Volumetric Breast Density Estimation From Three-Dimensional Reconstructed Digital Breast Tomosynthesis Images Using Deep Learning

Vinayak S. Ahluwalia, BSE[1,2,3]; Nehal Doiphode, MSE[4]; Walter C. Mankowski, PhD[4]; Eric A. Cohen, MS[1]; Sarthak Pati, MS[5]; Lauren Pantalone, MS[1]; Spyridon Bakas, PhD[6]; Ari Brooks, MD[6]; Celine M. Vachon, PhD[7]; Emily F. Conant, MD[8]; Aimilia Gastounioti, PhD[9]; and Despina Kontos, PhD[4]

DOI https://doi.org/10.1200/CCI.24.00103

ABSTRACT

PURPOSE Breast density is a widely established independent breast cancer risk factor. With the increasing utilization of digital breast tomosynthesis (DBT) in breast cancer screening, there is an opportunity to estimate volumetric breast density (VBD) routinely. However, current available methods extrapolate VBD from two-dimensional (2D) images acquired using DBT and/or depend on the existence of raw DBT data, which is rarely archived by clinical centers because of storage constraints.

METHODS We retrospectively analyzed 1,080 nonactionable three-dimensional (3D) reconstructed DBT screening examinations acquired between 2011 and 2016. Reference tissue segmentations were generated using previously validated software that uses 3D reconstructed slices and raw 2D DBT data. We developed a deep learning (DL) model that segments dense and fatty breast tissue from background. We then applied this model to estimate %VBD and absolute dense volume (ADV) in cm^3 in a separate case-control sample (180 cases and 654 controls). We created two conditional logistic regression models, relating each model-derived density measurement to likelihood of contralateral breast cancer diagnosis, adjusted for age, BMI, family history, and menopausal status.

RESULTS The DL model achieved unweighted and weighted Dice scores of 0.88 (standard deviation [SD] = 0.08) and 0.76 (SD = 0.15), respectively, on the held-out test set, demonstrating good agreement between the model and 3D reference segmentations. There was a significant association between the odds of breast cancer diagnosis and model-derived VBD (odds ratio [OR], 1.41 [95 % CI, 1.13 to 1.77]; $P = .002$), with an AUC of 0.65 (95% CI, 0.60 to 0.69). ADV was also significantly associated with breast cancer diagnosis (OR, 1.45 [95% CI, 1.22 to 1.73]; $P < .001$) with an AUC of 0.67 (95% CI, 0.62 to 0.71).

CONCLUSION DL-derived density measures derived from 3D reconstructed DBT images are associated with breast cancer diagnosis.

ACCOMPANYING CONTENT

Appendix

Accepted October 14, 2024
Published December 9, 2024

JCO Clin Cancer Inform
8:e2400103

INTRODUCTION

Breast density is a widely established independent breast cancer risk factor and is also associated with increased risk of tumor masking.[1-5] The American College of Radiology Breast Imaging Reporting and Data System (BI-RADS)[6] is currently the most used method for assessing breast density in the clinic; however, it is a subjective, qualitative method that carries large inter- and intra-reader variability.[7-10] Fully automated quantitative methods are needed to improve accuracy in breast density estimations and optimize screening protocols. Most current fully automated quantitative methods rely on two-dimensional (2D) digital mammography (DM) images,[11-14] but in the United States, DM has rapidly been replaced by a quasi-three-dimensional (3D) x-ray imaging modality, digital breast tomosynthesis (DBT).[15] DBT creates reconstructed quasi-3D (henceforth referred to as 3D) images of breasts from multiple 2D raw projection images, offering the unique opportunity to quantify dense breast tissue volumetrically, which may improve accuracy and improve risk calculations.[16,17] Approximately 91% of mammography facilities in the United States have incorporated DBT as of July 2024.[18]

Current US Food and Drug Administration (FDA)–cleared methods for breast density estimation exist but possess

CONTEXT

Key Objective

Can we develop a deep learning (DL) algorithm to estimate volumetric breast (VBD) density using three-dimensional (3D) reconstructed, for-presentation digital breast tomosynthesis (DBT) examinations, given that raw projection images are rarely archived by clinical centers?

Knowledge Generated

To our knowledge, we developed the first DL algorithm that can estimate VBD and absolute dense tissue volume from 3D DBT reconstructed slices. The breast density measurements produced by this algorithm are superior to those produced by previous algorithms that rely on two-dimensional raw projections.

Relevance *(J.L. Warner)*

Breast density is a well-known risk factor associated with breast carcinoma and creates challenges with standard screening and diagnostic imaging techniques. This study shows that a DL algorithm can be used to augment standard diagnostics and should be evaluated next in prospective fashion.*

*Relevance section written by *JCO Clinical Cancer Informatics* Editor-in-Chief Jeremy L. Warner, MD, MS, FAMIA, FASCO.

limitations. 2D tools can only approximate breast tissue volume because they extrapolate volumetric breast density (VBD) from 2D images acquired using DBT.[11,14] Current 3D tools can more routinely estimate VBD from DBT but still rely on the existence of 2D raw projection images to refine dense tissue segmentation. These raw projection images are rarely archived by clinical centers because of storage constraints on their picture archive and communication system (PACS).[16,17,19-22] As a result, this limits the ability to use DBT images to perform retrospective cohort studies. The 3D DBT reconstructed data, however, are readily available in many clinical centers.

This study aims to harness deep learning (DL) to develop a computational tool for VBD assessment based solely on 3D reconstructed, for-presentation, DBT images and assess its segmentation performance on a held-out test set. Then, we evaluate the model's efficacy for estimating the likelihood of contralateral breast cancer diagnosis using an independent case-control sample.

METHODS

Study Data Set

This study was performed in a Health Insurance Portability and Protection Act–compliant and institutional review board–approved manner under waiver of consent. To develop our DL model, we retrospectively analyzed 1,080 nonactionable (BI-RADS 1 or 2 and at least a 1-year nonactionable follow-up) DBT screening examinations (Selenia Dimensions, Hologic Inc, Marlborough, MA) obtained between 2011 and 2016 at the Hospital of the University of Pennsylvania (HUP). Both 2D raw projection images and 3D reconstructed slices from the standard acquisition angles (craniocaudal [CC] and mediolateral oblique [MLO]) or views acquired using DBT were available for each examination, with some examinations requiring multiple acquisitions to capture the full extent of their breast tissue. All images were anonymized deterministically by accession number.

3D reference-standard tissue segmentations were generated using previously validated software[21] that uses both 3D reconstructed slices and 2D raw projection images acquired with DBT. Briefly, the software works by first treating each 2D raw projection image as a low-dose DM and uses an adapted version of a previously published DM tissue segmentation algorithm[10] for these images. It then uses these segmented 2D images to build a statistical model of the breast tissue segmentation in 3D while also correcting for blurring effects. Voxels are segmented to be one of three classes (background/nonbreast, fatty breast tissue, dense breast tissue). Given resource constraints, reference tissue segmentations were not able to be refined by board-certified radiologists. Before model training, all 3D reconstructed DBT slices and corresponding reference tissue segmentations were padded to ensure that each image was at least 2,048 voxels in both the x- and y-directions.

DL Model Development

We used a DL-based approach, leveraging neural networks to perform image segmentation.[23] Using the open-source Generally Nuanced Deep Learning Framework (GaNDLF; v0.0.19; MLCommons),[24] we trained a convolutional neural network topology[25] that leverages weights from ImageNet[26] to pretrain encoders followed by a UNet[27]-style decoder, known as FlexiNet. The DL model performs three-label image segmentation on 3D reconstructed DBT reconstructed slices (ie, background/nonbreast, fatty breast

tissue, dense breast tissue). The data set was randomly split into training, validation, and held-out test subsets (80%, 10%, and 10%, respectively) such that all views from any single patient were restricted to a single subset. Each training batch, consisting of 30 patches of size 2048 voxels by 2048 voxels, was uniformly sampled from each image volume. Each batch was normalized such that each voxel value was within the range (0-1).

To evaluate the segmentation accuracy of our DL model, we used a common metric, the Dice similarity coefficient (Dice). Dice is a measure of overlap between the DL model's tissue segmentation and the reference tissue segmentation, with zero indicating absolutely no overlap and one indicating perfect overlap.[28,29] With a three-class segmentation problem, Dice is typically calculated as an unweighted average of the Dice scores of each class. To properly evaluate how the model performs with respect to the rarest class (dense breast tissue), we used a weighted Dice score,[30] in which the Dice score of each class was inversely weighted by the class prevalence in the reference tissue segmentations. A schematic of the experimental methodology is shown in Figure 1.

Additional training parameters included a batch size of 2, a learning rate of 0.0005, the Adam optimizer, weighted Dice loss function,[30] and maximum training epochs of 50. Training was stopped once the validation loss failed to decrease for two consecutive epochs. At the end of every epoch, weighted Dice was calculated for the validation set. We selected the final model weights as the weights corresponding to the epoch that demonstrated the highest weighted Dice on the validation set. Model segmentation performance was evaluated on the held-out test set (10% of the original data set as specified previously) using unweighted and weighted Dice scores.

Class-specific Dice scores are also reported as unweighted Dice scores using the class of interest as the foreground and the other two classes together as the background. For unweighted Dice, weighted Dice, background only, fatty breast tissue only, and dense breast tissue only, we compared Dice scores between CC images and MLO images via two-tailed Student's *t*-test with unequal variances.

VBD Model Evaluation in an Independent Case-Control Sample

After DL model development, the resulting DL model was applied to a case-control sample (n = 834)[19] that was independent of the training, validation, or test sets. The resulting density measures were used in regression models of breast cancer incidence to relate the DL-derived density measures to cancer status. The evaluation procedure was like that described by Gastounioti et al,[19] but on a subset of the case-control data that they used. Cases were selected from women who had biopsy-confirmed unilateral invasive breast

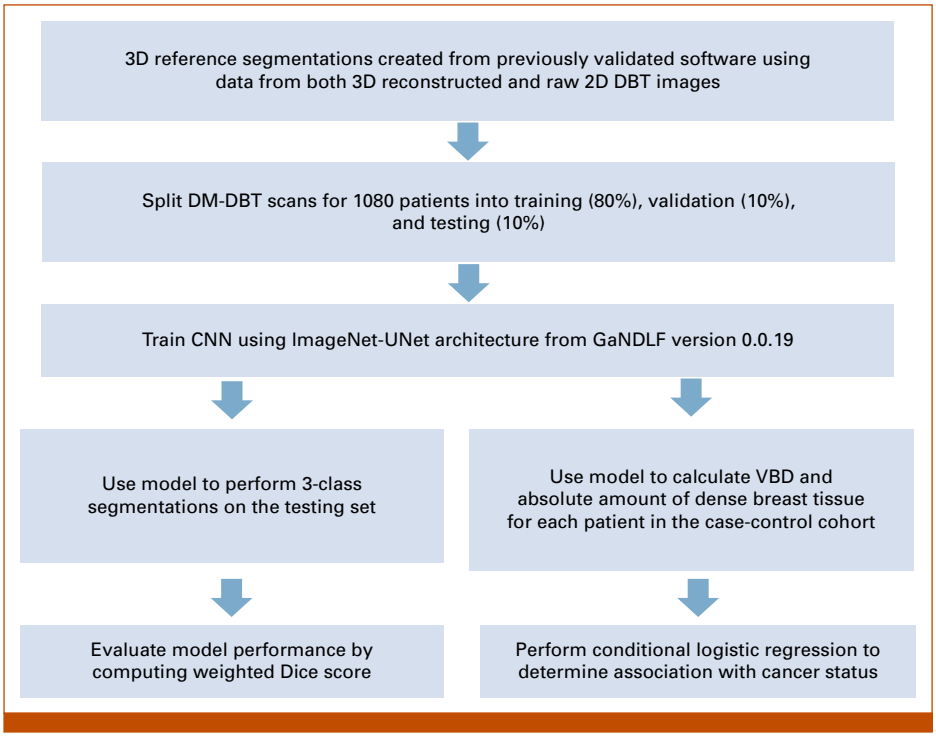

FIG 1. Schematic representing the methodology for reference tissue segmentation construction, DL model training, and DL model evaluation. 2D, two-dimensional; 3D, three-dimensional; CNN, convolutional neural network; DL, deep learning; DM-DBT, digital mammography-digital breast tomosynthesis; VBD, volumetric breast density.

cancer (biopsy- and state registry–confirmed) after dual DM/DBT screening at the HUP between 2012 and 2015. For each case, four age- and race-matched controls were selected from patients who received initial DM/DBT screening during the same period and had nonactionable follow-up screening at least 1 year later. Within each matched group, the age for all women was in the same 5-year range. Individual women were excluded from the regression model if age at screening and/or BMI were not available. All women in a matched group were excluded if no case or no controls were present because of the exclusion criteria.

For each patient in a matched group, DL-derived %VBD was calculated using image views contralateral to the side from which the respective case received a diagnosis of breast cancer.[19,31] If multiple contralateral CC and/or MLO views were available, %VBD was averaged among all contralateral views.[32] Total absolute dense volume (ADV), the sum of voxels segmented as dense tissue multiple by the voxel size (cm^3), was calculated for each patient using the same procedure as %VBD.

We had access to predetermined radiologist-assigned BI-RADS density categories (A, B, C, D) for patients in this sample, where A indicates lowest breast density and D indicates highest breast density. BI-RADS classifications were determined from both breasts together, per the BI-RADS 2013 definition.[33,34] We stratified DL-derived %VBD by the BI-RADS density classifications using box plots and compared the groups using one-way analysis of variance (ANOVA) with Bonferroni correction for multiple comparisons.

To assess the association between each VBD measure and breast cancer, we fit two conditional logistic regression models, one using DL-derived %VBD and the other using DL-derived ADV. Each model was adjusted for age, BMI, number of family members with a history of breast cancer, and menopausal status. Each continuous covariate parameter was normalized to have zero mean and unit standard deviation. Fitting of the conditional logistic regression model yielded log odds per unit change in standard deviation (henceforth referred to as odds ratios [ORs]), with 95% CIs relating the odds of having breast cancer to the respective density measurement. AUC was calculated with respect to the ability of the conditional logistic regression model to correctly classify cancer status, using 1,000-sample bootstraps for AUC CIs.

We performed supplementary analyses stratifying by race and image view. First, we compared unweighted and weighted Dice among patients who identify as White, Black, or Other Race using ANOVA with a Bonferroni correction. We also calculated ORs and AUCs for regression models that were first using only White or only Black participants and then calculated *P* values for the difference in the respective AUCs. Within strata of image view (CC, MLO), we plotted DL-derived %VBD versus reference %VBD and calculated Pearson's correlation coefficients (*r*) to quantify the strength of the linear relationships between the DL-derived VBDs and the reference segmentations.

Our final supplementary analysis compared our DL algorithm against previously published 2D breast density estimation methods for DM such as LIBRA 1.0.4.[34] Using a subset of the case-control sample for which both 3D DBT reconstructed slices and for-processing (ie, raw) 2D DM images were available, we fit four additional conditional logistic regression models using (1) % DL-derived VBD from DBT, (2) DL-derived ADV from DBT, (3) LIBRA-derived area percent density (%ADV) from DM, and (4) LIBRA-derived absolute dense area (ADA) from DM. AUC for these four conditional logistic regression models was calculated in a similar manner to the analyses described above. Finally, we compared AUCs for regression models with regard to (1) % VBD versus %ABD and (2) ADV versus ADA; these comparisons were also made using 1,000 bootstraps.

DL model training using GaNDLF was performed in Python 3.10.0. Logistic regression modeling and other statistical analyses were performed using Stata (version 18.0).

RESULTS

DL Evaluation on the Held-Out Test Set

The data set used for DL model training, validation, and testing included 1080 DBT screening examinations (41.2% Black, 54.2% White, 4.6% Other; mean age ± standard deviation [SD], 57.0 ± 11.3 years; mean BMI ± SD, 28.7 ± 7.1 kg/m^2). The training, validation, and test sets contained 6,294, 791, and 780 3D volumes of DBT reconstructed slices, respectively.

Dice scores showed generally good agreement for background and fatty tissue segmentations, whereas the agreement was somewhat lower for the dense breast tissue segmentation. On the held-out test set, the model achieved the overall unweighted and weighted Dice of 0.88 and 0.76, respectively (Table 1). Performance was best when segmenting the background (Dice = 0.99), indicating that the algorithm can accurately segment breast from nonbreast. There was more agreement between model-predicted segmentations and reference segmentations for CC-view images compared with MLO-view images (unweighted Dice 0.89 and 0.87, respectively ($P < .0001$), with the weighted Dice being 0.80 and 0.72, respectively ($P < .0001$)). In Appendix Table A1, there was similar performance among women who identify as White, Black, or Other Race in terms of unweighted Dice (0.89, 0.88, and 0.88, respectively; $P = .053$). Weighted Dice was lower for Black patients (0.74, $P = .004$) than for White patients (0.78) or patients of another race (0.76). There was moderate correlation between reference %VBD and DL-derived %VBD for the MLO view ($r = 0.67$) with weaker correlation in the CC view ($r = 0.34$; Appendix, Fig A1).

3D DBT reconstructed slices with corresponding reference tissue segmentations and DL model–predicted tissue

TABLE 1. Results on the Held-Out Test Set Using Dice Score

Dice Metric	Entire Test Set (n = 780)	CC-Only (n = 400)	MLO-Only (n = 380)	*P*
Unweighted Dice	0.88 (0.08)	0.89 (0.08)	0.87 (0.07)	<.0001
Weighted Dice	0.76 (0.15)	0.80 (0.12)	0.72 (0.17)	<.0001
Background	0.99 (0.01)	>0.99 (<0.01)	0.99 (0.01)	<.0001
Fatty breast tissue	0.93 (0.08)	0.93 (0.11)	0.93 (0.05)	.488
Dense breast tissue	0.72 (0.17)	0.76 (0.16)	0.69 (0.19)	<.0001

NOTE. Dice and standard deviations in parentheses. *P* values are derived from comparisons between CC-view and MLO-view Dice scores for each specified subgroup using Student's *t* test with unequal variances.
Abbreviations: CC, craniocaudal; MLO, mediolateral oblique.

segmentations are shown in Figure 2. The images shown in the third and fourth columns illustrate how the pectoralis muscle is present in the 3D DBT reconstructed slice in the MLO view, but our DL model successfully segments this region as background or nonbreast in its segmentation prediction.

Independent Case-Control Evaluation

After applying exclusion criteria (Appendix Fig A2), 180 matched groups (654 controls, 180 cases) were included in our independent case-control data set. There was no evidence of a statistically significant difference between cases and controls with respect to race, age, or BIRADS category although the breast cancer group had a statistically significantly higher BMI (Table 2). Mean DL-derived %VBD was not statistically significantly different between the cancer cases and controls (14.6% *v* 13.7%, $P = .19$). However, mean DL-derived ADV was higher in the cancer cases than the control (152.3 cm^3 *v* 127.4 cm^3) with $P < .0001$. The distribution of DL-derived %VBD had a mean of 13.9% with a right skew, indicating a long tail with few high %VBD estimations (Appendix Fig A3). This is in line with the standard distributions for VBD in the literature.[35] The mean DL-derived %VBD increased with the successive BI-RADS density category (7.1%, 10.9%, 21.6%, and 34.6% for categories A, B, C, and D,

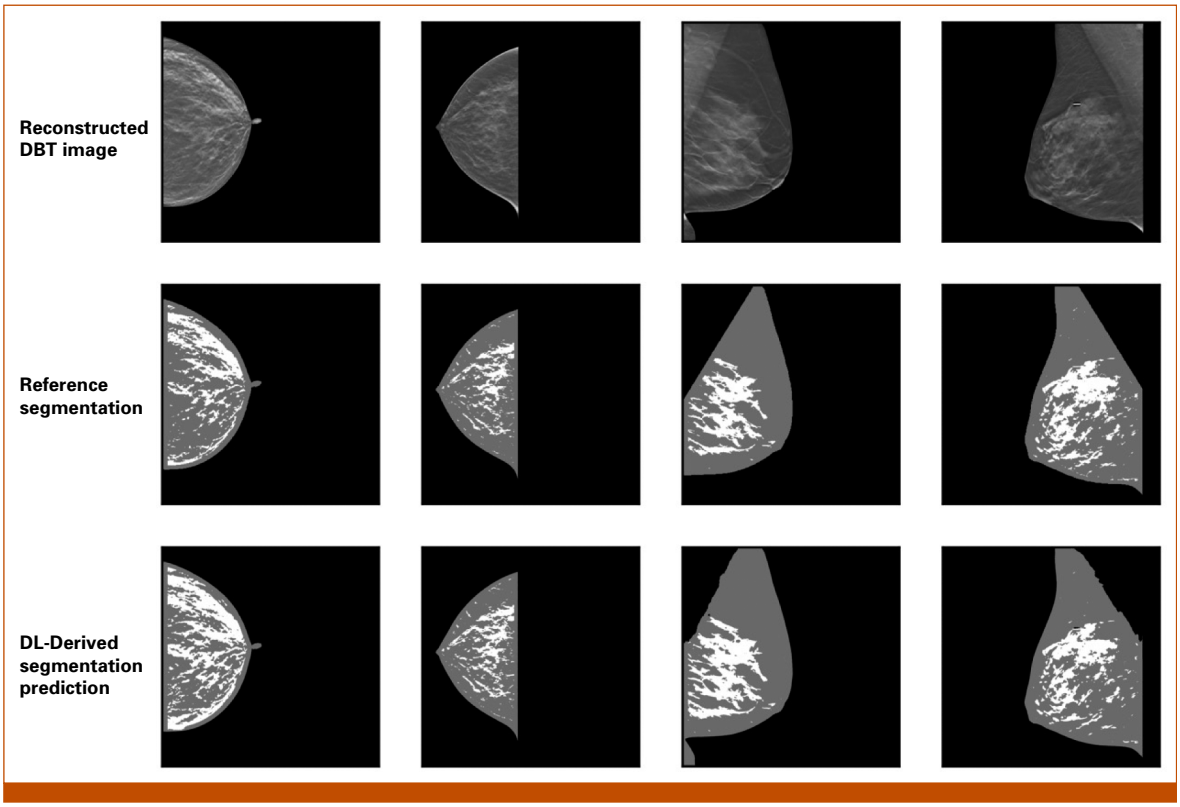

FIG 2. DBT 3D reconstructed slices (top row), corresponding reference segmentation (middle row), and DL-derived segmentation predictions (bottom row) from CC and MLO views in the test set. Black = background; gray = fatty breast tissue; white = dense breast tissue. CC, craniocaudal; DBT, digital breast tomosynthesis; DL, deep learning; MLO, mediolateral oblique.

TABLE 2. Demographics and Other Characteristics of the Case-Control Data Set, Consisting of 180 Groups Matched by Age and Race

Characteristic	Breast Cancer Group (n = 180)	Control Group (n = 654)	*P*
DL-derived %VBD	14.6 ± 8.7	13.7 ± 8.0	.19
DL-derived ADV, cm^3	152.3 ± 77.5	127.4 ± 61.8	.0001
Age, years	60.6 ± 11.0	61.0 ± 10.8	.65
BMI, kg/m^2	29.8 ± 7.1	28.8 ± 6.8	.07
Self-identified race, No.			.89
White	91	322	
Black	83	313	
Other Race or unknown	6	19	
Radiologist-determined BI-RADS density category, No.			.06
A	15	86	
B	96	379	
C	66	182	
D	3	7	

NOTE. *P* values for age and BMI refer to *t* tests with unequal variances; *P* values for race and BI-RADS density were derived from chi-squared tests.
Abbreviations: ADV, absolute dense volume; DL, deep learning; VBD, volumetric breast density.

respectively). As such, there was a clear relationship between DL-derived %VBD and a measure of breast density determined by radiologists (Fig 3).

There was a statistically significant association between breast cancer status and the DL-derived %VBD (OR, 1.41 [95% CI, 1.13 to 1.77]; $P = .002$), age at screening (OR, 4.71 [95% CI, 1.38 to 16.05]; $P = .013$), and BMI (OR, 1.46 [95% CI, 1.17 to 1.81]; $P = .001$). The AUC for this model was 0.65 (95% CI, 0.60 to 0.69). Similarly, there was statistically significant association between breast cancer status and the DL-derived ADV (OR, 1.47 [95% CI, 1.23 to 1.76]; $P < .001$) and age at screening (OR, 4.60 [95% CI, 1.34 to 15.81]; $P = .015$), with an AUC of 0.67 (95% CI, 0.62 to 0.72). Menopausal status, family history, and BMI were not significantly associated with breast cancer status in this model (Table 3). When we fit conditional logistic regression models after stratifying by self-identified race (Appendix Table A2), there

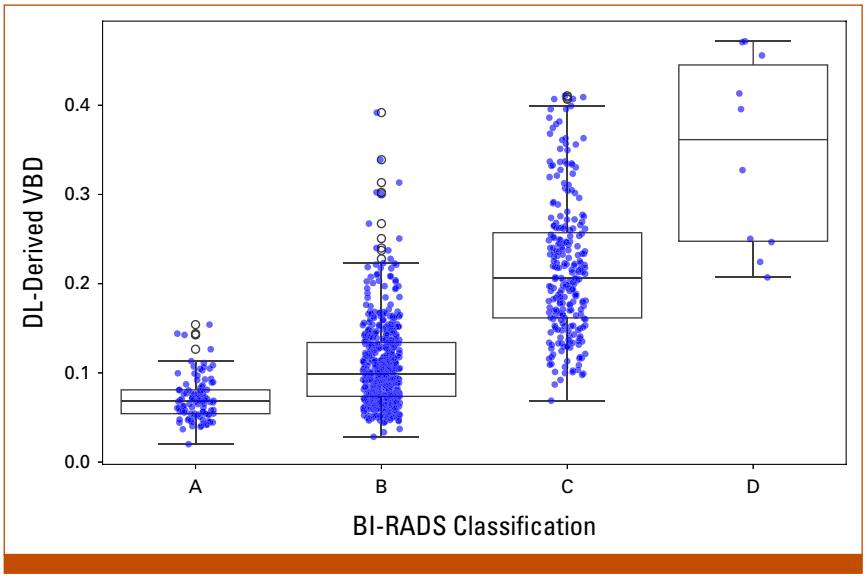

FIG 3. Delineation of DL-derived %VBD by radiologist-determined BI-RADS density category on the case-control set. One-way ANOVA analysis comparing mean DL-derived %VBD for each BI-RADS category was significant after Bonferroni correction for multiple comparisons ($P < .001$). The mean DL-derived %VBD was 7.1%, 10.9%, 21.6%, and 34.6% for BI-RADS density categories A (n = 101), B (n = 475), C (n = 248), and D (n = 10), respectively. ANOVA, analysis of variance; BI-RADS, Breast Imaging-Reporting and Data System; DL, deep learning; VBD, volumetric breast density.

TABLE 3. Odds Ratios Specifying Log Odds of Breast Cancer per Unit Change in Standard Deviation for the Two Conditional Logistic Regression Models

Patient Parameter	Log Odds per Unit Change (95% CI)	*P*	AUC (95% CI)
DBT %VBD-based model			0.65 (0.60 to 0.69)
DL-derived %VBD	1.41 (1.13 to 1.77)	.002	
Age	4.71 (1.38 to 16.05)	.013	
BMI	1.46 (1.17 to 1.81)	.001	
No. of relatives with breast cancer	1.16 (0.99 to 1.37)	.065	
Postmenopausal status	0.89 (0.66 to 1.19)	.437	
DBT ADV–based model			0.67 (0.62 to 0.71)
DL-derived ADV	1.45 (1.22 to 1.73)	<.001	
Age	4.63 (1.35 to 15.90)	.015	
BMI	1.13 (0.93 to 1.36)	.219	
No. of relatives with breast cancer	1.17 (0.99 to 1.38)	.060	
Menopausal status	0.87 (0.65 to 1.16)	.341	

NOTE. AUC was calculated using 1,000 bootstraps.
Abbreviations: ADV, absolute dense volume; DBT, digital breast tomosynthesis; DL, deep learning; VBD, volumetric breast density.

was no statistically significant difference in cancer classification AUC between White and Black patients when using ADV (P = .26) or %VBD (P = .17).

Additional analyses were conducted to determine if our 3D model outperforms previously published 2D methods. A total of 785 patients across 146 matched groups had both 2D for-processing DM available and 3D DBT reconstructed slices (Appendix Table A3). Replacing LIBRA-derived %ABD with DL-derived %VBD increased the AUC from 0.61 (95% CI, 0.55 to 0.65) to 0.63 (95% CI, 0.58 to 0.68). Similarly, replacing LIBRA-derived ADA with DL-derived ADV increased the AUC from 0.59 (95% CI, 0.53 to 0.64) to 0.66 (95% CI, 0.61 to 0.71); LIBRA-derived ADA did not have a statistically significant association with breast cancer diagnosis (OR, 1.17 [95% CI, 0.93 to 1.46]). The difference between the AUC for %VBD and ADA was not statistically significant (P = .101), but the superiority of the AUC for ADV over ADA was statistically significant (P < .001; Appendix Table A4).

DISCUSSION

Here, to our knowledge, we present the first DL model that can estimate continuous VBD measures from 3D DBT reconstructed slices rather than 2D raw DBT projections. The DL model demonstrated a good overlap with reference tissue segmentations on the basis of Dice score. There was an association between density measures estimated from the DL model and breast cancer status, justifying the clinical utility of such a model. The AUC values derived from the conditional logistic regression models were comparable with the AUC reported by Gastounioti et al[19] (0.62 [95% CI, 0.57 to 0.68]) in their work that used raw projections, rather than the 3D reconstructed DBT images. These AUC values were superior to those obtained by models that use %ABD or ADA extracted from DM, with the AUC for ADV differing from ADA at a statistically significant level. These predictions could conceivably be used to conduct population-level–based personal risk assessments. Moreover, the FDA now requires that women be informed of their breast density[36]; automated methods for calculating %VBD can quickly provide this information to women.

Given the abundance of literature describing the role of racial biases in machine learning algorithms[37,38] and the racial disparities in cancer care,[39,40] we must ensure that new machine learning models can remain robust to racial biases. Evaluation of the DL model on the held-out test set highlights that performance is comparable among patients of different self-identified races in terms of unweighted Dice although weighted Dice is lower among self-identified Black patients than White patients. Classification performance for the %VBD and ADV was slightly higher for White patients than Black patients, but this difference was not statistically significant. In our specific patient population at HUP, women who identify as Black tend to have larger breasts, so multiple tiled images are required in each projection to visualize all tissues. When there are more projections, it is more likely to have an overlap of visualized tissues between the projections, which can lead to errors in segmentation. This may explain the lower segmentation agreement and AUC for self-identified Black patients in this specific population compared with self-identified White patients.

Anecdotally, at one of our institutions, raw projection images are only stored in PACS for approximately 3 months. So, while previous algorithms that use raw projection images could be used immediately after the examination, only an algorithm that uses 3D reconstructed images, such as ours, could be used to facilitate large retrospective density estimations. We postulate that if clinical centers no longer need to retain 2D raw projection images, this may lead to

decreased information technology (IT) costs although future studies are needed to confirm this. Fortunately, the framework which we used to develop our DL model (GaNDLF) possesses postoptimization strategies to eliminate the need for a graphics processing unit and decrease random access memory needs by 50%.[24,41]

We would also like to highlight that glandular tissue in the axillary tail of the breast frequently overlaps with the pectoralis muscle on 2D imaging. This has led to issues with pectoral removal with computer-aided breast segmentation and has necessitated the development of specific algorithms whose sole aim is the correct segmentation of the pectoral muscle.[42] The DL algorithm was able to successfully segment out the pectoralis and chest wall in the MLO view while simultaneously distinguishing dense breast tissue from fatty breast tissue, eliminating the need for a separate pectoral segmentation algorithm.

This work has limitations. While the reference segmentations were generated from software, which was shown to be quite effective at segmenting raw DBT projection images,[16] the absence of clinician confirmation might have affected reference tissue segmentations. The correlations between the estimated %VBD and ADV and the reference segmentations were only moderately correlated in the CC view although we are comparing our model's density measures with those of a previously published algorithm, rather than with a radiologist's assessment. The case-control analysis used imaging taken at the time of diagnosis for cases, rather than previous imaging. This study also used a single DBT vendor system, so it is not necessarily generalizable to multiple vendor systems.

Our immediate next steps are to establish a multi-institutional cohort that will allow us to have sufficient statistical power to determine our algorithm's risk estimation capabilities by predicting future cancer cases from negative screening examinations. Extension to include 3D segmentation of breast magnetic resonance imaging (MRI) and comparison with that of quasi-3D DBT imaging may prove to be useful. Finally, leveraging strategies such as transfer learning with systemic data curation from multiple DBT vendor systems can increase this method's generalizability.

In conclusion, to our knowledge, we have developed the first DL model that calculates VBD from 3D reconstructed DBT images, which are readily archived by clinical centers in PACS, in contrast to raw 2D DBT projections, which are not. In the clinical setting, this tool can predict personalized breast cancer risk and guide management, as well as perform large epidemiologic studies of breast cancer risk.

AFFILIATIONS

[1]Center for Biomedical Image Computing and Analytics (CBICA), University of Pennsylvania, Philadelphia, PA
[2]Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA
[3]Leonard Davis Institute of Health Economics, University of Pennsylvania, Philadelphia, PA
[4]Department of Radiology, Columbia University Irving Medical Center, New York, NY
[5]Department of Pathology and Laboratory Medicine, Indiana University School of Medicine, Indianapolis, IN
[6]Department of Surgery, Division of Endocrine and Oncologic Surgery, Hospital of the University of Pennsylvania, Philadelphia, PA
[7]Department of Quantitative Health Sciences, Division of Epidemiology, Mayo Clinic, Rochester, MN
[8]Department of Radiology, Hospital of the University of Pennsylvania, Philadelphia, PA
[9]Mallinckrodt Institute of Radiology, Washington University School of Medicine, St Louis, MO

CORRESPONDING AUTHOR

Despina Kontos, PhD; e-mail: dk3360@cumc.columbia.edu.

EQUAL CONTRIBUTION

A.G. and D.K. contributed equally to this work as cosenior authors.

PRIOR PRESENTATION

Presented at AACR Annual Meeting 2022, New Orleans, LA, April 8-13, 2022; Abstract published only at 2022 ASCO Annual Meeting, Chicago, IL, June 3-7, 2022; and Presented at 2022 RSNA Annual Meeting, Chicago, IL, November 27-December 1, 2022.

DATA SHARING STATEMENT

Code is available online at https://www.columbiaradiology.org/research/research-labs/computational-biomarker-imaging-group-cbig/software. For any further inquiries, please contact dk3360@cumc.columbia.edu.

AUTHOR CONTRIBUTIONS

Conception and design: Vinayak S. Ahluwalia, Nehal Doiphode, Spyridon Bakas, Ari Brooks, Aimilia Gastounioti, Despina Kontos
Financial support: Spyridon Bakas
Administrative support: Sarthak Pati, Spyridon Bakas
Provision of study materials or patients: Emily F. Conant
Collection and assembly of data: Vinayak S. Ahluwalia, Walter C. Mankowski, Sarthak Pati, Lauren Pantalone, Spyridon Bakas, Celine M. Vachon, Emily F. Conant, Aimilia Gastounioti, Despina Kontos
Data analysis and interpretation: Vinayak S. Ahluwalia, Walter C. Mankowski, Eric A. Cohen, Lauren Pantalone, Spyridon Bakas, Celine M. Vachon, Emily F. Conant, Aimilia Gastounioti, Despina Kontos
Manuscript writing: All authors
Final approval of manuscript: All authors
Accountable for all aspects of the work: All authors

AUTHORS' DISCLOSURES OF POTENTIAL CONFLICTS OF INTEREST

The following represents disclosure information provided by authors of this manuscript. All relationships are considered compensated unless otherwise noted. Relationships are self-held unless noted. I = Immediate Family Member, Inst = My Institution. Relationships may

not relate to the subject matter of this manuscript. For more information about ASCO's conflict of interest policy, please refer to www.asco.org/rwc or ascopubs.org/cci/author-center.
Open Payments is a public database containing information reported by companies about payments made to US-licensed physicians (Open Payments).

Vinayak S. Ahluwalia
Research Funding: Flatiron Health (Inst)

Spyridon Bakas
Patents, Royalties, Other Intellectual Property: In vivo detection of EGFR mutation in glioblastoma via MRI signature consistent with deep peritumoral infiltration (Inst)

Ari Brooks
Stock and Other Ownership Interests: UE Life Sciences
Speakers' Bureau: Cianna Medical
Research Funding: UE Life sciences (Inst), OncoNano Inc (Inst), Intuitive Surgical (Inst), Anixa Diagnostics (Inst), Integro Theranostics Inc (Inst)
Patents, Royalties, Other Intellectual Property: I have co-authored three issued patents, all are held by Drexel University. Two are in the area of quantum dot technology for tissue staining, one is for Piezo-electric finger technology
Travel, Accommodations, Expenses: Intuitive Surgical
Uncompensated Relationships: UE Life Sciences
Open Payments Link: https://openpaymentsdata.cms.gov/physician/75370/summary

Celine M. Vachon
Research Funding: GRAIL (Inst)
Patents, Royalties, Other Intellectual Property: Breast Density software

Emily F. Conant
Leadership: iCAD, Inc, Hologic
Honoraria: Medality
Consulting or Advisory Role: Hologic, ICAD
Speakers' Bureau: Medscape, iiCME, Inc
Research Funding: Hologic (Inst), iCAD, Inc (Inst)

Aimilia Gastounioti
Research Funding: Whiterabbit AI, Inc (Inst)

Despina Kontos
Research Funding: Hologic (Inst), GenMab Inc (Inst), iCAD (Inst), Calico Inc (Inst)

No other potential conflicts of interest were reported.

ACKNOWLEDGMENT

The authors acknowledge support from the Susan G. Komen for the Cure Breast Cancer Foundation (CCR231011994) and the National Cancer Institute at the National Institutes of Health (R01 Research Projects [2R01CA161749-05, R01CA286120, and R01CA177150]).

REFERENCES

1. Boyd NF, Martin LJ, Bronskill M, et al: Breast tissue composition and susceptibility to breast cancer. J Natl Cancer Inst 102:1224-1237, 2010
2. McCormack VA, dos Santos Silva I: Breast density and parenchymal patterns as markers of breast cancer risk: A meta-analysis. Cancer Epidemiol Biomarkers Prev 15:1159-1169, 2006
3. Brentnall AR, Cuzick J, Buist DSM, et al: Long-term accuracy of breast cancer risk assessment combining classic risk factors and breast density. JAMA Oncol 4:e180174, 2018
4. Mandelson MT, Oestreicher N, Porter PL, et al: Breast density as a predictor of mammographic detection: Comparison of interval- and screen-detected cancers. J Natl Cancer Inst 92:1081-1087, 2000
5. Boyd NF, Guo H, Martin LJ, et al: Mammographic density and the risk and detection of breast cancer. N Engl J Med 356:227-236, 2007
6. Breast Imaging Reporting & Data System. https://www.acr.org/Clinical-Resources/Reporting-and-Data-Systems/Bi-Rads
7. Sprague BL, Conant EF, Onega T, et al: Variation in mammographic breast density assessments among radiologists in clinical practice: A multicenter observational study. Ann Intern Med 165:457-464, 2016
8. Conant EF, Sprague BL, Kontos D: Beyond BI-RADS density: A call for quantification in the breast imaging clinic. Radiology 286:401-404, 2018
9. Ekpo EU, McEntee MF: Measurement of breast density with digital breast tomosynthesis—A systematic review. Br J Radiol 87:20140460, 2014
10. Tari DU, Santonastaso R, De Lucia DR, et al: Breast density evaluation according to BI-RADS 5th edition on digital breast tomosynthesis: AI automated assessment versus human visual assessment. J Pers Med 13:609, 2023
11. Highnam R, Brady SM, Yaffe MJ, et al: Robust breast composition measurement—VolparaTM, in Martí J, Oliver A, Freixenet J, et al (eds): Digital Mammography. Lecture Notes in Computer Science, Volume 6136. Heidelberg, Germany, Springer, 2010, pp 342-349
12. Eng A, Gallant Z, Shepherd J, et al: Digital mammographic density and breast cancer risk: A case–control study of six alternative density assessment methods. Breast Cancer Res 16:439, 2014
13. Keller BM, Nathan DL, Wang Y, 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. Med Phys 39:4903-4917, 2012
14. Hartman K, Highnam R, Warren R, et al: Volumetric assessment of breast tissue composition from FFDM images, in Krupinski EA (eds): Lecture Notes in Computer Science, Volume 5116. Berlin, Heidelberg, Springer, 2008, pp 33-39
15. Center for Devices and Radiological Health: 2023 Scorecard Statistics. FDA, 2024. https://www.fda.gov/radiation-emitting-products/mqsa-insights/2023-scorecard-statistics
16. Sechopoulos I: A review of breast tomosynthesis. Part I. The image acquisition process. Med Phys 40:014301, 2013
17. Sechopoulos I: A review of breast tomosynthesis. Part II. Image reconstruction, processing and analysis, and advanced applications. Med Phys 40:014302, 2013
18. Center for Devices and Radiological Health. MQSA National Statistics. FDA. 2024. https://www.fda.gov/radiation-emitting-products/mqsa-insights/mqsa-national-statistics
19. Gastounioti A, Pantalone L, Scott CG, et al: Fully automated volumetric breast density estimation from digital breast tomosynthesis. Radiology 301:561-568, 2021
20. Conant EF: Clinical implementation of digital breast tomosynthesis. Radiol Clin North Am 52:499-518, 2014
21. Pertuz S, McDonald ES, Weinstein SP, et al: Fully automated quantitative estimation of volumetric breast density from digital breast tomosynthesis images: Preliminary results and comparison with digital mammography and MR imaging. Radiology 279:65-74, 2016
22. Lee CI, Lehman CD: Digital breast tomosynthesis and the challenges of implementing an emerging breast cancer screening technology into clinical practice. J Am Coll Radiol 13:R61-R66, 2016
23. Kotu V, Deshpande B: Chapter 10—Deep learning, in Kotu V, Deshpande B (eds): Data Science (ed 2). Burlingon, MA, Morgan Kaufmann, 2019, pp 307-342
24. Pati S, Thakur SP, Hamamcı İE, et al: GaNDLF: The generally nuanced deep learning framework for scalable end-to-end clinical workflows. Commun Eng 2:23-17, 2023
25. Baheti B, Pati S, Menze B, et al: Leveraging 2D deep learning ImageNet-trained models for native 3D medical image analysis, in Bakas S, Crimi A, Baid U, et al (eds): Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries. Cham, Switzerland, Springer Nature, 2023, pp 68-79
26. Deng J, Dong W, Socher R, et al: ImageNet: A large-scale hierarchical image database. 2009 IEEE Conference on Computer Vision and Pattern Recognition, Miami, FL, June 20-25, 2009
27. Ronneberger O, Fischer P, Brox T: U-Net: Convolutional networks for biomedical image segmentation, in Navab N, Hornegger J, Wells WM, et al (eds): Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015. Munich, Germany, Springer International Publishing, 2015, pp 234-241
28. Zijdenbos AP, Dawant BM, Margolin RA, et al: Morphometric analysis of white matter lesions in MR images: Method and validation. IEEE Trans Med Imaging 13:716-724, 1994
29. Maier-Hein L, Reinke A, Godau P, et al: Metrics reloaded: Recommendations for image analysis validation. Nat Methods 21:195-212, 2024
30. Sugino T, Kawase T, Onogi S, et al: Loss weightings for improving imbalanced brain structure segmentation using fully convolutional networks. Healthcare 9:938, 2021
31. Gastounioti A, Desai S, Ahluwalia VS, et al: Artificial intelligence in mammographic phenotyping of breast cancer risk: A narrative review. Breast Cancer Res 24:14, 2022
32. Gastounioti A, Conant EF, Kontos D: Beyond breast density: A review on the advancing role of parenchymal texture analysis in breast cancer risk assessment. Breast Cancer Res 18:91, 2016
33. Melnikow J, Fenton JJ, Whitlock EP, et al: Table 1, BI-RADS Breast Density Descriptions, 2016. https://www.ncbi.nlm.nih.gov/books/NBK343794/table/ch1.t1/
34. Haji Maghsoudi O, Gastounioti A, Scott C, et al: Deep-LIBRA: An artificial-intelligence method for robust quantification of breast density with independent validation in breast cancer risk assessment. Med Image Anal 73:102138, 2021

35. Yaffe MJ, Boone JM, Packard N, et al: The myth of the 50-50 breast. Med Phys 36:5437-5443, 2009
36. Office of the Commissioner: FDA Updates Mammography Regulations to Require Reporting of Breast Density Information and Enhance Facility Oversight. FDA, 2023. https://www.fda.gov/news-events/press-announcements/fda-updates-mammography-regulations-require-reporting-breast-density-information-and-enhance
37. Rajpurkar P, Chen E, Banerjee O, et al: AI in health and medicine. Nat Med 28:31-38, 2022
38. Ratajczak R, Cockerill RG: Artificial intelligence in violence risk assessment: Addressing racial bias and inequity. J Psychiatr Pract 29:239-245, 2023
39. Coughlin SS: Social determinants of breast cancer risk, stage, and survival. Breast Cancer Res Treat 177:537-548, 2019
40. Zavala VA, Bracci PM, Carethers JM, et al: Cancer health disparities in racial/ethnic minorities in the United States. Br J Cancer 124:315-332, 2021
41. Thakur SP, Pati S, Panchumarthy R, et al: Optimization of deep learning based brain extraction in MRI for low resource environments, in Crimi A, Bakas S (eds): Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries. Singapore, Springer International Publishing, 2022, pp 151-167
42. Chen S, Bennett DL, Colditz GA, et al: Pectoral muscle removal in mammogram images: A novel approach for improved accuracy and efficiency. Cancer Causes Control 35:185-191, 2024

APPENDIX

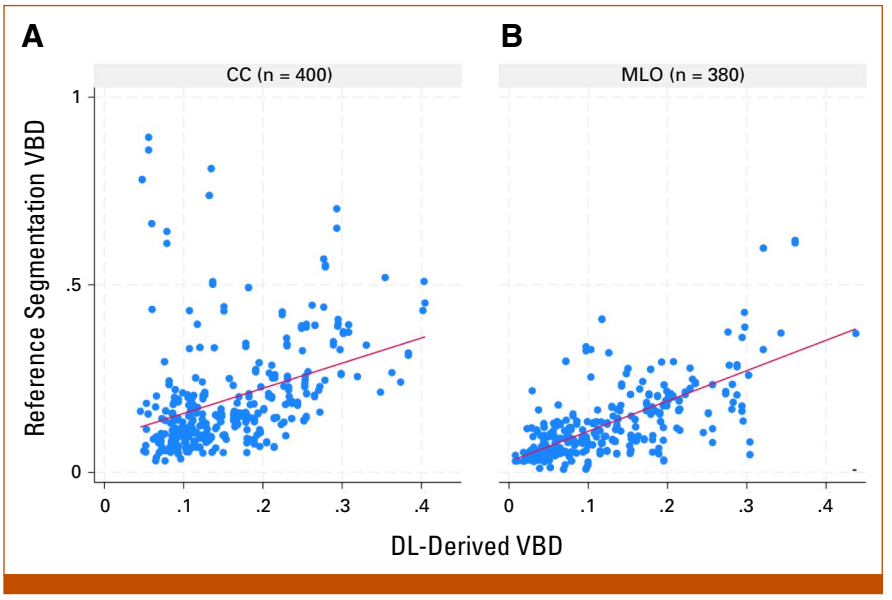

FIG A1. Scatter plot illustrating the relationship between (A) DL-derived VBD and reference VBD in the test set when stratifying by image view. CC, craniocaudal; DL, deep learning; MLO, mediolateral oblique; VBD, volumetric breast density.

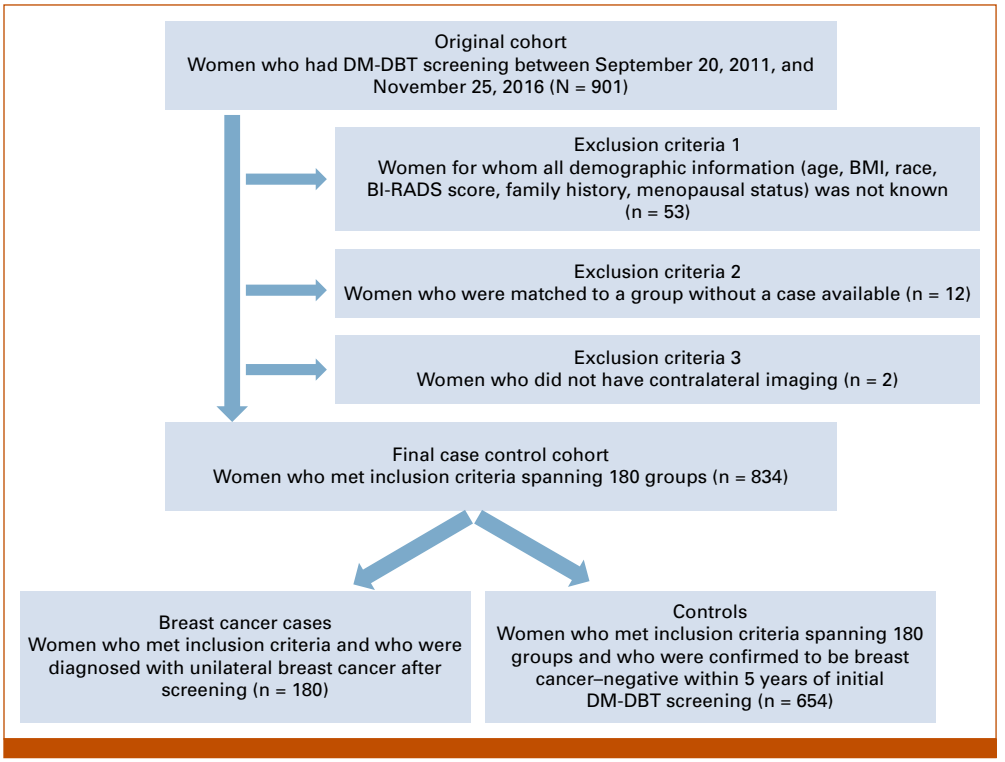

FIG A2. Inclusion and exclusion criteria for case-control sample selection. BI-RADS, Breast Imaging-Reporting and Data System; DM-DBT, digital mammography-digital breast tomosynthesis.

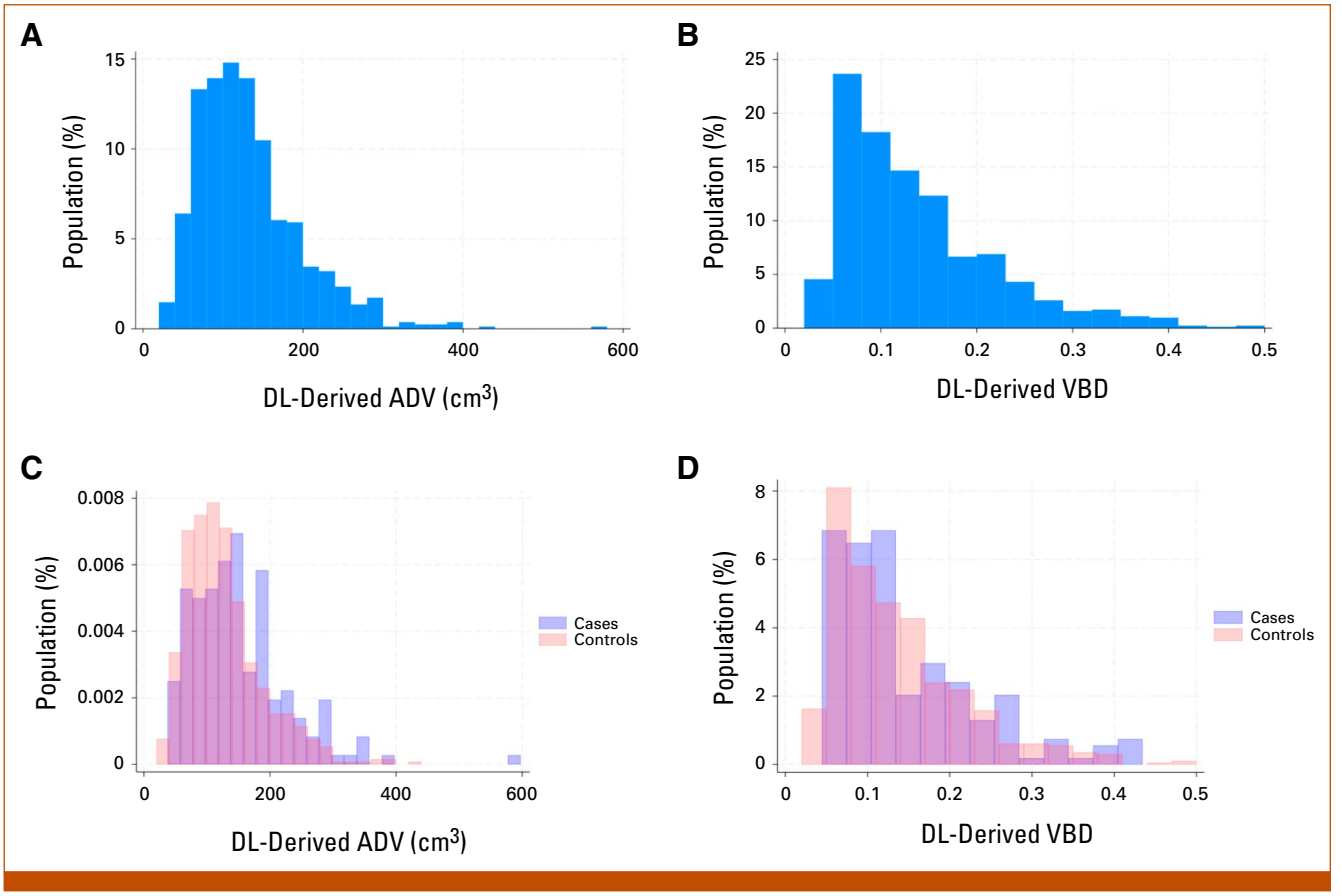

FIG A3. Distribution of DL-derived ADV and %VBD on the case-control data set. (A) ADV for the entire case-control cohort; (B) %VBD for the entire case-control cohort; (C) ADV for the entire case-control cohort with cases and controls separately overlaid; (D) %VBD for the entire case-control cohort with cases and controls separately overlaid. ADV, absolute dense volume; DL, deep learning; VBD, volumetric breast density.

TABLE A1. Results on the Held-Out Test Set Using Dice Score

	Black Race (n = 340)	White Race (n = 404)	Other Race (n = 36)	*P*
Unweighted Dice	0.88 (0.08)	0.89 (0.08)	0.88 (0.07)	.053
Weighted Dice	0.74 (0.15)	0.78 (0.15)	0.76 (0.16)	.004

NOTE. Dice and standard deviations in parentheses. *P* values are derived from one-way analysis of variance comparisons between self-identified racial groups after Bonferroni correction for multiple comparisons.

TABLE A2. AUCs for Conditional Logistic Regression Models When Trained on Only One Self-Identified Racial Group

Model	Case, No.	Control, No.	Self-Identified Race	AUC (95% CI)	*P*
DBT ADV	83	313	White	0.70 (0.63 to 0.76)	.255
	91	322	Black	0.64 (0.57 to 0.71)	
DBT %VBD	83	313	White	0.69 (0.63 to 0.76)	.171
	91	322	Black	0.63 (0.56 to 0.69)	

NOTE. *P* value reflects the difference between AUC of regression models fit on White versus Black participants. AUC was calculated using 1,000 bootstraps.
Abbreviations: ADV, absolute dense volume; DBT, digital breast tomosynthesis; VBD, volumetric breast density.

TABLE A3. Demographics and Other Characteristics of the Case-Control Sample Subset for Which Raw for-Processing 2D DM Images and 3D DBT Reconstructed Slices Were Available

Characteristic	Breast Cancer Group (n = 146)	Control Group (n = 639)	*P*
DL-derived %VBD	15.0 ± 9.0	13.7 ± 7.9	.10
DL-derived %ADV	158.7 ± 78.7	127.2 ± 61.6	<.0001
LIBRA-derived ABD	23.6 ± 12.2	20.3 ± 11.9	.01
LIBRA-derived ADA	13.9 ± 9.5	14.5 ± 9.5	.46
Age, years	60.2 ± 10.4	61.1 ± 10.8	.36
BMI (kg/m^2)	29.8 ± 7.4	28.7 ± 6.8	.13
Self-identified race, No.			.73
White	69	316	
Black	71	304	
Other Race or unknown	6	19	
Radiologist-determined BI-RADS density category, No.			.04
A	12	84	
B	75	370	
C	56	178	
D	3	7	

NOTE. *P* values for age and BMI refer to *t* tests with unequal variances; *P* values for race and BI-RADS were derived from chi-squared tests.
Abbreviations: ADA, absolute dense area; ADV, absolute dense volume; DBT, digital breast tomosynthesis; DL, deep learning; DM, digital mammography; VBD, volumetric breast density.

TABLE A4. Odds Ratios Specifying Log Odds of Breast Cancer per Unit Change in Standard Deviation for Conditional Logistic Regression Models Using %ABD, %VBD, ADA, and ADV on the Subset of the Case-Control Sample

Patient Parameter	Log Odds per Unit Change (95% CI)	Covariate *P* Value	AUC (95% CI)	AUC Comparison *P* Value
2D ABD-based model			0.61 (0.55 to 0.65)	.101
LIBRA-derived %ABD	1.23 (1.03 to 1.47)	.021		
Age	2.81 (0.71 to 11.11)	.139		
BMI	1.27 (1.04 to 1.55)	.022		
No. of relatives with breast cancer	1.15 (0.96 to 1.37)	.137		
Menopausal status	0.83 (0.60 to 1.14)	.250		
3D VBD-based model			0.63 (0.58 to 0.68)	
DL-derived %VBD	1.56 (1.21 to 2.02)	.001		
Age at initial screening	3.03 (0.77 to 11.95)	.114		
BMI	1.60 (1.26 to 2.04)	<.001		
No. of relatives with breast cancer	1.13 (0.95 to 1.36)	.175		
Menopausal status	0.88 (0.63 to 1.22)	.443		
2D ADA-based model			0.59 (0.53 to 0.64)	<.001
LIBRA-derived 2D ADA	1.17 (0.93 to 1.46)	.174		
Age at initial screening	2.91 (0.74 to 11.39)	.125		
BMI	1.35 (1.09 to 1.68)	.007		
No. of relatives with breast cancer	1.14 (0.96 to 1.37)	.137		
Menopausal status	0.83 (0.61 to 1.15)	.265		
3D ADV-based model			0.66 (0.61 to 0.71)	
DL-derived 3D ADV	1.63 (1.33 to 2.01)	<.001		
Age at initial screening	2.15 (0.78 to 12.64)	.106		
BMI	1.15 (0.93 to 1.43)	.187		
No. of relatives with breast cancer	1.15 (0.95 to 1.39)	.139		

NOTE. Only patients for whom 2D DM and 3D reconstructed DBT imaging were available were included in this subset analysis. AUC was calculated using 1,000 bootstraps. Covariate *P* value refers to *P* value reported from the fitting of the conditional logistic regression model. AUC comparison *P* value refers to the *P* value reported from comparing the AUC of the listed regression models.

Abbreviations: 2D, two-dimensional; 3D, three-dimensional; ADA, absolute dense area; ADV, absolute dense volume; DBT, digital breast tomosynthesis; DL, deep learning; DM, digital mammography; VBD, volumetric breast density.