Journal of Translational Medicine

https://doi.org/10.1186/s12967-025-06844-1

Article in Press

3-D ordinal regression CNN for opportunistic breast density grading on chest CT

Received: 27 Feb 2025

Accepted: 08 Jul 2025

Published online: 23 July 2026

Cite this article as: Wysoczanski, A., Angelini, E., Jambawalikar, S. *et al.* 3-D ordinal regression CNN for opportunistic breast density grading on chest CT. *J Transl Med* (2026). https://doi.org/10.1186/s12967-025-06844-1

Artur Wysoczanski, Elsa Angelini, Sachin Jambawalikar, Mary Salvatore & Andrew Laine

3-D Ordinal Regression CNN for Opportunistic Breast Density Grading on Chest CT

Artur Wysoczanski[1], Elsa D. Angelini[1,2], Sachin R. Jambawalikar[1], Mary M. Salvatore[3] and Andrew F. Laine[1]

[1]Department of Biomedical Engineering, Columbia University, New York, NY, USA

[2]Telecom Paris LTCI, Institut Polytechnique de Paris, Paris, France

[3] Department of Radiology, Columbia University Irving Medical Center, New York, NY, USA

ABSTRACT

Background: Breast cancer has the highest incidence and second-highest mortality burden of all cancers in women. Breast density, the extent of fibroglandular tissue within the breast, is strongly associated with breast cancer risk, and is visually graded on a fourlevel ordinal scale on imaging studies including chest CT. However, current automated methods for CT breast density assessment are limited, relying on semi-automated breast segmentation followed by intensity thresholding, and excluding higher-order image features. In this work, we train and evaluate a 3-D convolutional neural net (3D-CNN) to predict breast density grade on CT.

Methods: IRB-approved chart review identified female patients from 2010-19 with both mammography and chest CT acquired within one year. CT breast density was visually assessed following the BI-RADS density criteria by a subspecialist faculty radiologist. CT scans were downsampled to isotropic 2.5 mm resolution, and volumes of interest (VOIs) of size 80x80x80 voxels were extracted for each imaged breast.

Our 3D-CNN model consists of consecutive 3-D convolutional layers with ReLU activation, convolutional block attention modules (CBAM), and batch normalization and pooling operations. Rank-consistent ordinal regression was applied to the final layer to predict the density grade of each VOI, with models optimized to minimize ordinal crossentropy. Patient-level predictions were generated on test cases by averaging over the right- and left-breast VOIs.

Results: 503 scans matched our inclusion criteria; 86 (17.1%) were assigned breast density grade 1 (i.e. mostly fatty), 244 (48.5%) were grade 2 (scattered dense), 131

(26.0%) were grade 3 (heterogeneously dense), and 42 (8.3%) were grade 4 (extremely dense). Our optimized 3D-CNN model achieved ROC AUC of 0.935, and Cohen kappa of 0.787, for classification of high (grades 3-4) versus low breast density, comparable to literature-reported inter-radiologist agreement. Both ordinal regression and attention mechanisms improved model performance over standard architectures. In test cases where both breasts were within the CT image, ROC AUC and Cohen kappa improved to 0.955 and 0.873, respectively.

Conclusions: We have developed a 3-D CNN model for prediction of breast density grade on chest CT which achieves test-set performance comparable to prior literature reports of inter-radiologist agreement on CT breast density.

BACKGROUND

Breast cancer is the single most common malignancy and the second-leading cause of cancer mortality in women (1). Breast density, defined as the extent of fibroglandular tissue in the breast, is an independent determinant of breast cancer risk (24). On digital mammography or volumetric imaging such as chest computed tomography (CT), the abundance of dense breast tissue is assessed semi-quantitatively on a fourgrade scale (Figure 1) following the BI-RADS criteria, from mostly fatty (grade 1) to completely fibrotic (grade 4) (5, 6).

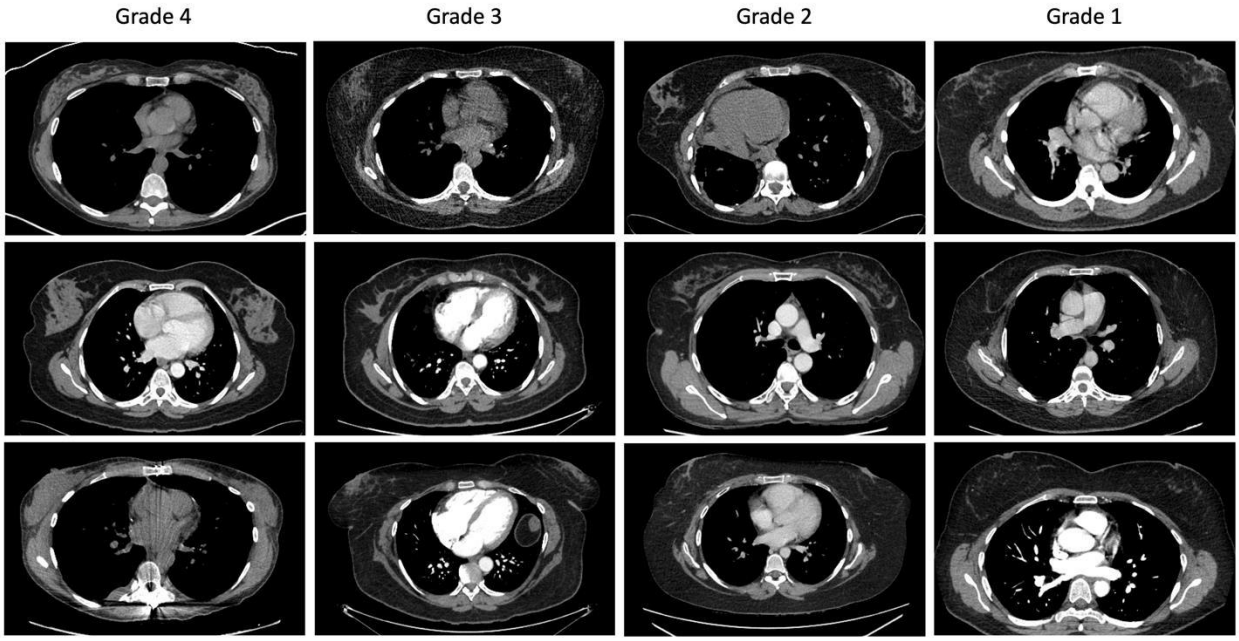

Figure 1: Representative axial slices of chest CT scans from a clinical dataset acquired at Columbia University Irving Medical Center, arranged in columns by breast density grade from highest (grade 4) to lowest (grade 1).

Density grades visually assigned by radiologists on chest CT and mammography show a high degree of concordance, measured by both intra-reader agreement in breast density grading between modalities and inter-reader agreement on either CT or mammography (7). Prior studies have shown that the highest visual breast density grade is associated with a several-fold increase in breast cancer risk compared to the lowest grade (2, 3); accordingly, although practice guidelines are evolving, a determination of high breast density (i.e., grade of 3 or 4) may trigger supplementary screening including ultrasound or MRI, in addition to lifestyle or pharmacologic intervention (8, 9). Due to its implications for both breast cancer risk and sensitivity of screening mammography for breast cancer (10), United States federal regulations require breast density to be

disclosed on mammographic reports (11). However, while visual breast density assessment is equally feasible and *more reproducible* on chest CT than on

mammography (7), it is not yet examined or reported as part of the standard workflow for that exam (8). While CT is not an element of routine breast cancer screening, over 13 million chest CTs are acquired annually in the United States (12), and therefore millions of CT volumes are acquired each year in patients eligible for routine mammographic screening, representing a significant opportunity for incidental density reporting and risk stratification in this population, especially among the subgroup of patients without a recent screening mammogram (13, 14).

Automated quantitative density grading on screening mammography is robust and widely deployed (15), but assessment on CT remains limited by comparison. Prior methods for breast density assessment on chest CT generally relied on semi-automated, rule-based (or deep-learned) methods for whole-breast segmentation, followed by binary segmentation of low-and high-density tissue regions (7, 16-20). Such methods, however, suffer from various drawbacks: density assessment is dependent on accurate segmentation or region-of-interest selection, which may or may not require manual quality control; they are furthermore unsuited for the frequent cases where the breast falls partially outside the scan field of view. In addition, quantitative density grades assigned to individual CT scans - based on the measured percentage of the breast occupied by high-density tissue and application of the BI-RADS-defined cutoffs (<25%, <50%, <75%, and >75% dense tissue) (5) – tend to significantly undercall breast density relative to clinicians reviewing the same scans, using the same grading criteria (7), suggesting that naïve percent-density approaches to grading may not fully explain the grades assigned in

clinical practice. A model for quantitative breast density grading on chest CT that is fully automated, robust to the image field of view, explainable, and *concordant with clinician-assigned density grades* would be a significant improvement on the current state of the art.

To this end, we have designed a 3-D convolutional neural network (CNN) model to predict the radiologist visually-assigned breast density grade, on both the BI-RADS four-grade and binary (high-vs.-low) scales, from a retrospective dataset of chest CT scans of patients imaged at Columbia University Irving Medical Center. Similar to the adaptation of CNN models for a wide variety of image classification tasks in diagnostic imaging (21), we propose to learn image features salient to the density visual grading task directly from the input image, without user pre-selection, nor requiring minimal preprocessing of the input CT scan. Our model is optimized for agreement with *radiologistassigned visual labels* as the evaluation gold-standard, rather than volumetric percent density, to directly learn the grading assigned in clinical practice without the use of percent density as a proxy. In this process, we will evaluate two embellishments on the baseline CNN architecture well-suited for the current task: (1) the conditional-rank-logit (CORAL) rank-consistent ordinal regression (CORN) layers and ordinal cross-entropy loss (22), which provide robust and mathematically-consistent frameworks for ordinal regression problems (like assigning semi-quantitative density grades), and (2) convolutional-block attention modules (CBAM) (23), which can be incorporated into CNNs to more robustly identify the image *features* and *regions* important to the training task, and qualitatively improve models' focus on salient image regions and the explainability of their predictions.

A fast, robust, and fully-automated estimator of CT breast density would be extremely valuable in both clinical and translational domains, facilitating the incorporation of breast density evaluation into standard chest CT workflow at scale, with greater ease of use and broader applicability than comparable methods in the existing literature (1620). Moreover, estimating breast density grade *independently* of threshold-based volumetric percent density is a direct precursor to AI capacities on chest CT scans to improve image-based breast cancer risk stratification.

METHODS

CT Dataset

A retrospective, IRB-approved chart review (Protocol #AAAU4972) identified female patients at Columbia University Irving Medical Center with records of a mammogram and chest CT acquired within one year of one another from January 1, 2010 through March 2019. N=707 patients (aged 30 to 92 years at the time of CT, with an average age of 65 years) matched the search criteria. N=306 of 707 cases had contrast-enhanced CT performed, 31 had undergone mastectomy prior to imaging, and 88 had pathologically confirmed primary breast malignancy. De-identified DICOM tags were searched to identify chest CT series with soft-tissue reconstruction, which yielded 614 scans for further analysis. Images were acquired with tube voltage of 120 kVp and tube current in the range [30, 800] mA, and manufacturer-specific soft-tissue reconstruction kernels (standard kernel (GE), one of (B20f, B26f, B30f, B30s, I30f, I31f) (Siemens), or one of (FC08, FC18) (Toshiba)). In-plane resolution was in the range [0.53, 1.02] mm, and slice thickness in the range [0.5, 5] mm.

All cases were subsequently reviewed by one faculty radiologist (M.M.S.) with fellowship training in both chest (20 years experience) and breast (5 years experience) imaging. CT breast density grades were assigned by visual assessment following the BIRADS criteria as described in previous publications (7, 8, 24), assigning scans on an ordinal scale to grade 1 (fatty, 0-25% dense parenchyma), 2 (scattered fibroglandular, 25-50% dense parenchyma), 3 (heterogeneously dense, 50-75% dense parenchyma), or 4 (extremely dense, 75-100% dense parenchyma). CT scans were additionally evaluated for the presence and laterality of mastectomy, implants or prostheses, and follow-up recommendations (interval imaging vs. biopsy) following the BI-RADS criteria were noted.

Image Pre-Processing

Volumes of interest (VOIs) were identified by isolating CT sub-volumes containing the imaged left and right breasts from each scan, in a fully automated fashion and without the use of any hard-coded image landmarks. Lung segmentations were first generated using a publicly available, pre-trained deep learning model (25), and each CT scan was cropped along the craniocaudal axis to the superior and inferior limits of the segmented lungs. Volumes were subsequently cropped in the axial plane to the bounding box of the body (identified by thresholding each axial slice at -200 Hounsfield units, followed by taking the convex hull of the largest connected component). Any scans where the segmented lung volume fell below 1L were presumed not to include the whole thorax and were excluded. The final dataset consisted of 503 patients who had a CT scan matching the selection criteria, and had no history of (1) mastectomy, (2) implants, or (3) primary breast malignancy (Figure 2).

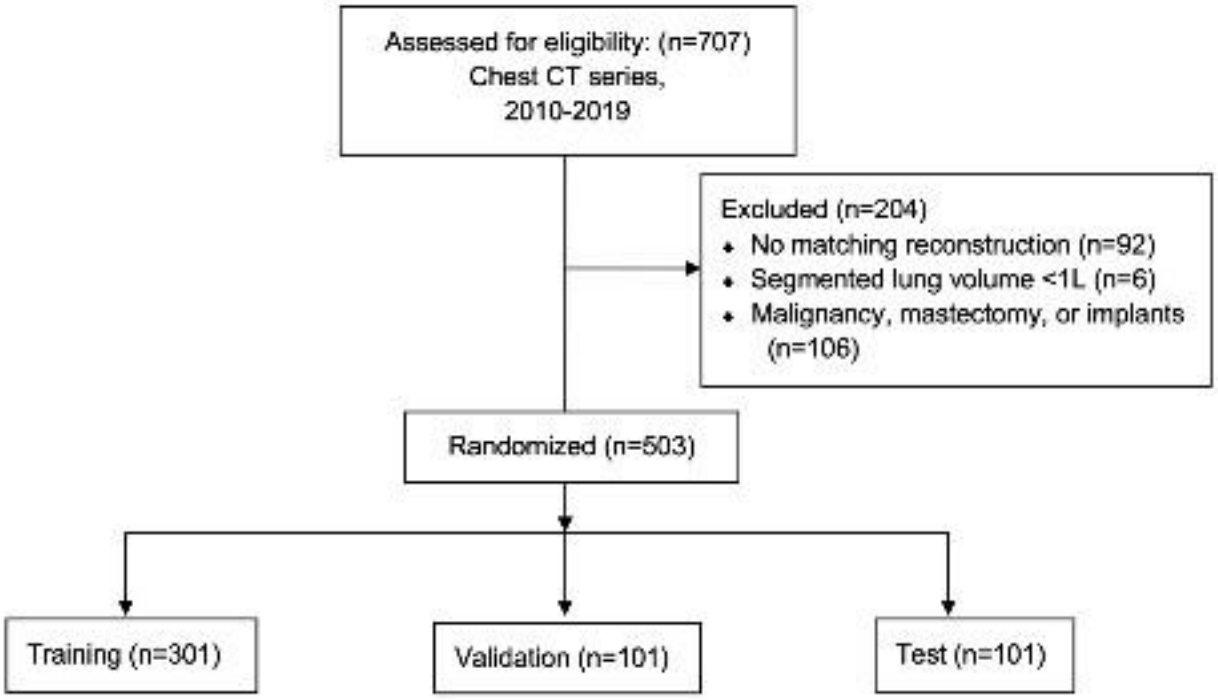

Figure 2: Flow chart detailing sample size, exclusion criteria, and randomization to training, validation and test sets.

The resulting 3D volumes were resampled to isotropic resolution of 2.5 mm using cubic interpolation, with image intensities clipped to the range [-1024, 1024] Hounsfield units (HU), and the 80 central axial slices were selected from the resampled volume. Two volumes of interest (VOIs) of size 80x80x80 voxels were then generated from each scan, originating from counting from the anterior-left and anterior-right corners of the CT 3D patches to contain the left and right breasts, respectively. Representative VOIs are depicted in Figure 3 below.

All VOIs were then randomly allocated at the participant level to training, validation, and test sets in a 3:1:1 ratio (N = 301, 101, 101 cases each for the training, validation, and test sets, respectively), with stratification by density grade.

For all participants in the training set, we also performed 1:1 offline augmentation for each input 3D patch by repeating the same pre-processing pipeline to generate VOIs while masking out intrathoracic structures from the image before resampling and VOI extraction, with the aim of aiding identification of relevant anatomy without requiring any further pre-processing during inference. For each axial slice in a CT scan, the convex hull of the lung mask was computed and voxel intensity values in the union of the convex hull from that slice and all slices superior to it were set to -1024 HU before continuing with the VOI extraction pipeline as above. Model validation- and test-set performance was assessed on CT volumes without masking. All VOIs were rescaled from the intensity range [-400, +400] HU to [0,1] to facilitate model training.

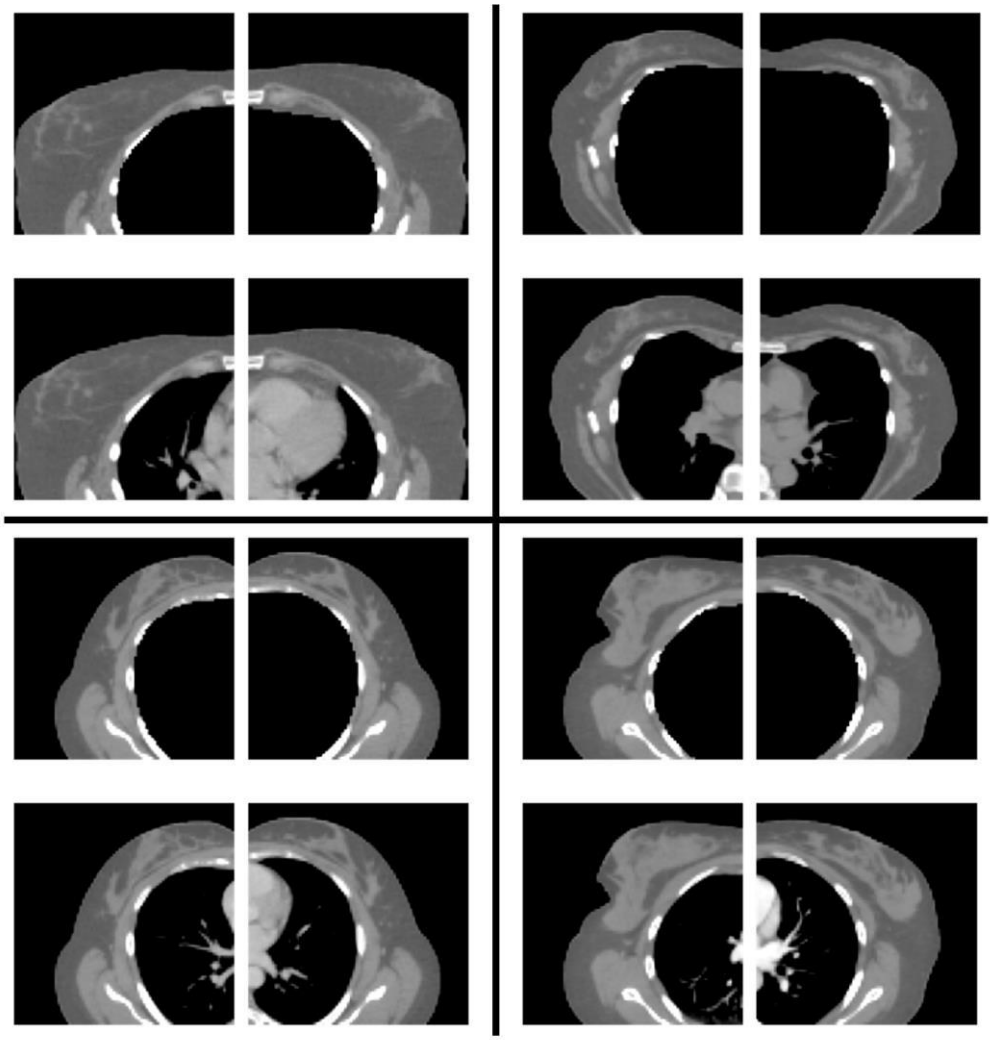

Figure 3: Axial slices from four representative CT scans after 80x80x80 voxel VOI selection and post-processing. Each panel contains the VOIs for the left and right breast at the same axial level, with (top row) and without (bottom row) masking of the thoracic cavity.

Network Architecture and Training

Our network (Figure 4) consists of five blocks each composed of a 3-D convolutional layer with ReLU activation and batch normalization; twofold max-pool downsampling follows the first four convolution blocks; the fifth is followed by a global average pooling (GAP)

operation and three fully-connected layers. The design of the final fully-connected layer is dependent on the objective function, with two output nodes and softmax activation in the case of binary classification, four output nodes and softmax activation for four-class classification, and CORAL or CORN output layers for ordinal regression models. Each model included either a CBAM block, a spatial attention block, or neither between each batch-normalization and downsampling layer. All pairs of objective functions (binary, categorical, CORAL and CORN) and attention mechanisms (spatial attention, CBAM, or none) were tested individually.

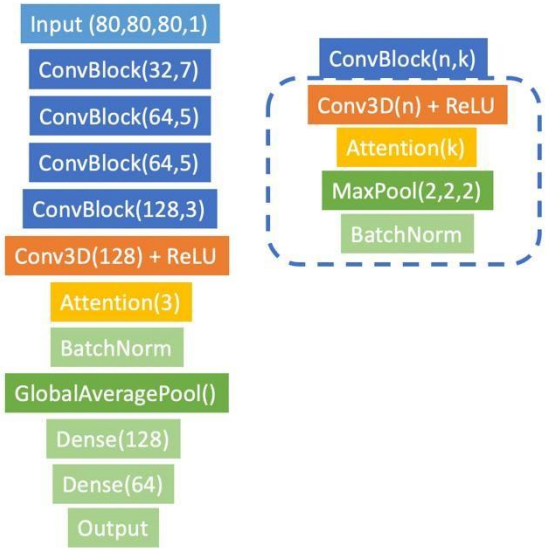

Figure 4: Architecture of the proposed network. Networks consist of repeated convolutional blocks (dashed component, on the right), followed by a global average pooling operation and three fully-connected layers. Output layer design is dependent on the objective function used (binary, categorical, CORAL or CORN cross-entropy); parameters *n* and *k* define the size of the convolution kernels in the 3-D convolutional layer and spatial attention block, respectively. Variants for each network were trained with a CBAM attention block, with a spatial attention block alone, or neither.

Breast density is assessed on a four-grade scale, and density assignment can be approached as a multi-class classification task, or as a binary (grades 1 or 2 vs. grades 3 or 4) to simplify to the most clinically-relevant thresholds. However, given the natural ordering of the density grades, this is fundamentally an ordinal regression task, and we hypothesize that an ordinal objective function which reflects this relationship between grades may show superior performance to a standard classifier. To this end, in addition to a binary and multi-class classifier, we investigate the performance of ordinal regression networks using the COnsistent RAnk Logits (CORAL) (22) and rank-Consistent Ordinal RegressioN (CORN) (26) objective functions. The CORAL framework ensures rankconsistent prediction on ordinal regression tasks, by reformulating ordinal regression as a nested series of binary classification tasks; CORN operates on a similar principle, offering greater network capacity by estimating conditional probabilities rather than class membership at each step along the ordinal scale.

The Convolutional Block Attention Module (CBAM) (23) incorporates two complementary attention modules: one addressing inter-channel relationships (“channelbased” attention), facilitating adaptive recalibration of feature maps across distinct channels, and the other attending to intra-channel dependencies, capturing pertinent spatial information within each channel. We hypothesized that since visual features relevant to breast density grade are restricted to a relatively small region of the VOI, CBAM and its spatial attention component in particular may significantly improve network performance over that of a baseline CNN model.

All models were trained for up to 1,000 epochs with early stopping conditioned on validation cross-entropy loss. An epoch was defined as thirty training steps; each training

batch consisted of 20 samples balanced across density grades, with five examples selected randomly without replacement for each of the four grades. Train-time image augmentation was performed using the Volumentations package (27), with left-right flipping, rotation about all axes by ± 15 degrees, and additive Gaussian noise with variance sampled uniformly from the interval $[0,10^{-3}]$, each applied with 50% probability. Each model was trained using the ADAM optimizer with an initial learning rate of 10^{-4} and L_2 regularization on all kernel weights. Model optimization was performed with objective functions specific to the training task, i.e. standard cross-entropy for the binary and categorical models, and ordinal cross-entropy for the CORAL and CORN frameworks.

Evaluation

Test-time accuracy is evaluated on individual VOIs, and also at the patient-level by averaging the output predictions over the two breasts. Test-set accuracy, receiveroperating characteristic area under the curve (ROC AUC) and Cohen's kappa were computed for all models on prediction of low vs. high density; the decision threshold for the binary task was optimized for the test-set Cohen statistic. For the categorical, CORAL and CORN models, which accommodated a four-class grading scheme, we additionally perform ROC analysis for all three density cutoffs (i.e. density grade > 1, grade > 2, and grade > 3). Additionally, we qualitatively evaluated the visual saliency of image regions used to perform the classification inference using class activation maps (28).

RESULTS

Descriptive Statistics

Of 503 scans which met the criteria for VOI selection, 86 (17.1%) were assigned breast density grade 1 (i.e. mostly fatty), 244 (48.5%) were assigned grade 2 (i.e. scattered dense), 131 (26.0%) were assigned grade 3 (i.e. heterogeneously dense, and 42 (8.3%) were assigned grade 4 (i.e. extremely dense). In 134 (26.6%) scans, one or both breasts were partially out of the scanner field of view, with the ground-truth values assessed by the reviewing clinician on the basis of the imaged tissue only.

Model Performance

Table 1 : Test-set (N=101) ROC AUC, Cohen's kappa, and overall accuracy for trained models using different density grading loss functions (rows) and attention mechanisms (columns). All models are evaluated on the binary prediction task (high vs. low breast density). Metrics reflect performance on the class probabilities *averaged* over the left and right breasts for each participant. Best configuration is highlighted in bold.

Metric	Model	Attention Mechanism		
		None	Spatial	CBAM
AUC	Binary	0.900	0.896	**0.903**
	Categorical	0.901	0.887	**0.911**
	CORAL	0.878	0.880	**0.935**
	CORN	0.858	0.888	**0.903**
□	Binary	**0.681**	0.672	0.664
	Categorical	**0.727**	0.661	**0.727**
	CORAL	0.626	0.730	**0.787**
	CORN	0.599	0.718	**0.681**
Accuracy	Binary	**0.861**	**0.861**	0.841
	Categorical	**0.881**	0.851	**0.881**
	CORAL	0.832	0.881	**0.901**
	CORN	0.832	**0.871**	0.861

Model performance across grading methods and attention mechanisms are summarized in Table 1 above. Across the three primary performance metrics – ROC AUC, Cohen's kappa, and overall accuracy – we observe that incorporation of channel-wise or spatial attention mechanisms improve test-set performance over models with no attention modules. Among models incorporating CBAM into the network architecture, we note that the ordinal regression frameworks – CORAL and CORN – outperform both binary (lowvs. high-density) or multi-class classification.

The top-performing model configuration (CORAL/CBAM) achieved an ROC AUC of 0.935 on the binary classification task, with tuned Cohen kappa of 0.787 and overall accuracy of 90.1%. Inter-rater agreement between the deep-learned predictions and radiologist-assigned ground truth is comparable to between-radiologist reproducibility as reported in the prior literature on CT breast density: the kappa statistic achieved on the test dataset is equivalent to the reported inter-rater agreement between breast radiologists evaluating chest CT (0.79) (7), and is competitive with the agreement between general radiologists and expert-assigned labels (0.61-0.88) (24).

Confusion-matrix analysis of test-set performance of the four-class density grading problem (Figure 5) indicates that at the tuned threshold, 68 of 101 scans in the test set (67.3%) were assigned the correct breast density grade on the four-level BI-RADS scale; 99 of 101 (98.0%) of test cases were assigned the correct density grade or off by only one grade. t-SNE visualization of the latent-space feature distribution of training examples (Figure 6) shows the input images arranged on a continuum from low to high density, with appreciable separation in feature space by ground-truth density grade.

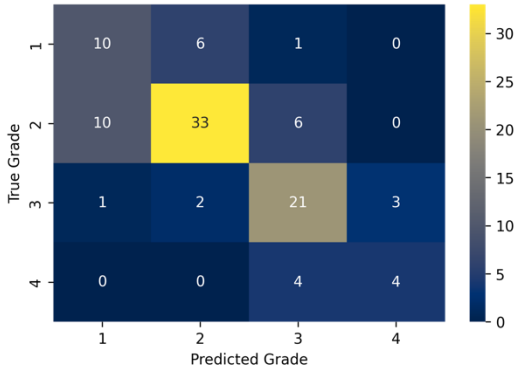

Figure 5: Test-set confusion matrix for the top-performing model (CORAL/CBAM) on the four-class density grading problem.

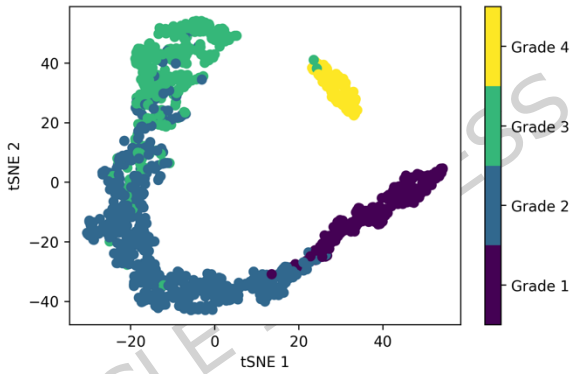

Figure 6: t-SNE projection of the bottleneck feature representation of training examples in the top-performing model (CORAL/CBAM), color-coded by density grade.

Patient-level breast density estimates showed modestly improved ROC AUC over the per-VOI predictions for all four density grades (Figure 7, top row), including an increase in AUC from 0.901 to 0.935 on the binary task. For both the VOI-level and patient-level prediction, model performance for predicting high density (grades 3 and 4) or extremely high density (grade 4) outperformed prediction of lowest density (grade 1). Model performance where both breasts were fully within the scan field of view (N=80 of 101 cases) was generally higher than on the full test set: ROC AUC for prediction of density grade greater than one improved from 0.822 to 0.870, while ROC AUC for

prediction of density grade greater than two improved from 0.935 to 0.955 (Figure 7, bottom), while the associated Cohen kappa statistic improved from 0.787 to 0.873.

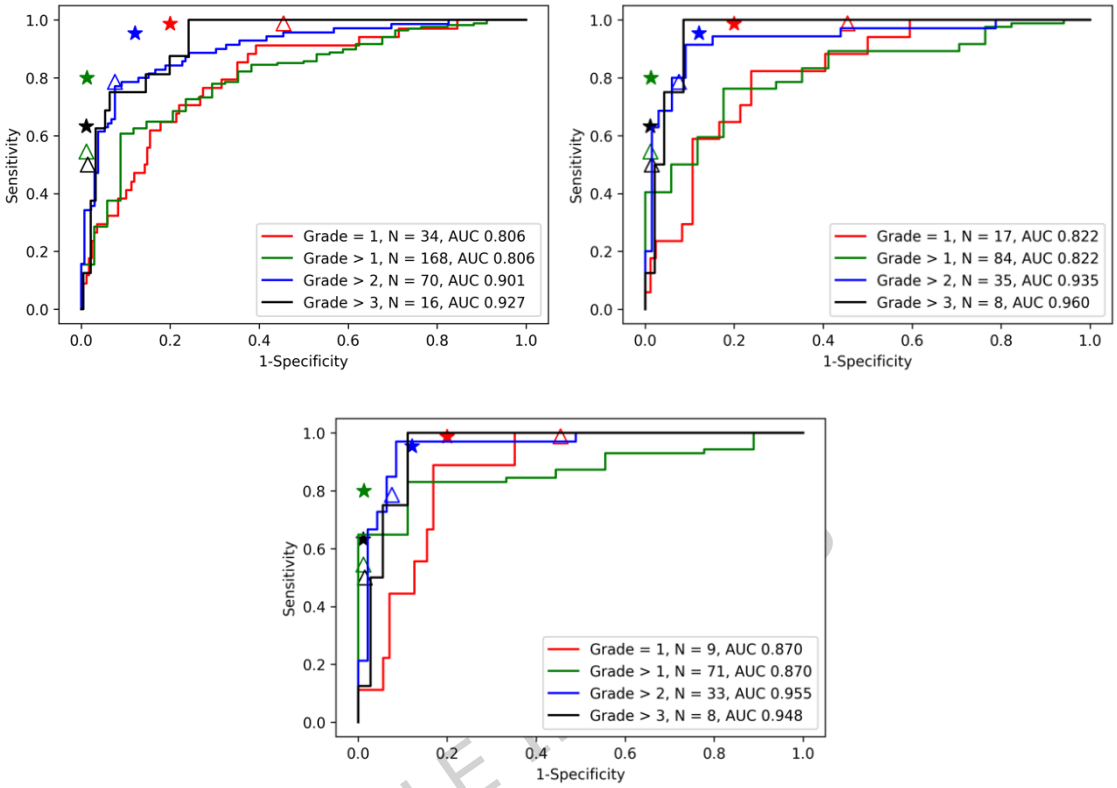

Figure 7: ROC curves for the top-performing model (CORAL/CBAM), per density grade threshold. Stars and triangles indicate equivalent inter-rater performance among radiologists on CT and mammography, respectively, from previous reports (7). Top left: ROC curves for predictions obtained on the left and right breasts individually; top right: ROC curves for the *ensemble* prediction averaging the class probabilities for the left and right breast; bottom: ROC curve for the *ensemble prediction* including only cases where the breasts were fully within the scan field of view.

Grad-CAM, or gradient-weighted class activation mapping, is a visualization technique which identifies the image regions that most strongly contribute to CNN predictions. Briefly, for a given input image and output class (e.g., high breast density), the gradients

of the output class score with respect to the activations of a chosen convolutional layer are averaged across the image, as a measure of the relative importance of each layer to the final prediction. The sum of the feature maps, weighted by their averaged gradients, are combined to produce a heatmap, with higher magnitude indicating greater saliency of that image region to the final network output. Visualization of representative test-set VOIs with correct binary density prediction (Figure 8) demonstrates the saliency of different image regions with respect to *low* and *high* breast density, at the level of the third and the fifth convolutional blocks. We note that the activation map on low-density cases focuses on low-attenuation tissue and the air-tissue boundary in the deepest layer, and that at intermediate depth the model discriminates between low- and high-density tissue within the breast. In high-density cases (grades 3 and 4), high-attenuation breast tissue is identified at shallow layers and isolated with greater specificity at the final convolutional block. Some spurious regions are highlighted in the activation maps for both low- and high-density cases, including the heart, scapula, and spine.

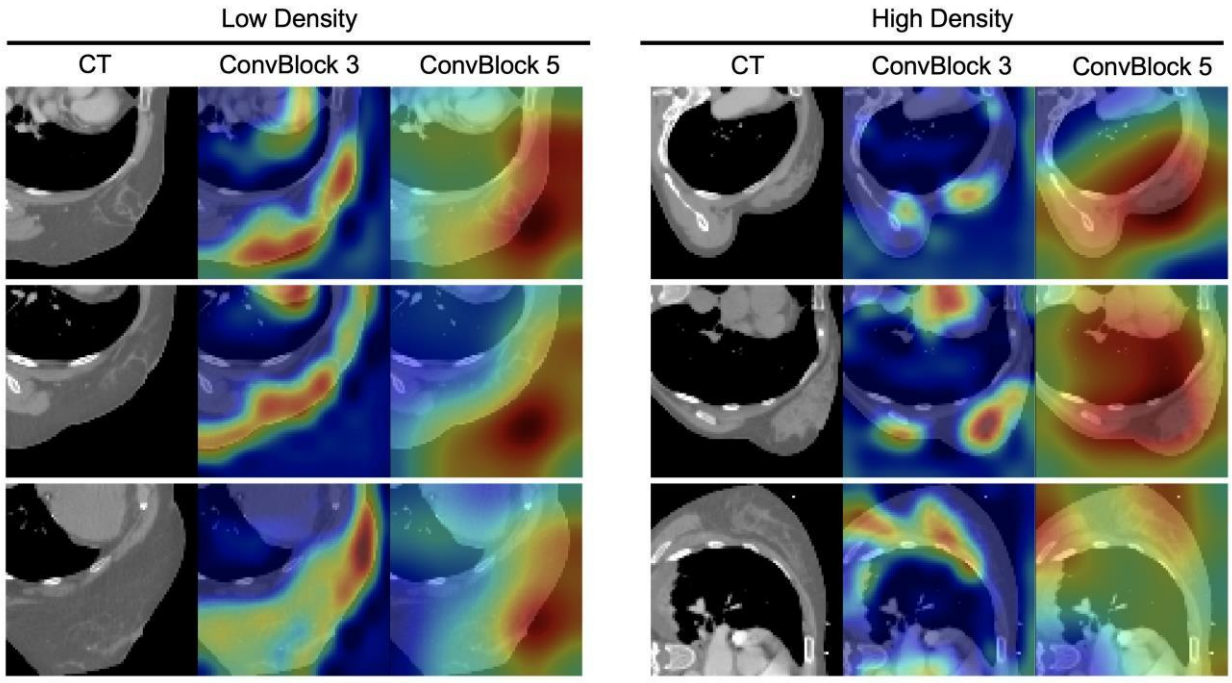

Figure 8: Grad-CAM saliency mapping for low (left) and high (right) density classes at the level of the third and fifth convolutional blocks, for representative CT images.

DISCUSSION

In this study, we have introduced a fully automated, end-to-end model for assessing breast density grade on chest CT scans, which is independent of breast segmentation or tissue intensity thresholding, and which demonstrates concordance with the clinician-assigned ground truth which is equivalent to the inter-rater agreement between board-certified breast radiologists. On the primary task of identifying cases with high breast density (grades 3 and 4), our top-performing model attains an ROC AUC of 0.935, Cohen kappa of 0.787, and overall accuracy exceeding 90%. Our work further suggests that an ordinal regression framework, encoding the intrinsic relationship between density grades, improves learning from radiologist-assigned labels over a naïve classification scheme; similarly, incorporation of channel-wise and spatial attention mechanisms improves density assessment from the input VOI.

Our analysis is robust to volumes of interest encompassing the breast and adjacent tissues, both with and without contrast, and tolerates severalfold downsampling (to 2.5mm isotropic voxel size) from the native resolution of the CT scans. Latent representation of input volumes demonstrates separation of samples in a low-dimensional subspace ordered by density grade, and saliency mapping using grad-CAM suggests that the model is able to identify relevant high- and low-density breast regions in the absence of explicit segmentation. Unlike approaches reliant on volumetric percent density, our algorithm avoids time-intensive segmentation quality control, and predicts density grades directly

from learned image features correlated with the radiologist-assigned labels, rather than from a single associated measure. Preprocessing and density grade prediction are automated and utilize existing freely-available Python packages, allowing for prospective application of the pipeline to future datasets.

It is of interest to note that the model is considerably better at distinguishing between high-density tissue grades than between grades 1 and 2 (e.g., ROC AUC 0.935 for predicting density grade >2 vs. 0.822 for predicting density grade >1); such an effect is not observed among clinicians in prior studies, although a direct comparison of interrater reproducibility is not available on this dataset. Similarly, performance on scans where the full breast is present still exceeds performance on scans with only partial visualization of breast tissue. Promising directions for further optimization include finetuning VOI pre-processing, augmentation, and model attention mechanisms to further refine network activation (Figure 8) and improve robustness to both of these cases.

The diversity of scanner models and acquisition protocols in the current dataset also present potential challenges. CT scans used were acquired with and without contrast, slice thickness varied by an order of magnitude, scans were acquired across three scanner manufacturers, with varying reconstruction kernels, and there is severalfold variability in the total mAs exposure, leading to significant variation in both image sharpness and signal-to-noise ratio.

The current work suggests multiple future avenues for further validation of our model. Since high breast density is a strong risk factor for breast cancer, we would expect a robust deep-learned estimator of breast density to preserve the association between (predicted) density grade and risk of breast cancer over multi-year follow-up. A

quantitative comparison of breast cancer risk stratification between deep-learned and radiologist-assigned breast density will be a valuable confirmatory experiment; the test set used in this work is underpowered for such an analysis due to its small sample size (N=101) relative to the incidence rate of breast cancer in the general population (estimated at approximately 130 per 100,000 women per annum) (29, 30). Future analyses in a larger cohort may similarly examine model robustness to repeat imaging. While breast density notably changes with factors including age, weight, and hormone exposure (31), a cohort of patients who have short-term follow-up CT available may provide a sample with minimal longitudinal change in the underlying breast density, and serve as a measure of intra-rater reproducibility of the CNN model.

Chest CT allows for the opportunistic screening of breast tissue as well as for density grading in mammography-eligible patient populations, particularly where screening participation varies widely due to factors including resource limitations and barriers to care (32, 33). Beyond direct applications of our model for density grading following current clinical practice, our work also serves as a proof-of-concept for a translational deep-learning framework to leverage chest CT to improve breast cancer risk stratification and identify deep-learned image markers associated with elevated cancer risk. Breast density is a strong predictor of prospective risk, but density alone does not account for all cases of breast cancer: women with low breast density might develop breast cancer, and most women with high breast density never will. Genetic, environmental, and hormonal factors all contribute to lifetime breast cancer risk (8); however, there is growing evidence that image features *beyond* volumetric percent density are also informative of cancer risk:

deep-learned image analysis breast cancer risk stratification on screening mammography is an area of active research (34-41).

Chest CT, in contrast to mammography, is a *quantitative, three-dimensional* modality, which demonstrably provides a more reproducible assessment of breast density (7) and retrospectively has shown greater sensitivity and specificity for breast cancer detection than mammography (42); furthermore, quantitative features of CT-resolved breast tissue may be markers for tissue changes mechanistically linked to cancer onset (43). We hypothesize that higher-order, CT-resolved features of breast tissue morphology, including texture, quantitative radiodensity and spatial distribution, may be predictive of the risk for and location of incident breast cancer. The architecture and training objective presented in this work are amenable to prediction of density as a proxy for breast cancer risk, or to estimation of incident cancer on multi-year follow-up, directly from the input image; we hypothesize that related architectures may achieve effective risk stratification from chest CT, with performance equal or superior to that reported in the existing literature on screening mammograms.

CONCLUSION

In this work, we have designed, trained, and validated a 3-D CNN model for automated ordinal regression of breast density on chest CT, using a clinical CT dataset acquired at Columbia University Irving Medical Center, *without* relying on explicit segmentation of the breast or on intensity thresholding coupled with volumetric percent density. Our results demonstrate that predictive performance is improved both by the adoption of an ordinal regression objective over binary or multi-class classification, and by the adoption of

channel-based and spatial attention mechanisms. On the identification of CT scans with high breast density, the model performance with respect to the radiologist-assigned ground truth is equivalent to the literature-reported inter-rater reproducibility of a sample of trained radiologists.

DECLARATIONS

Ethics approval and consent to participate: This retrospective study was approved by the IRB at Columbia University (IRB No. AAAU4972).

Consent for publication: Not applicable

Availability of data and materials: The data that support the findings of this study are available from the corresponding author upon reasonable request.

Competing interests: Mary M. Salvatore—Grants from Genentech and Boehringer Ingelheim. Consulting for AbbVie, Bioclinica, LungLife AI. Speaking engagements for PeerView and the France Foundation. The remaining authors have no conflicts to disclose.

Funding: Not applicable

Authors' contributions: AW, EA, and AFL contributed to design and analysis of the ordinal regression model. MS interpreted CT data. AW, EA, SJ, MS contributed to manuscript writing. All authors read and approved the final manuscript.

Acknowledgements: Not applicable

Authors' information: Not applicable

REFERENCES

1. Siegel RL, Miller KD, Wagle NS, Jemal A. Cancer statistics, 2023. CA Cancer J Clin. 2023;73(1):17-48.
2. Boyd NF, Dite GS, Stone J, Gunasekara A, English DR, McCredie MR, et al. Heritability of mammographic density, a risk factor for breast cancer. N Engl J Med. 2002;347(12):886-94.
3. Boyd NF, Guo H, Martin LJ, Sun L, Stone J, Fishell E, et al. Mammographic density and the risk and detection of breast cancer. N Engl J Med. 2007;356(3):227-36.
4. Vachon CM, Brandt KR, Ghosh K, Scott CG, Maloney SD, Carston MJ, et al. Mammographic breast density as a general marker of breast cancer risk. Cancer Epidemiol Biomarkers Prev. 2007;16(1):43-9.
5. Radiology ACo. Breast imaging reporting and data system. BI-RADS. 2003.
6. Radiology ACo. Breast imaging reporting and data system. BI-RADS. 2013.
7. Salvatore M, Margolies L, Kale M, Wisnivesky J, Kotkin S, Henschke CI, et al. Breast density: comparison of chest CT with mammography. Radiology. 2014;270(1):67-73.
8. Beizavi Z, Desperito E, Capaccione KM, Patrizio R, Salvatore MM. Reporting breast density on chest computed tomography. Translational Breast Cancer Research. 2023;4.
9. Mann RM, Athanasiou A, Baltzer PAT, Camps-Herrero J, Clauser P, Fallenberg EM, et al. Breast cancer screening in women with extremely dense breasts recommendations of the European Society of Breast Imaging (EUSOBI). Eur Radiol. 2022;32(6):4036-45.
10. Lynge E, Vejborg I, Andersen Z, von Euler-Chelpin M, Napolitano G. Mammographic Density and Screening Sensitivity, Breast Cancer Incidence and Associated Risk Factors in Danish Breast Cancer Screening. J Clin Med. 2019;8(11).
11. FDA Updates Mammography Regulations to Require Reporting of Breast Density Information and Enhance Facility Oversight. US Food and Drug Administration [press release]. 2023.
12. Mahesh M, Ansari AJ, Mettler FA, Jr. Patient Exposure from Radiologic and Nuclear Medicine Procedures in the United States and Worldwide: 2009-2018. Radiology. 2023;307(1):e221263.
13. Bandi P, Star J, Mazzitelli N, Nargis N, Islami F, Siegel RL, et al. Prevalence and Review of Major Modifiable Cancer Risk Factors, HPV Vaccination, and Cancer Screenings in the United States: 2025 Update. Cancer Epidemiology, Biomarkers & Prevention. 2025:OF1-OF14.
14. Sprague BL, Nowak SA, Ahern TP, Herschorn SD, Kaufman PA, Odde C, et al. Long-term mammography screening trends and predictors of return to screening after the COVID-19 pandemic: results from a statewide registry. Radiology: Imaging Cancer. 2024;6(3):e230161.
15. Lamb LR, Lehman CD, Gastounioti A, Conant EF, Bahl M. Artificial Intelligence (AI) for Screening Mammography, From the AJR Special Series on AI Applications. AJR Am J Roentgenol. 2022;219(3):369-80.

16. Chen JH, Chan S, Lu NH, Li Y, Tsai YC, Huang PY, et al. Opportunistic Breast Density Assessment in Women Receiving Low-dose Chest Computed Tomography Screening. Acad Radiol. 2016;23(9):1154-61.
17. Liu S, Margolies L, Xie Y, Yankelevitz D, Henschke C, Reeves A. Fully automated breast density assessment from low-dose chest CT: SPIE; 2017.
18. Moon WK, Lo CM, Goo JM, Bae MS, Chang JM, Huang CS, et al. Quantitative analysis for breast density estimation in low dose chest CT scans. J Med Syst. 2014;38(3):21.
19. Qureshi TA, Veeraraghavan H, Sung JS, Kaplan JB, Flynn J, Tonorezos ES, et al. Automated Breast Density Measurements From Chest Computed Tomography Scans. J Med Syst. 2019;43(8):242.
20. Zhou X, Kano T, Koyasu H, Li S, Zhou X, Hara T, et al. Automated assessment of breast tissue density in non-contrast 3D CT images without image segmentation based on a deep CNN: SPIE; 2017.
21. Litjens G, Kooi T, Bejnordi BE, Setio AAA, Ciompi F, Ghafoorian M, et al. A survey on deep learning in medical image analysis. Med Image Anal. 2017;42:60-88. 22. Cao W, Mirjalili V, Raschka S. Rank consistent ordinal regression for neural networks with application to age estimation. Pattern Recognition Letters. 2020;140:32531.
23. Woo S, Park J, Lee J-Y, Kweon IS. CBAM: Convolutional Block Attention Module2018 July 01, 2018:[arXiv:1807.06521 p.]. Available from: https://ui.adsabs.harvard.edu/abs/2018arXiv180706521W.
24. Margolies L, Salvatore M, Eber C, Jacobi A, Lee IJ, Liang M, et al. The general radiologist's role in breast cancer risk assessment: breast density measurement on chest CT. Clin Imaging. 2015;39(6):979-82.
25. Hofmanninger J, Prayer F, Pan J, Rohrich S, Prosch H, Langs G. Automatic lung segmentation in routine imaging is primarily a data diversity problem, not a methodology problem. Eur Radiol Exp. 2020;4(1):50.
26. Shi X, Cao W, Raschka S. Deep Neural Networks for Rank-Consistent Ordinal Regression Based On Conditional Probabilities2021 November 01, 2021:[arXiv:2111.08851 p.]. Available from: https://ui.adsabs.harvard.edu/abs/2021arXiv211108851S.
27. Solovyev R, Kalinin AA, Gabruseva T. 3D convolutional neural networks for stalled brain capillary detection. Comput Biol Med. 2022;141:105089.
28. Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D. Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization2016 October 01, 2016:[arXiv:1610.02391 p.]. Available from: https://ui.adsabs.harvard.edu/abs/2016arXiv161002391S.
29. SEER*Explorer: An interactive website for SEER cancer statistics: Surveillance Research Program, National Cancer Institute; [Available from: https://seer.cancer.gov/statistics-network/explorer/.
30. Giaquinto AN, Sung H, Newman LA, Freedman RA, Smith RA, Star J, et al. Breast cancer statistics 2024. CA: a cancer journal for clinicians. 2024;74(6):477-95.
31. Mokhtary A, Karakatsanis A, Valachis A. Mammographic density changes over time and breast cancer risk: a systematic review and meta-analysis. Cancers.

2021;13(19):4805.
32. Newman LA. Breast cancer screening in low and middle-income countries. Best Practice & Research Clinical Obstetrics & Gynaecology. 2022;83:15-23.
33. Funaro K, Niell B. Screening mammography utilization in the United States. Journal of Breast Imaging. 2023;5(4):384-92.
34. Arefan D, Mohamed AA, Berg WA, Zuley ML, Sumkin JH, Wu S. Deep learning modeling using normal mammograms for predicting breast cancer risk. Med Phys.
2020;47(1):110-8.
35. Gastounioti A, Desai S, Ahluwalia VS, Conant EF, Kontos D. Artificial intelligence in mammographic phenotyping of breast cancer risk: a narrative review. Breast Cancer Res. 2022;24(1):14.
36. Ha R, Chang P, Karcich J, Mutasa S, Pascual Van Sant E, Liu MZ, et al. Convolutional Neural Network Based Breast Cancer Risk Stratification Using a Mammographic Dataset. Acad Radiol. 2019;26(4):544-9.
37. Kontos D, Winham SJ, Oustimov A, Pantalone L, Hsieh MK, Gastounioti A, et al. Radiomic Phenotypes of Mammographic Parenchymal Complexity: Toward Augmenting Breast Density in Breast Cancer Risk Assessment. Radiology. 2019;290(1):41-9.
38. Warner ET, Rice MS, Zeleznik OA, Fowler EE, Murthy D, Vachon CM, et al. Automated percent mammographic density, mammographic texture variation, and risk of breast cancer: a nested case-control study. NPJ Breast Cancer. 2021;7(1):68. 39. Yala A, Lehman C, Schuster T, Portnoi T, Barzilay R. A Deep Learning Mammography-based Model for Improved Breast Cancer Risk Prediction. Radiology.
2019;292(1):60-6.
40. Yala A, Mikhael PG, Strand F, Lin G, Smith K, Wan YL, et al. Toward robust mammography-based models for breast cancer risk. Sci Transl Med. 2021;13(578).
41. Zhu X, Wolfgruber TK, Leong L, Jensen M, Scott C, Winham S, et al. Deep Learning Predicts Interval and Screening-detected Cancer from Screening Mammograms: A Case-Case-Control Study in 6369 Women. Radiology.
2021;301(3):550-8.
42. Desperito E, Schwartz L, Capaccione KM, Collins BT, Jamabawalikar S, Peng B, et al. Chest CT for Breast Cancer Diagnosis. Life (Basel). 2022;12(11).
43. Tarchi SM, Pernia Marin M, Hossain MM, Salvatore M. Breast stiffness, a risk factor for cancer and the role of radiology for diagnosis. J Transl Med. 2023;21(1):582.