10th International Breast Density & Cancer Risk Assessment Workshop. June 7-9, 2023 | Kailua-Kona, Hawai'i

Check out our other research!

Artificial Intelligence Predicts Mammographic Breast Density in Clinical Breast Ultrasound Images

Arianna Bunnell[12], Dustin Valdez[12], Thomas Wolfgruber[2], Brandon Quon[2], Lambert Leong[12], Jami Fukui[2], Brenda Hernandez[2], Yurii Shvetsov[2], Peter Washington[12], Peter Sadowski[12], John A. Shepherd[2]

University of Hawai'i at Mānoa, Honolulu, HI, USA[1], University of Hawai'i Cancer Center, Honolulu, HI, USA[2]

ARTIFICIAL INTELLIGENCE PRECISION HEALTH INSTITUTE
UNIVERSITY OF HAWAI'I

Introduction

- Advanced-stage breast cancer (stages III and IV) rates in the U.S.-Affiliated Pacific Islands (USAPI) are much higher than in the continental United States. Examples include Guam (60% advanced-stage cancer rate), American Samoa (74%), and the Federated States of Micronesia (81%) [1]. The USAPI needs accessible methods of cancer detection and risk assessment to reduce the advanced-stage cancer rate.
- The USAPI suffers from a lack of trained radiologists and poor access to screening mammography. Breast ultrasound (BUS) is a feasible substitute for mammography in low-resource areas.
- Our overall hypothesis is that portable BUS systems coupled with an AI detection algorithm operated by a general healthcare worker will reduce advanced-stage cancer rates.
- In this study, we ask if AI can determine categorical BI-RADS breast density from clinically-acquired breast ultrasound images for breast cancer risk assessment in the USAPI.

Methods

Study Design: Prospective case-control study of 101,437 women who received screening or diagnostic breast ultrasound imaging from 2009 to 2021 in the Hawai'i and Pacific Islands Mammography Registry (NIH R01CA263491 and U54CA143728)
Participant Selection: Participants had to have a negative screening mammography and BUS visits (BI-RADS 1&2) within 1 year of each other. Both imaging dates had to occur before diagnosis for cases. Cases were defined as women who developed breast cancer within 10 years. Controls were matched 10:1 to cases on birth year and BUS machine.
Mammographic Density Labels: BI-RADS mammographic breast density was estimated from mammograms using a published AI model [3].
Dataset: Images were split into training (60%), validation (20%), and testing (20%) by woman, stratified by AI-derived density labels.
BUS Image Processing: Doppler, elastography, and invalid scans were excluded. Scan images were cleaned according to methods adapted from [2], including automatic cropping and artifact removal. A copy of the validation set (clean validation set) and the testing set additionally had text annotations and lesion markers removed. Images were cropped to 224 x 224.
AI Model Design: We designed a convolutional network architecture for our AI model. Model architecture, hyperparameters, and augmentation strategy were tuned through Optuna [4]. Model predictions were mean-pooled per woman to create a single prediction.
Benchmark Model Design: We additionally compared the performance of our AI model to logistic regression and multi-layer perceptron (MLP) benchmark models based on gray-level bins defined by [5]. Model predictions were mean-pooled per woman to create a single prediction.
Performance Metrics: The ability of the models to classify patients into the density categories was compared by measuring the one-vs-rest Area under the Receiver Operating Characteristic curve (AUROC) on the held-out testing set.

Table 2: Image- and patient-level counts along cancer status and AI-derived breast density lines. Note that the patients contained in the clean validation set are a proper subset of the patients contained in the dirty validation set.

	Overall		Train		Validation (Dirty/Clean)		Test	
	Cases	Controls	Cases	Controls	Cases	Controls	Cases	Controls
Women, N	378	3,722	226	2,226	77	742/737	75	754
Women with fatty/A breasts	3	60	2	37	0	13	1	10
Women with scattered/B breasts	194	1,932	110	1,133	48	393/390	36	406
Women with heterogeneous/C breasts	166	1,606	107	983	24	309/307	35	314
Women with dense/D breasts	15	124	7	73	5	27	3	24
Images, N	11,273	93,692	7,061	56,406	2,409/1,863	19,776/16,474	1,803	17,510
Images with fatty/A breasts	32	1,387	11	804	0	334/251	21	249
Images with scattered/B breasts	5,521	45,295	3,409	27,255	1,400/1,093	9,458/7,862	712	8,582
Images with heterogeneous/C breasts	5,245	43,577	3,487	26,144	742/589	9,297/7,802	1,016	8,136
Images with dense/D breasts	475	3,433	154	2,203	267/181	687/559	54	543

Results

The 4,100 women which met our inclusion criteria with non-Doppler/elastography imaging had a total of 104,965 BUS images (average of 26 images/woman) and a mean age of 59.36. **Figure** 1 shows the selection process with patient counts. The distribution of AI-derived breast density categories among selected women was approximately 1.5% fatty/A, 51.9% scattered/B, 43.2% heterogeneous/C, and 3.4% dense/D. **Table 2** shows a comprehensive breakdown of patient counts by breast density category and cancer status per data split.

Table 1 shows AUROC values for the baseline gray-level models (logistic regression and MLP) as well as our AI solution (CNN). **Figure** 1 shows ROC plots. The deep-learning approach outperformed both gray-level approaches on identifying high and low density breasts, with an AUROC of 0.87 for classifying patients with low density breasts for the CNN as compared to 0.55 and 0.50 for the MLP and logistic regression, respectively. All methods had modest and similar performance for breast density categories B and C.

Table 1: Image- and patient-level modeling results from the gray-level baseline models (logistic regression and MLP) and the CNN model we developed.

Density	Logistic Regression		MLP		CNN	
	Image	Patient	Image	Patient	Image	Patient
A	0.53 (0.50, 0.57)	0.50 (0.31, 0.69)	0.54 (0.50, 0.57)	0.55 (0.39, 0.70)	0.73 (0.70, 0.76)	0.87 (0.79, 0.95)
B	0.59 (0.58, 0.59)	0.60 (0.56, 0.64)	0.64 (0.63, 0.64)	0.66 (0.62, 0.70)	0.66 (0.65, 0.67)	0.72 (0.69, 0.76)
C	0.57 (0.56, 0.57)	0.58 (0.54, 0.62)	0.62 (0.61, 0.63)	0.65 (0.61, 0.69)	0.65 (0.64, 0.66)	0.71 (0.67, 0.74)
D	0.70 (0.68, 0.72)	0.74 (0.65, 0.84	0.74 (0.71, 0.76)	0.80 (0.72, 0.88)	0.78 (0.77, 0.80)	0.89 (0.84, 0.94)

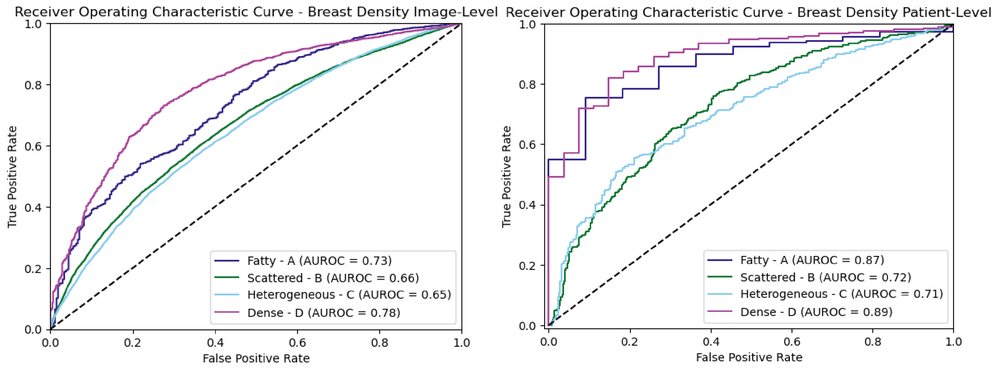

Figure 2: Receiver Operating Characteristic curves for our CNN model for both image- and patient-level predictions of mammographic breast density. Image-level predictions were mean-pooled over each patient to create patient-level predictions of categorical breast density.

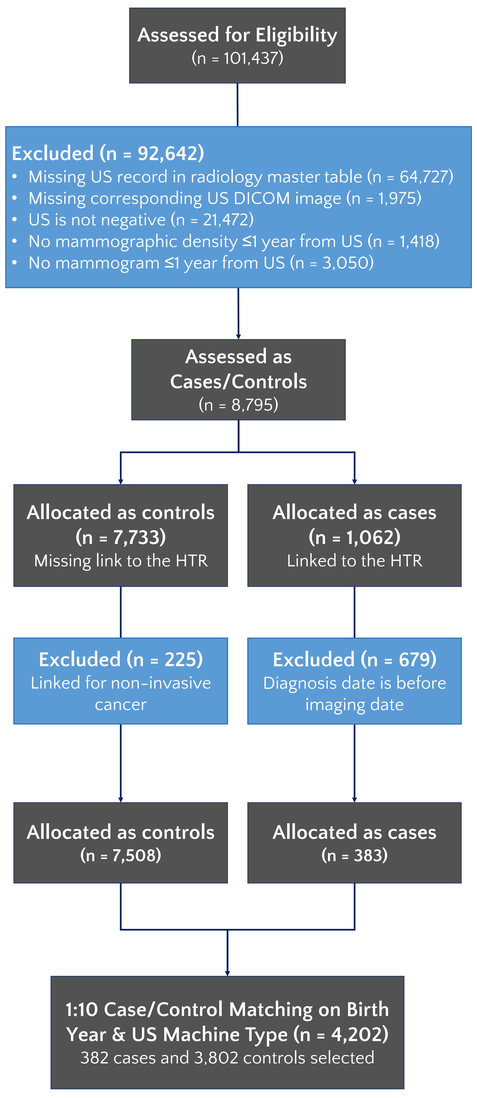

Figure 1: CONSORT-style flowchart showing the selection of patient cases and controls from available data in the HIPIMR.

Conclusion

Accurate high risk (high breast density), as well as low risk (low breast density) classifications are possible from clinically-acquired breast ultrasound images. An AI model trained on images outperforms models trained only on tabular gray-level features.

References

[1] Cancer in the U.S. Affiliated Pacific Islands 2007-2018. Pacific Regional Cancer Registry, 2021.
[2] Shamout FE, Shen Y, Witowski JS, Oliver JR, Kannan K, Wu N, Park J, Beatriu, Reig, Moy L, Heacock L, Geras KJ, editors. The NYU Breast Ultrasound Dataset v1.02021
[3] Wu N, Geras KJ, Shen Y, Su J, Kim SG, Kim E, Wolfson S, Moy L, Cho K, editors. Breast Density Classification with Deep Convolutional Neural Networks. 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP); 2018 2018-04-01: IEEE.
[4] Akiba T, Sano S, Yanase T, Ohta T, Koyama M, editors. Optuna: A Next-generation Hyperparameter Optimization Framework. International Conference on Knowledge Discovery and Data Mining; 2019: ACM.
[5] Jud SM, Häberle L, Fasching PA, Heusinger K, Hack C, Faschingbauer F, Uder M, Wittenberg T, Wagner F, Meier-Meitinger M, Schulz-Wendtland R, Beckmann MW, Adamietz BR. Correlates of mammographic density in B-mode ultrasound and real time elastography. European Journal of Cancer Prevention. 2012;21(4):343-9.