The 11th International Breast Density Workshop, June 4–6, 2025 | Līhuʻe, Kauaʻi Island, Hawaiʻi

Mammography AI Models and Radiomic Features for Breast Cancer Risk Prediction: A Matched Case-Control Study in an Ethnically-Diverse Cohort

Arianna Bunnell[1,2], Thomas K. Wolfgruber[1], Brandon Quon[1], Cade Kane[2], Dustin Valdez[1], Brenda Y. Hernandez[1], Karla Kerlikowske[3], Peter Sadowski[2], and John A. Shepherd[1]
[1]University of Hawaiʻi Cancer Center, [2]University of Hawaiʻi at Mānoa, [3]University of California, San Francisco

Introduction

The age-adjusted incidence of invasive breast cancer is higher in Hawaiʻi (139.6 per 100,000) than in the United States overall (126.9 per 100,000) as of 2018 [1]. The burden of advanced-stage breast cancer varies considerably among Hawaiʻi's Asian, Native Hawaiian, and Pacific Islander (NHPI) ethnic minorities. Among Asian and NHPI women, the proportion of cases diagnosed at an advanced stage is lowest in Japanese (17.2%) and highest in Native Hawaiian (24.6%) women [1]. AI- and radiomic-empowered risk models working from mammography (MG) imaging have shown good performance but are likely estimating breast cancer risk for the population majority. It is unclear how they perform for breast cancer risk in an ethnically-diverse Asian and NHPI population. In this study we ask how AI vs. conventional radiomic features perform in a disaggregated, multi-ethnic population.

Methods

Study Design: Prospective study of all women who received mammograms from 2009 to 2021 at clinical sites participating in the Hawaiʻi and Pacific Islands Mammography Registry (NIH R01CA263491 and U54CA143728). Cases were identified through linkage to the Hawaiʻi Tumor Registry, a SEER registry. Participants had to have a negative (BI-RADS 0/1/2) 2D mammography, all four mammographic views, have known race/ethnicity, be between 40-79 years of age and have known clinical breast density. Cases additionally needed to been imaged ≥6 months before their diagnosis date and have known tumor stage at diagnosis. Controls were matched 3:1 to cases on age, race/ethnicity, first visit date, and MG model/manufacturer.

Radiomic/AI Risk Models: The following AI models/radiomic features were included: 3 radiomic feature sets [2–4], 3 density models [5–6] (Transpara, ScreenPoint Medical, NL), 3 academic AI models [7–9], and 2 commercial AI scores (ProFound AI, iCAD Inc., USA and Transpara, ScreenPoint Medical, NL). A total of 401 conventional radiomics and AI features were generated.

Statistical Analysis: Radiomic families and AI risk models with multiple outputs were reduced in-family through correlation analysis. Features with Pearson's r > |0.7| were removed. Ranked features were considered in an interleaved ordering for removal. Univariate analysis of radiomic features and multivariate analysis with clinical risk factors was done through conditional logistic regression with woman-level clustered standard errors. Radiomic features/AI model outputs were standardized, age was rescaled to decades, and density was represented with base level B.

Racial/Ethnic Subgroup Analysis: For subgroup analysis, case/control groups were limited to those cases who had at least one control whose race/ethnicity *matched exactly*. Cases with no controls of their same race/ethnicity and controls not matching *exactly* were excluded. Racial/ethnic groups with <100 case/control groups remaining do not have subgroup models.

Performance Metrics: The ability of the models to discriminate between cases and controls was compared by measuring the Area under the Receiver Operating Characteristic curve (AUC) on the dataset as well as odds ratio values. Benjamini-Hochberg-adjusted [10] p-values were examined to determine variable significance. AUC 95% confidence intervals were calculated using DeLong's method [11].

Results

- A total of 1,283 cases and 3,159 matched controls met our inclusion criteria. Included women had a mean age of 57.6 years at their MG exam, a mean age of 62.3 at diagnosis, and a mean time from exam to diagnosis of 5.2 years. **Table 1** shows characteristics of included cases and controls. For racial/ethnic group specific models, the following counts were observed. Japanese: 315 cases and 476 matched controls; NHPI: 272 case and 772 matched controls; Filipina: 122 cases and 178 matched controls; and White 182 cases and 506 matched controls.
- From correlation analysis, the number of features from CaPTk [4], OpenBreast [3], Malkov [2], and Mirai [7] were reduced from 328, 32, 11, and 5 features to 32, 9, 4, and 1 features, respectively.
- Baseline models were constructed with only age, clinical breast density, and menopausal status. Baseline model performance as AUC (95% confidence interval) was the following. All women: 0.55 (0.54–0.56); White: 0.57 (0.54–0.59); Japanese: 0.78 (0.77–0.80); NHPI 0.61 (0.59–0.63); and Filipina 0.79 (0.77–0.82). **Figure 2** shows receiver operating characteristic curves for baseline and fully-adjusted radiomic models for all racial/ethnic subgroups.
- **Figure 1** displays forest plots showing the odds ratios and AUC values for overall and subgroup models for radiomic models and clinical risk factors. Adding age into the models attenuates the predictive power of most radiomic features. Conversely, the addition of age and clinical breast density does not appear to have much effect on radiomic predictive power. The lack of effect of breast density for AI features supports the hypothesis that MG features such as breast density are already implicitly encoded into the AI risk score.

Conclusion

AI and radiomic risk models appear to add differing benefit for different racial/ethnic groups, especially when adjusted for common clinical risk factors. Future work will explore multivariate modeling with radiomics, combination of traditional radiomics (OpenBreast, CaPTk, Malkov) with AI model features, as well as expansion into other populations, such as in the San Francisco Mammography Registry.

Table 1: Patient-level counts along cancer status, risk factor, and demographics.

		Cases	Controls
	Women, N	1,283	3,159
Age at Exam	40-49	324	720
	50-59	418	975
	60-69	395	1,062
	70-79	146	402
Race/Ethnicity	NHPI	272	773
	White	182	508
	Chinese	99	138
	Filipina	161	178
	Japanese	444	476
	Hispanic	62	174
	Other/NOS Asian	47	869
	Other	16	43
Breast Density	Fatty (A)	34	189
	Scattered (B)	610	1,438
	Heterogeneous (C)	515	1,272
	Extremely dense (D)	124	260
	Menopausal	904	2,314
	Pre-menopausal	379	845

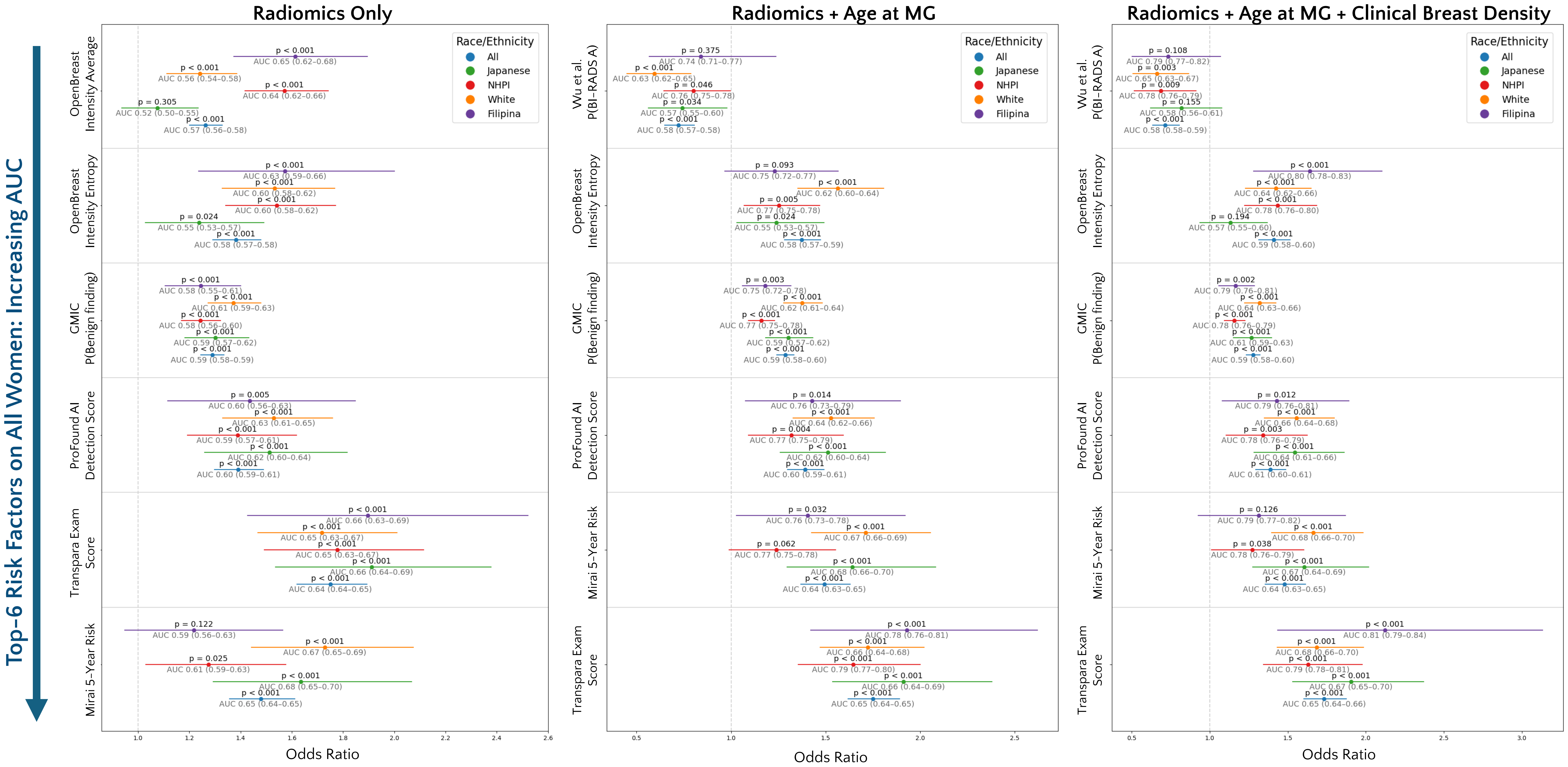

Figure 1: Forest plot showing AUCs (95% confidence interval), odds ratios (line representing 95% confidence interval) for models for all women as well as racial/ethnic subgroup models.

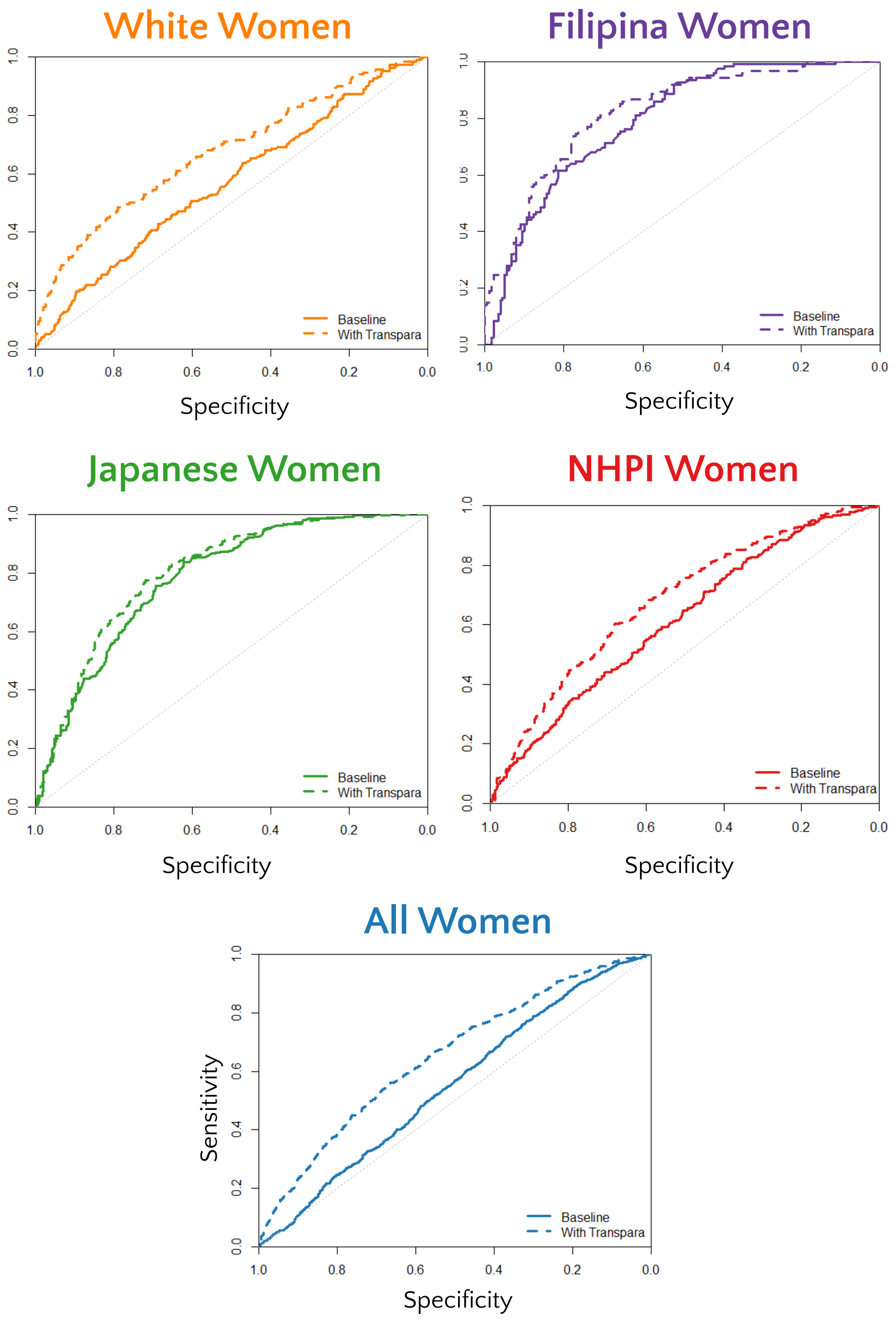

Figure 2: Receiver operating characteristic curves of baseline vs. fully-adjusted radiomic models with using the Transpara Exam Score.

References

1. Cancer at a Glance 2014-2018, Hawaiʻi Tumor Registry, 2022. 2. Malkov S et al. Mammographic texture and risk of breast cancer by tumor type and estrogen receptor status. Breast Cancer Research. 2016;18:1-11. 3. Pertuz S et al. Open framework for mammography-based breast cancer risk assessment. 2019 IEEE EMBS International Conference on Biomedical & Health Informatics (BHI); 2019: IEEE. 4. Zheng Y et al. Parenchymal texture analysis in digital mammography: A fully auctomated pipeline for breast cancer risk assessment. Medical physics. 2015;42(7):4149-60. 5. Wu N et al. Breast Density Classification with Deep Convolutional Neural Networks. 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP); 2018 2018-04-01: IEEE. 6. Keller BM et al. Estimation of breast percent density in raw and processed full field digital mammography images via adaptive fuzzy c-means clustering and support vector machine segmentation. Medical physics. 2012;39(8):4903-17. 7. Yala A et al. Toward robust mammography-based models for breast cancer risk. Science Translational Medicine. 2021;13(578):eaba4373. 8. Shen Y et al. Globally-aware multiple instance classifier for breast cancer screening. Machine Learning in Medical Imaging: 10th International Workshop, MLMI 2019, Held in Conjunction with MICCAI 2019, Shenzhen, China, October 13, 2019, Proceedings 10; 2019: Springer. 9. Zhu X et al. Deep Learning Predicts Interval and Screening-detected Cancer from Screening Mammograms: A Case-Case-Control Study in 6369 Women. Radiology. 2021;301(3):550-8. doi: 10.1148/radiol.2021203758. PubMed PMID: 34491131. 10. Hochberg, Y., and Benjamini, Y.: 'More powerful procedures for multiple significance testing', Statistics in medicine, 1990, 9, (7), pp. 811-818. 11. DeLong, E.R., DeLong, D.M., and Clarke-Pearson, D.L.: 'Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach', Biometrics, 1988, pp. 837-845