

Strasbourg, France · October 1, 2026

Personalized Federated Learning for Equitable Breast Cancer Detection in Underrepresented Pacific Islander Populations

Lydia J. Sollis^{1,2} · Patricia M. Young³ · Arianna Bunnell^{1,2} · Brandon Quon¹ · Brenda Y. Hernandez¹ · Thomas K. Wolfgruber¹ · John A. Shepherd¹

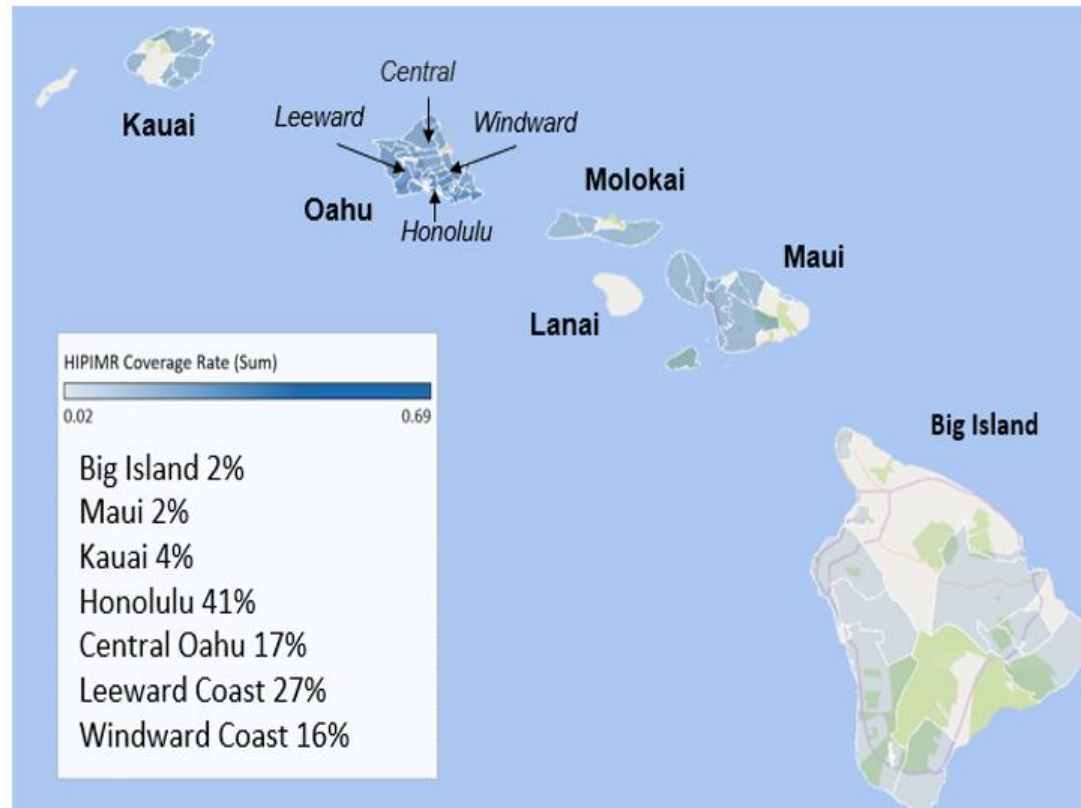
University of Hawai'i Cancer Center · University of Hawai'i at Mānoa · Hawai'i Pacific University

lsollis@cc.hawaii.edu

● Why Federated Learning Matters for Hawai‘i

The University of Hawai‘i Cancer Center is the only NCI-designated cancer center in the Pacific, serving a geographically isolated and diverse population across Hawai‘i and the wider Pacific.

Federated learning could allow Pacific sites and larger external cancer centers to train models without pooling patient data, while respecting Indigenous data sovereignty.



HIPIMR imaging coverage across the Hawaiian Islands, by home-address region (2009–present).

HIPIMR at a Glance

Registry: prospective cohort of >125,000 women imaged 2009–present, linked annually to the Hawai‘i Tumor Registry.

Imaging volume: 568,429 total visits — 354,028 mammograms, 116,675 tomosynthesis, 77,352 ultrasound, 6,158 MRI.

Confirmed cancers: 6,571 invasive (8,375 any), via the same registry linkage.

<https://hipimr.shepherdresearchlab.org/> · R01CA263491

● The Problem: Federation Does Not Guarantee Equal Benefit

- Federated learning in mammography has focused largely on breast-density classification; cancer-detection studies remain limited, and prior work has not evaluated personalized FL in Pacific Islander populations [20–23].

Research Question

Does personalized FL improve worst-site performance and reduce cross-site variation compared with shared-global FL in a heterogeneous mammography federation?

Federated learning (FL) enables multi-site training without sharing patient data,

but standard FedAvg [13] produces one global model that can underperform at sites that differ in population, case mix, and acquisition.

Personalized FL instead lets each site keep a model adapted to its own distribution while still benefiting from shared learning.

● A Heterogeneous Three-Site Federation

	Site A HIPIMR (Hawai'i)	Site B RSNA-US	Site C RSNA-AU
Exams (cancer)	2,188 (520)	996 (249)	936 (234)
NHPI representation	16.2%	0.8%	n/a
Scanner(s)	Hologic (~96%)	7 units (Hologic)	3 units
Hardware	V100 (DGX-1)	GCP VM, T4	RTX A4000

Labeling Across Sites

- Site A (HIPIMR): controls matched 3:1 to registry-confirmed invasive + in-situ cancers on age, density, and scanner — Hawai'i population.
- Sites B/C (public RSNA dataset [24]): screen-detected, biopsy-confirmed cancers, trimmed to the same 3:1 control-to-case ratio — mainland U.S. (Site B) and Australian (Site C) population.

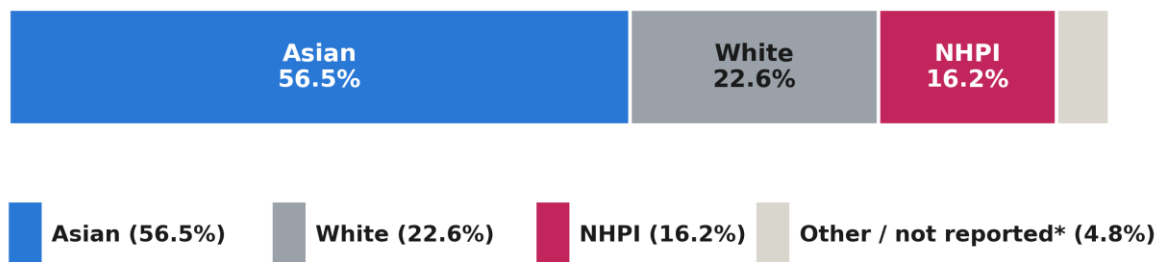


Fig. 1. Site A (HIPIMR) test-set composition by self-reported race/ethnicity, $n=878$ breasts (this study).

*Includes small-n groups pooled for privacy.

● FL Strategies Compared

Centralized *	Single model on pooled 3-site data (pooled-data benchmark).
Local-only *	Each site trains GMIC independently with only site-local data
FedAvg [13]	Dataset-size-weighted averaging into one shared global model.
FedProx [14]	FedAvg + proximal penalty ($\mu=0.1$) limiting client drift.
FedBN [15]	FedAvg with batch-normalization kept local per site.
Ditto [16]	Global model + regularized per-site personalized model vk.
Module-wise Ditto	Ditto with a separate λ per GMIC module (global / local / fusion).

* Local (non-federated) training strategies

● How Ditto Personalizes

Ditto [16] jointly trains a shared global model w^* (via FedAvg [13]) and a personalized model v_k per site:

minimize over v_k : $F_k(v_k) + (\lambda/2) \cdot \|v_k - w^*\|^2$

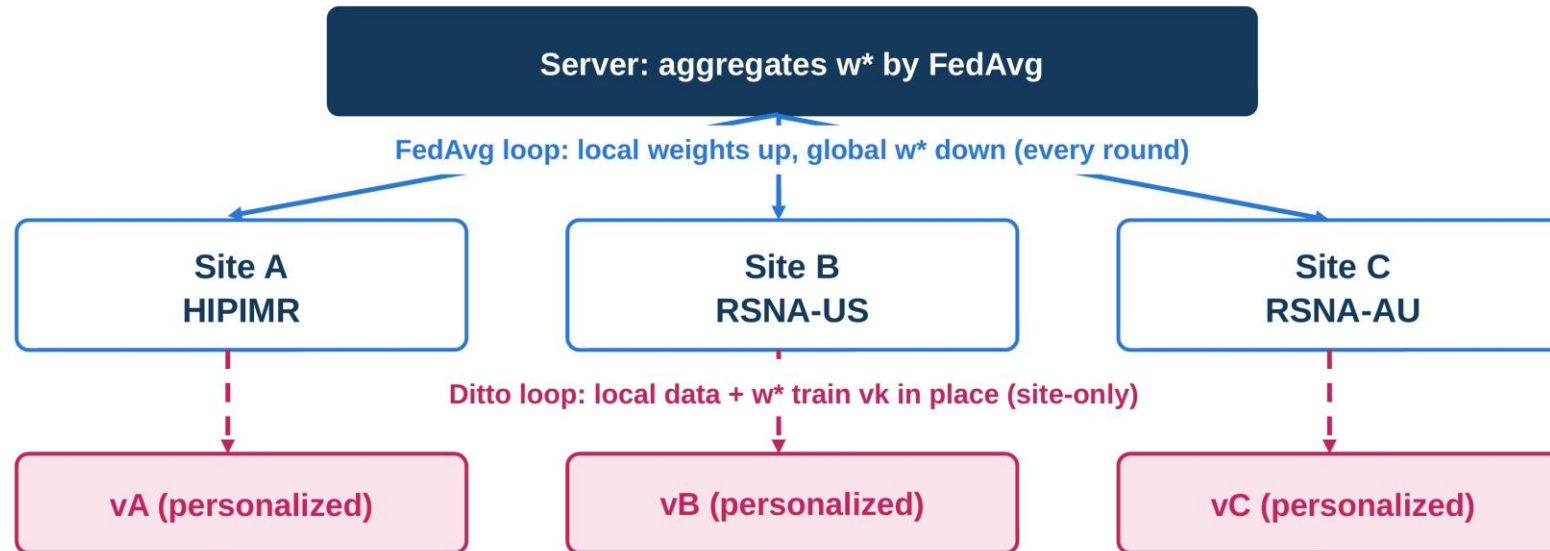


Fig. 2. Ditto's two training loops. Blue: the shared FedAvg loop through the server, each round. Magenta: a loop that runs only at each site, fine-tuning a personalized v_k toward w^* on that site's own data; v_k is never sent to the server.

Larger λ pulls a site's model back toward the shared global model; smaller λ allows more local adaptation. A single λ , swept over $\{0.01, 0.05, 0.1, 0.5, 1, 2\}$ on validation data and chosen to maximize worst-site AUC, is applied at every site; each site deploys its personalized v_k , not the global model.

GMIC Architecture & Fine-Tuning

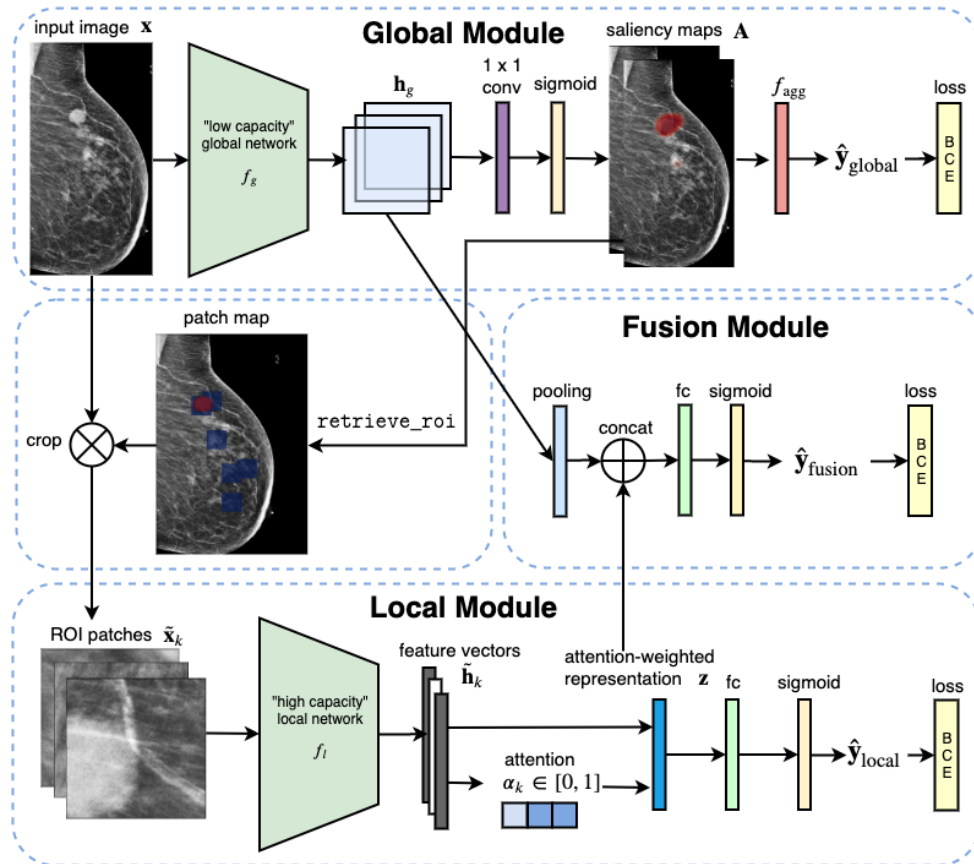


Fig. 3. GMIC's three-module architecture at 2944×1920 resolution: a global module scans the whole image for saliency, a local module examines the $K=6$ highest-attention ROI patches at full capacity, and a fusion module combines both. Adapted from [2,3].

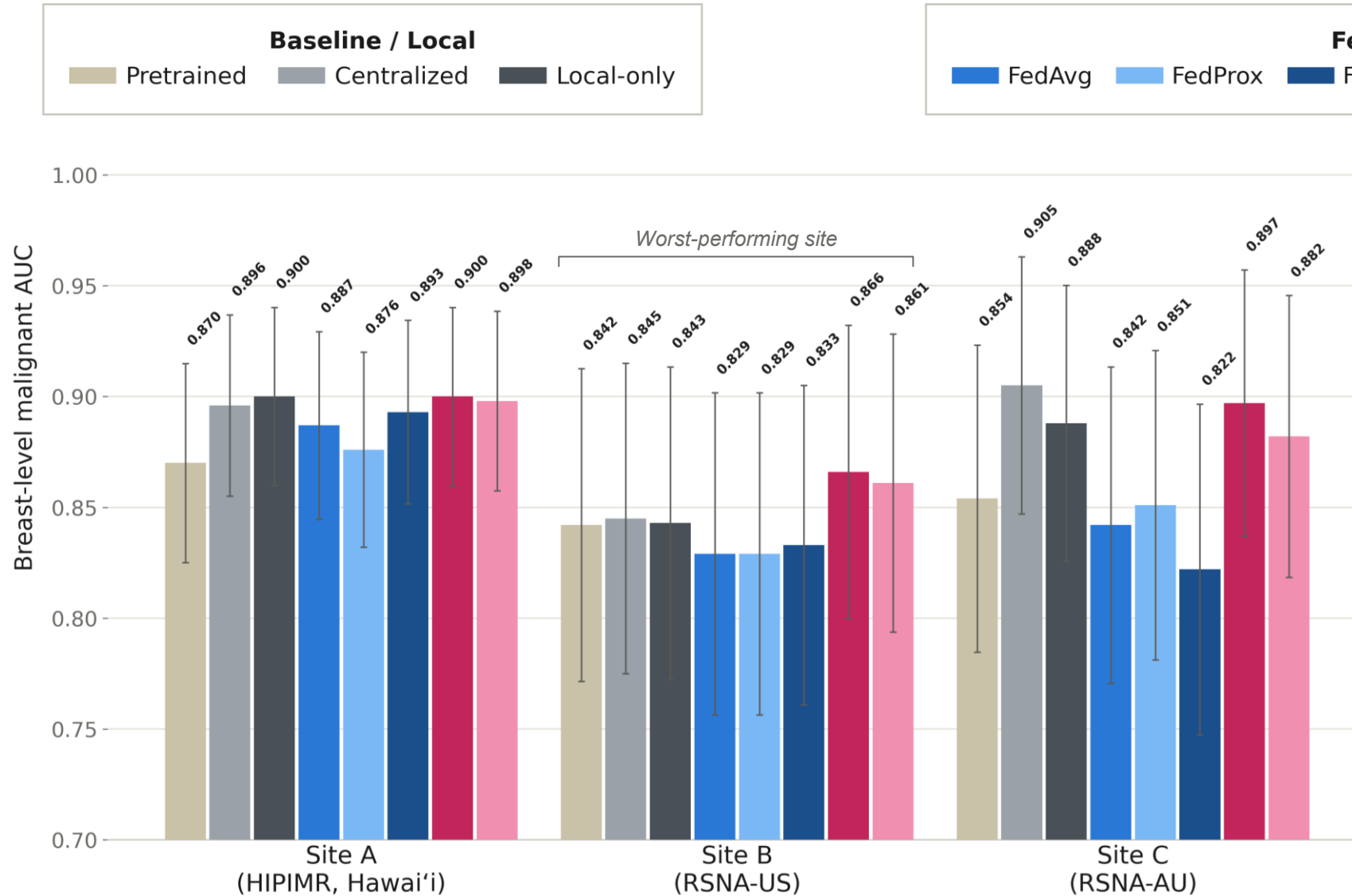
Fine-Tuning (based on published pretrained weights)

- Train/validation/test split 70/10/20%; class imbalance (7.2:1 view-level pos:neg) handled via class-balanced sampler, positive-class weight 1.0
- Adam: lr 3×10^{-5} (heads) / 3×10^{-7} (backbone), weight decay 1×10^{-6} , 2-epoch initial backbone freeze.
- Local models: up to 40 epochs, batch size 32, early stopping on validation AUC (patience 6).
- Federated: 60 rounds $\times$ 1 local epoch/round from the same pretrained init; FedProx $\mu=0.1$, Ditto $\lambda = 0.05$, module-wise Ditto λ (global/local/fusion) = 0.01, 0.5, 0.0; NVFLARE 2.5 used for federated training

Evaluation & Cross-Site Performance Variation

- Breast-level AUC, averaging CC + MLO view predictions per breast, with DeLong 95% confidence intervals [26].
- Deployed round chosen by validation AUC: pooled for shared-global models, site-specific for personalized ones.
- Cross-site AUC spread = sample SD of site-level AUC; lower values indicate more uniform site performance (not a demographic-equity measure by itself).
- Site A subgroup analysis at fixed operating points: sensitivity at 90% specificity, specificity at 90% sensitivity.

Personalization Improved Worst-Site AUC



Ditto is the only federated method that simultaneously lifted Site B above the pretrained baseline, reduced cross-site spread from 0.030 (FedAvg) to 0.019, and matched the pooled AUC (0.889 Ditto vs. 0.887 centralized).

● Exploratory Site A (HIPIMR) Subgroup Analysis

Group	Pretrained AUC	Sn@Sp90	Sp@Sn90	Ditto AUC	Sn@Sp90	Sp@Sn90
Overall (878, 104)	0.870 (0.825–0.915)	0.683	0.532	0.899 (0.859–0.939)	0.760	0.672
NHPI (142, 20)†	0.839 (0.727–0.951)	0.600	0.393	0.898 (0.805–0.991)	0.800	0.787
Asian (496, 66)	0.882 (0.828–0.936)	0.667	0.605	0.900 (0.849–0.951)	0.758	0.593
White (198, 11)†	0.825 (0.672–0.978)	0.727	0.492	0.870 (0.733–1.000)	0.818	0.690

n = (group size, positive breasts). AUC columns show 95% CIs (Hanley-McNeil). NHPI (highlighted) pretrained AUC was 0.839 and increased to 0.898 after fine-tuning. †small-*n* groups are exploratory.

At fixed 90% specificity, Ditto raised NHPI sensitivity 0.60 to 0.80; at fixed 90% sensitivity it raised NHPI specificity 0.39 to 0.79, the largest operating-point gain of any group.

NHPI performance improved after fine-tuning, but similar gains occurred across federated strategies; this analysis does not isolate a personalization effect.

● Takeaways & Limitations

Key Takeaways

- Federated methods did not consistently outperform local fine-tuning
- Ditto produced the strongest worst-site and overall federated performance, matching centralized non-federated training in point estimates
- Module-wise Ditto did not improve standard Ditto; Site A subgroup gains were not specific to personalization

Limitations

- Single training runs, fine-tuning vs training from scratch, and modest site counts produced wide uncertainty
- 1 epoch per round may have limited algorithm variability
- Individual race/ethnicity data available only at the largest, best-served site

Future Work

- Rerun with:
 - Harmonized demographic labels at all sites
 - HIPIMR as smallest site
 - Unique λ per dataset
 - Additional model architectures/types
 - More epochs per round
- Explore more advanced personalization methods

Personalized FL may be particularly useful when heterogeneous sites do not benefit equally from one shared global model.

Thank you · Questions welcome

Lydia J. Sollis · lsollis@cc.hawaii.edu · Shepherd Research Lab, University of Hawai'i Cancer Center

Code: github.com/shepherd-lab/GMIC-FL

References

- [2] Shen, Y., Wu, N., Phang, J., Park, J., Liu, K., Tyagi, S., Heacock, L., Kim, S.G., Moy, L., Cho, K., Geras, K.J.: An interpretable classifier for high-resolution breast cancer screening images utilizing weakly supervised localization. *Medical Image Analysis* 68, 101908 (2021)
- [3] Shen, Y., Wu, N., Phang, J., Park, J., Kim, G., Moy, L., Cho, K., Geras, K.J.: Globally-aware multiple instance classifier for breast cancer screening. In: *Machine Learning in Medical Imaging (MLMI)*. LNCS, vol. 11861, pp. 18–26. Springer (2019)
- [4] McKinney, S.M., Sieniek, M., Godbole, V., Godwin, J., Antropova, N., Ashrafian, H., Back, T., Chesus, M., Corrado, G.S., Darzi, A., et al.: International evaluation of an AI system for breast cancer screening. *Nature* 577(7788), 89–94 (2020)
- [5] Miyawaki, I.A., Banerjee, I., Batalini, F., Jorge, C.A.C., Celi, L.A., Cobanaj, M., Dee, E.C., Gichoya, J.W., Kaffey, Z., Lloyd, M.R., et al.: Global disparities in artificial intelligence-based mammogram interpretation for breast cancer: A scientometric analysis of representation, trends, and equity. *European Journal of Cancer* 220, 115394 (2025)
- [6] Taparra, K., Miller, R.C., Deville, C.: Navigating native hawaiian and pacific islander cancer disparities from a cultural and historical perspective. *JCO Oncology Practice* 17(3), 130–134 (2021)
- [7] Yoshikawa, G.T., Miyazaki, K.S.Y., Acoba, J.D., Fujii, T.: Racial and survival disparities in inflammatory breast cancer (IBC) and non-IBC: a population-based study focused on native hawaiians and other pacific islanders. *Frontiers in Oncology* 14, 1390080 (2024)
- [8] Nguyen, A., Duckworth, E., Pascua, D., Coles, B., Galiano, R.: Disparities in breast cancer treatment and reconstruction among native hawaiian and pacific islander women: systematic review and meta-analysis. *Journal of Surgical Oncology* 131(6), 1002–1012 (2025)
- [9] U.S. Census Bureau: Hawaii state profile, race and ethnicity. American Community Survey 5-year estimates, 2019–2023 (2024)
- [10] Kerlikowske, K., Bissell, M.C.S., Sprague, B.L., Tice, J.A., Tossas, K.Y., Bowles, E.J.A., Ho, T.Q.H., Keegan, T.H.M., Miglioretti, D.L.: Impact of BMI on prevalence of dense breasts by race and ethnicity. *Cancer Epidemiology, Biomarkers & Prevention* 32(11), 1524–1530 (2023)
- [11] Loo, L.W.M., Williams, M., Hernandez, B.Y.: The high and heterogeneous burden of breast cancer in Hawaii: A unique multiethnic U.S. population. *Cancer Epidemiology* 58, 71–76 (2019)
- [12] Giaquinto, A.N., Sung, H., Newman, L.A., Freedman, R.A., Smith, R.A., Star, J., Jemal, A., Siegel, R.L.: Breast cancer statistics 2024. *CA: A Cancer Journal for Clinicians* 74(6), 477–495 (2024)
- [13] McMahan, B., Moore, E., Ramage, D., Hampson, S., Arcas, B.A.y.: Communication-efficient learning of deep networks from decentralized data. In: *Proc. 20th Int. Conf. on Artificial Intelligence and Statistics (AISTATS)*. pp. 1273–1282 (2017)
- [14] Li, T., Sahu, A.K., Zaheer, M., Sanjabi, M., Talwalkar, A., Smith, V.: Federated optimization in heterogeneous networks. *Proceedings of Machine Learning and Systems* 2, 429–450 (2020)
- [15] Li, X., Jiang, M., Zhang, X., Kamp, M., Dou, Q.: FedBN: Federated learning on non-iid features via local batch normalization. *arXiv preprint arXiv:2102.07623* (2021)
- [16] Li, T., Hu, S., Beirami, A., Smith, V.: Ditto: Fair and robust federated learning through personalization. In: *International Conference on Machine Learning (ICML)*. pp. 6357–6368. PMLR (2021)
- [17] Roth, H.R., Cheng, Y., Wen, Y., Yang, I., Xu, Z., Hsieh, Y.T., Kersten, K., Harouni, A., Zhao, C., Lu, K., et al.: NVIDIA FLARE: Federated learning from simulation to real-world. *arXiv preprint arXiv:2210.13291* (2022)
- [20] Roth, H.R., Chang, K., Singh, P., Neumark, N., Li, W., Gupta, V., Gupta, S., Qu, L., Ihsani, A., Bizzo, B.C., et al.: Federated learning for breast density classification: A real-world implementation. In: *DART/DCL, MICCAI Workshops*. pp. 181–191. LNCS, Springer (2020)
- [21] Jiménez-Sánchez, A., Tardy, M., González Ballester, M.A., Mateus, D., Piella, G.: Memory-aware curriculum federated learning for breast cancer classification. *Computer Methods and Programs in Biomedicine* 229, 107318 (2023)
- [22] Tzortzis, I.N., Gutierrez-Torre, A., Sykiotis, S., Agulló, F., Bakalos, N., Doulamis, A., Doulamis, N., Berral, J.L.: Towards generalizable federated learning in medical imaging: A real-world case study on mammography data. *Computational and Structural Biotechnology Journal* 28, 106–117 (2025)
- [23] Xing, H., Sun, R., Ren, J., Wei, J., Feng, C.M., Ding, X., Guo, Z., Wang, Y., Hu, Y., Wei, W., Ban, X., Xie, C., Tan, Y., Liu, X., Cui, S., Duan, X., Li, Z.: Achieving flexible fairness metrics in federated medical imaging. *Nature Communications* 16(1) (2025)
- [24] Trivedi, H.M., Vazirabad, M., Kitamura, F.C., Chen, Y., Frazer, H.: Open-source dataset for the RSNA screening mammography cancer detection challenge. *Radiology: Artificial Intelligence* p. e250375 (2026)
- [26] DeLong, E.R., DeLong, D.M., Clarke-Pearson, D.L.: Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics* 44(3), 837–845 (1988)