Artificial Intelligence in Breast Imaging:
Image Interpretation and Clinical Implementation

Connie Lehman MD PhD
Breast Cancer: Most Frequent Cancer in Women Worldwide

Every Year:
- Of 3.8 billion women in the world, > 2 million diagnosed with breast cancer each year
- > 40,000 deaths in the US alone
- > 600,000 deaths in the world
Precision Medicine/Risk Assessment Supports All Levels of Care Pathway

- **Therapy**: Informing and guiding targeted Rx
- **Prevention and Screening**: Detection of first cancer, Detection of recurrent ca
- **Diagnosis**: B9 vs MG, Staging
Our Challenge

Screening/early detection is key to cure

- Effective screening programs require:
  - accurate risk assessment tools
  - effective screening tests
Mammography as a Screening Examination in Breast Cancer

JOHN N. WOLFE, M.D.


Supported by grants from the Michigan Cancer Foundation and Woman’s Hospital Research Fund.

2 Associate Radiologist, Woman’s Hospital, Detroit, Mich.

The tedious task of examining about 250 women to detect one cancer seems relatively unrewarding unless it is realized that the cancer found is most likely to be in a curable stage. If left until it is clinically evident, the likelihood of salvage diminishes rapidly.
AI and Screening Mammography

• Problems to address
  – No risk assessment models that predict individual risk with any accuracy
  – Human variation in interpretation (quality)
  – Lack of human breast imaging specialists to support screening mammography expansion (access)
Our Challenge

- In order for screening tests to be effective, essential to screen an **at-risk population**
- False positives are decreased when prevalence is increased through **risk assessment**

![Graph showing prevalence vs. PPV and NPV](attachment:graph.png)

National Expenditure For False-Positive Mammograms And Breast Cancer Overdiagnoses Estimated At $4 Billion A Year

By Mei-Sing Ong and Kenneth D. Mandl

DOI: 10.1377/hlthaff.2014.1087
HEALTH AFFAIRS 34, No. 4 (2015): 576-583
©2015 Project HOPE—The People-to-People Health Foundation, Inc.
Impact of False High Risk Assessment on Patients and Systems

- Anxiety, unnecessary tests, interventions
  - MRI or US screening
  - Chemoprevention
  - Mastectomy
  - Costs
American Cancer Society 2007

“Based on the evidence from studies of MR screening high risk women, and the limitations of mammography and CBE alone, the American Cancer Society recommends annual MR screening in conjunction with mammography in women at significantly increased risk of breast cancer.”
Rapid Increase in Breast Magnetic Resonance Imaging Use Trends From 2000 to 2011

Breast MRI use grows, but does it benefit the right women?

Oncology/Hematology
Breast MRI Use Up Despite Lack of Indications

A. Age-specific rates. Each bar within an age group represents a calendar year. B. Indication-specific rates for 4 primary indications: screening, diagnostic, staging or treatment, and surveillance.
75% of all screening MRIs performed were in women with less than 20% lifetime risk.

Of women at greater than 20% lifetime risk, less than 2% had received an MRI.
Classical Risk Models

- Age
- Family History
- Prior Breast Procedure
- Parity
- Breast Density

Risk

AUC: 0.631

AUC: 0.607 without Density

---


**Prospective breast cancer risk prediction model for women undergoing screening mammography.**

Barlow WE¹, White E, Ballard-Barbash R, Vacek PM, Titus-Ernstoff L, Carney PA, Tice JA, Buist DS, Geller BM, Rosenberg R, Yankaskas BC, Kerlikowske K.
Screening Mammography Interpretation and AI

- Breast Density?
- Normal or Not?
A Study of Breast Parenchyma by Mammography in the Normal Woman and Those with Benign and Malignant Disease

JOHN N. WOLFE, M.D.

All normal and abnormal parenchymal elements were noted, but the main emphasis was on assessment of the alveolar tissue and ducts; their presence or absence, amount, and distribution. This material was coded and later subjected to analysis by computer.
Breast Composition

• “visually estimated content of fibroglandular-density within the breasts”

<table>
<thead>
<tr>
<th>Breast Composition Categories</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. The breasts are almost entirely fatty</td>
</tr>
<tr>
<td>b. There are scattered areas of fibroglandular density</td>
</tr>
<tr>
<td>c. The breasts are heterogeneously dense, which may obscure small masses</td>
</tr>
<tr>
<td>d. The breasts are extremely dense, which lowers the sensitivity of mammography</td>
</tr>
</tbody>
</table>
Advocacy efforts to inform women

- Are You Dense? FACT #5
  While mammogram detects 98% of cancers in women with fatty breasts, it finds ONLY 48% in women with the densest breasts.

- Early Matters
  stage of tumor at discovery influences prognosis
  Be informed about your breast density

- Are You Dense? FACT #1
  Breast density is one of the strongest predictors of the failure of mammography screening to detect cancer.
Breast Density Law

• Diagnosed: 2003, stage III
• Her last mammogram was false negative
• She lobbied for supplemental screening law in Connecticut
• The law was enacted in 2005
New federal law requires mammography providers to send breast density notifications

February 19, 2019 | Michael Walter | Policy

When President Donald Trump signed a federal funding bill into law on Feb. 15, it included text that said that all mammography providers must include updated information about breast density in reports sent to both patients and their physicians.

The notifications sent out to patients will inform them about their own personal breast density and explain the importance of that information. More than 30 states currently require such information to be shared with patients after they undergo a mammogram, a number that has been rising steadily for years.
Breast Cancer Surveillance Consortium data from over 3.8 million screening mammograms in U.S. community practice: over 50% of women told they have dense tissue.
Wide Variation in Radiologists’ Assessment of Mammograms as “Dense”

83 radiologists: 6% to 85% of large (>500) number of mammograms read as “dense”
Screening Mammography Interpretation and AI

- Breast Density?
- Normal or Not?
Interpretation: Normal or Not?
Challenges

• Our imaging screening tests depend on highly specialized human expertise
  – Human variation in performance of tasks
Advances in imaging technology have outpaced human performance in interpreting mammograms accurately.
DBT Reveals Occult ILC

2D FFDM

Lobular Carcinoma

Images courtesy of Drs. Di Maggio & G Gennaro,
Istituto Oncologico Veneto I.R.C.C.S. - Padova, Italia
Women Younger than 50 Yr

Digital (AUC, 0.84 ± 0.03)

Film (AUC, 0.69 ± 0.05)

Sensitivity vs. 1 – Specificity

P < 0.002
National Performance Benchmarks for Modern Screening Digital Mammography: Update from the Breast Cancer Surveillance Consortium

Constance D. Lehman, MD, PhD
Robert F. Arao, MPH
Brian L. Sprague, PhD
Janie M. Lee, MD, MSc
Diana S. M. Buist, PhD, MPH
Karla Kerlikowske, MD
Louise M. Henderson, PhD, MSPH
Tracy C.

**Purpose:** To establish performance benchmarks for modern screening digital mammography and assess performance trends over time in U.S. community practice.

**Materials and Methods:** This HIPAA-compliant, institutional review board-approved study measured the performance of digital screening mammography interpreted by 350 radiologists across 95 facilities.

---

*Acceptable region: 5–12%*

**Recall Rate**

<table>
<thead>
<tr>
<th>Number of radiologists</th>
</tr>
</thead>
<tbody>
<tr>
<td>34 min</td>
</tr>
</tbody>
</table>

*Acceptable region: >= 2.5*

**Cancer Detection Rate per 1,000 exams**

<table>
<thead>
<tr>
<th>Number of radiologists</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.7 min</td>
</tr>
</tbody>
</table>
More than 40% of radiologists have AIRs outside the recommended ranges, and more than 37% fall below recommended ranges for specificity.

Lehman et al *Radiology* April 2017
Modern technology is better but wide variation across radiologists.
Performance of screening test influenced by group (> 1 million cases)

Yankaskas et al., 2005

![Graph showing recall and specificity](Image)
“No Comparison Mammogram” strongest predictor of “harms”
Putting data in the hands of doctors

Computer scientist Regina Barzilay empowers cancer treatment with machine learning.
Knowledge of effective strategies for clinical implementation essential

- Breast density DL platform in place now at MGH and implemented in routine clinical care
  - 50,000 screening mammograms/year performed/processed
- 1 (triage), 2 and 5 year risk assessment DL model platform in place at MGH and under evaluation for performance

- Rigorous peer reviewed original scientific publications
Culture and Resistance to Change

"I'll be happy to give you innovative thinking. What are the guidelines?"
Brief History of Past Traditional CAD Methods in Mammography
Overview

- CAD applied to mammography approved by FDA in 1998
- With reimbursement, use rapidly increased across the U.S.
- Multiple study designs in early phases: retrospective, reader studies, prospective small single site, etc. with mixed results on impact of CAD on accuracy of mammographic interpretation
Background

- 1998-2002 at 43 BCSC facilities (GHC Seattle, New Hampshire, Colorado)
- Conducted early in adoption (7 of 43 facilities implemented CAD during the study)
Influence of Computer-Aided Detection on Performance of Screening Mammography

Data source: BCSC

Study Limitations

- Data from early years of CAD integration (1998-2002)
- Didn’t control for learning curve (weeks to a year to learn to use CAD)
- Outdated “obsolete” technology (film screen CAD)
- Low numbers (25k CAD exams)
Challenges addressed by BCSC:
No improvement of digital mammography performance with CAD

Study Strengths
- Current performance 2003-09
- Only digital mammography with CAD
- Learning curve addressed
- > 569k CAD exams

Odds ratio for CAD vs. No CAD adjusted for site, age, race, time since prior mammogram and calendar year of exam using mixed effects model with random effect for exam reader and varying with CAD use found no significant difference in sensitivity, specificity or recall rate.
Intra-radiologist analysis:
Mammography performance not improved with CAD
—sensitivity trended to worse with CAD

Odds ratios comparing CAD use versus no CAD, both overall and intra-radiologist

110/271 radiologists read with and without CAD
Drivers of Practice: Science and Reimbursement

- 1998. FDA approves CAD
- 2002 CMS payment
- 2005 NEJM DMIST
- 2007 NEJM CAD


- Plain Film: $90.50
- Digital Mammography: $133.58
- Approved Increment for CAD: $17.74
- Digital Mammography CAD Plus: $151.32

Source: Centers for Medicare and Medicaid Services (2002).
88% binary accuracy on previous logs
97% agreement with an expert radiologist

In clinical implementation in first year at MGH:

**Human Agreement: 94%**

>40K mammograms read by the machine
<table>
<thead>
<tr>
<th>MRN</th>
<th>Age</th>
<th>Study Date</th>
<th>Accession</th>
<th>TC Life</th>
<th>NCI Life</th>
<th>BRCA Life</th>
<th>MGH/MIT 1yr</th>
<th>MGH/MIT 2yr</th>
<th>MGH/MIT 5yr</th>
</tr>
</thead>
</table>


<table>
<thead>
<tr>
<th>TRIAGE hi risk &gt;.0156</th>
<th>count</th>
<th>percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>TRIAGE low risk &lt;.00156</td>
<td>206</td>
<td>19%</td>
</tr>
<tr>
<td>&gt;.0125 5 year hi risk</td>
<td>6</td>
<td>8%</td>
</tr>
<tr>
<td>5 year intermedi</td>
<td>306</td>
<td>23%</td>
</tr>
<tr>
<td>&lt;.005 5 year low risk</td>
<td>645</td>
<td>56%</td>
</tr>
<tr>
<td>0.05 2 year hi risk</td>
<td>107</td>
<td>10%</td>
</tr>
<tr>
<td>2 year intermediate</td>
<td>530</td>
<td>43%</td>
</tr>
<tr>
<td>0.01 2 year low</td>
<td>402</td>
<td>42%</td>
</tr>
</tbody>
</table>

| MRN | Age | Study Date | Accession | TC Life | NCI Life | BRCA Life | MGH/MIT 1yr | MGH/MIT 2y | MGH/MIT 5yr | INIT BIRADS | FINAL BIRADS | Highest order SCREEN | TP | FP | FN |
AI and Breast Cancer: Phase 1

• Problem to address
  – No risk assessment models that predict individual risk with any accuracy
  – Human variation in interpretation (quality)
  – Lack of human breast imaging specialists to support screening mammography expansion (access)

• Large quality databases with known outcomes
  – > 250,000 modern digital consecutive mammograms at MGH linked to tumor registries
  – Partnerships with other institutions outside MGH

• AI expertise: MIT

• Clinical expertise and engagement: MGH
Future

- Machine Learning is a tool to address our greatest challenges for our patients worldwide and amplify our impact
  - Workflow
  - Image acquisition
  - Risk assessment
  - Image interpretation
  - Lesion and patient management
- Clinical implementation of discoveries critical
Thank you
Integration of DBT at MGH

- 2D: 2009 - 2011
  - N = 78,29
  - 8

- 3D: 2012 - 2015
  - N = 76,987