Radiologists’ Agreement on Mammographic Density from FFDM Images and Validation of a Fully Automated Algorithm that Generates Reliably Reproducible Breast Density Measures

M Abdolell¹,²,³, K Tsuruda², G Schaller¹,³, J Payne¹,²,⁴,⁵, S Iles¹,², C Lightfoot¹,², M McQuaid¹,², J Caines¹,²,⁴

¹ Department of Diagnostic Radiology, Dalhousie University, Halifax, Nova Scotia, Canada
² Department of Diagnostic Imaging, Capital District Health Authority, Halifax, Nova Scotia, Canada
³ Division of Medical Education/Informatics, Dalhousie University, Halifax, Nova Scotia, Canada
⁴ Nova Scotia Breast Screening Program, Halifax, Nova Scotia, Canada
⁵ Department of Community Health and Epidemiology, Dalhousie University, Halifax, Nova Scotia, Canada
Disclosure

- Nothing to disclose
Background

- Mammographic density is a major risk factor for breast cancer
  - Cancer grows in fibrous and glandular tissue
  - Fibroglandular tissue appears white on a mammogram, as do cancers, thus obscuring detection
Density & Br Ca$^+$ Risk

- Women with dense breasts are at 4-6 times increased risk of developing breast cancer
Current Practice

- Subjective visual assessment using the ACR 4-level BIRADS lexicon
- ACR states visual assessment of breast density is “not reliably reproducible”
- Time consuming to report accurately/precisely
Objectives

• Evaluate the agreement between mammographers visually assessing breast density

• Determine if a fully-automated breast density measurement algorithm agrees with expert radiologist assessments of density
Methods

Visual Assessments

• 137 final DICOM full-field digital mammograms spanned the density range from 0-100%

• 5 radiologists visually assessed density in 5% increments

• Median density measure from the radiologists was used as the reference standard to validate algorithm performance
Methods
Breast Density Algorithm

- Automated FFDM density program assessed density in 1% increments, from 0% to 100%
Methods

Agreement & Validation

• Intra-class correlation coefficient (ICC) used to measure agreement

• Bland-Altman plot used to assess agreement

• Box-and-whisker plot used to display distributions
Results
Radiologist Agreement

- 7-way agreement between rads:
  ICC = 0.88, 95% CI (0.86, 0.91)

- Pairwise agreement between rads:
  ICC range = 0.84 – 0.96

<table>
<thead>
<tr>
<th></th>
<th>Rad1</th>
<th>Rad2</th>
<th>Rad3</th>
<th>Rad4</th>
<th>Rad5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rad1</td>
<td>1</td>
<td>0.90</td>
<td>0.87</td>
<td>0.96</td>
<td>0.87</td>
</tr>
<tr>
<td>Rad2</td>
<td>-</td>
<td>1</td>
<td>0.89</td>
<td>0.93</td>
<td>0.88</td>
</tr>
<tr>
<td>Rad3</td>
<td>-</td>
<td>-</td>
<td>1</td>
<td>0.87</td>
<td>0.84</td>
</tr>
<tr>
<td>Rad4</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>1</td>
<td>0.90</td>
</tr>
<tr>
<td>Rad5</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>1</td>
</tr>
</tbody>
</table>
Results

Algorithm Performance

The agreement between the algorithm and the median radiologist as measured by the ICC was 0.87, falling within the 95% confidence limits of the radiologists’ agreement.

The red line is the line of best fit determined by linear least squares regression.

The blue line indicates perfect agreement.
Results

Algorithm Performance

Bland-Altman plot showing agreement between the algorithm and the reference standard.

The absolute values of the upper and lower agreement levels are approximately the span of a single density category in the BI-RADS lexicon.

There is a small, positive bias in the reference standard densities compared to the algorithm densities (bias = 3.90%).
Results

Algorithm Performance

Box-and-whisker plot displaying the distribution of algorithm-generated densities falling into the radiologist’s classifications designated by the BI-RADS density lexicon.

Ideally each of the boxes and whiskers would be entirely contained in their respective BI-RADS level.
Conclusion

• Expert mammographers demonstrate excellent agreement in reporting breast density from FFDM images.

• The automated algorithm demonstrates excellent agreement with the reference standard radiologist assessment of breast density.

• As such, it shows promise in its ability to provide reliably reproducible automated mammographic density measurements that can be adjunctively used to inform clinical practice.
Thank you

(Atlantic Region)