



Centro di Riferimento per l'Epidemiologia  
e la Prevenzione Oncologica in Piemonte

[www.cpo.it](http://www.cpo.it)



# AI e screening mammografico

Dott.ssa Elisa Camussi  
SSD Epidemiologia e Screening

PROGRAMMA REGIONALE DI SCREENING PER IL TUMORE DELLA  
MAMMELLA PREVENZIONE SERENA – WORKSHOP 2021

Torino 3 dicembre 2021

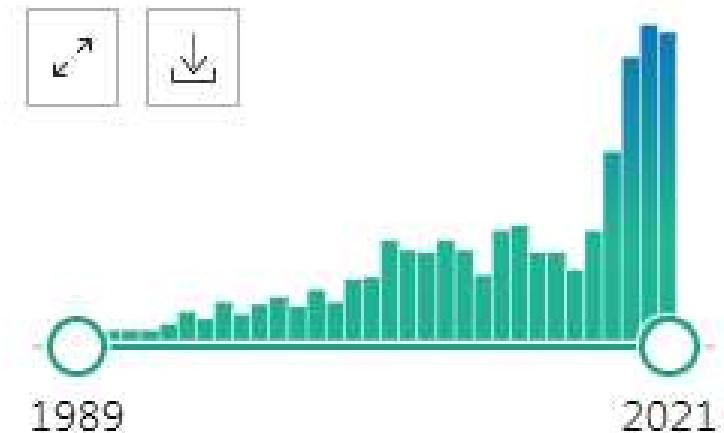
# Intelligenza artificiale e mammografia

- Crescente interesse in ambito scientifico
- Oltre 3.200 risultati su PubMed, considerando lo screening: oltre 2.000
- Marcato incremento pubblicazioni dal 2015

**Area di innovazione** per lo sviluppo di AI in campo medico:

- ✓ Digital Mammography Challenge (2019)
- ✓ Google Deepmind Health
- ✓ NHS trust
- ✓ Cancer Research UK
- ✓ Universities
- ✓ Start-up

RESULTS BY YEAR



## 1970s-1980s: "Old AI"

Low computing power, small data sets. Tools developed for specific, local decision support tasks. Developers, with clinicians, manually select patient attributes to include. Algorithms easy to explain, modest performance.

## 1990s-2000s

"AI winter" including in healthcare. Promise of AI seems limited.

## 2006-Now: acceleration of Machine Learning

Processing power, cloud computing, availability of digital data (e.g. medical images, medical literature, electronic health records) rapidly increase. Privately-funded developers rush to build 'black box' algorithms that automatically extract relevant attributes, often without clinical collaboration. Algorithms are hard to explain, performance apparently greatly improved.

# CAD – computer aided diagnosis

- Lo sviluppo di sistemi CAD per la mammografia risale a **fine anni '60**
- Obiettivo: assistere il radiologo nell'identificazione di lesioni sospette, evidenziando aree ad aumentata densità e micro-calcificazioni.
- Primi software con **approvazione FDA** nel 1998
- Studi di efficacia, **valore clinico** incerto per alto tasso **FP+**

**Table 1: Summary of Landmark Decisions or Studies on CAD**

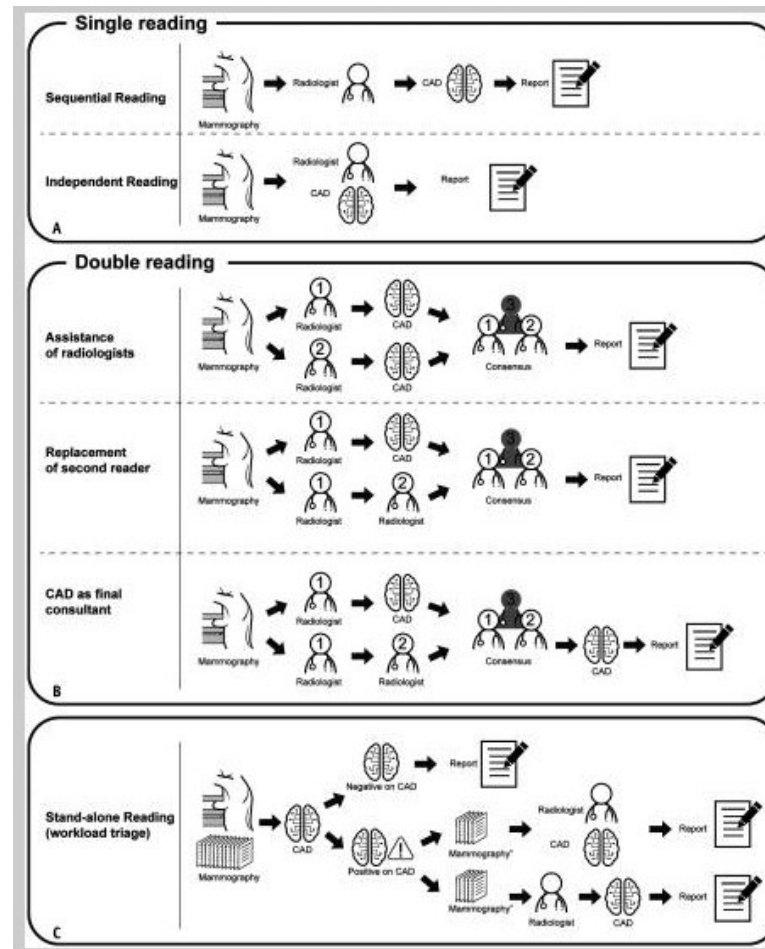
| Landmark and Year   | Reference                                 | Key Findings  |
|---|---|---|
| First studies on automated analysis of mammograms, 1967   | Winsberg et al (20)                       | Density within scanned mammograms could be automatically characterized  |
| U.S. FDA approval for a CAD application (R2 Image Checker; Hologic, Marlborough, Mass) to detect cancers, 1998                                  | U.S. FDA premarket approval database (21) | A computer system could identify and mark regions of interest on routine screening mammograms                             |
| Studies showing that CAD may detect cancers missed by radiologists  |   |   |
| 2000  | Warren Burhenne et al (22)                | CAD could reduce the false-negative rate by 77% without increasing the recall rate  |
| 2001  | Freer and Ulissey (23)                    | CAD led to a 19.5% increase in the number of cancers detected   |
| 2001  | Birdwell et al (24)                       | CAD marked 77% of cancers missed at screening mammography   |
| Studies showing potential for CAD as an independent second reader of screening mammograms   |   |   |
| 2004  | Destounis et al (25)                      | CAD could potentially decrease the false-negative rate at double reading by 39%   |
| 2008  | Gilbert et al (65)                        | Single reading with CAD has equal sensitivity at slightly higher recall (3.9% vs 3.4%) compared with double reading       |
| Studies showing that the large amount of false-positive findings generated by classic CAD may have a negative impact on radiologist performance |   |   |
| 2007  | Fenton et al (27)                         | Use of CAD only led to nonsignificant increase in sensitivity and overall lower accuracy (AUC = 0.87 vs 0.92; $P < .01$ ) |
| 2015  | Lehman et al (28)                         | Use of CAD significantly reduces sensitivity and is not associated with any improved performance parameters               |

Note.—AUC = area under the receiver operating characteristic curve, CAD = computer-aided diagnosis, FDA = Food and Drug Administration.

# Uso AI nell'ambito del tumore mammario

✓ Reti neurali e deep-learning

## AMBITI di APPLICAZIONE



# Efficacia

- Ambito in continua evoluzione
- Una recente **revisione (2019)**: accuratezza dei modelli AI da impiegarsi nell'ambito dello screening è variabile (**range: 69,2%-97,8%**, mediana AUC: 88,2%).
- Performance superiori a sistemi CAD classici
- Criticità:
  - Numerosità bassa
  - Dati retrospettivi, insufficiente follow-up
  - Dataset con un numero sproporzionatamente elevato di lesioni tumorali (>25%).
  - Mancanza di real world setting

## Detection of Breast Cancer with Mammography: Effect of an Artificial Intelligence Support System

Alejandro Rodríguez-Ruiz, MSc • Elizabeth Krupinski, PhD • Jan-Jurre Mordang, MSc • Kathy Schilling, MD •  
Sylvia H. Heywang-Köbrunner, MD, PhD • Ioannis Sechopoulos, PhD • Ritse M. Mann, MD, PhD

From the Department of Radiology and Nuclear Medicine, Radboud University Medical Center, PO Box 9101, 6500 HB Nijmegen, Geert Grooteplein 10, 6525 GA, Post 766, Nijmegen, the Netherlands (A.R.R., I.S., R.M.M.); Department of Radiology & Imaging Sciences, Emory University, Atlanta, Ga (E.K.); ScreenPoint Medical BV, Nijmegen, the Netherlands (J.J.M.); Lynn Women's Health & Wellness Institute, Boca Raton Regional Hospital, Boca Raton, Fla (K.S.); Referenzzentrum Mammographie München, Brustdiagnostik München and FFB, Munich, Germany (S.H.H.); and Dutch Expert Centre for Screening, Nijmegen, the Netherlands (I.S.). Received June 10, 2018; revision requested July 30; final revision received September 21; accepted September 28. Address correspondence to R.M.M. (e-mail: [Ritse.Mann@radboudumc.nl](mailto:Ritse.Mann@radboudumc.nl)).

Conflicts of interest are listed at the end of this article.

See also the editorial by Bahl in this issue.

Radiology 2019; 00:1–10 • <https://doi.org/10.1148/radiol.2018181371> • Content code: **BR**



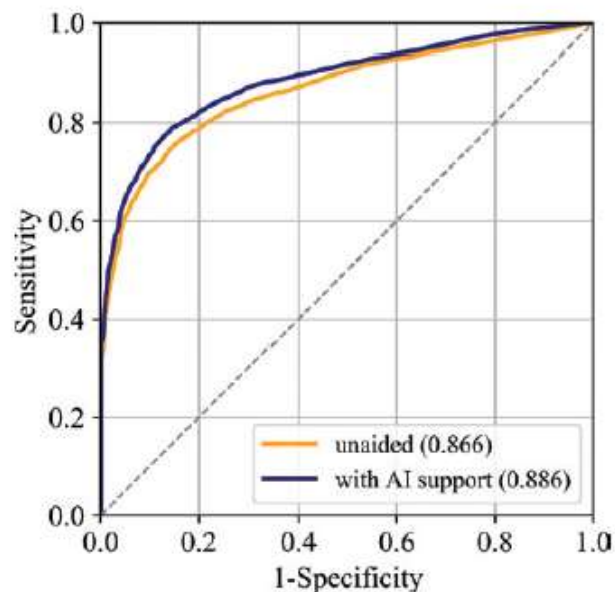
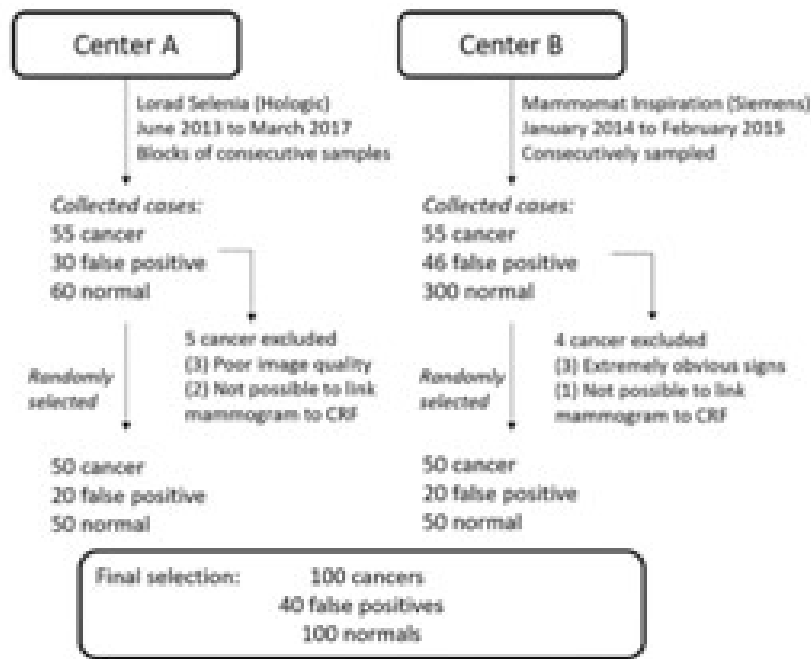
**Purpose:** To compare breast cancer detection performance of radiologists reading mammographic examinations unaided versus supported by an artificial intelligence (AI) system.

**Materials and Methods:** An enriched retrospective, fully crossed, multireader, multicase, HIPAA-compliant study was performed. Screening digital mammographic examinations from 240 women (median age, 62 years; range, 39–89 years) performed between 2013 and 2017 were included. The 240 examinations (100 showing cancers, 40 leading to false-positive recalls, 100 normal) were interpreted by 14 Mammography Quality Standards Act–qualified radiologists, once with and once without AI support. The readers provided a Breast Imaging Reporting and Data System score and probability of malignancy. AI support provided radiologists with interactive decision support (clicking on a breast region yields a local cancer likelihood score), traditional lesion markers for computer-detected abnormalities, and an examination-based cancer likelihood score. The area under the receiver operating characteristic curve (AUC), specificity and sensitivity, and reading time were compared between conditions by using mixed-models analysis of variance and generalized linear models for multiple repeated measurements.

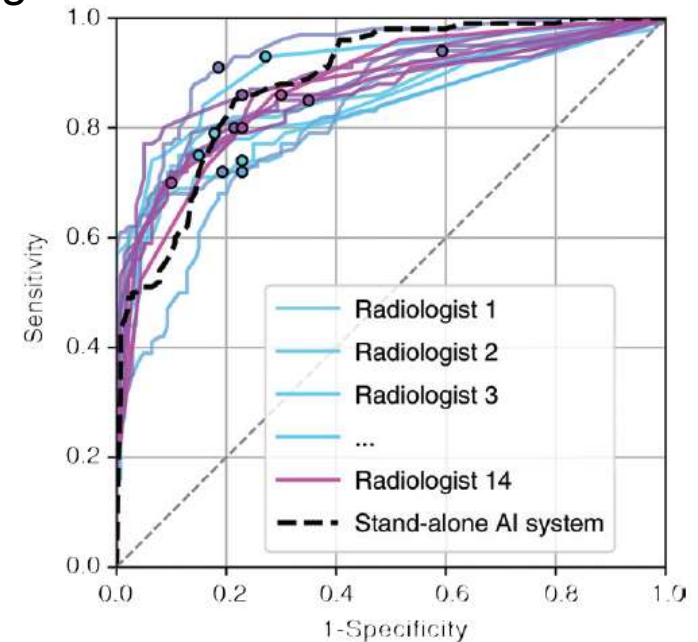
**Results:** On average, the AUC was higher with AI support than with unaided reading (0.89 vs 0.87, respectively;  $P = .002$ ). Sensitivity increased with AI support (86% [86 of 100] vs 83% [83 of 100];  $P = .046$ ), whereas specificity trended toward improvement (79% [111 of 140] vs 77% [108 of 140];  $P = .06$ ). Reading time per case was similar (unaided, 146 seconds; supported by AI, 149 seconds;  $P = .15$ ). The AUC with the AI system alone was similar to the average AUC of the radiologists (0.89 vs 0.87).

**Conclusion:** Radiologists improved their cancer detection at mammography when using an artificial intelligence system for support, without requiring additional reading time.

Published under a CC BY 4.0 license.



- I radiologi migliorano la loro performance con il supporto della AI ( $p=0.002$ ), con miglioramento per 12 radiologi su 14, con miglioramento anche in analisi per sottogruppi.
- Miglioramento più marcato per radiologi meno esperti.
- La performance della AI da sola è risultata simile alla media dei radiologi.



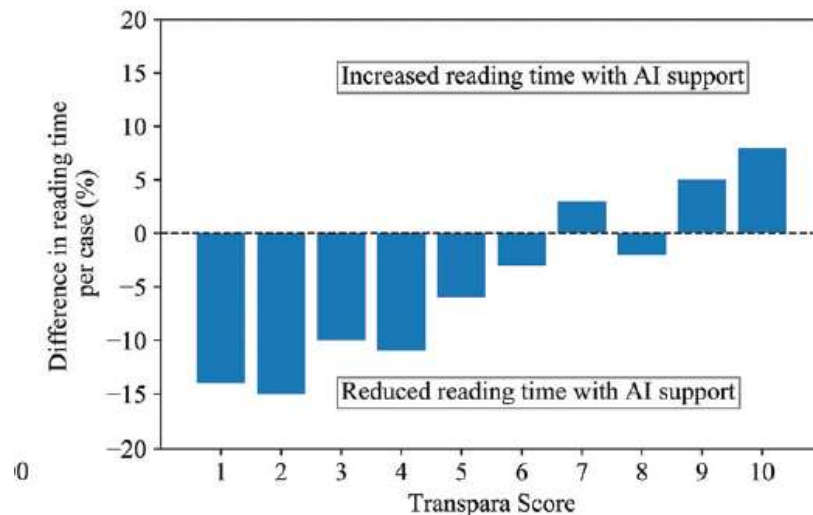
**Figure 5:** Receiver operating characteristic curves for radiologists reading mammograms unaided and stand-alone artificial intelligence (AI) computer system (Transpara; Screenpoint, Nijmegen, the Netherlands). Circles indicate the radiologists' operating points at Breast Imaging Reporting and Data System category 3 thresholds. (Reprinted, with permission, from reference 82.)



# Supporto alla lettura radiologica

Rodriguez-Ruiz et al.: un AI che conferisce un **punteggio da 1 a 10** per definire la probabilità di lesioni maligne risulta associata positivamente con una riduzione dei tempi di lettura radiologici. In particolare, i radiologi **riducono il tempo di lettura** in caso di punteggi bassi, con una riduzione complessiva del tempo di lettura del 5% per un data set di screening.

- Possibili vantaggi superiori perché nello studio tempi in generale superiori a quelli della pratica clinica.
- Incremento della sicurezza del radiologo, con possibilità di velocizzare la lettura delle immagini chiaramente “normali”, consentendo tempi superiori in caso di immagini sospette.
- I risultati promettenti, suggeriscono il **possibile ruolo di AI come II lettore indipendente** + eventuale II radiologo che interviene solo un caso di discrepanze, con vantaggio in termini di carico di lavoro. L'effetto su tasso di richiamo e VPP è ancora conosciuto.



b.

Article | [Published: 01 January 2020](#)

# International evaluation of an AI system for breast cancer screening

[Scott Mayer McKinney](#) , [Marcin Sieniek](#), [...] [Shravya Shetty](#) 

*Nature* **577**, 89–94 (2020) | [Cite this article](#)

**69k** Accesses | **493** Citations | **3582** Altmetric | [Metrics](#)

- Uso della **classificazione AI** per ridurre il carico di lavoro radiologico nel processo di doppia lettura in **UK**, preservando gli standard di qualità dello screening.
- Scenario di simulazione di **AI come II lettore: performance equivalenti** allo standard, con un **risparmio di tempo dell'88%** dello sforzo del secondo lettore.

a

|                          | Sensitivity (%)<br>( <i>n</i> = 414) | Specificity (%)<br>( <i>n</i> = 25,422) | Simulated reduction<br>of second reader<br>workload (%) |
|--------------------------|--------------------------------------|---|---|
| AI as second reader (UK) | 66.66                                | 96.26                                   | 87.98   |
| Existing workflow (UK)   | 67.39                                | 96.24                                   | -   |
| 95% CI on the difference | (-2.68, 1.23)                        | (-0.13, 0.17)                           | -   |

# Valutazione degli errori di performance

- ✓ Processo a **doppio senso**
- ✓ Comparazione degli errori AI e errori lettura radiologica
- ✓ La maggior parte dei casi in cui solo AI ha individuato la lesione si sono rilevati essere tumori invasivi.
- ✓ Nei casi in cui solo il radiologo ha individuato la lesione tumorale, si ripartiscono equamente tra carcinomi invasivi e carcinomi in situ.
- ✓ Non si evidenziano associazioni significative con altre caratteristiche tumorali: dimensione, grado, marker molecolari.



## Can artificial intelligence reduce the interval cancer rate in mammography screening?

Kristina Lång<sup>1,2</sup> · Solveig Hofvind<sup>3,4</sup> · Alejandro Rodríguez-Ruiz<sup>5</sup> · Ingvar Andersson<sup>1,2</sup>

Received: 11 September 2020 / Revised: 18 December 2020 / Accepted: 12 January 2021 / Published online: 23 January 2021  
© The Author(s) 2021

- Studio retrospettivo mostra che AI è in grado di identificare il **19% dei tumori intervallo** alla mammografia di screening precedente, che in analisi successive mostrano almeno segni minimi di malignità.
- Si evidenzia in questi casi una **corretta localizzazione** da parte AI, ottimizzando le modalità diagnostiche supplementari.
- AI è in grado di riconoscere **soprattutto lesioni aggressive**.
- Si evidenzia una correlazione tra il punteggio conferito da AI e la classificazione come veri negativi, falsi negativi o segni minimi.

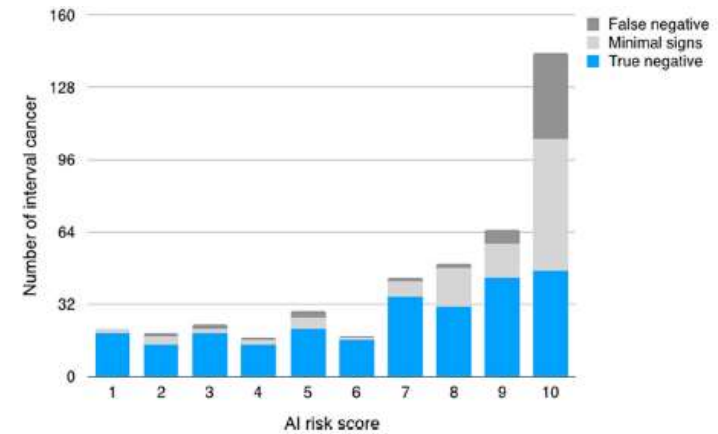


Fig. 1 Distribution of interval cancer and classification groups of interval cancer by AI risk score

# Possibile ruolo AI in diagnosi tumore mammario e confronto con CAD tradizionale

**Table 2: Potential Use of CAD for Cancer Diagnosis**

| Task  | Conventional CAD  | AI-based CAD  | Advantages of AI   |
|---|---|---|--|
| Prevention of overlook errors                         | Yes, prompts potential abnormalities                                  | Yes, provides prompts or heat map to show most suspicious abnormalities | Fewer false-positive findings  |
| Interactive decision support                          | Possible, but not commonly used                                       | Possible  | Provides feedback on areas not deemed abnormal enough to prompt  |
| Determination of cancer likelihood                    | No, only lesion-based likelihoods of malignancy are provided          | Yes, case-based scores show likelihood of cancer presence               | Confirmation of normality may increase confidence and speed of evaluation  |
| Independent second reading                            | No, CAD findings lead to improved detection but also increased recall | Potentially, as sensitivity is on par with that of radiologists         | Strong reduction of workload when double reading is standard, but clinical validation is still required; may increase recall |
| Dismissal of normal cases (independent first reading) | No, absence of malignancy is unreliable                               | Potentially, as sensitivity is on par with that of radiologists         | May lead to dismissal of some human-detected cancers, potential is dependent on cost-benefit evaluation                      |

Note.— AI = artificial intelligence, CAD = computer-aided diagnosis.

# Oltre la mammografia...

**Table 5: Status of Current AI Algorithms**

| AI Use              | Current Performance   | Clinical Role  | Experimental Strategies  |
|---------------------|---|--|--|
| Mammography         | On par with human readers for lesion detection and classification | In use as an aid for concurrent reading of mammograms to improve human performance                           | Preselection of normal mammograms; independent second reader of screening mammograms; risk prediction for the development of breast cancer   |
| DBT                 | On par with human readers for lesion detection and classification | In use as an aid for concurrent reading of mammograms to improve human performance and decrease reading time | Independent reading of DBT examinations; creation of “intelligent” synthetic mammograms that highlight potential abnormalities   |
| Handheld US         | On par with human readers for lesion classification               | In use to aid in lesion classification detected by human readers   | Incorporation of multiparametric data, including Doppler and elastography; extraction of quantitative data for lesion classification   |
| Automated breast US | Below human performance for lesion detection and classification   | In use to prevent overlook errors and decrease reading time  | Improvement of detection and classification algorithms; creation of intelligent summary images   |
| MRI                 | Below human performance for lesion detection and classification   | Experimental only  | Incorporation of multiparametric data, including dynamics, T2-weighted imaging, and DWI; improvement of detection and classification algorithms; enhanced image reconstruction; creation of intelligent summary images |

Note.—AI = artificial intelligence, DBT = digital breast tomosynthesis, DWI = diffusion-weighted imaging.

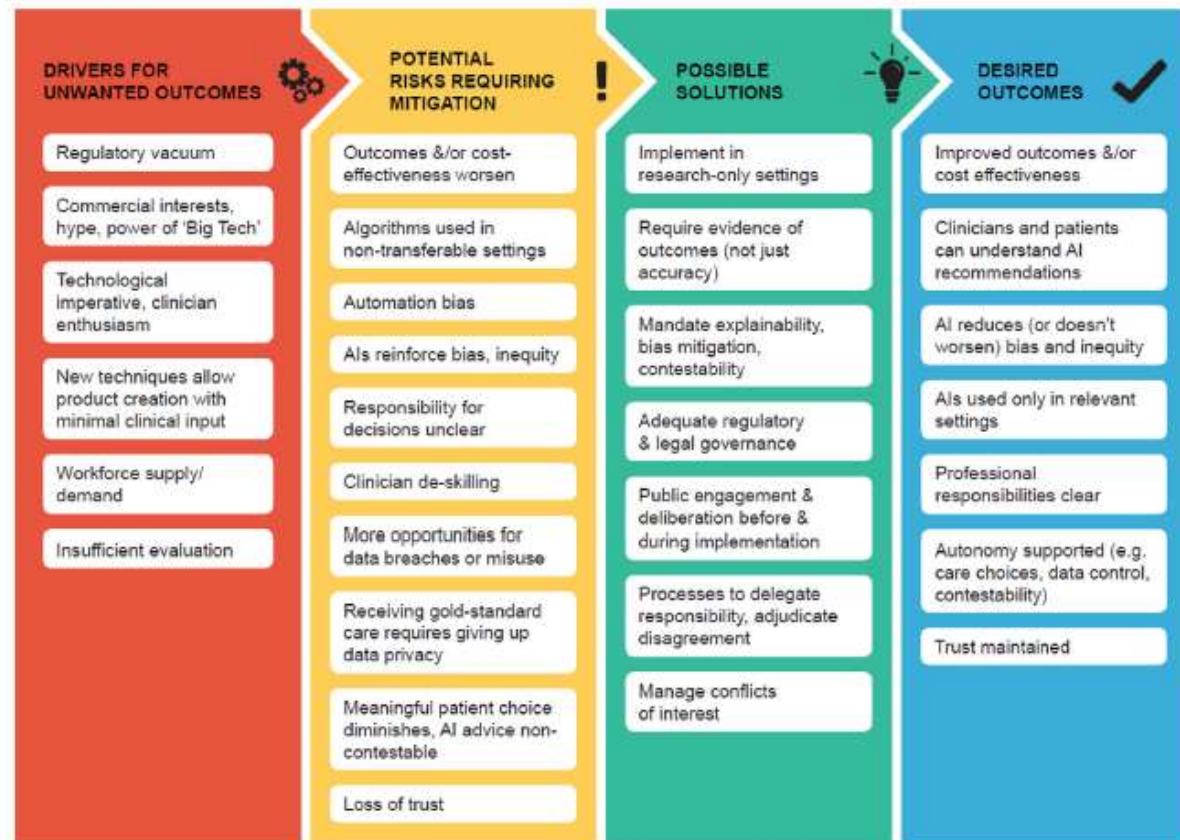
# Un focus sulla tomosintesi

| Issues with DBT  | Potential AI Solution   |
|--|---|
| Radiation  | Deep learning–based reconstruction of synthetic images may lead to a reduction in radiation dose                                |
| More images  | Generation of AI-enhanced synthetic mammograms optimized for a specific task  |
| Calcifications not always clearly visible  | Enhance the conspicuity of calcifications   |
| Longer reading times compared with digital mammography   | Confirmation of absence of lesions; help localize a finding detected on multiple images   |
| Reduces the effect of tissue superimposition and detects more benign and malignant abnormalities | Improved lesion classification with the three-dimensional information available in the DBT volume                               |
| Limited sensitivity in dense breast tissue   | Deep learning algorithms may remove normal fibroglandular tissue from composite images to “see through” the dense breast tissue |

Note.—AI = artificial intelligence, DBT = digital breast tomosynthesis.

# ...al di là dell'efficacia

- Implicazioni **etiche, legali e sociali**
- Prese di posizione a livello governativo, intergovernativo, professionale e industriale
  - Interesse
  - Preoccupazione riguardo il bilancio rischi/benefici





# Explainability

- Effetto “**scatola nera**”
- È possibile per l'operatore conoscere **come funziona l'algoritmo**, e quali “**valori**” codifica
  - Priorità a minimizzare FN o FP
  - Differenze di performance in base a caratteristiche del seno
  - Saliency maps
- Area attiva di ricerca per l'integrazione AI nella pratica clinica, continuo processo di QA

# Effetti su esiti paziente e sistema sanitario

- Studi di **outcome clinico**
- Predominanza studi comparativi (*AI vs. clinico*)
- Necessità di studi di **fattibilità in *real-world setting***

Obiettivo attuale: migliorare accuratezza e efficienza del processo di screening

In futuro: combinare immagini, risultati AP, dati clinici e genetici per consentire al sistema di discriminare lesioni tumorali clinicamente significative e sovra-diagnosi

# Trasferibilità

- Dataset di **addestramento e validazione**
- Rischio di fallire in **setting differenti**
  - Ruolo **trasparenza dei dati** (per test indipendenti)
  - **Pre-test in nuova coorte e setting differenti**
- Necessità di continuo **controllo qualità e versatilità** con evoluzione per nuovi dati.
- **Equità**

# Privacy

- Proprietà dei dati, consenso, protezione dati.
- Anonimizzazione
- Espansione uso dei dati incrementa le possibilità di data-leakage e re-identificazione
- Il GDPR UE enfatizza l'importanza di implementare stringenti regolamenti su privacy e sicurezza dei dati in AI-CAD, incorporando elementi di explainability dell'AI e interpretabilità delle decisioni.

# Rischio legale e responsabilità

- Vuoto regolatorio
- GDPR 2016/79
- **Collaborazione enti regolatori e sviluppatori** per assicurare adeguata protezione, favorendo al contempo sviluppo AI
- Accettabilità Radiologo
  - Responsabilità professionale
  - Implicazioni in variazioni attività e capacità professionali
- Paziente

## Effetto su pratica clinica

- ***Anchoring effect***: eccessiva fiducia in lettura automatizzata
- **Bandwagon effect**: tendenza del lettore ad allinearsi al risultato AI
- Ruolo formazione
- ***Human-in-the-loop***: offrire ai clinici l'opportunità di interagire con il modello decisionale al fine di migliorare la performance di sistema e agire come salvaguardia da possibili errori.



SOLICITED REVIEW / Breast imaging

## Artificial intelligence and breast screening: French Radiology Community position paper



I. Thomassin-Naggara<sup>a,b,\*</sup>, C. Balleyguier<sup>c</sup>,  
L. Ceugnart<sup>d</sup>, P. Heid<sup>e</sup>, G. Lenczner<sup>f</sup>, A. Maire<sup>g</sup>,  
B. Séradour<sup>d</sup>, L. Verzaux<sup>e</sup>, P. Taourel<sup>c</sup>, and Conseil  
national professionnel de la radiologie et imagerie  
médicale (G4)

### AI Research issues

- To validate artificial intelligence (AI) models in a French population.
- To test the ability of AI systems to integrate comparisons with previous mammograms.
- To test the performance of AI algorithms on combined two-view MG analysis.
- To compare two readings versus one reading + AI software.
- To test AI algorithms on 2DMG according to the different vendors.

## Cosa aspettarsi da AI?

- Ottimizzazione dosi rX
- Qualità acquisizione immagini
- Controllo qualità
- Lettura densità
- Supporto alla lettura radiologica

### AI Research issues

- Organization issue
  - To identify reasons for non-participation in the breast screening program in the French national medical database.
  - To analyze differences in the cancer detection rates at a local level (town, districts) to better select women at higher risk.
  - To establish prediction parameters to improve cost effectiveness of organized breast screening programs.
- Reading issues
  - To predict comorbidities (vascular calcification correlated with heart attack).
  - To improve screening performance of radiologists with different levels of experience.
  - To better select patients for ultrasonographic evaluation.

## Il futuro?

### AI research issues

- To develop an AI model that improves the definition of breast risk assessment to select patients for different screening modalities (HR+ patients, breast density, B3 lesions) – better define intermediate-risk patients.
- To build an AI model that integrates all epidemiological and clinical data (physical activity) to improve personalized screening programs.

Documento SIRM

## Intelligenza Artificiale in Radiologia

Giovanni Pasceri, Corrado Bibbolino, Michele Stasi,  
Marco Foracchia, Gigliola Pirotta, Francesca Coppola,  
Emanuele Neri, Roberto Grassi

2020

✓ Documento della **Società Italiana Radiologia Medica e Interventistica**

✓ Obiettivo: fornire indicazioni utili a garantire la qualità nell'erogazione di prestazioni sanitarie rese utilizzando l'intelligenza artificiale, e ad esplicitare la portata e i limiti di queste applicazioni.

✓ Cooperazione AI e radiologo nel processo decisionale

✓ Responsabilità professionale

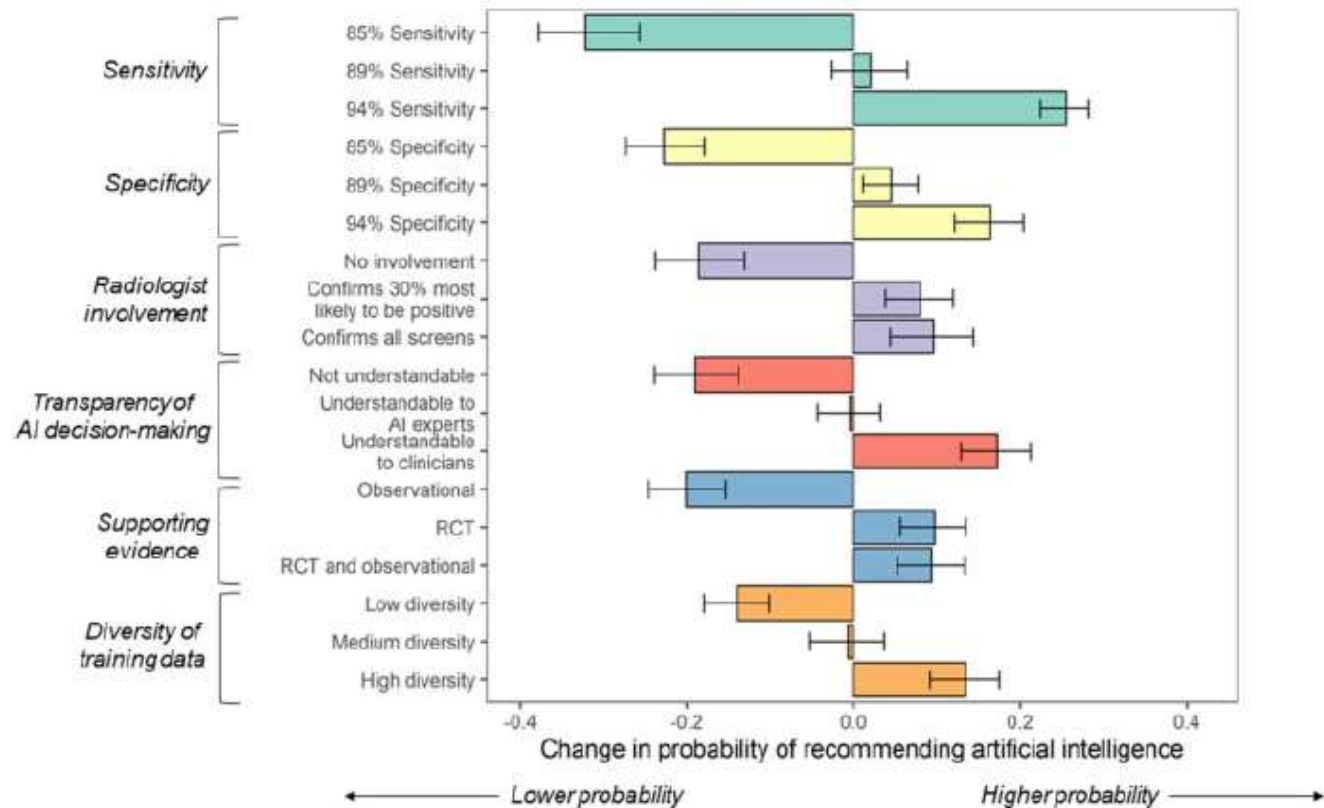


Research and Applications

### Artificial intelligence in breast cancer screening: primary care provider preferences

Nathaniel Hendrix<sup>1</sup>, Brett Hauber<sup>1,2</sup>, Christoph I. Lee<sup>3,4,5</sup>, Aastha Bansal<sup>1</sup>, and David L. Veenstra<sup>1</sup>

- Indagine su preferenze dei provider sanitari in sviluppo AI per screening
- La maggior parte degli intervistati (75%) accetterebbe l'uso di AI come triage per esami negativi in ambito screening



# Accettabilità paziente/popolazione

- Survey (2019) USA:
  - Maggiore supporto a diffusione AI in ambito clinico in soggetti sani, maschi, titolo di studio elevato, buona conoscenza in ambito tecnologico
  - Bassa consapevolezza in generale dei potenziali rischi AI
- Implicazioni non previste?

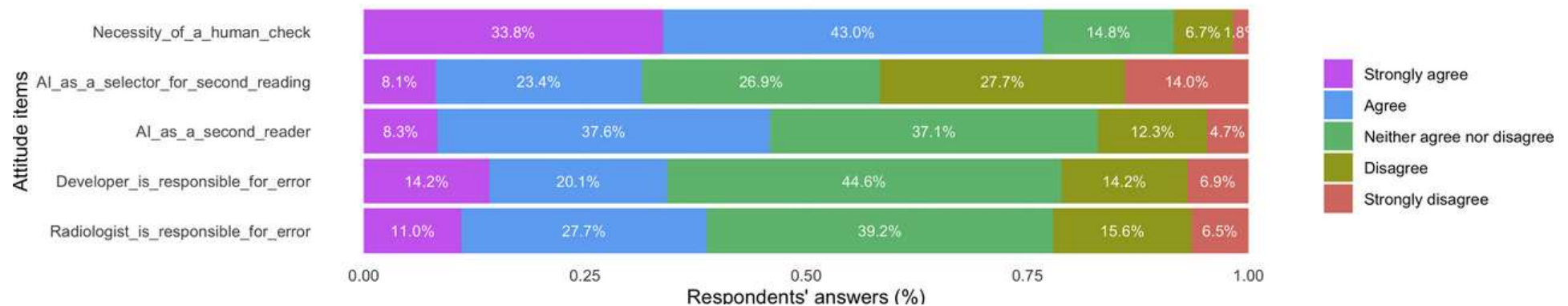


Original Article  
Data Science

## Artificial Intelligence in Screening Mammography: A Population Survey of Women's Preferences

Yfke P. Ongena PhD<sup>a</sup>, Derya Yakar MD, PhD<sup>b</sup>, Marieke Haan PhD<sup>c</sup>, Thomas C. Kwee MD, PhD<sup>d</sup>

- Survey su **997 donne olandesi (range: 16-75 anni)**
- Valutare l'atteggiamento della popolazione sull'uso di AI per l'interpretazione delle mammografie di screening
- Valutazione su **scala Lickert a 5 punti** di: necessità controllo umano, AI come triage per identificare immagini che richiedono una seconda lettura, AI come secondo lettore, responsabilità di eventuali errori (sviluppatore vs. radiologo)



In analisi multivariata emerge il **ruolo di educazione** per fiducia nell'uso AI e su responsabilità.

Altri fattori rilevanti: importanza discussione risultati, fiducia AI in ambito clinico, età.



# Conclusioni

- ✓ L'uso di AI in radiologia e in particolare nello screening mammografico presenta **notevoli potenzialità** nel campo del miglioramento della sicurezza del paziente e esiti clinici e dell'organizzazione dello screening (maggiore efficienza di processo).
- ✓ La mammografia è un **ambito di ricerca promettente** per le AI dato il problema clinico rilevante, l'organizzazione del flusso di lavoro, la disponibilità di ampi dataset per lo sviluppo e la validazione.
- ✓ Permangono delle **criticità e ambiti di ricerca** da approfondire
- ✓ Il **miglior risultato deriva dall'integrazione dei vantaggi del lavoro "umano" e della "macchina"** → necessità di studi di validazione e fattibilità che tengano conto del flusso di lavoro e dell'interazione con il clinico e il paziente.