Mammographic Breast Density Classification by a Deep Learning Approach

Aly Mohamed, PhD
Robert Nishikawa, PhD
Wendie A. Berg, MD, PhD
David Gur, ScD
Shandong Wu, PhD

Department of Radiology, University of Pittsburgh
Mammography is the standard screening examination for breast cancer.

Breast density is a measure used to describe the proportion of fibroglandular tissue in a woman’s breast.
Breast density is an established risk marker for breast cancer.

Breast density is routinely assessed by radiologists in digital mammogram image reading using the BI-RADS qualitative assessment.

BI-RADS: Breast Imaging-Reporting And Data System
BI-RADS density categories

- Qualitative
- Subjective
- Reader variability

CC view

CC view

MLO view

A-Fatty

B- Scattered density

C- Heterogeneously dense

D- Extremely dense
Density and breast cancer risk

4 - 6 fold higher risk
“Dense vs. non-dense” breasts

Non-dense breasts

A- Fatty
B- Scattered density

Dense breasts

C- Heterogeneously dense
D- Extremely dense
Highly variable categories

A- Fatty  
B- Scattered density  
C- Heterogeneously dense  
D- Extremely dense
Breast density notification legislation

Need: Consistent assessment

- Recommendations for supplemental screening and risk management may vary by breast density.

- In the clinic it is highly desirable to have **consistent assessment** of breast density.
Trend: quantitative assessment

- Calculate an area- or volume-based quantitative breast density measures
- These software tools lack clinical validation or limited to specific setting (e.g., work only on “raw” images).
- BI-RADS density categories: current clinical standard
To investigate a deep learning-based classifier to consistently distinguish the two most common and most variably assigned breast density categories.

- B - “scattered density”
- C - “heterogeneously dense”
Materials and Methods
Compliant with HIPAA and approved by IRB.

A single-institutional retrospective study of 1427 women who undergo standard digital mammography screening from 2005-2016, and a large dataset of total 22,000 breast-cancer-free digital mammogram images.

The truths of the density categories were based on routine clinical assessment made by board-certified breast imaging radiologists.
Both the mediolateral oblique (MLO) and craniocaudal (CC) views of both breasts are used.

We used balanced numbers (ranging from 500-7000 images) for each category/class for training.

A separate unseen set of 1850 images (925 for each category/class) for testing.
We used a deep learning model based on **Convolutional Neural Networks (CNNs)**.

We constructed a **two-class** CNN model, aiming at classifying the two BI-RADS breast density categories.

The CNN used an improved version of the **AlexNet** model.

The network was implemented using the **Caffe** platform.
Data analysis

- We used receiver operating characteristic (ROC) analysis with the area under the curve (AUC) to measure the performance of the classifier.

- We evaluated the effects of transfer learning.

- We performed several robustness analyses.
We used a **pre-trained** AlexNet model learned on a very large existing **non-medical** imaging dataset.

Followed by a **fine-tuning** process with our own mammogram images.

- 1,000 object classes (categories).
- Images:
  - 1.2 M train
  - 100k test.
Robustness analysis (2)

- Due to **reader variability**, we tested the model’s robustness by removing the “**potentially inaccurately labeled**” images.

- Doing so by calculating a **quantitative breast density (PD%)** and comparing to the **BI-RADS-based categories**.
Robustness analysis (3)

B- Scattered density

<table>
<thead>
<tr>
<th>Case</th>
<th>BreastDensity(%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Case1</td>
<td>36.689</td>
</tr>
<tr>
<td>Case2</td>
<td>27.7539</td>
</tr>
<tr>
<td>Case3</td>
<td>33.69588</td>
</tr>
<tr>
<td>Case4</td>
<td>15.38029</td>
</tr>
<tr>
<td>Case5</td>
<td>12.35513</td>
</tr>
<tr>
<td>Case6</td>
<td>11.74867</td>
</tr>
<tr>
<td>Case7</td>
<td>4.90261</td>
</tr>
<tr>
<td>Case8</td>
<td>4.687749</td>
</tr>
<tr>
<td>Case9</td>
<td>7.289479</td>
</tr>
</tbody>
</table>

Average PD% (15.0%)

C- Heterogeneously dense

<table>
<thead>
<tr>
<th>Case</th>
<th>BreastDensity(%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Case1</td>
<td>41.895</td>
</tr>
<tr>
<td>Case2</td>
<td>18.7258</td>
</tr>
<tr>
<td>Case3</td>
<td>37.84276</td>
</tr>
<tr>
<td>Case4</td>
<td>12.1868</td>
</tr>
<tr>
<td>Case5</td>
<td>23.6157</td>
</tr>
<tr>
<td>Case6</td>
<td>22.64162</td>
</tr>
<tr>
<td>Case7</td>
<td>14.34989</td>
</tr>
<tr>
<td>Case8</td>
<td>9.2762</td>
</tr>
<tr>
<td>Case9</td>
<td>18.08623</td>
</tr>
</tbody>
</table>

Average PD% (29.5%)
Robustness analysis (4)

- Effects of the **MLO vs. CC view** in the density assessment made by radiologists.

- Effects of distinguishing the four categories by classifying “dense (category C&D) vs. non-dense (category A&B)” breasts.
Results
Direct training with our own mammogram dataset

AUC: 0.9421 (7000 training samples)

AUC: 0.9081 (500 training samples)
Effects of transfer learning

AUC: **0.9243** (7000 training samples)

AUC: **0.9265** (500 training samples)
Robustness analysis (Cleaner dataset)

Removing “potentially inaccurately labeled” images

<table>
<thead>
<tr>
<th>Density Type</th>
<th>Original dataset size</th>
<th>“Potentially inaccurately labeled”</th>
<th>After removal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scattered density</td>
<td>7,925</td>
<td>867</td>
<td>7,058</td>
</tr>
<tr>
<td>Heterogeneously dense</td>
<td>14,075</td>
<td>2,286</td>
<td>11,789</td>
</tr>
<tr>
<td>Total</td>
<td>22,000</td>
<td></td>
<td>18,847</td>
</tr>
</tbody>
</table>
Robustness analysis (Cleaner dataset)

Comparison to training = 6000 images/class

<table>
<thead>
<tr>
<th></th>
<th>Before Removal</th>
<th>After Removal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Without transfer learning</td>
<td>0.9421</td>
<td>0.9692</td>
</tr>
<tr>
<td>With transfer learning</td>
<td>0.9243</td>
<td>0.9726</td>
</tr>
</tbody>
</table>
Robustness analysis
(MLO vs. CC view)

https://radiologykey.com/mammography-3/
Robustness analysis (MLO vs. CC view)
Robustness analysis ("dense" vs. "non-dense")

- **A-** Fatty
- **B-** Scattered density
- **C-** Heterogeneously dense
- **D-** Extremely dense

ROC curve showing the performance of different views (MLO, CC, CC + MLO) with AUC values.

#SIIM17
In this study we present a new deep learning-based method to distinguish between “scattered density” and “heterogeneously dense” categories in clinical breast density assessment.

We collected a large (22,000) mammogram imaging dataset and showed that overall, the CNN-based classifier can achieve the highest AUC of 0.94 / 0.97.
We observed that the transfer learning can achieve a comparable classification performance to without.

In transfer learning test, classification seems to be not that sensitive to the size of the fine-tuning samples.

Fine-tuning is necessary, because otherwise, the standard pre-trained model was not able to make any meaningful classification (AUC=0.5).
Original AUC is 0.94; it boosted to 0.97 after removing “potentially inaccurately-labeled” images, with or without transfer learning.

The high AUCs in both cases showed the deep learning-based classifier of breast density is robust to real-world clinical dataset.
Limitations

- This is a single-center retrospective study.

- The studied images were read by many radiologists and we did not track which radiologist interpreted which images.

- Future work: comparing the deep-learning based method to traditional feature engineering-created descriptors.
Our study showed encouraging classification performance by a CNN-based deep learning model in distinguishing the breast density categories of “scattered density” vs. “heterogeneously dense”.

We anticipate that our approach will provide a promising toolkit to help enhance current clinical assessment of breast density.

This work adds a new example of applying deep learning and transfer learning in analyzing a large clinical breast imaging dataset.
Acknowledgements

- NIH/NCI R01 (#1R01CA193603)
- RSNA Scholar Grant (#RSCH1530)
- UPMC CMRF Grant (#20151068)
- UPCI-IPM Pilot Award (#MR2014-77613)

Dr. Shandong Wu
wus3@upmc.edu