

EVALUATING THE PERFORMANCE OF AUTOMATED BREAST DENSITY ALGORITHMS – WHEN CORRELATION IS NECESSARY BUT NOT SUFFICIENT

Mohamed Abdolell^{1,2}, Peter Brown^{1,2}, Kaitlyn Tsuruda^{1,2}, Jennifer Payne^{1,3}, Judy Caines^{1,2}, Sian Iles^{1,2}
¹ Dalhousie University, ² Nova Scotia Health Authority, ³ Nova Scotia Breast Screening Program

1. INTRODUCTION

- Evaluation of automated breast density algorithms are typically done by comparing density results with expert Radiologists' assessments.
- Linear correlation between algorithms and a consensus of expert radiologists is often used to demonstrate algorithm performance
- Yet, linear correlation can be very high while the agreement is very poor.

3. METHODOLOGY

PMD Assessments

Sample size: 757 digital mammograms
Views: CC and MLO
Vendors: Hologic (n=344) and GE (n=413)

Assessment Methods

1. Three expert radiologists
2. Densitas (DM-Density)
3. LIBRA

Statistical Analyses

Linear Association Statistic:

- Pearson Correlation Coefficient (PCC)

Inter-rater Agreement Statistics:

- Intraclass Correlation Coefficient (ICC)
- Cohen's Kappa Coefficient (κ)

5. CONCLUSIONS

- The PCC is a measure of linear association and thus alone is not sufficient to establish face validity nor reliability, potentially leading to erroneous conclusions regarding algorithm performance.
- Validity and reliability of breast density algorithms should be evaluated using measures of agreement including the ICC for continuous density scales, and Kappa statistic for categorical density scales.

2. OBJECTIVE

To evaluate the association and agreement mammographic density assessments between a consensus of expert radiologists' and two fully automated algorithms, Densitas and LIBRA.

4. RESULTS

Table 1. Pearson Correlation Coefficient (PCC) comparing raters on a percent mammographic density scale from 0 – 100%

PCC Values	GE		Hologic	
	CC	MLO	CC	MLO
Expert Radiologists vs. Densitas	0.891	0.876	0.933	0.922
Expert Radiologists vs. LIBRA	0.743	0.797	0.761	0.831
LIBRA vs. Densitas	0.748	0.755	0.774	0.854

The reported PCC values indicate that all raters are well correlated to one another. Densitas had the highest correlation with Expert Radiologists.

Table 2. Intraclass Correlation Coefficient (ICC) comparing raters on a percent mammographic density scale from 0 – 100%

ICC Values	GE		Hologic	
	CC	MLO	CC	MLO
Expert Radiologists vs. Densitas	0.874	0.871	0.923	0.921
Expert Radiologists vs. LIBRA	0.282	0.214	0.283	0.319
LIBRA vs. Densitas	0.261	0.161	0.270	0.333

LIBRA has poor agreement with the Expert Radiologists and with Densitas. Densitas has almost perfect agreement with Expert Radiologists.

Table 3. Cohen's Kappa Coefficient (κ) comparing raters; using BIRADS 4 density scale.

κ Values	GE		Hologic	
	CC	MLO	CC	MLO
Expert Radiologists vs. Densitas	0.796	0.795	0.857	0.852
Expert Radiologists vs. LIBRA	0.342	0.334	0.309	0.343
LIBRA vs. Densitas	0.396	0.285	0.325	0.391

Similar results are demonstrated using the Kappa statistic as was shown using the ICC in Table 2 above.

DISCLOSURE: M. Abdolell is CEO of Densitas Inc.; K. Tsuruda was an employee of Densitas Inc. at the time of the inception of the study.