and clinical BI-RADS breast density were sourced from the HIPIMR. We compared the clinical breast density to our BUS AI derived density by inputting the values into a Tyrer-Cuzick risk model as well as a logistic regression model to compare the AUC performance in predicting cancer.

Inclusion Criteria: Participants had to have a BUS (screening or diagnostic) and radiologist BI-RADS breast density score. Cancer labels were collected from the Hawaii Tumor Registry.

AI Model Design: A fully convolutional neural network analysis of breast ultrasound images were used to derive a predicted BI-RADS breast density score. Patients were split into training (70%), validation (20%), and test (10%).

Tyrer-Cuzick Model: Clinical breast density and AI BUS predicted values were input into the IBIS RiskEvaluator v8b. Only breast density and age were input values while other fields remained default. 10-Year risk estimates were used for all models.

Logistic Regression Model: Logistic regression was used to compare clinical breast density and AI BUS predicted density to predict the probability of developing breast cancer in 10 years.

Statistics: To compare the performance of clinical radiologist breast density and our AI BUS predicted density for risk assessment we calculated AUC values, 95% confidence intervals, ROC plots, and Pearson correlation.

Results

Tyrer Cuzick ROC Curves- Clinical Density, BUS Density and No Density

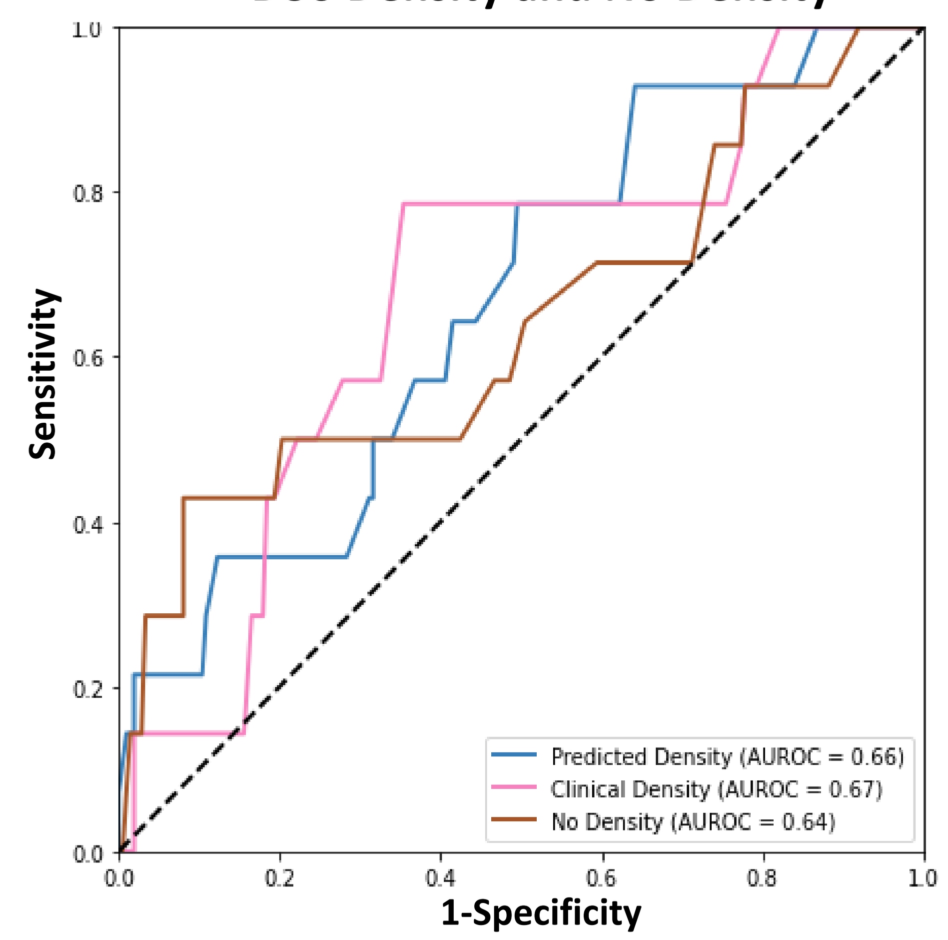

Figure 1. ROC curves between Clinical and AI BUS predicted density and no density input into the Tyrer-Cuzick model to predict breast cancer risk.

AI Predicted ROC Curves-BUS BI-RADS Density

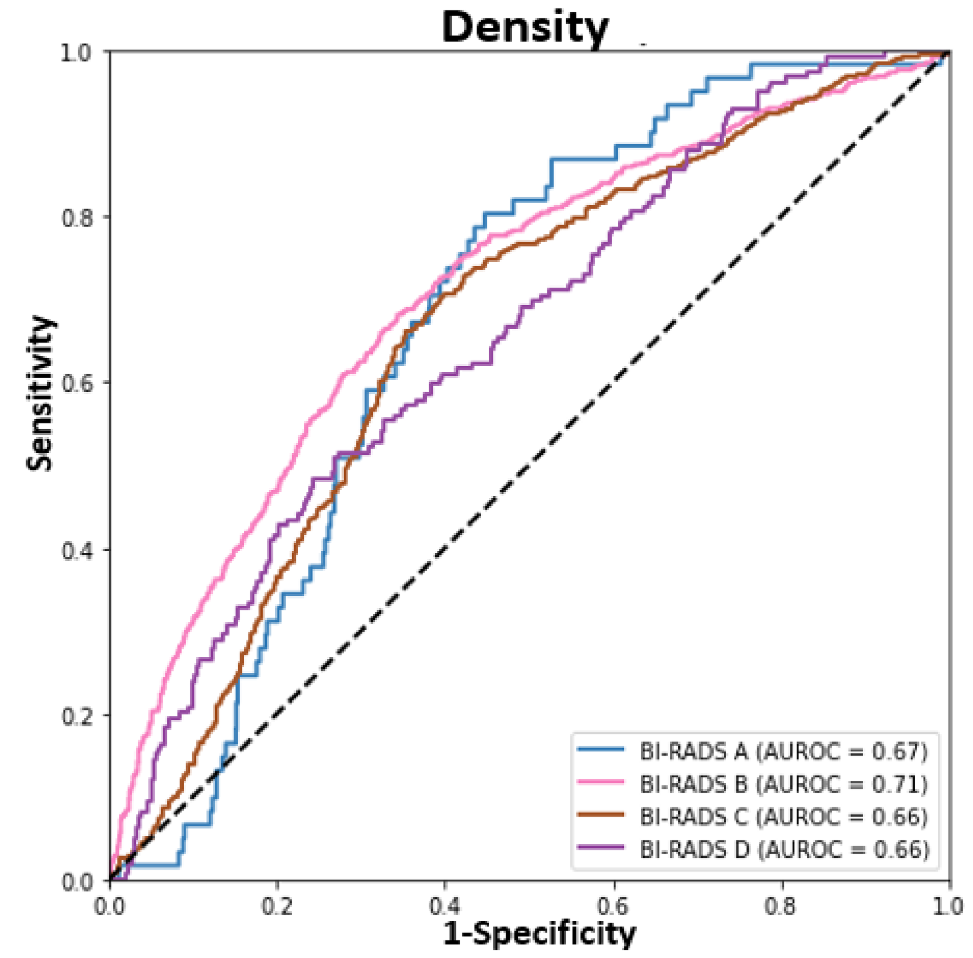

Figure 2. AI BUS predicted ROC curves between BI-RADS density categories.

Logistic Regression Curve- Clinical vs BUS Density for Mean Age

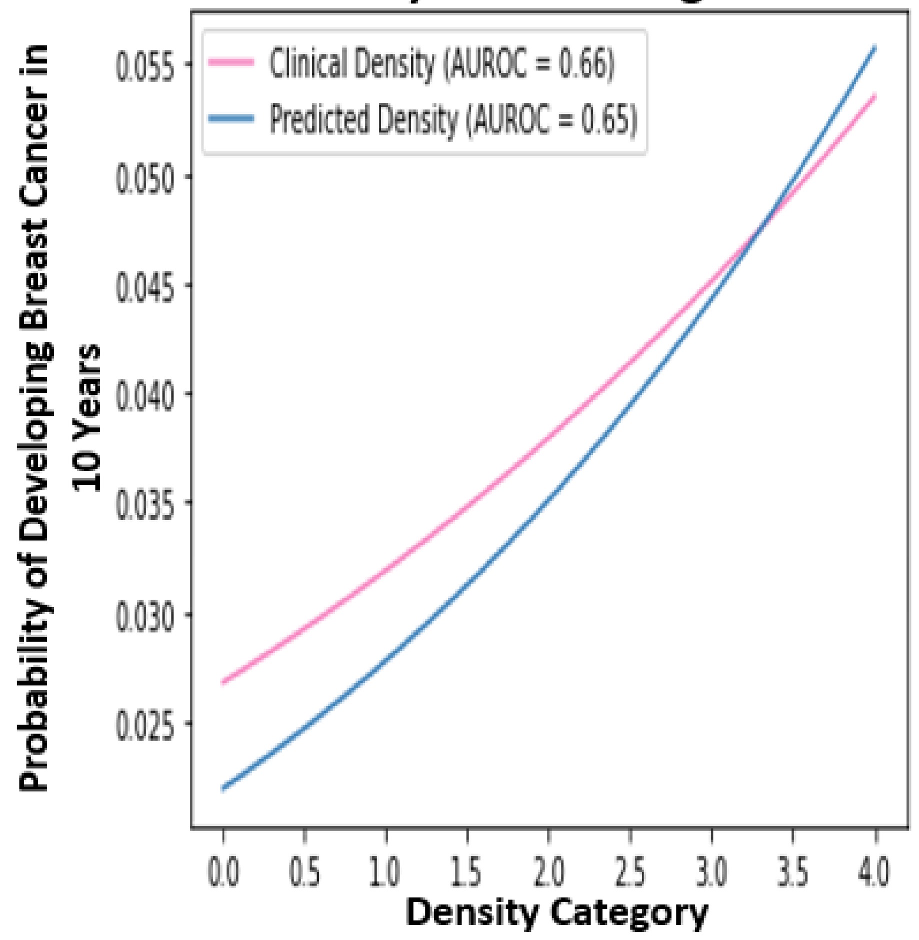

Figure 3. ROC Curves between Clinical and BUS Density for a logistic regression model that predicts probability of developing cancer in 10 years.

Demographic Summary of Study Population

Characteristic, unit	Overall	Training	Validation	Testing
Cancer Status, N	1317	895	274	148
Patients with benign findings, N	1248	850	261	137
Patients with malignant findings, N	69	45	13	11
Age, mean years (SD)	55.53	55.65	55.12	55.59
< 40 yrs old, N (%)	104	65	27	12
40 − 49 years old, N (%)	354	237	73	44
50 − 59 years old, N (%)	344	244	54	46
60 − 69 years old, N (%)	240	165	53	22
≥70 years old, N (%)	194	139	41	14
Unknown	81	45	26	10
Mammographic Density Category				
A (breasts are almost entirely fatty), N (%)	70	47	14	9
B (scattered areas of fibroglandular density), N (%)	512	339	115	58
C (breasts are heterogeneously dense), N (%)	634	438	124	72
D (the breasts are extremely dense), N (%)	115	80	24	11
Images, N	13227	8799	2630	1798
Images with benign findings, N	12413	8375	2419	1619
Images with malignant findings, N	814	424	211	179
Average no. of images per patient, N	10.04	9.83	9.6	12.15

Table 1. Summary of demographic patient data including cancer status, age, breast density and total number of images.

Results (continued)

Table 2. Summary of AUC and 95% confidence intervals for Tyrer-Cuzick, Logistic regression model and individual BI-RADS categories between breast AI Predicted ultrasound density and radiologist density.

	AUC	95% CI
Tyrer-Cuzick Results		
Predicted Density (Ultrasound)	0.66	0.52-0.81
Clinical Density (Radiologist)	0.67	0.53-0.82
No Density	0.64	0.46-0.82
Logistic Regression		
Predicted Density (Ultrasound)	0.64	0.46-0.82
Clinical Density (Radiologist)	0.66	0.49-0.82
Predicted Density		
Predicted Density (BI-RADS A)	0.67	0.62-0.72
Predicted Density (BI-RADS B)	0.71	0.69-0.74
Predicted Density (BI-RADS C)	0.66	0.64-0.69
Predicted Density (BI-RADS D)	0.66	0.61-0.70
Pearson Correlation (Ultrasound and Radiologist)	0.32 (moderate correlation)	

Over the 10-year study period, 1317 had matched mammograms and BUS images and 69 went on to develop breast cancer. Using the test set, the Pearson's correlation between breast density from mammography and BUS was 0.32 (moderate correlation). The AUC for TC 10-year personal risk was higher when breast density from mammograms was used 0.66 (95% CI=0.52-0.81) versus BUS images 0.67 (95% CI=0.53-0.82). The age used to calculate the logistic regression was the mean of 56 years old with an AUC for clinical density was 0.66 and predicted BUS density at 0.65. When looking at each BI-RADS category separately A, C, D had similar AUC of around 0.66 while BI-RADS B had a slightly higher AUC of 0.71.

Conclusion

- Breast cancer risk performance was similar when breast density was derived from either mammograms or BUS. The performance of our BUS breast density model is expected to improve further when more BUS training data becomes available.
- Breast cancer screening programs exclusively using BUS imaging may be able to provide equivalent risk modeling to clinics using mammography.

Future Work

A 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 is currently in progress.

References

1. Tyrer J, Duffy SW, Cuzick J (2004). A breast cancer prediction model incorporating familial and personal risk factors. Stat Med. 2004 Apr 15;23(7):1111-30. doi: 10.1002/sim.1668. Erratum in: Stat Med. 2005 Jan 15;24(1):156. PMID: 15057881.

2. Wu, N., K. J. Geras, Y. Shen, J. Su, S. G. Kim, E. Kim, S. Wolfson, L. Moy and K. Cho (2018). Breast Density Classification with Deep Convolutional Neural Networks. 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE.