# Supplementary Material

**A. Definitions of segmentation performance evaluation measures**

Given two sets, $y_t$ (ground-truth) and $y_p$ (prediction), the Dice coefficient is defined as:

$$Dice = \frac{2|y_t \cap y_p|}{|y_t| + |y_p|}$$

where |.| indicates the cardinality of a set. This measure is applied to the target class by ignoring the background class. We extend this definition to the weighted Dice coefficient, that considers both classes (target and background) and weights based on the number of pixels assigned to each class as follows:

$$Weighted\ Dice = \frac{2 \times W_1|y_{t1} \cap y_{p1}|}{(|y_{t1}| + |y_{p1}|) \times (W_1 + W_2)} + \frac{2 \times W_2|y_{t2} \cap y_{p2}|}{(|y_{t2}| + |y_{p2}|) \times (W_1 + W_2)}$$

where weights can be calculated by:

$$W_1 = \frac{|y_{t2}|}{|y_{t1}| + |y_{t2}|}; W_2 = \frac{|y_{t1}|}{|y_{t1}| + |y_{t2}|}$$

Sensitivity is defined as follows:

$$Sensitivity = \frac{TP}{FN + TP}$$

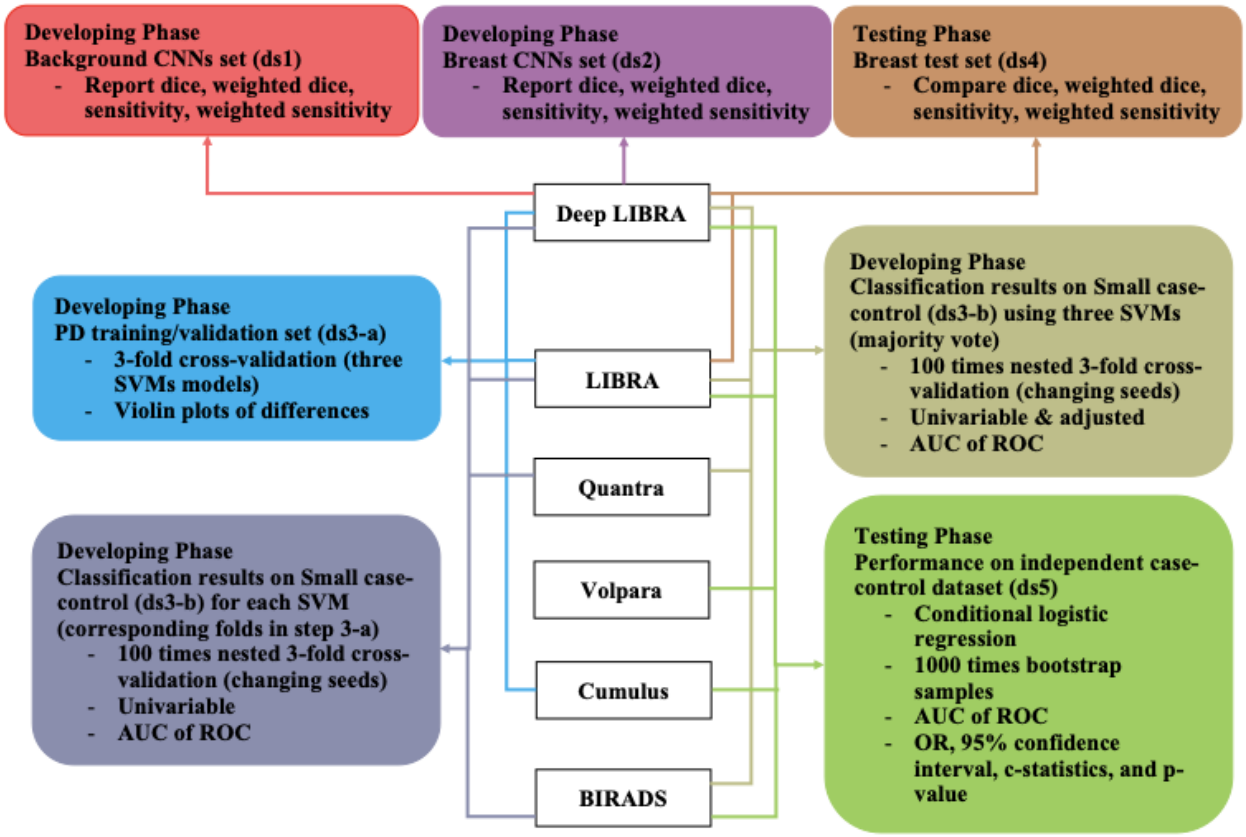
The terms TP and FN refer to the number of pixels which are correctly classified as target and the number of pixels which are incorrectly classified as background, respectively. The weighted sensitivity considers both classes (target and background) and weights based on the number of pixels assigned to each class as follows:

$$Weighted\ Sensitivity = \frac{W_1 \times TP_1}{(TP_1 + FP_1) \times (W_1 + W_2)} + \frac{W_2 \times TP_1}{(TP_2 + FP_2) \times (W_1 + W_2)}$$

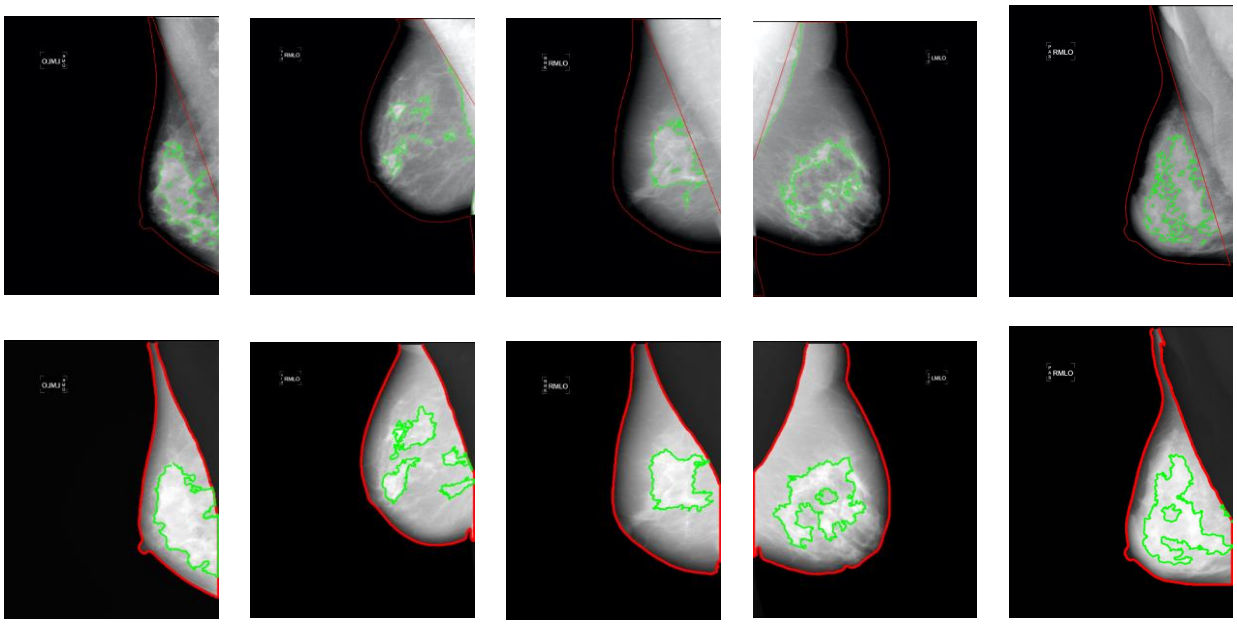
$$W_1 = \frac{|P_2|}{|P_1| + |P_2|}; W_2 = \frac{|P_1|}{|P_1| + |P_2|}$$

**B. Abdominal effect removal**

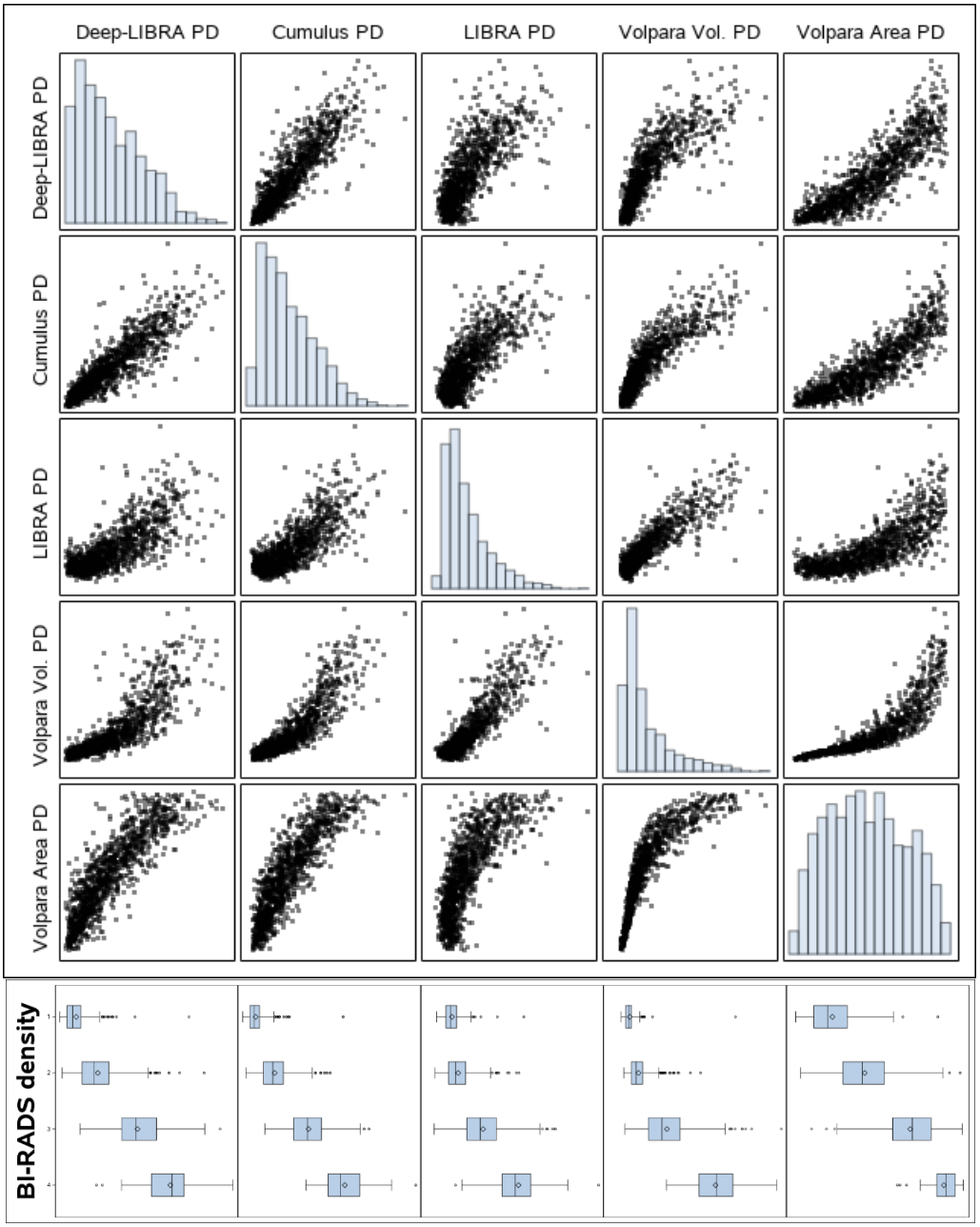
Large breasts often cover a large part of the FFDM image. A negative image intensity gradient in the horizontal direction indicates that the breast curve gets closer to the body (i.e., to the left side of the image when breast orientation is standardized). We find the farthest point in the horizontal direction $(X_{max}, Y_{max})$, where $X$ and $Y$ are the coordinates in the horizontal and vertical directions, respectively. Any pixels above $1.25 \times Y_{max}$ are removed to simplify the process to assure being 25% far from gradient changes around the breast nipple. Then, we find the contour for the remaining points and smooth the resulted coordinates signal. Finally, we check if the negative gradient in the horizontal direction changes to positive sign (higher than 0.1 to avoid noise affecting our results). The positive gradient in the horizontal direction indicates an abdominal bump. Accordingly, the points between the contour and the left side of image, considered as the abdominal region, are removed. If an image has a positive gradient in the vertical direction, this method is not applied. In other words, the condition to remove the abdominal part is to have a change in the gradient in the horizontal direction from negative to positive, while the gradient in the vertical direction remains negative. The reason for such a condition is to make sure that the breast was not larger than FFDM field of view.




**Supplementary Figure 1.** Evaluation experiments and datasets involved in the developing and testing phases of Deep-LIBRA. The middle boxes in white color correspond to different breast density scores that are available for our six study datasets. These scores are linked to evaluation experiments using the colored lines associated with each evaluation.

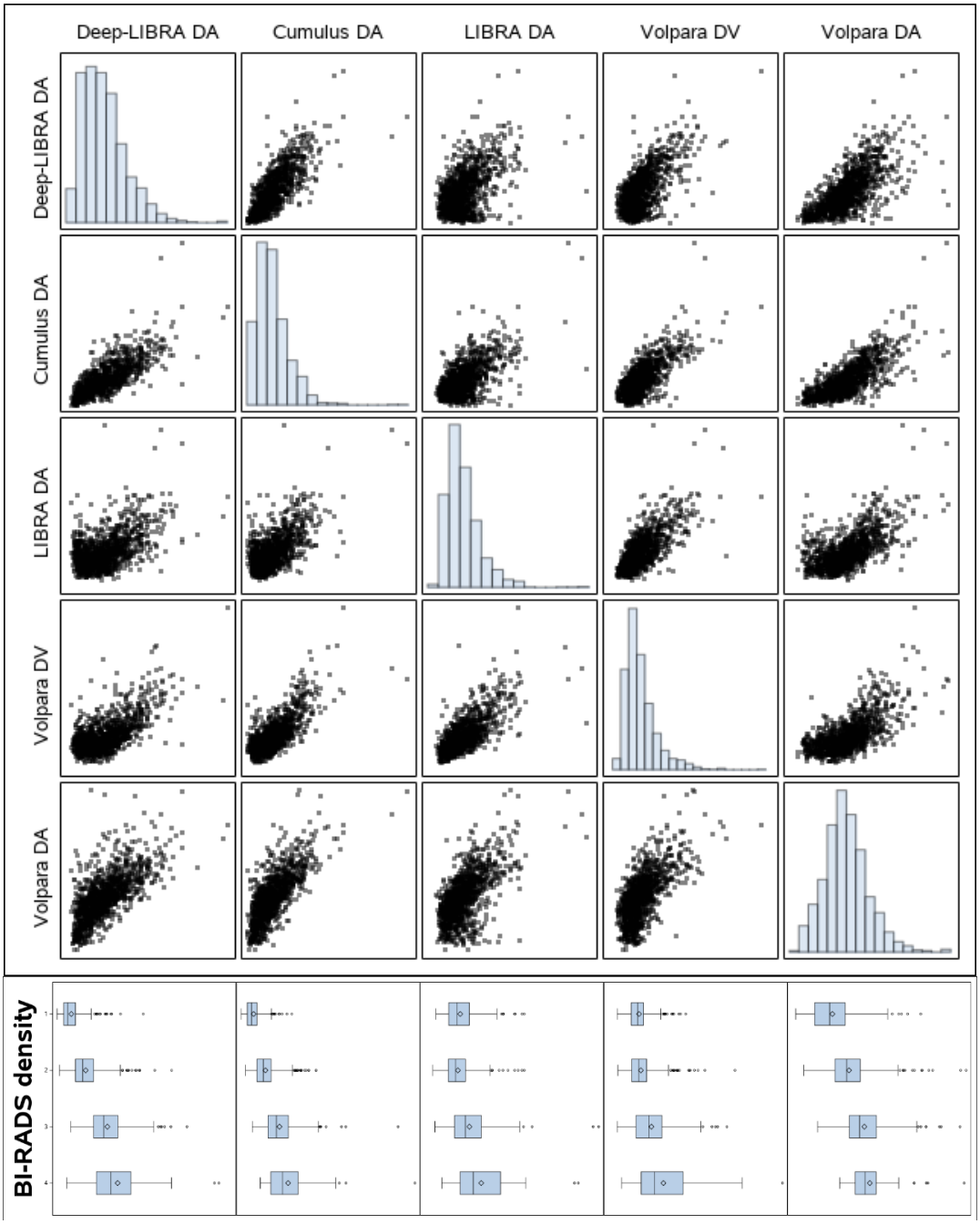


**Supplementary Figure 2.** Examples of breast segmentations by Deep-LIBRA (bottom row) and LIBRA (top row) on ds4. The red lines indicate the breast outline and the green lines are the segmented dense issue areas. LIBRA assumes a line for pectoralis muscle which often causes inaccurate segmentation of breast area. In addition, Deep-LIBRA can eliminate the abdominal region under the breast, as shown in the third and fourth columns.




| Density measure | Deep-LIBRA PD | Cumulus PD | LIBRA PD | Volumetric Volpara PD | Area Volpara PD | BI-RADS density |
|---|---|---|---|---|---|---|
| **Deep-LIBRA PD** | 1 | 0.90 | 0.76 | 0.89 | 0.89 | 0.80 |
| **Cumulus PD** | | 1 | 0.76 | 0.91 | 0.88 | 0.84 |
| **LIBRA PD** | | | 1 | 0.85 | 0.80 | 0.67 |
| **Volpara Volumetric PD** | | | | 1 | 0.96 | 0.80 |
| **Volpara Area PD** | | | | | 1 | 0.77 |
| **BI-RADS density** | | | | | | 1 |

**Supplementary Figure 3.** (Top row) Scatter plot of different continuous percent density (PD) measures against each other, (Second row) box-whisker plot of each the continuous PD measures by BI-RADS density category and (Bottom row) correlation coefficients between different density measures, on ds5. The figure includes controls only, and continuous PD measures have been averaged over the four breast views for each woman.




| Density measure | Deep-LIBRA DA | Cumulus DA | LIBRA DA | Volpara DV | Volpara DA | BI-RADS density |
|---|---|---|---|---|---|---|
| **Deep-LIBRA DA** | 1 | 0.79 | 0.44 | 0.52 | 0.71 | 0.71 |
| **Cumulus DA** | | 1 | 0.53 | 0.66 | 0.74 | 0.66 |
| **LIBRA DA** | | | 1 | 0.69 | 0.58 | 0.27 |
| **Volpara DV** | | | | 1 | 0.62 | 0.31 |
| **Volpara DA** | | | | | 1 | 0.52 |
| **BI-RADS density** | | | | | | 1 |

**Supplementary Figure 4.** (Top row) Scatter plot of continuous absolute density measures against each other, (Second row) box-whisker plot of each the continuous absolute density measures by BI-RADS density category and (Bottom row) correlation coefficients between different density measures, on ds5. The figure includes controls only, and continuous absolute density measures have been averaged over the four breast views for each woman. DA: Dense tissue area; DV: Dense tissue volume.

**Supplementary Table 1.** Breast cancer risk factors and demographic characteristics by case-control status for ds3-b. For age and BMI, data corresponds to mean and standard deviation. For race and BI-RADS density, data corresponds to number of women and percentage in parentheses. *P-values from two-sample t-tests for age and BMI, and from Pearson chi- squared tests for ethnicity and BI-RADS density.

| Variable | Breast cancer cases ($n$ = 115) | Controls ($n$ = 460) | p-value* |
|---|---|---|---|
| Age | 59.02 ± 11.7 | 56.7 ± 11.5 | 0.049 |
| BMI (kg/m$^2$) | 29.7 ± 6.9 | 29.5 ± 7.6 | 0.799 |
| Race | | | 0.20 |
| Caucasian/White | 54 (47%) | 216 (47%) | |
| African American/Black | 61 (53%) | 244 (53%) | |
| BI-RADS density | | | 0.075 |
| A | 9 (7.8%) | 54 (11.9%) | |
| B | 61 (53.0%) | 279 (60.7%) | |
| C | 38 (33.0%) | 123 (26.7%) | |
| D | 3 (2.6%) | 3 (0.7%) | |
| Missing | 4 (3.5%) | 1 (0.2%) | |

**Supplementary Table 2.** Breast cancer risk factors and demographic characteristics by case-control status for ds5. For continuous covariates, data corresponds to median values with interquartile ranges in brackets. For categorical covariates, data corresponds to number of women and percentage in parentheses. *P-values from Wilcoxon rank-sum tests for continuous covariates and from Pearson chi-squared tests for categorical covariates.

| Variable | Breast cancer cases ($n$ = 414) | Controls ($n$ = 1178) | *p-value |
|---|---|---|---|
| Age | 59.4 [50.4, 69.0] | 59.5 [35, 90] | 0.98 |
| BMI (kg/m$^2$) | 28.0 [24.2, 31.9] | 27.4 [24.1, 32.2] | 0.51 |
| Race | | | 0.001 |
| Caucasian/White | 390 (94.2%) | 1140 (96.8%) | |
| Other | 24 (5.8%) | 38 (3.2%) | |
| Years to breast cancer diagnosis for cases | 4.7 [4.1, 5.1] | n/a | n/a |
| Family history of breast cancer | 103 (25%) | 256 (22%) | 0.001 |
| Menopausal Status | | | 0.025 |
| Pre-menopausal | 100 (24%) | 335 (28%) | |
| Post-menopausal | 297 (72%) | 819 (70%) | |
| Unknown | 17 (4%) | 24 (2%) | |

**Supplementary Table 3.** Evaluation of breast segmentation performance of Deep-LIBRA on the independent testing dataset ds4. LIBRA, which also provides breast segmentation masks, is included in the evaluation. Data corresponds to mean and standard deviation of each evaluation measure.

| | **Deep-LIBRA** | **LIBRA** |
|---|---|---|
| **Dice** | 92.49 ± 2.37 | 83.37 ± 5.31 |
| **Weighted Dice** | 92.87 ± 2.49 | 84.71 ± 6.27 |
| **Sensitivity** | 92.19 ± 2.24 | 81.89 ± 5.76 |
| **Weighted Sensitivity** | 94.11 ± 1.76 | 82.93 ± 6.13 |

**Supplementary Table 4.** Associations of percent density (PD) measures with breast cancer and case-control discriminatory performance on ds5, using unadjusted logistic regression models.

| Density measure | OR (95% CI) | AUC (95% CI) |
|---|---|---|
| **Deep-LIBRA PD (4 views)** | 1.39 (1.22, 1.58) | 0.598 (0.571, 0.627) |
| **Deep-LIBRA PD (CC)** | 1.40 (1.23, 1.59) | 0.606 (0.578, 0.634) |
| **Deep-LIBRAPD (MLO)** | 1.33 (1.17, 1.52) | 0.573 (0.545, 0.602) |
| **Cumlus PD (CC)** | 1.36 (1.19, 1.54) | 0.600 (0.572, 0.628) |
| **LIBRA PD (4 views)** | 1.16 (1.03, 1.31) | 0.555 (0.527, 0.583) |
| **LIBRA PD (CC)** | 1.13 (1.00, 1.28) | 0.550 (0.521, 0.578) |
| **LIBRA PD (MLO)** | 1.16 (1.03, 1.31) | 0.552 (0.524, 0.581) |
| **Volumetric Volpara PD (4 views)** | 1.28 (1.13, 1.45) | 0.566 (0.538, 0.595) |
| **Volumetric Volpara PD (CC)** | 1.25 (1.10, 1.42) | 0.561 (0.533, 0.589) |
| **Volumetric Volpara PD (MLO)** | 1.30 (1.15, 1.48) | 0.571 (0.553, 0.600) |
| **Area Volpara PD (4 views)** | 1.26 (1.11, 1.43) | 0.573 (0.545, 0.602) |
| **Area Volpara PD (CC)** | 1.22 (1.08, 1.39) | 0.572 (0.543, 0.600) |
| **Area Volpara PD (MLO)** | 1.28 (1.12, 1.45) | 0.563 (0.535, 0.591) |
| **BI-RADS density** | 1.38 (1.20, 1.59) | 0.571 (0.543, 0.600) |

**Supplementary Table 5.** Associations of absolute density measures with breast cancer and case-control discriminatory performance on ds5, using logistic regression models adjusted for age and BMI. DA: Dense tissue area; DV: Dense tissue volume.

| Density measure | OR (95% CI) | AUC (95% CI) |
|---|---|---|
| **Deep-LIBRA DA (4 views)** | 1.62 (1.42, 1.86) | 0.642 (0.615, 0.669) |
| **Deep-LIBRA DA (CC)** | 1.64 (1.43, 1.87) | 0.640 (0.613, 0.668) |
| **Deep-LIBRA DA (MLO)** | 1.50 (1.32, 1.72) | 0.630 (0.602, 0.657) |
| **Cumlus DA (CC)** | 1.59 (1.40, 1.81) | 0.636 (0.609, 0.664) |
| **LIBRA DA (4 views)** | 1.34 (1.19, 1.51) | 0.580 (0.551, 0.608) |
| **LIBRA DA (CC)** | 1.27 (1.13, 1.44) | 0.573 (0.544, 0.601) |
| **LIBRA DA (MLO)** | 1.34 (1.19, 1.51) | 0.576 (0.548, 0.604) |
| **Volumetric Volpara DV (4 views)** | 1.64 (1.44, 1.88) | 0.627 (0.600, 0.655) |
| **Volumetric Volpara DV (CC)** | 1.57 (1.37, 1.79) | 0.600 (0.573, 0.629) |
| **Volumetric Volpara DV (MLO)** | 1.68 (1.47, 1.92) | 0.631 (0.604, 0.659) |
| **Area Volpara DA (4 views)** | 1.46 (1.30, 1.65) | 0.614 (0.586, 0.642) |
| **Area Volpara DA (CC)** | 1.38 (1.23, 1.55) | 0.594 (0.566, 0.622) |
| **Area Volpara DA (MLO)** | 1.48 (1.31, 1.67) | 0.625 (0.598, 0.653) |
| **BI-RADS density** | 1.54 (1.30, 1.81) | 0.596 (0.568, 0.624) |

**Supplementary Table 6.** Associations of absolute density measures with breast cancer and case-control discriminatory performance on ds5, using unadjusted logistic regression models. DA: Dense tissue area; DV: Dense tissue volume.

| Density measure | OR (95% CI) | AUC (95% CI) |
|---|---|---|
| **Deep-LIBRA DA (4 views)** | 1.56 (1.38, 1.78) | 0.638 (0.609, 0.664) |
| **Deep-LIBRA DA (CC)** | 1.58 (1.39, 1.80) | 0.621 (0.593, 0.649) |
| **Deep-LIBRA DA (MLO)** | 1.46 (1.29, 1.66) | 0.603 (0.575, 0.631) |
| **Cumlus DA (CC)** | 1.59 (1.40, 1.80) | 0.646 (0.619, 0.673) |
| **LIBRA DA (4 views)** | 1.34 (1.18, 1.50) | 0.588 (0.560, 0.616) |
| **LIBRA DA (CC)** | 1.27 (1.13, 1.43) | 0.575 (0.547, 0.603) |
| **LIBRA DA (MLO)** | 1.34 (1.19, 1.51) | 0.570 (0.542, 0.598) |
| **Volumetric Volpara DV (4 views)** | 1.60 (1.41, 1.82) | 0.603 (0.575, 0.631) |
| **Volumetric Volpara DV (CC)** | 1.53 (1.35, 1.73) | 0.602 (0.574, 0.630) |
| **Volumetric Volpara DV (MLO)** | 1.64 (1.44, 1.87) | 0.624 (0.597, 0.652) |
| **Area Volpara DA (4 views)** | 1.47 (1.30, 1.66) | 0.605 (0.577, 0.633) |
| **Area Volpara DA (CC)** | 1.39 (1.23, 1.56) | 0.591 (0.563, 0.619) |
| **Area Volpara DA (MLO)** | 1.47 (1.31, 1.66) | 0.619 (0.591, 0.645) |
| **BI-RADS density** | 1.38 (1.20, 1.59) | 0.571 (0.543, 0.600) |

**Supplementary Table 7.** Associations of percent density (PD) measures with breast cancer and case-control discriminatory performance on ds5, when density measures from different density estimation approaches are combined in logistic regression models adjusted for age and BMI.

| Density measure | OR (95% CI) | AUC (95% CI) |
|---|---|---|
| **Deep-LIBRA PD (4 views)** | 1.61 (1.37, 1.88) | 0.612 (0.584, 0.640) |
| **Deep-LIBRA PD (4 views)**<br>**Cumulus PD (CC)** | 1.32 (1.00, 1.73)<br>1.29 (0.97, 1.72) | 0.614 (0.586, 0.642) |
| **Deep-LIBRA PD (4 views)**<br>**LIBRA PD (4 views)** | 1.72 (1.40, 2.11)<br>0.91 (0.75, 1.10) | 0.616 (0.589, 0.644) |
| **Deep-LIBRA PD (4 views)**<br>**Volumetric Volpara PD (4 views)** | 1.52 (1.17, 1.96)<br>1.08 (0.82, 1.42) | 0.611 (0.583, 0.639) |
| **Deep-LIBRA PD (4 views)**<br>**Area Volpara PD (4 views)** | 1.69 (1.29, 2.20)<br>0.94 (0.71, 1.25) | 0.614 (0.586, 0.642) |
| **Deep-LIBRA PD (4 views)**<br>**BI-RADS density** | 1.45 (1.18, 1.79)<br>1.17 (0.94, 1.47) | 0.606 (0.579, 0.634) |

**Supplementary Table 8.** Associations of absolute density measures with breast cancer and case-control discriminatory performance on ds5, when density measures from different density estimation approaches are combined in logistic regression models adjusted for age and BMI. DA: Dense tissue area; DV: Dense tissue volume.

| Density measure | OR (95% CI) | AUC (95% CI) |
|---|---|---|
| **Deep-LIBRA DA (4 views)** | 1.62 (1.42, 1.86) | 0.642 (0.615, 0.669) |
| **Deep-LIBRA DA (4 views)**<br>**Cumulus DA (CC)** | 1.29 (1.05, 1.60)<br>1.32 (1.08, 1.61) | 0.635 (0.608, 0.663) |
| **Deep-LIBRA DA (4 views)**<br>**LIBRA DA (4 views)** | 1.55 (1.32, 1.82)<br>1.08 (0.94, 1.24) | 0.638 (0.610, 0.665) |
| **Deep-LIBRA DA (4 views)**<br>**Volumetric Volpara DV (4 views)** | 1.30 (1.08, 1.55)<br>1.39 (1.17, 1.65) | 0.641 (0.614, 0.669) |
| **Deep-LIBRA DA (4 views)**<br>**Area Volpara DA (4 views)** | 1.44 (1.20, 1.74)<br>1.17 (0.99, 1.38) | 0.645 (0.617, 0.672) |
| **Deep-LIBRA DA (4 views)**<br>**BI-RADS density** | 1.55 (1.31, 1.84)<br>1.10 (0.89, 1.36) | 0.640 (0.612, 0.667) |
| **Deep-LIBRA DA (4 views)**<br>**Deep-LIBRA PD (4 views)** | 1.84 (1.36, 2.47)<br>0.85 (0.60, 1.20) | 0.641 (0.614, 0.669) |

**List 1.** List of radiomic features used for dense versus fatty tissue classification. Features with the term "seg" are estimated at the superpixel-level and the rest are global features estimated at the image-level.

| | |
|---|---|
| LBP_avg_R1_P8 | RunLengthNonUniformity |
| LBP_std_R1_P8 | RunLengthNonUniformityNormalized |
| LBP_ske_R1_P8 | RunPercentage |
| LBP_kur_R1_P8 | RunVariance |
| LBP_avg_R3_P24 | ShortRunEmphasis |
| LBP_std_R3_P24 | ShortRunHighGrayLevelEmphasis |
| LBP_ske_R3_P24 | ShortRunLowGrayLevelEmphasis |
| LBP_kur_R3_P24 | GrayLevelNonUniformity.2 |
| 10Percentile | GrayLevelNonUniformityNormalized.1 |
| 90Percentile | GrayLevelVariance.2 |
| Energy | HighGrayLevelZoneEmphasis |
| Entropy | LargeAreaEmphasis |
| InterquartileRange | LargeAreaHighGrayLevelEmphasis |
| Kurtosis | LargeAreaLowGrayLevelEmphasis |
| Maximum | LowGrayLevelZoneEmphasis |
| MeanAbsoluteDeviation | SizeZoneNonUniformity |
| Mean | SizeZoneNonUniformityNormalized |
| Median | SmallAreaEmphasis |
| Minimum | SmallAreaHighGrayLevelEmphasis |
| Range | SmallAreaLowGrayLevelEmphasis |
| RobustMeanAbsoluteDeviation | ZoneEntropy |
| RootMeanSquared | ZonePercentage |
| Skewness | ZoneVariance |
| TotalEnergy | Seg_LBP_avg_R1_P8 |
| Uniformity | Seg_LBP_std_R1_P8 |
| Variance | Seg_LBP_ske_R1_P8 |
| Autocorrelation | Seg_LBP_kur_R1_P8 |
| ClusterProminence | Seg_LBP_avg_R3_P24 |
| ClusterShade | Seg_LBP_std_R3_P24 |
| ClusterTendency | Seg_LBP_ske_R3_P24 |
| Contrast | Seg_LBP_kur_R3_P24 |
| Correlation | Seg_10Percentile |
| DifferenceAverage | Seg_90Percentile |
| DifferenceEntropy | Seg_Energy |
| DifferenceVariance | Seg_Entropy |
| Id | Seg_InterquartileRange |
| Idm | Seg_Kurtosis |
| Idmn | Seg_Maximum |
| Idn | Seg_MeanAbsoluteDeviation |
| Imc1 | Seg_Mean |
| Imc2 | Seg_Median |

| | |
|---|---|
| InverseVariance | Seg_Minimum |
| JointAverage | Seg_Range |
| JointEnergy | Seg_RobustMeanAbsoluteDeviation |
| JointEntropy | Seg_RootMeanSquared |
| MCC | Seg_Skewness |
| MaximumProbability | Seg_TotalEnergy |
| SumAverage | Seg_Uniformity |
| SumEntropy | Seg_Variance |
| SumSquares | Seg_Autocorrelation |
| Busyness | Seg_ClusterProminence |
| Coarseness | Seg_ClusterShade |
| Complexity | Seg_ClusterTendency |
| Contrast.1 | Seg_Contrast |
| Strength | Seg_Correlation |
| DependenceEntropy | Seg_DifferenceAverage |
| DependenceNonUniformity | Seg_DifferenceEntropy |
| DependenceNonUniformityNormalized | Seg_DifferenceVariance |
| DependenceVariance | Seg_Id |
| GrayLevelNonUniformity | Seg_Idm |
| GrayLevelVariance | Seg_Idmn |
| HighGrayLevelEmphasis | Seg_Idn |
| LargeDependenceEmphasis | Seg_Imc1 |
| LargeDependenceHighGrayLevelEmphasis | Seg_Imc2 |
| LargeDependenceLowGrayLevelEmphasis | Seg_InverseVariance |
| LowGrayLevelEmphasis | Seg_JointAverage |
| SmallDependenceEmphasis | Seg_JointEnergy |
| SmallDependenceHighGrayLevelEmphasis | Seg_JointEntropy |
| SmallDependenceLowGrayLevelEmphasis | Seg_MCC |
| GrayLevelNonUniformity.1 | Seg_MaximumProbability |
| GrayLevelNonUniformityNormalized | Seg_SumAverage |
| GrayLevelVariance.1 | Seg_SumEntropy |
| HighGrayLevelRunEmphasis | Seg_SumSquares |
| LongRunEmphasis | |
| LongRunHighGrayLevelEmphasis | |
| LongRunLowGrayLevelEmphasis | |
| LowGrayLevelRunEmphasis | |
| RunEntropy | |