Deep Neural Networks Improve Radiologists’ Performance in Breast Cancer Screening（201903）

Author： Nan Wu , Jason Phang, Jungkyu Park, Yiqiu Shen, Zhe Huang

李衡

20190524

1 Background

- In 2014： over **39 million** screening and diagnostic mammography exams
- In 2015 ： 232,000 women were diagnosed with breast cancer
- approximately 40,000 died from it

- The vast majority of the 10–15% of women asked to return to exams.
- 10–20% are recommended to undergo a needle biopsy
- 20–40% yield a diagnosis of cancer

- Traditional computer-aided detection (CAD) :

  →assist with image interpretation

  →do not improve their diagnostic performance

- Deep convolutional networks (DCN):

  →without investigating the fundamental differences between medical and natural images

2 DATA

- includes 229,426 digital screening mammography exams

  1,001,093 images

  141,473 patients

- Four images:

  R-CC (right craniocaudal), L-CC (left craniocaudal),

  R-MLO (right mediolateral oblique) and L-MLO (left mediolateral oblique).

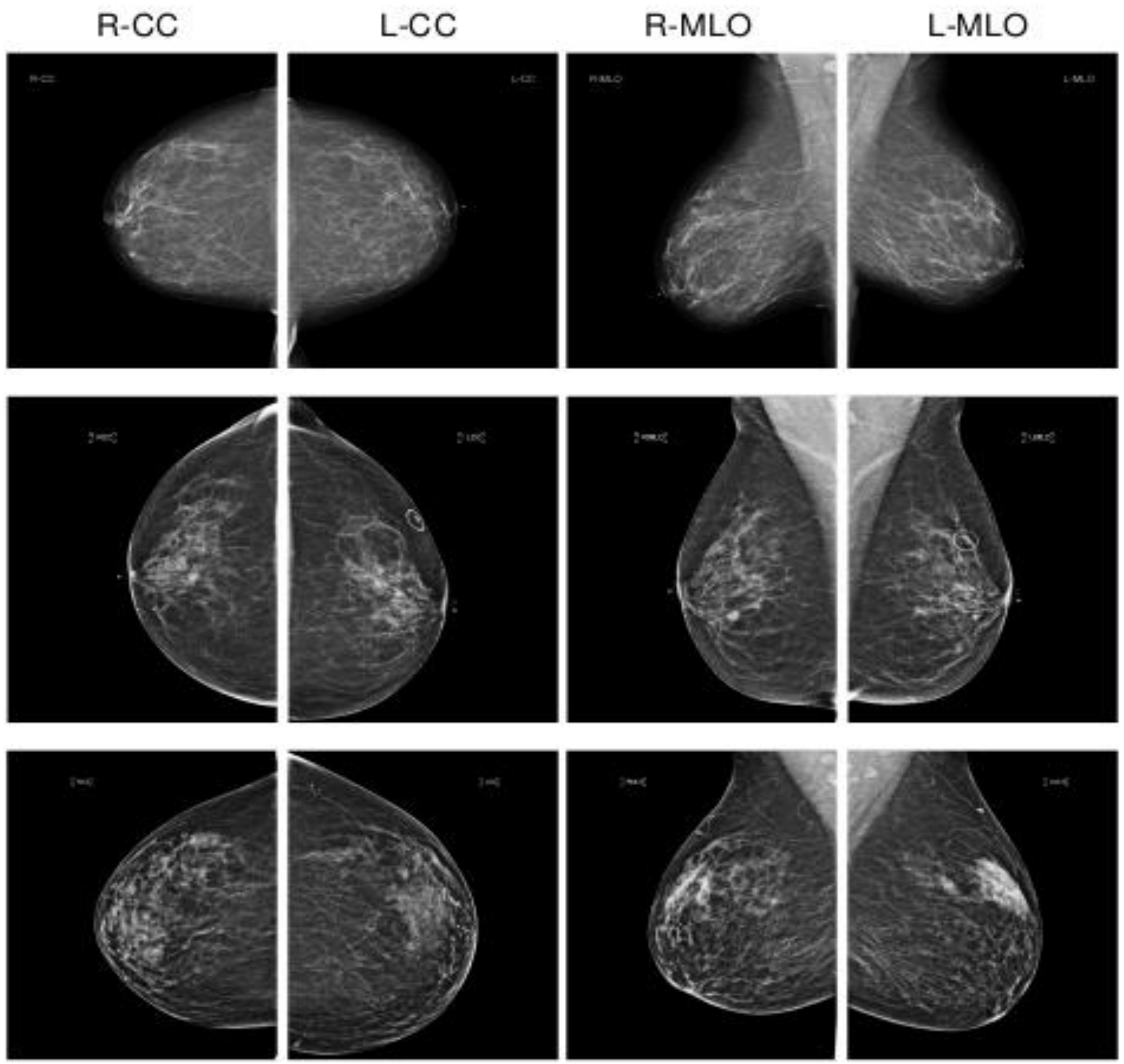


Fig. 1. Examples of breast cancer screening exams. First row: both breasts without any findings; second row: left breast with no findings and right breast with a malignant finding; third row: left breast with a benign finding and right breast with no findings.

3 MODEL

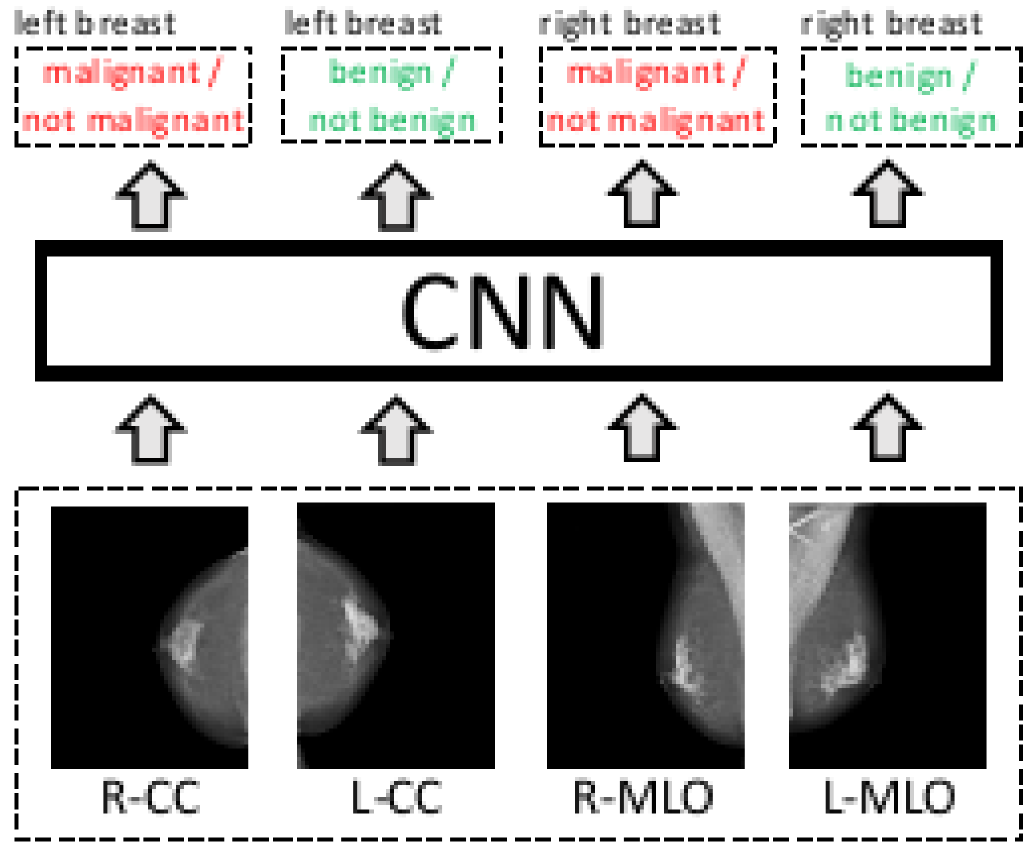


Fig. 3. A schematic representation of how we formulated breast cancer exam classification as a learning task.

3 MODEL

- deep multi-view CNN of architecture

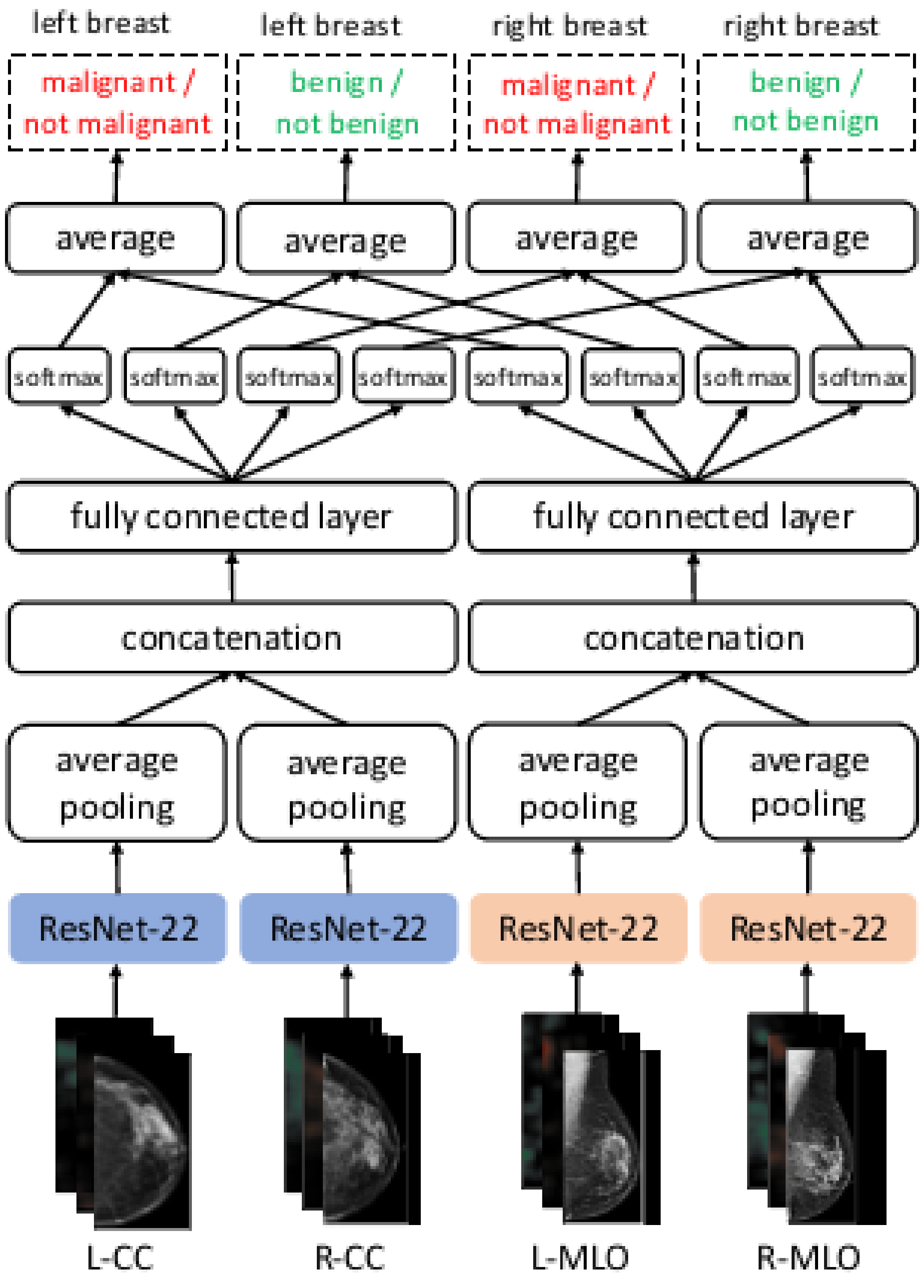
  (i) four view-specific columns:

  →each based on the ResNet architecture

  →outputs a fixed-dimension hidden representation for each mammography view,

  (ii) two fully connected layers :

  →map the computed hidden representations to the output predictions.



left breast
left breast
right breast
right breast
malignant / not malignant
benign / not benign
malignant / not malignant
benign / not benign
average
average
average
average
softmax
softmax
softmax
softmax
softmax
softmax
softmax
softmax
fully connected layer
fully connected layer
concatenation
concatenation
average pooling
average pooling
average pooling
average pooling
ResNet-22
ResNet-22
ResNet-22
ResNet-22
L-CC
R-CC
L-MLO
R-MLO

Auxiliary network

- Auxiliary patch-level classification :

→classify 256 × 256-pixel patches of mammograms

→the presence or absence of malignant and benign findings

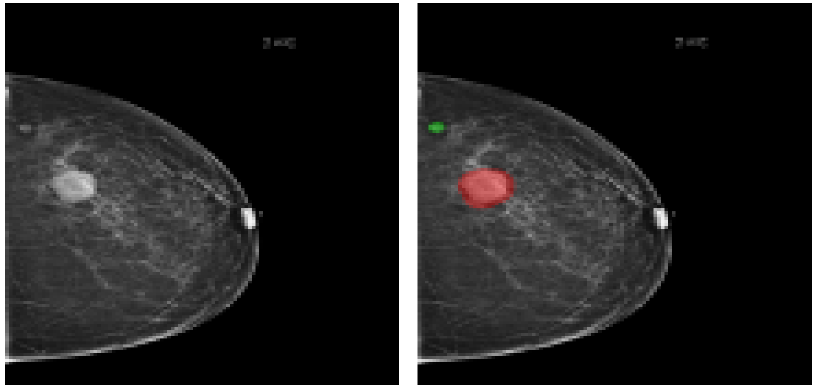


Fig. 2. An example of a segmentation performed by a radiologist. Left: the original image. Right: the image with lesions requiring a biopsy highlighted. The malignant finding is highlighted with red and benign finding with green.

- apply auxiliary network to the full to create two heatmaps

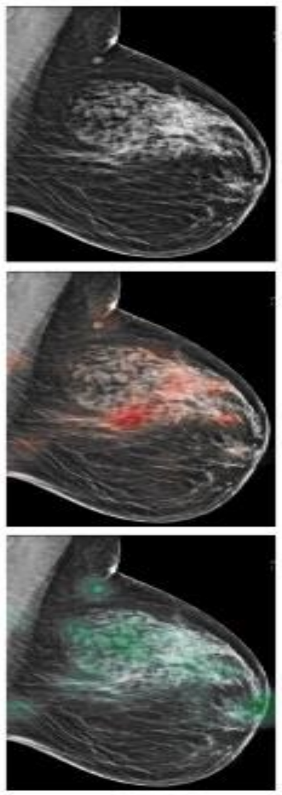


Fig. 5. The original image, the 'malignant' heatmap over the image and the 'benign' heatmap over the image.

→ containing an estimated probability of a malignant finding for each pixel

4 Experiments

- 1 Screening population:
- 2 Biopsied subpopulation.

Table 1. AUCs of our models on screening and biopsied populations.

	single		5x ensemble	
	malignant	benign	malignant	benign
screening population				
image-only	0.827±0.008	0.731±0.004	0.840	0.743
image-and-heatmaps	**0.886±0.003**	**0.747±0.002**	**0.895**	**0.756**
biopsied subpopulation				
image-only	0.781±0.006	0.673±0.003	0.791	0.682
image-and-heatmap	**0.843±0.004**	**0.690±0.002**	**0.850**	**0.696**

- Results across ages and breast densities.

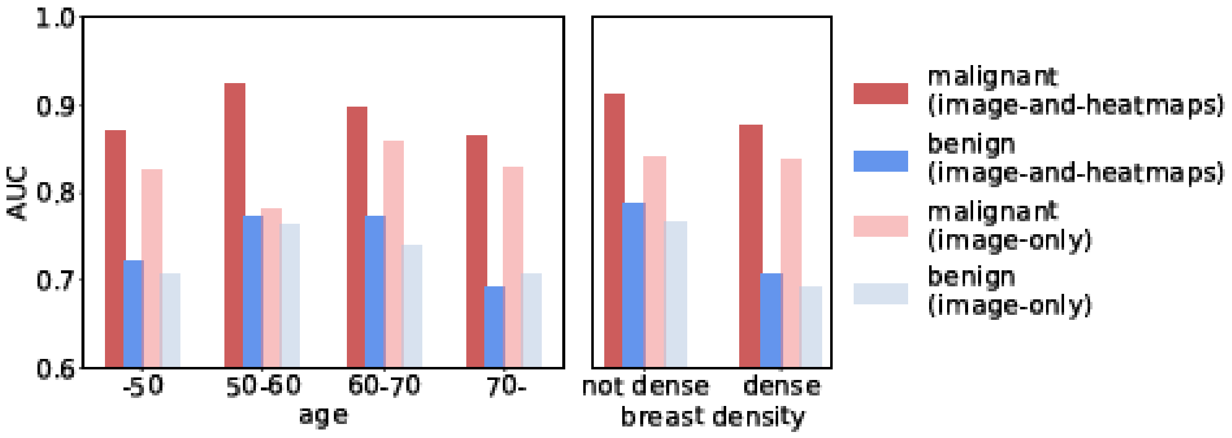


Fig. 6. AUCs for patients grouped by age and by breast density.

- Reader study

  → 14 readers

  → each reading 740 exams from the test set

  → Readers were asked to provide a probability estimate of malignancy on a 0%-100% scale for each breast in an exam.

- 混合：

$$y_{hy} = \lambda y_{people} + (1 - \lambda) y_{model}$$

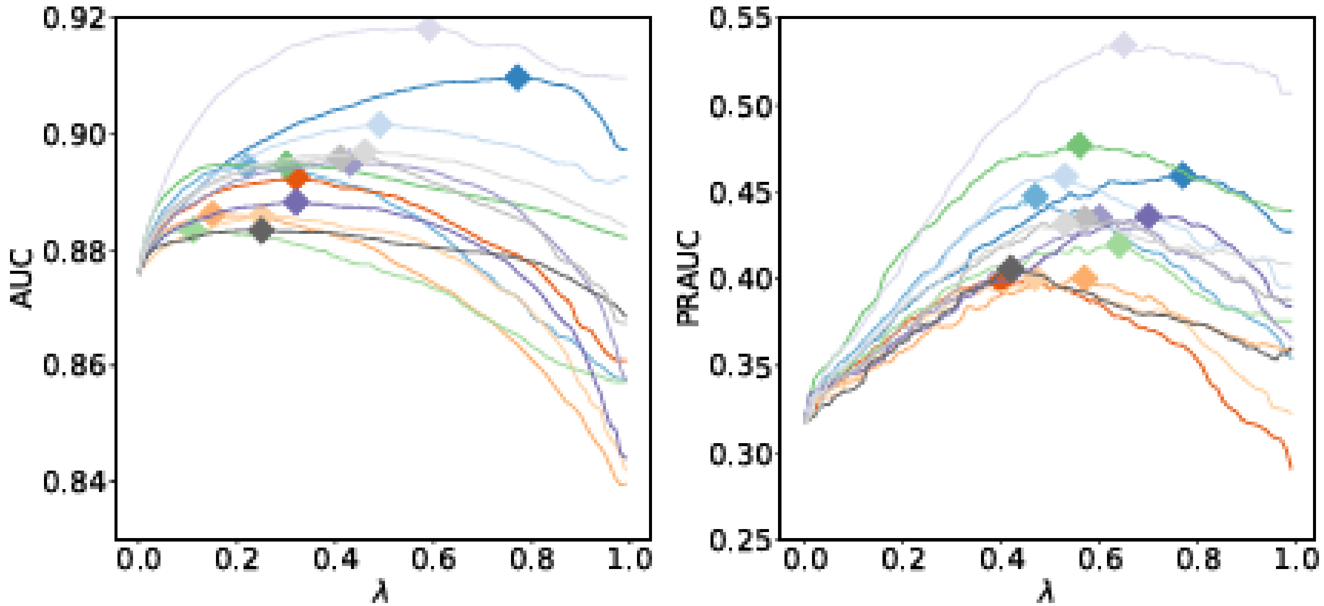


Fig. 8. AUC (left) and PRAUC (right) as a function of $\lambda \in [0, 1)$ for hybrids between each reader and our image-and-heatmaps ensemble. Each hybrid achieves the highest AUC/PRAUC for a different λ (marked with $\diamond$).