Investigating the Potential Ability of AI to Recognize Patient Ethnicity In Mammograms

Yu-Ting Tseng, MSc[1]; Reece Walsh, MSc[1]; John Brandon Graham-Knight, PhD[1]; Patricia Lasserre, PhD[2]; Hua Shen, PhD[3]; Charlotte J. Yong-Hing, MD[4]; Rasika Rajapakshe, PhD[1,2]

[1]BC Cancer Kelowna, Kelowna, BC, Canada, [2]The University of British Columbia-Okanagan, Kelowna, BC, Canada, [3]University of Calgary, Calgary, BC, Canada, [4]BC Cancer Vancouver, Vancouver, BC, Canada

2026 JULY 19–22 VANCOUVER, BC
JOINT AAPM | COMP MEETING

ABSTRACT

Purpose:
To determine if Artificial Intelligence (AI) deep learning models can infer patient ethnicity from screening mammograms, and to identify the image features enabling this inference.

Methods:
Approximately 287,000 mammographic studies (2013–2019) from the BC Cancer Screening program were available, with a class-balanced subset of 18,000 patients(White, East/Southeast Asian, South Asian) used for training. EfficientNetB3 and DINOv2 models were trained on images at various input sizes. Performance was compared to a regression model using volumetric breast density, volume, and fibroglandular volume. External validation used the EMBED dataset.

Results:
The baseline model, EfficientNetB3 (512x512), achieved an overall accuracy of 0.71and a macro-AUC of 0.84. Class-specific F1-scores were 0.78 for East/Southeast Asian, 0.61 for South Asian, and 0.72 for White patients, indicating consistent predictive performance across groups, with the strongest signal in the East/Southeast Asian population. Ablation studies showed that segmentation masks alone retained substantial predictive ability for East/Southeast Asian patients (F1: 0.68), while tissue-only patches performed poorly (F1: 0.41). This suggests shape and size are highly informative, though internal texture also contributes. In contrast, a regression model using standard volumetric features achieved lower performance (F1: 0.61),confirming that deep learning models extract richer visual cues beyond conventional anatomical measurements. Performance generalized to external data (Asian F1: 0.67,White F1: 0.76) and remained robust to manufacturer exclusion.

Conclusion:
AI models can reliably infer patient ethnicity from mammograms by identifying complex visual patterns

CONTACT

Rasika Rajapakshe, PhD, FCCPM
BC Cancer Kelowna
399, Royal Avenue, Kelowna, BC
rrajapak@bccancer.bc.ca
+1 250 712 3915

INTRODUCTION

Disparities in breast cancer screening outcomes across different ethnic groups remain a persistent public health challenge. While artificial intelligence (AI) applications are being developed to improve diagnostic accuracy, there is a risk that these tools could inadvertently reinforce or even amplify existing inequities (1, 2). A significant concern stems from recent work showing that AI can identify demographic traits from medical images, including chest radiographs, CT scans, and even mammograms, that are invisible to the human eye (3-5). However, the existing studies on mammography have been limited to low-resolution images and lack details on which specific anatomical features carry ethnic signals.
Therefore, the purpose of our study was to determine if AI models can implicitly recognize ethnicity from screening mammograms, with a specific focus on distinguishing between Asian - East/Southeast, Asian - South, and White populations.

METHODS AND MATERIALS

Two distinct patient datasets from the BC Cancer Breast Screening Program were used for this study. The primary dataset, used for all deep learning experiments was assembled from a source of approximately 287,000 mammography studies performed between 2013 and 2019. We applied stratified random sampling to create a class-balanced set of 18,000 unique female patients, with 6,000 individuals for each of three self-reported ethnicity groups: Asian - East/Southeast, Asian - South, and White. Only one mammogram (Left Craniocaudal view) per patient was included.
A secondary dataset of same size was used for the radiomics-based and Volpara volumetric estimates - based regression benchmark models since it contained the volumetric breast measurements from Volpara software. We evaluated four distinct modeling approaches. A convolutional neural network (CNN) EfficientNetB3 was used for its balance of efficiency and performance. The model, pretrained on the ImageNet dataset, was fine-tuned for the mammogram-based ethnicity classification task. DINOv2, a Vision Transformer (ViT) model, pretrained on large dataset of unlabeled images was also evaluated. A radiomics-based classification model was developed using handcrafted 2D features extracted from processed mammogram images from the secondary dataset.

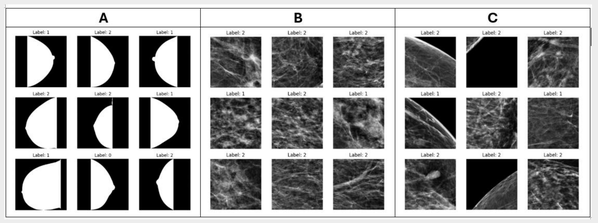

Figure 1. Example images from each ablation type, which includes samples from nine different patients and is for purposes only. (A) Shape- and size-only Analysis (segmentation map). (B) Texture-only Analysis (random tissue patching). (C) Context-free Analysis (random patching).

RESULTS

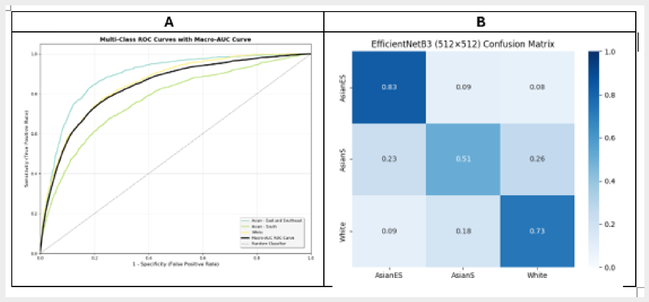

Figure 2. **(A)** AUC-ROC curves and **(B)** confusion matrix for EfficientNetB3 (512×512) on the test set.

The fine-tuned EfficientNetB3 model predicted patient ethnicity across all resolutions, achieving a macro-averaged AUC of 0.85 (95% CI: 0.80, 0.91) on full images at 512×512 pixels. The DINOv2 model also performed strongly with a macro-AUC of 0.84 (95% CI: 0.78, 0.89). Ablation studies revealed that shape and size-only masks retained strong performance (AUC 0.75; 95% CI: 0.68, 0.83), while performance on isolated tissue patches was substantially lower (AUC 0.60; 95% CI: 0.57, 0.63). The deep learning models substantially outperformed the radiomics (AUC 0.77; 95% CI: 0.75, 0.78) and Volpara volumetric estimates (AUC 0.68; 95% CI: 0.67, 0.68) models. Performance was maintained on external validation.

Table 1. Performance of deep learning models for ethnicity prediction.

Model	Image Size	Class	AUC	Sensitivity	Specificity
EfficientNetB3	224×224	Asian - East and Southeast	0.88 (0.87-0.89)	0.70 (0.68-0.73)	0.88 (0.86-0.89)
		Asian - South	0.73 (0.72-0.75)	0.50 (0.48-0.53)	0.81 (0.80-0.82)
		White	0.82 (0.81-0.83)	0.70 (0.68-0.72)	0.77 (0.75-0.78)
		Macro Average	0.81 (0.74-0.88)	0.64 (0.61-0.66)	0.82 (0.80-0.83)
	512×512	Asian - East and Southeast	0.90 (0.89-0.91)	0.81 (0.79-0.83)	0.85 (0.83-0.86)
		Asian - South	0.78 (0.77-0.80)	0.51 (0.49-0.54)	0.86 (0.85-0.87)
		White	0.88 (0.87-0.89)	0.75 (0.72-0.77)	0.82 (0.81-0.84)
		Macro Average	0.85 (0.80-0.91)	0.69 (0.67-0.71)	0.85 (0.83-0.86)
	1024×1024	Asian - East and Southeast	0.91 (0.90-0.92)	0.82 (0.80-0.84)	0.85 (0.84-0.87)
		Asian - South	0.80 (0.79-0.81)	0.53 (0.50-0.55)	0.88 (0.87-0.89)
		White	0.80 (0.88-0.90)	0.78 (0.76-0.80)	0.83 (0.82-0.85)
		Macro Average	0.86 (0.81-0.92)	0.71 (0.70-0.72)	0.85 (0.84-0.87)
DINOv2	224×224	Asian - East and Southeast	0.88 (0.86-0.89)	0.79 (0.77-0.81)	0.81 (0.80-0.83)
		Asian - South	0.74 (0.72-0.75)	0.58 (0.56-0.61)	0.76 (0.74-0.78)
		White	0.82 (0.81-0.83)	0.53 (0.50-0.56)	0.88 (0.87-0.89)
		Macro Average	0.81 (0.75-0.88)	0.63 (0.61-0.66)	0.82 (0.80-0.83)
	512×512	Asian - East and Southeast	0.89 (0.88-0.90)	0.75 (0.73-0.77)	0.87 (0.86-0.88)
		Asian - South	0.77 (0.75-0.78)	0.68 (0.65-0.70)	0.74 (0.72-0.75)
		White	0.85 (0.84-0.86)	0.57 (0.54-0.60)	0.89 (0.88-0.90)
		Macro Average	0.84 (0.78-0.89)	0.67 (0.64-0.69)	0.83 (0.82-0.85)
	1024×1024	Asian - East and Southeast	0.89 (0.88-0.90)	0.77 (0.75-0.79)	0.87 (0.85-0.88)
		Asian - South	0.77 (0.76-0.79)	0.50 (0.48-0.53)	0.87 (0.86-0.89)
		White	0.86 (0.85-0.87)	0.78 (0.76-0.80)	0.78 (0.77-0.80)
		Macro Average	0.84 (0.83-0.85)	0.68 (0.66-0.70)	0.84 (0.83-0.85)

Table 2. : Performance of ablation studies for feature contribution using EfficientNetB3 (512×512).

Model	Experiment Type	Class	AUC	Sensitivity	Specificity
AI Models (EfficientNetB3)	Full Image	Asian - East and Southeast	0.90 (0.89-0.91)	0.81 (0.79-0.83)	0.85 (0.83-0.86)
		Asian - South	0.78 (0.77-0.80)	0.51 (0.49-0.54)	0.86 (0.85-0.87)
		White	0.88 (0.87-0.89)	0.75 (0.72-0.77)	0.82 (0.81-0.84)
		Macro Average	0.85 (0.80-0.91)	0.69 (0.67-0.71)	0.85 (0.83-0.86)
	Shape- and size-only Analysis	Asian - East and Southeast	0.84 (0.83-0.85)	0.83 (0.81-0.85)	0.69 (0.67-0.70)
		Asian - South	0.67 (0.66-0.69)	0.37 (0.35-0.40)	0.82 (0.80-0.83)
		White	0.75 (0.73-0.76)	0.47 (0.44-0.49)	0.83 (0.82-0.84)
		Macro Average	0.75 (0.68-0.83)	0.56 (0.53-0.58)	0.78 (0.76-0.79)
	Texture-only Analysis	Asian - East and Southeast	0.59 (0.58-0.61)	0.36 (0.33-0.38)	0.74 (0.72-0.75)
		Asian - South	0.57 (0.55-0.59)	0.44 (0.41-0.46)	0.63 (0.61-0.65)
		White	0.64 (0.62-0.66)	0.45 (0.43-0.48)	0.75 (0.74-0.77)
		Macro Average	0.60 (0.57-0.63)	0.42 (0.39-0.44)	0.71 (0.69-0.72)
	Context-free Analysis	Asian - East and Southeast	0.81 (0.79-0.82)	0.69 (0.67-0.71)	0.78 (0.76-0.80)
		Asian - South	0.68 (0.66-0.70)	0.47 (0.45-0.50)	0.75 (0.74-0.77)
		White	0.74 (0.73-0.76)	0.51 (0.48-0.53)	0.80 (0.79-0.82)
		Macro Average	0.74 (0.68-0.80)	0.56 (0.53-0.58)	0.78 (0.76-0.79)
Conventional Regression Models	Radiomics-based	Asian - East and Southeast	0.83 (0.82-0.85)	0.75 (0.73-0.77)	0.77 (0.76-0.79)
		Asian - South	0.72 (0.71-0.74)	0.44 (0.42-0.47)	0.83 (0.82-0.84)
		White	0.74 (0.73-0.76)	0.55 (0.52-0.57)	0.77 (0.75-0.78)
		Macro Average	0.77 (0.76-0.78)	0.58 (0.57-0.59)	0.79 (0.78-0.80)
	Volumetric estimates-based	Asian - East and Southeast	0.76 (0.75-0.77)	0.71 (0.70-0.72)	0.70 (0.68-0.70)
		Asian - South	0.62 (0.62-0.63)	0.40 (0.39-0.42)	0.72 (0.72-0.73)
		White	0.65 (0.65-0.66)	0.36 (0.35-0.37)	0.82 (0.81-0.82)
		Macro Average	0.68 (0.67-0.68)	0.50 (0.49-0.50)	0.75 (0.74-0.75)

DISCUSSION

This study investigated whether artificial intelligence models can infer patient ethnicity from screening mammograms and identify the features enabling this capability. We demonstrated that a fine-tuned convolutional neural network predicted a patient's ethnicity (Asian - East/Southeast, Asian - South, or White) with a macro-averaged AUC of 0.85. This performance significantly surpassed that of models based on traditional volumetric measurements or engineered radiomics features. Ablation studies revealed that this predictive ability stems from a complex combination of breast shape and internal tissue texture, confirming that multiple visual features in mammograms carry a demographic signal.

The clinical and ethical implications of these findings are substantial. If AI models can implicitly learn and use ethnicity in their decision-making, it creates a pathway for hidden bias. An algorithm that performs a primary task, such as cancer detection, might use its latent knowledge of ethnicity to perform differently on certain groups, potentially leading to under- or over-diagnosis and exacerbating existing health disparities (1,2). The fact that a general-purpose "foundation model" like DINOv2 also demonstrated this capability highlights the pervasiveness of the issue; this risk is not limited to custom-trained models but may be inherent in the powerful pretrained backbones that are becoming the standard for building modern AI.

CONCLUSIONS

This research provides new evidence that AI can detect ethnicity from high-resolution breast images. Future work should focus on validating these findings in more diverse populations, using explainable AI techniques to further deconstruct the predictive signals, and developing methods to audit and correct for such biases in clinical AI tools to ensure they are developed and deployed responsibly.

REFERENCES

1. Obermeyer Z, Powers B, Vogeli C, Mullainathan S. Dissecting racial bias in an algorithm used to manage the health of populations. Science 2019;366(6464):447–453.
2. Seyyed-Kalantari L, Zhang H, Mcdermott MBA, Chen IY, Ghassemi M. Underdiagnosis bias of artificial intelligence algorithms applied to chest radiographs in under-served patient populations. Nat Med 2021;27(12):2176. doi:10.1038/s41591-021-01595-0
3. Gichoya JW, Banerjee I, Bhimireddy AR, et al. AI recognition of patient race in medical imaging: a modelling study. The Lancet Digital Health 2022;4(6):e406. doi:10.1016/s2589-7500(22)00063-2
4. Adleberg J, Wardeh A, Doo FX, et al. Predicting Patient Demographics From Chest Radiographs With Deep Learning. Journal of the American College of Radiology 2022;19(10):1151. doi:10.1016/j.jacr.2022.06.008
5. Banerjee I, Bhimireddy AR, Burns JL, et al. Reading race: AI recognises patient's racial identity in medical images. arXiv preprint arXiv:2107.10356 2021

© Poster Template by Genigraphics® 1.800.790.4001 www.genigraphics.com