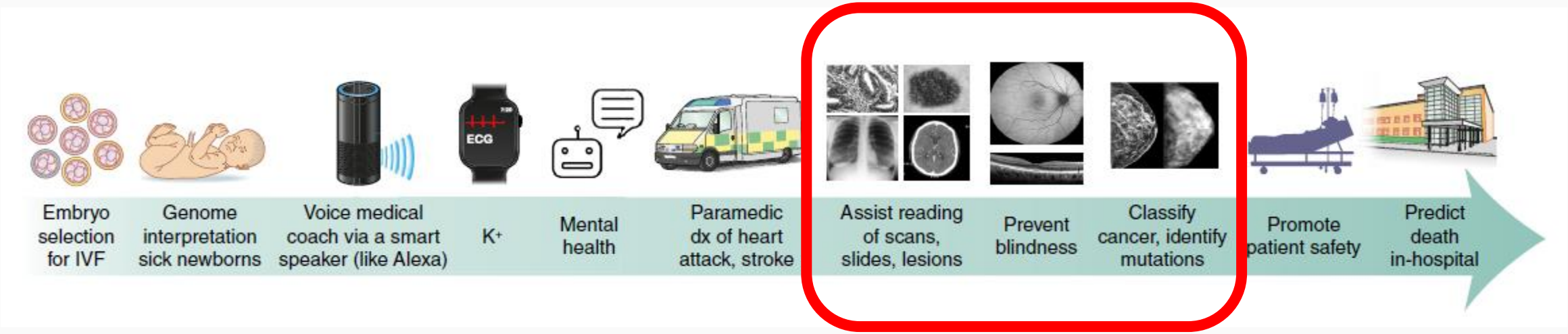
AI in Radiology : industry perspective

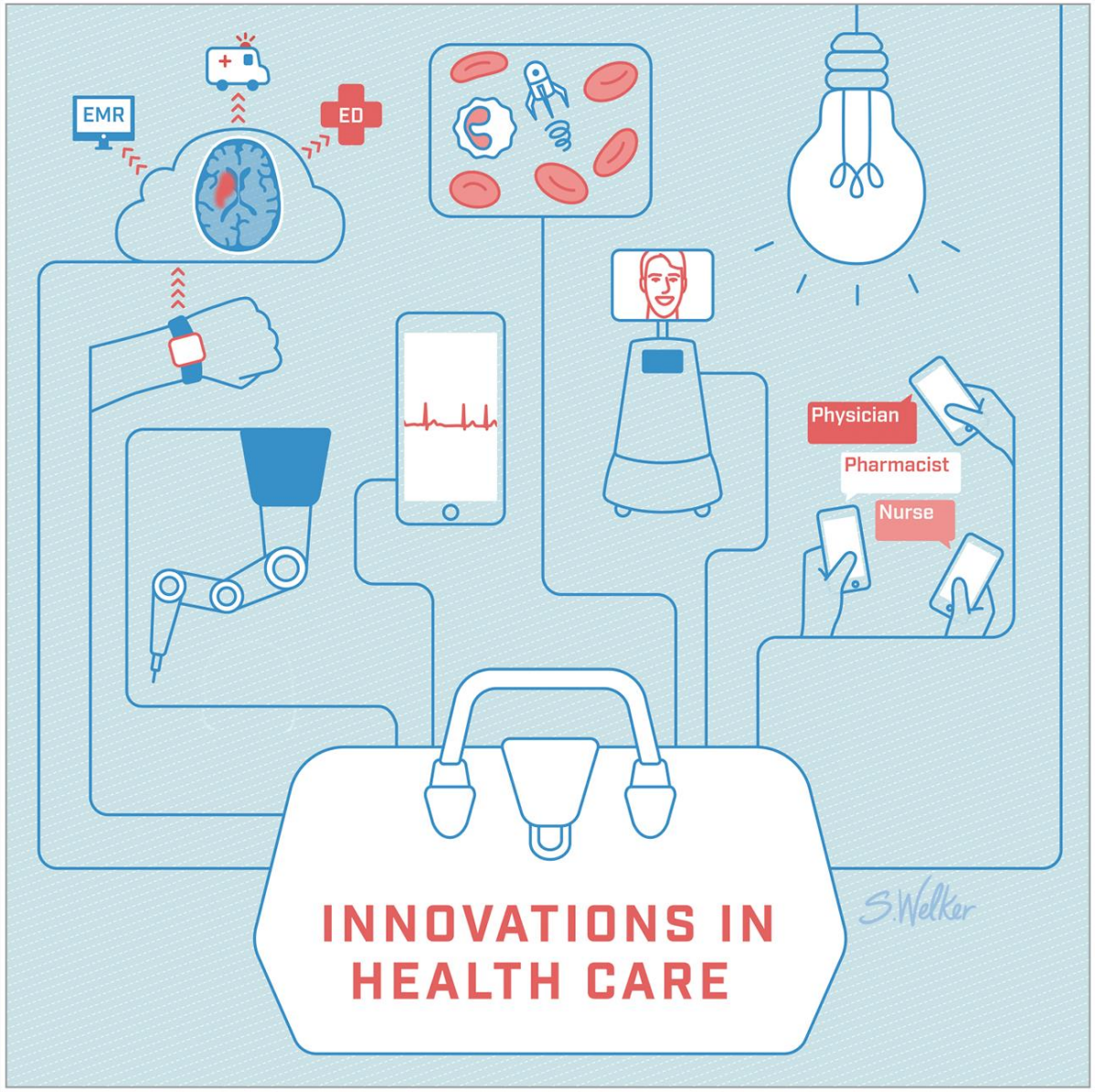
Ki Hwan Kim, MD, PhD
Chief Medical Officer
khkim@lunit.io
2021-04-16

Lunit

AI applications across the human lifespan



→ *Saving lives through AI*



Technically innovative

Clinically innovative

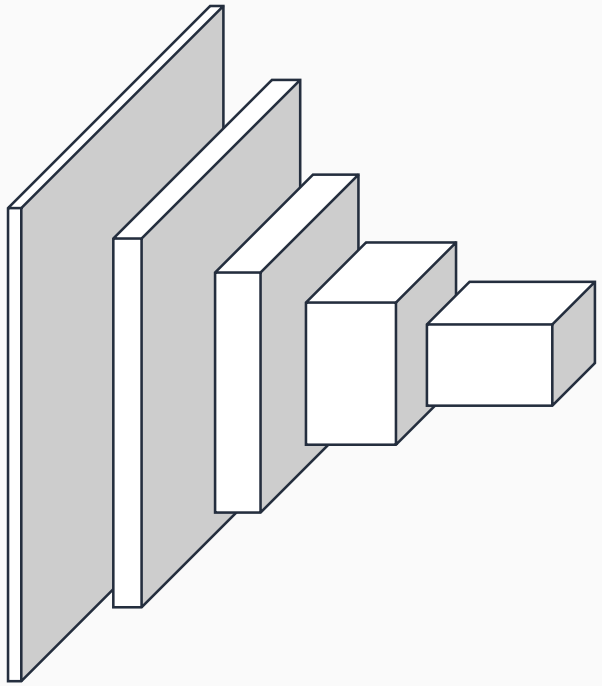
JAMA. 2018;320(5):427-429

Lunit INSIGHT

Lunit INSIGHT MMG for Breast Cancer Screening

Lunit INSIGHT MMG

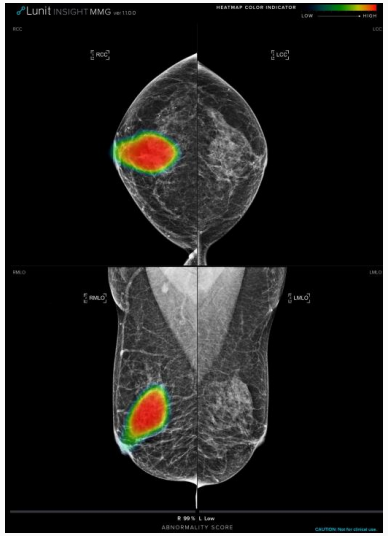
Large-scale mammograms with **diagnosis labels**



Cancer Prediction

Cancer

Non-cancer



Internal validation 2020, Excepted Bi-RADS score 0 and 6

Step 1: Exploring Clinical Unmet Needs

- Difficult image interpretation
- Poor performance of Traditional CAD
- Market already exists

Step 2: Development of AI

- Winning strategy
  - Data number – Breast cancer 40K, total 200K: AMC/SMC/YUHS/US/UK
  - Annotation – by breast radiologists
  - AI performance – top AI researchers

Step 2: Development of AI

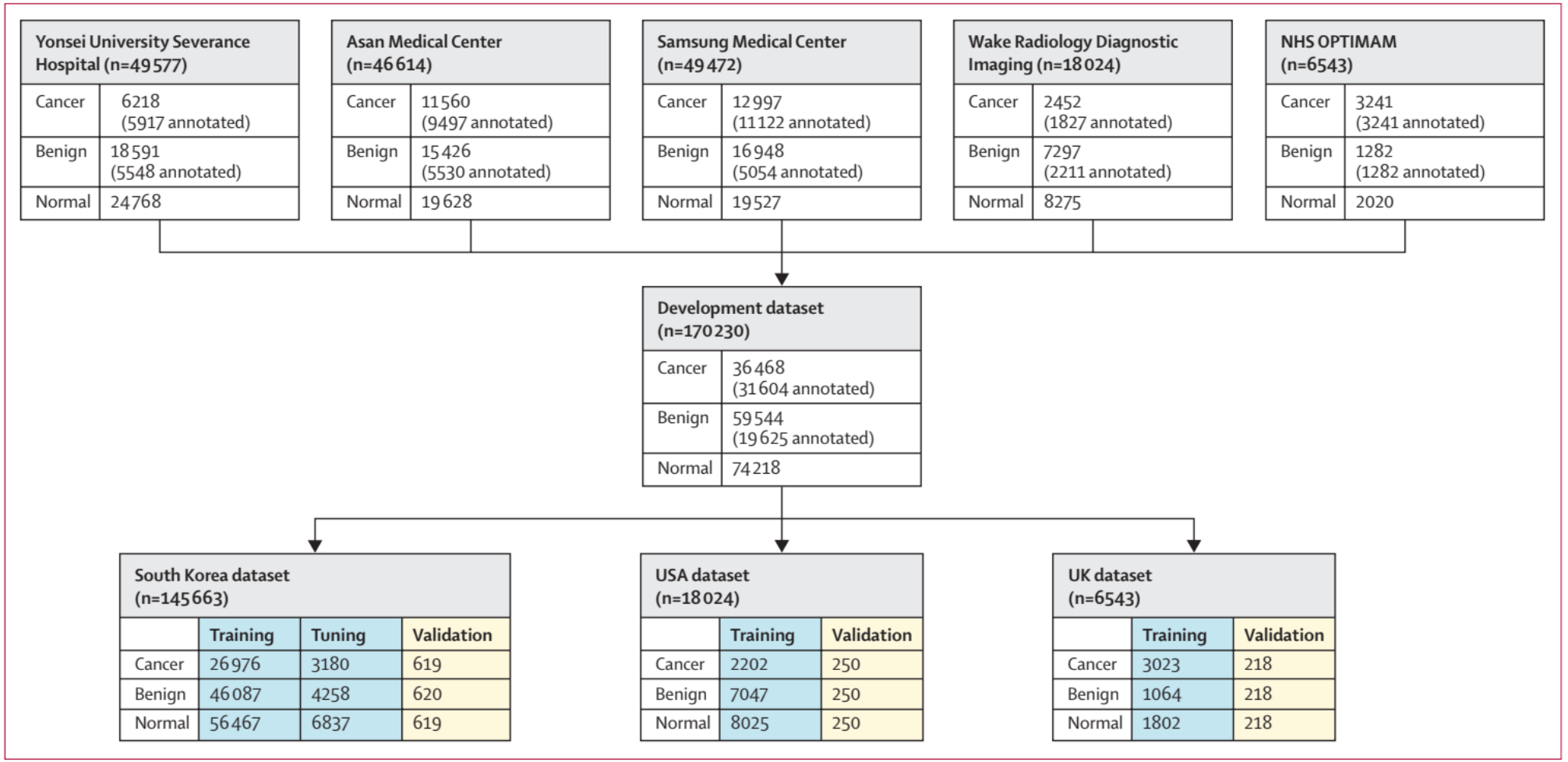


Figure 1: **Development dataset generation and partitioning**
All mammograms are four-view paired (left and right craniocaudal and mediolateral oblique). There was no overlap between categories (cancer, benign, and normal). NHS=National Health Service.

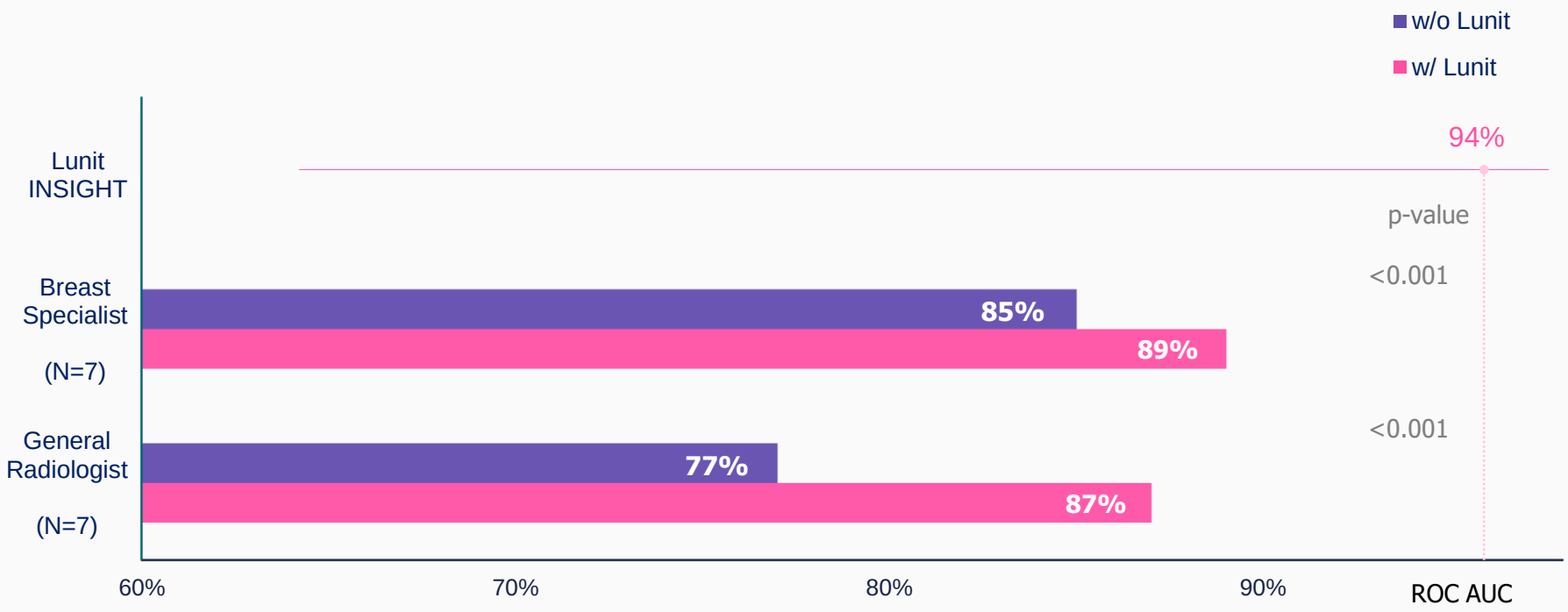
Step 3: Internal Validation

THE LANCET
Digital Health

	ROC AUC	Sensitivity	Specificity
Korea (n=1,858)	97%	90%	92%
US (n=750)	95%	94%	80%
Europe (n=654)	94%	92%	77%

Step 4: External Validation with Reader Study

THE LANCET Digital Health

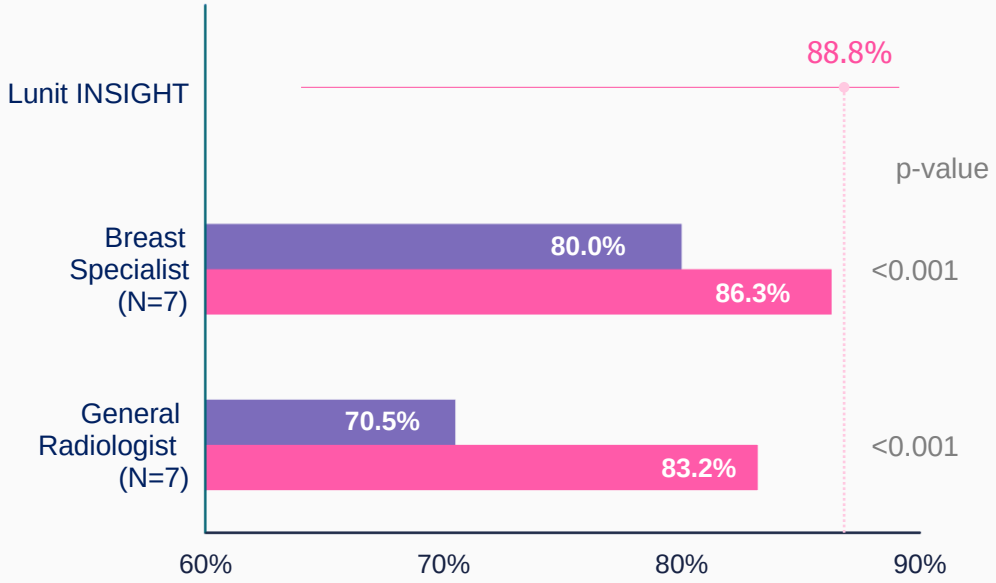


Yonsei University Severance Hospital & Soon Chun Hyang University Hospital, Feb. 2019
N=320

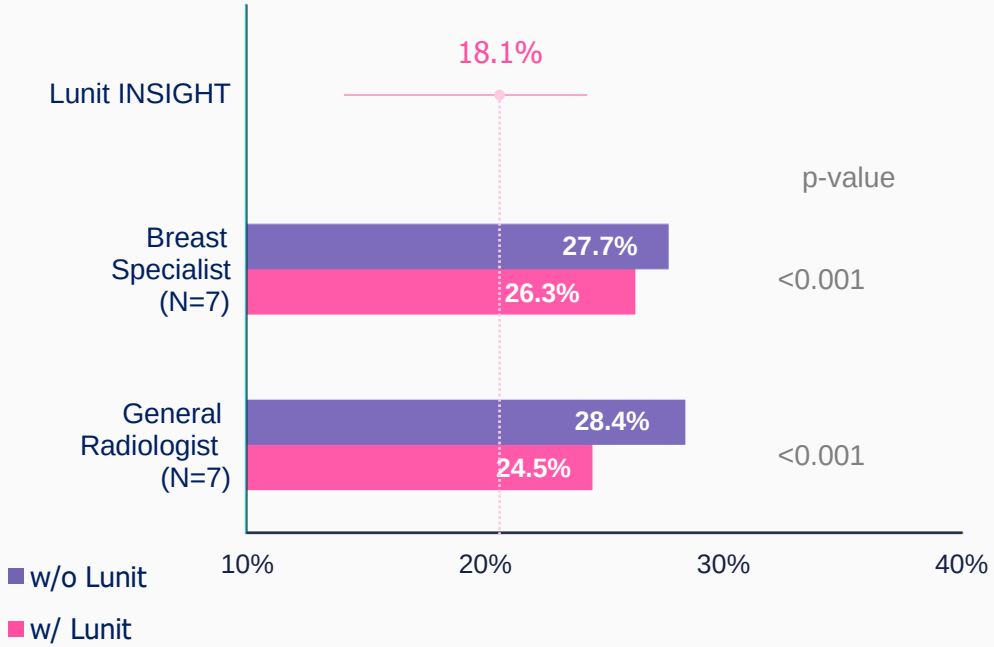
Step 4: External Validation with Reader Study

THE LANCET Digital Health

Sensitivity



False Positive Rate

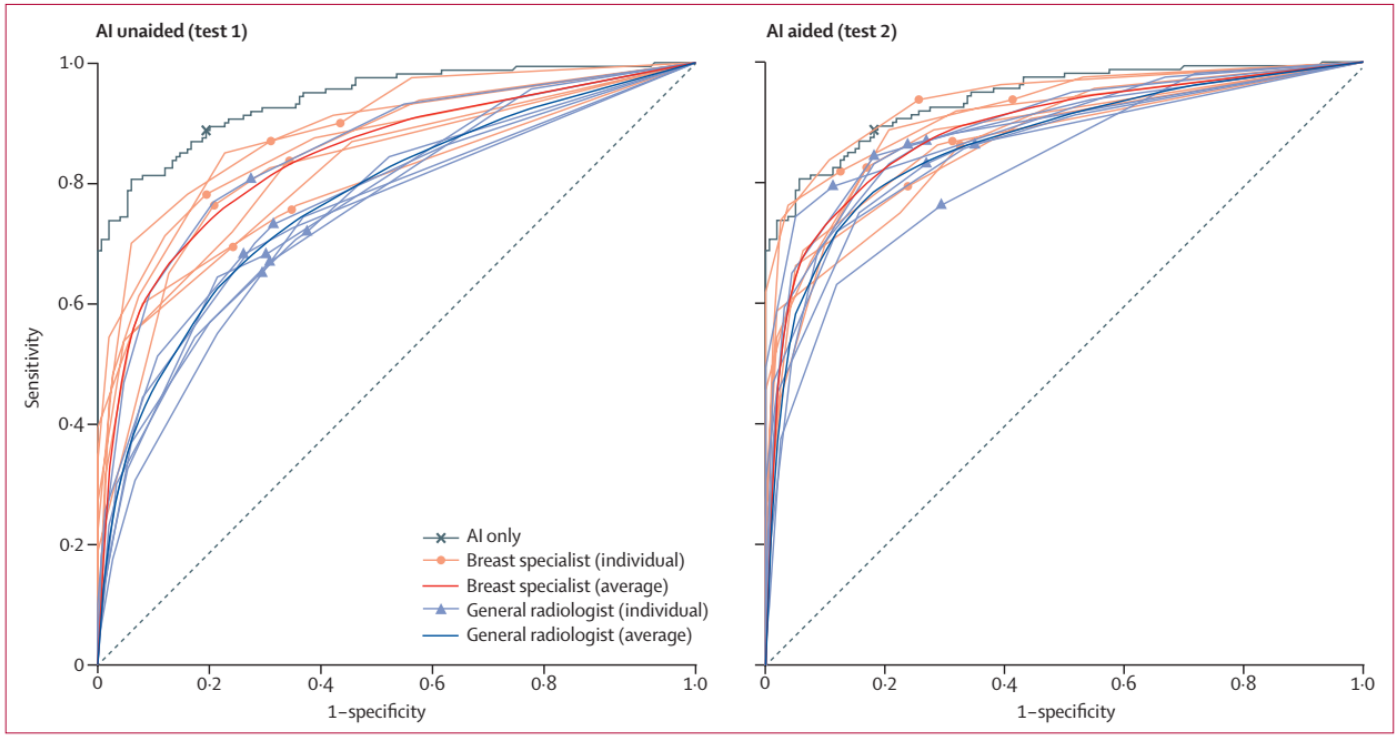


Yonsei University Severance Hospital & Soon Chun Hyang University Hospital, Feb. 2019
Cancer (n=260), non-cancer (n=160)

First peer-reviewed paper

Changes in cancer detection and false-positive recall in mammography using artificial intelligence: a retrospective, multireader study

Hyo-Eun Kim, Hak Hee Kim*, Boo-Kyung Han*, Ki Hwan Kim, Kyunghwa Han, Hyeonseob Nam, Eun Hye Lee, Eun-Kyung Kim*



Lancet Digital Health 2020; 2: e138–48

Published Online February 6, 2020

Figure 3: **ROC analysis for AI-unaided and AI-aided diagnosis**
Sensitivity and specificity of each individual (including AI standalone) are marked on each curve. AI=artificial intelligence. ROC=receiver operating characteristic.

Step 5: External Validation in Screening Cohorts

JAMA Oncology | Original Investigation

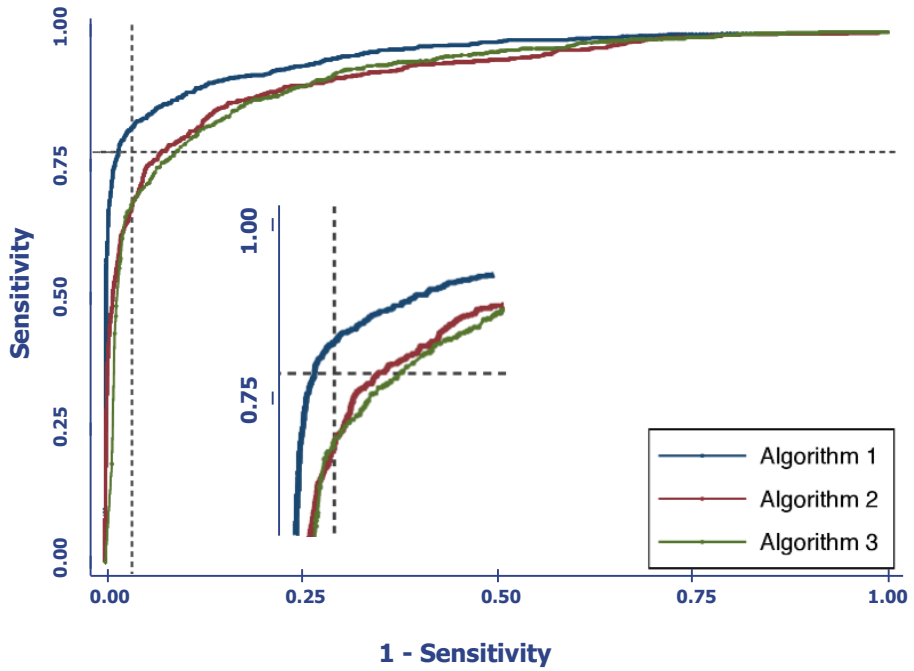
JAMA Oncology 2018 Journal Impact Factor: 22.416

External Evaluation of 3 Commercial Artificial Intelligence Algorithms for Independent Assessment of Screening Mammograms

Mattie Salim, MD; Erik Wåhlin, MSc; Karin Dembrower, MD; Edward Azavedo, MD, PhD; Theodoros Foukakis, MD, PhD; Yue Liu, MSc; Kevin Smith, MSc, PhD; Martin Eklund, MSc, PhD; Fredrik Strand, MD, PhD

Salim, et al. JAMA Oncol. 2020 Aug 27.

Karolinska Institutet



8 years of Breast Screening Cohort Retrospective Analysis: Comparison of Three Leading MMG AI Products

Screening Performance Benchmarkers for AI Algorithms in 739 Women Who Received a Diagnosis of Breast Cancer and 112,924 Healthy Women

Benchmark	Lunit INSIGHT MMG Algorithm 1	Algorithm 2	Algorithm 3
Accuracy (ROC AUC)	0.956 (0.948-0.965)	0.922 (0.910-0.934)	0.920 (0.909-0.931)
Sensitivity	81.9% (79-85%)	67.0% (64-70%)	67.4% (64-71%)
False Negative Rate	18.1% (15.4-21.1%)	33.0% (29.6-36.4%)	33.0% (29.6-36.4%)

Step 5: External Validation in Screening Cohorts

Table 2. Screening Performance Benchmarks for Artificial Intelligence Algorithms and for Radiologists in 739 Women Who Received a Diagnosis of Breast Cancer and 112 924 Healthy Women

	Benchmark point estimate (95% CI)[a]					
	Algorithm[b]			Reader		
Benchmark	1	2	3	First	Second	Consensus
Specificity, %	96.6 (96.5-96.7)	96.6 (96.5-96.7)	96.7 (96.6-96.8)	96.6 (96.5-96.7)	97.2 (97.1-97.3)	98.5 (98.4-98.6)
Sensitivity, %	81.9 (78.9-84.6)	67.0 (63.5-70.4)	67.4 (63.9-70.8)	77.4 (74.2-80.4)	80.1 (77.0-82.9)	85.0 (82.2-87.5)
Accuracy, %	96.5 (96.4-96.6)	96.4 (96.3-96.5)	96.5 (96.4-96.6)	96.5 (96.4-96.6)	97.1 (97.0-97.1)	98.4 (98.3-98.5)
PPV, %	13.6 (12.5-14.7)	11.4 (10.5-12.4)	11.8 (10.8-12.8)	13.0 (12.0-14.0)	15.9 (14.7-17.1)	27.2 (25.4-29.1)
AIR	39.1 (38.0-40.2)	38.1 (37.0-39.2)	37.3 (36.2-38.4)	38.8 (37.7-39.9)	32.8 (31.8-33.9)	20.3 (19.5-21.1)
CDR	5.32 (4.91-5.76)	4.36 (3.98-4.76)	4.38 (4.00-4.78)	5.03 (4.63-5.46)	5.21 (4.80-5.64)	5.53 (5.10-5.97)
FNR	0.181 (0.154-0.211)	0.330 (0.296-0.364)	0.330 (0.296-0.364)	0.226 (0.196-0.256)	0.177 (0.150-0.205)	0.150 (0.124-0.176)

Abbreviations: AIR, abnormal interpretation rate (per 1000 examinations); CDR, cancer detection rate (per 1000 examinations); FNR, false-negative rate (per cancer diagnosed within 12 months); PPV, positive predictive value.

[a] Benchmark estimates based on stratified bootstrapping to attain a proportion of women who received a diagnosis of breast cancer to healthy women similar to the source screening cohort (approximately 0.5%).

[b] The operating point of each algorithm was set at a specificity as close as possible to that of the first reader (96.6%).

Mattie Salim, et al. JAMA Oncology 2020

Step 5: External Validation in Screening Cohorts

Table 3. Number of Abnormal Interpretations and Cases Positive for Cancer Detected by Algorithms and Readers Alone and by Algorithms Combined With the Assessment of the First, Second, or Both Readers

	No. (% increase vs alone)				
	Algorithm			Reader	
Assessment	1	2	3	First	Second
Abnormal interpretation[a]					
Alone	4441	4331	4236	4408	3728
With first reader	7851 (77)	7998 (85)	7847 (85)	NA	5484 (47)
With second reader	7188 (62)	7260 (68)	7139 (69)	5484 (24)	NA
With both readers	8745 (97)	8885 (105)	8762 (107)	NA	NA
Cancer detected[b]					
Alone	605	495	498	572	592
With first reader	655 (8)	620 (25)	623 (25)	NA	640 (8)
With second reader	664 (10)	638 (29)	643 (29)	640 (12)	NA
With both readers	667 (10)	653 (32)	656 (32)	NA	NA

Abbreviation: NA, not applicable.

[a] Based on a total of 113 663 screenings. Observations of healthy women have been duplicated to attain a similar proportion as in the source screening cohort (0.5% with a diagnosis of cancer).

[b] Actual screen-detected cancer (n = 618); actual clinically detected cancer (n = 121).

Lunit

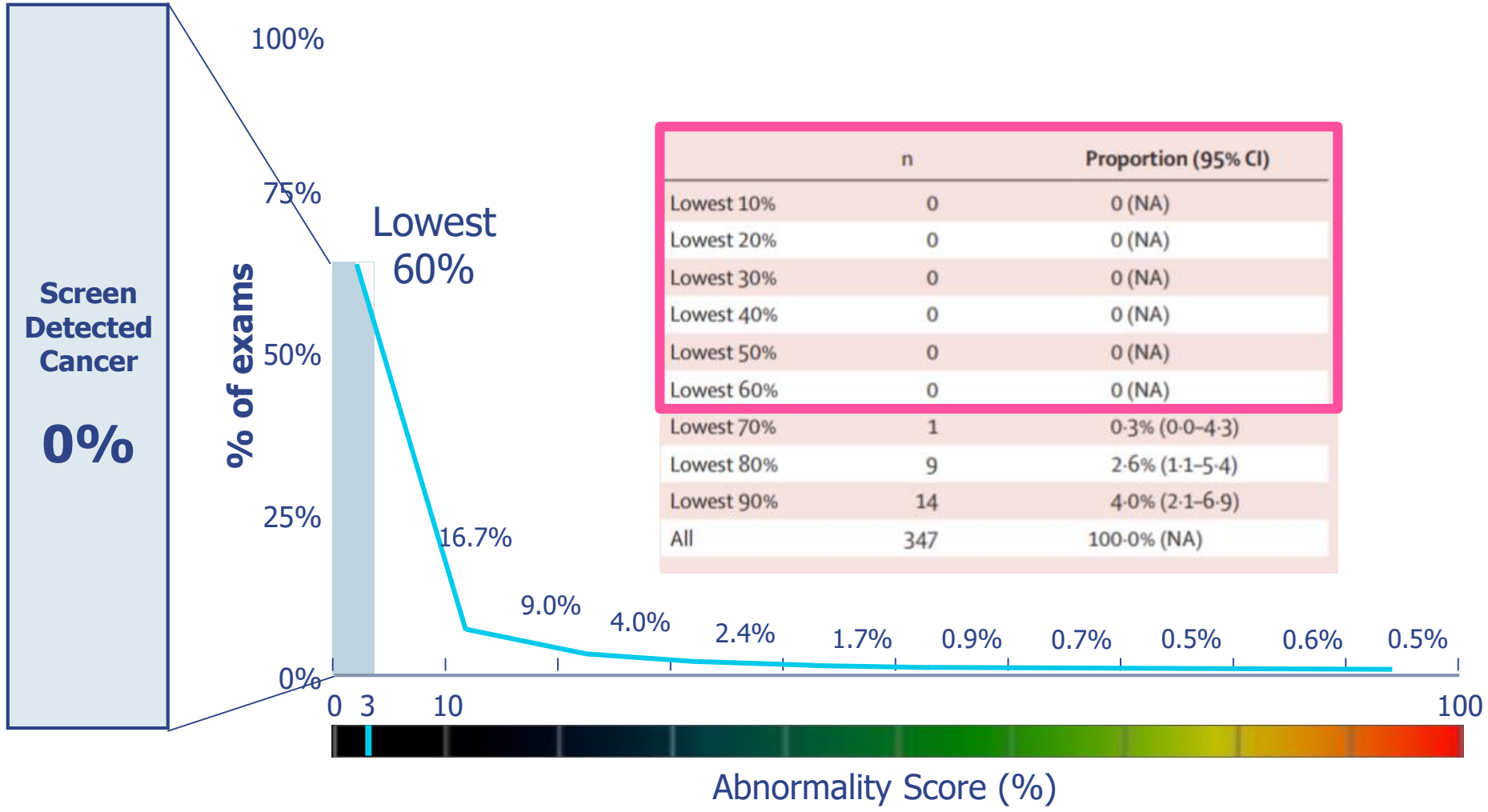
10-14%
검출 향상

AI 와 readers 는 상호 보완적이며, 서로 다른 특징을 보는 것으로 추측할 수 있음.

Mattie Salim, et al. JAMA Oncology 2020

Step 6: AI-powered Triaging for Breast Cancer Screening

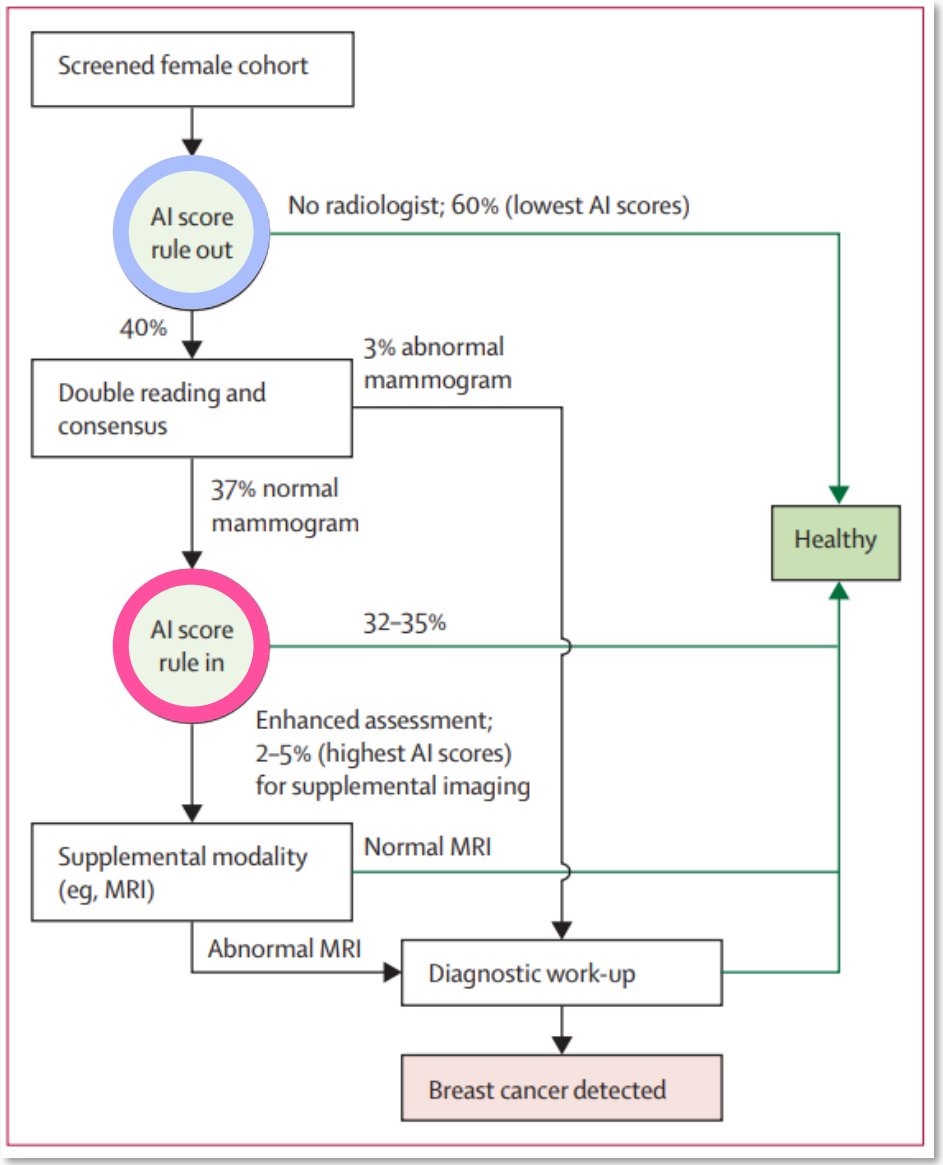
NPV 100%



	n	Proportion (95% CI)
Lowest 10%	0	0 (NA)
Lowest 20%	0	0 (NA)
Lowest 30%	0	0 (NA)
Lowest 40%	0	0 (NA)
Lowest 50%	0	0 (NA)
Lowest 60%	0	0 (NA)
Lowest 70%	1	0·3% (0·0–4·3)
Lowest 80%	9	2·6% (1·1–5·4)
Lowest 90%	14	4·0% (2·1–6·9)
All	347	100·0% (NA)

Lancet Digital Health 2020; 2: e468–74

AI-powered Triaging for Breast Cancer Screening



Rule Out

used to triage women into a **no radiologist work stream** when having a score below a rule-out threshold.

Rule In

an **enhanced assessment** when having a score above a rule-in threshold (after negative double reading by radiologists)

Lancet Digital Health 2020; 2: e468–74

Step 7: Prospective Study

Artificial Intelligence in Large-scale Breast Cancer Screening (ScreenTrustCAD)

The safety and scientific validity of this study is the responsibility of the study sponsor and investigators. Listing a study does not mean it has been evaluated by the U.S. Federal Government. Know the risks and potential benefits of clinical studies and talk to your health care provider before participating. Read our disclaimer for details.

ClinicalTrials.gov Identifier: NCT04778670

Recruitment Status : Not yet recruiting
First Posted : March 3, 2021
Last Update Posted : March 3, 2021

See **Contacts and Locations**

Sponsor:
Karolinska University Hospital

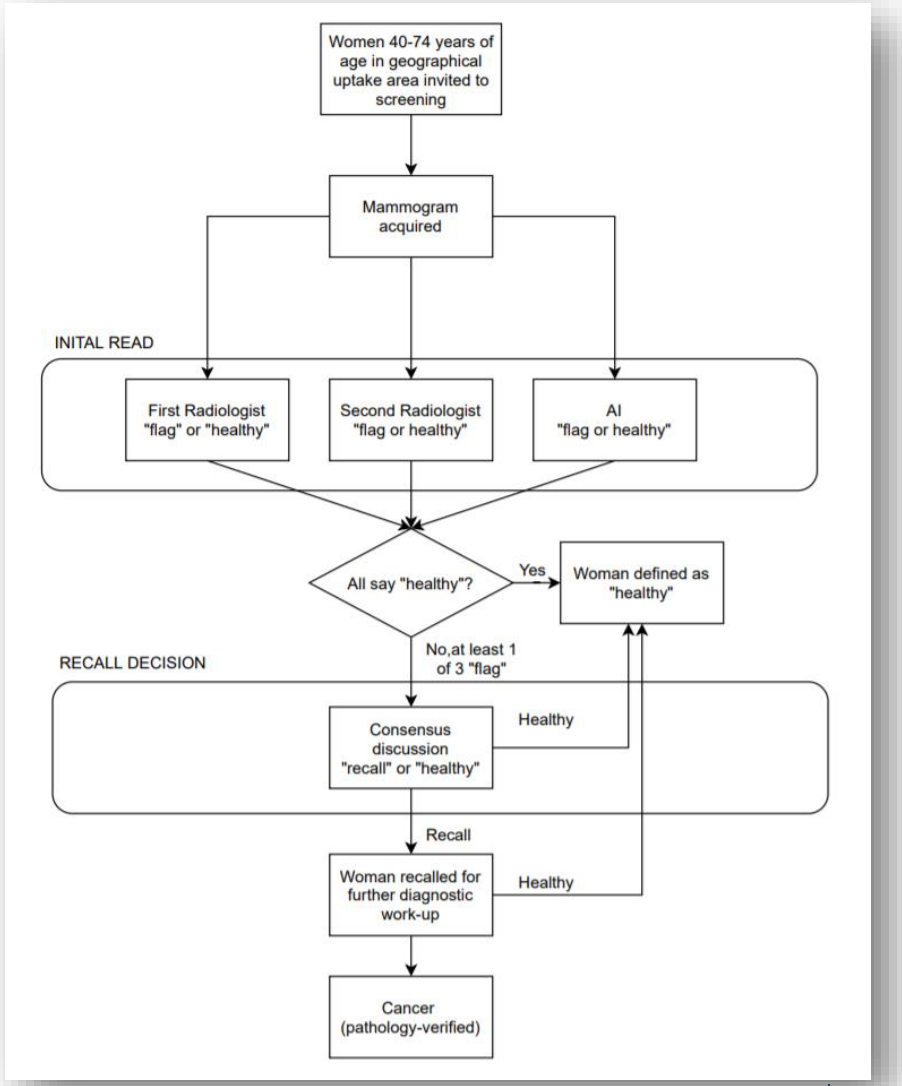
Collaborators:
Capio Sankt Görans Hospital
Lunit Inc.
Karolinska Institutet

Information provided by (Responsible Party):
Fredrik Strand, Karolinska University Hospital

Study Details | Tabular View | No Results Posted | Disclaimer | How to Read a Study Record

Study Description Go to

Brief Summary:
This is a prospective clinical trial following a paired screen-positive design, with the aims to assess the performance of an artificial intelligence (AI) computer-aided detection (CAD) algorithm as an independent reader, in addition to two radiologists, of screening mammograms in a true screening population. Since all decisions by individual readers will be recorded, it is possible to determine what the outcome would have been had one or two of the readers not been allowed to assess images, and to determine what the outcome would have been had the recall decision been performed by consensus decision (actual) compared to single reader arbitration of discordant cases.



https://clinicaltrials.gov/ct2/show/NCT04778670

Summary

Based on my experience in commercial AI-based image analysis SWs

- AI technology is a great opportunity.
- Clinical need is more important.
- A business strategy is the most important and AI is just a tool to solve that problem.

Thank you

khkim@lunit.io