MAGI
Mammografi & Kunstig Intelligens

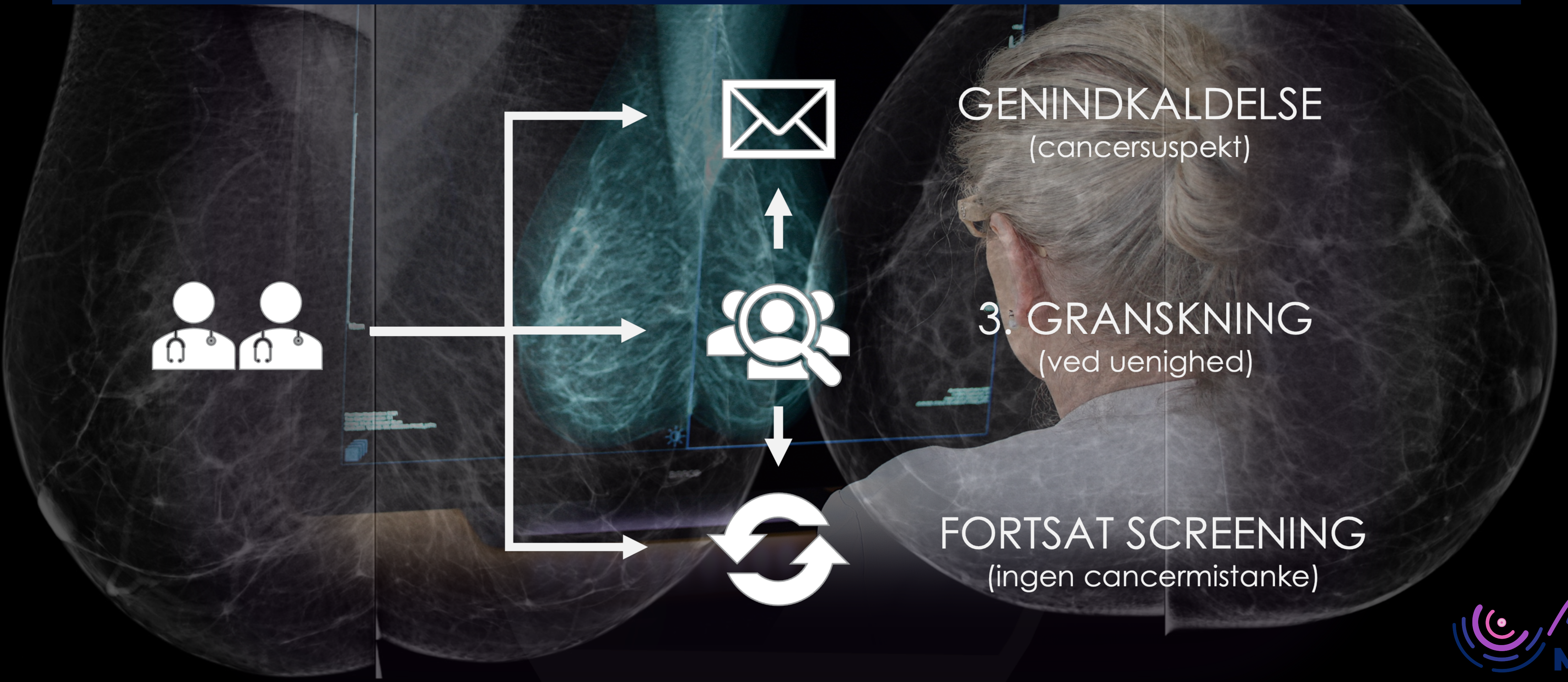
Fra forskning til implementering – Erfaringer fra MAGI projektet

Mohammad Talal Elhakim, Læge, Ph.d-studerende

Region Syddanmark
CIMT Center for Innovativ Medicinsk Teknologi
Centre for Clinical Artificial Intelligence
Radiologisk Forsknings- & Innovationsenhed Syddansk Universitet
OUH Odense Universitetshospital
KØBENHAVNS UNIVERSITET

Hvordan? Hvorfor? Hvor? Hvornår?

Nuværende screeningspraksis i Region Syddanmark



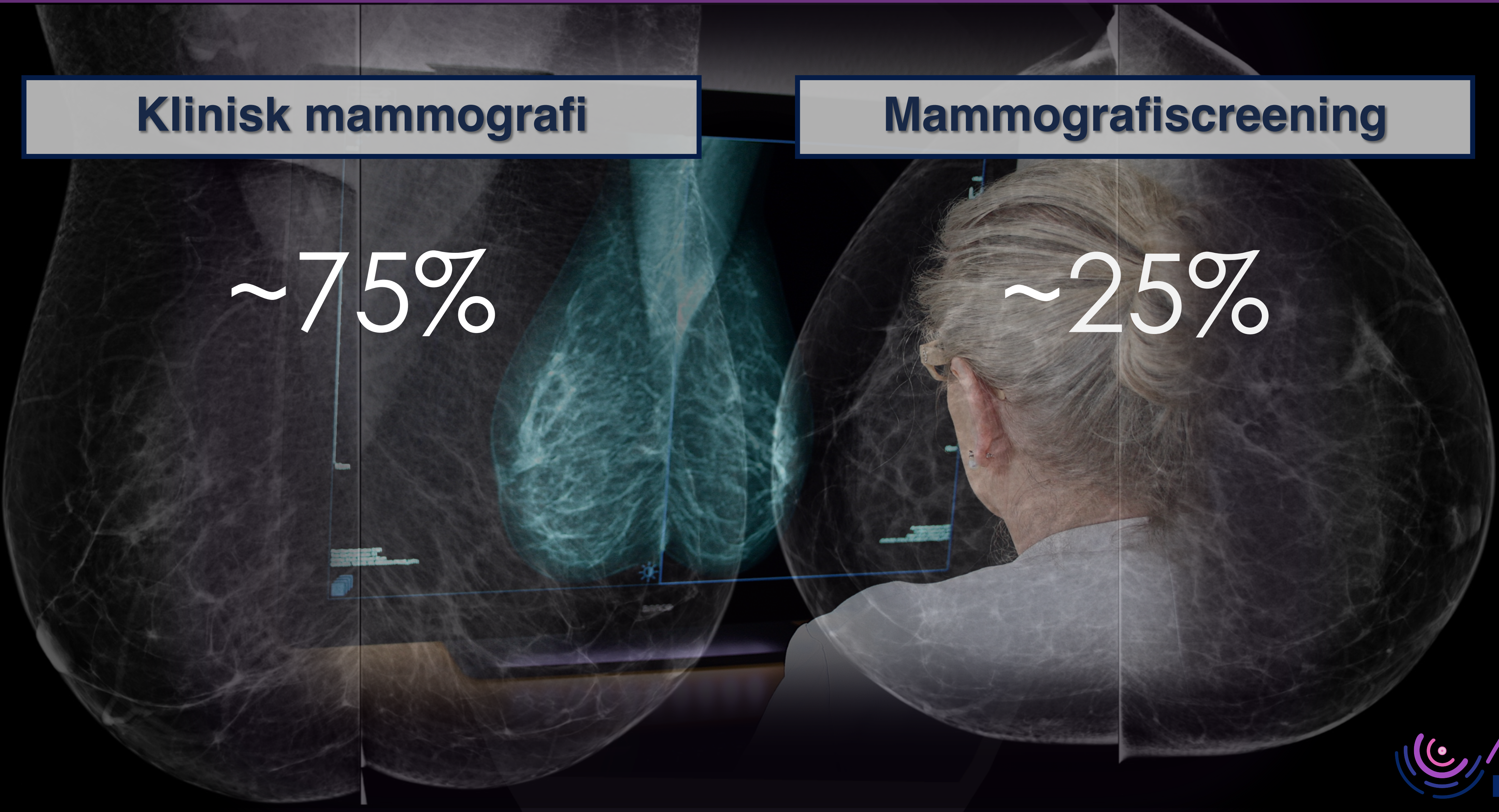
Hvordan? | Hvorfor? | Hvor? | Hvornår?

Klinisk mammografi

~75%

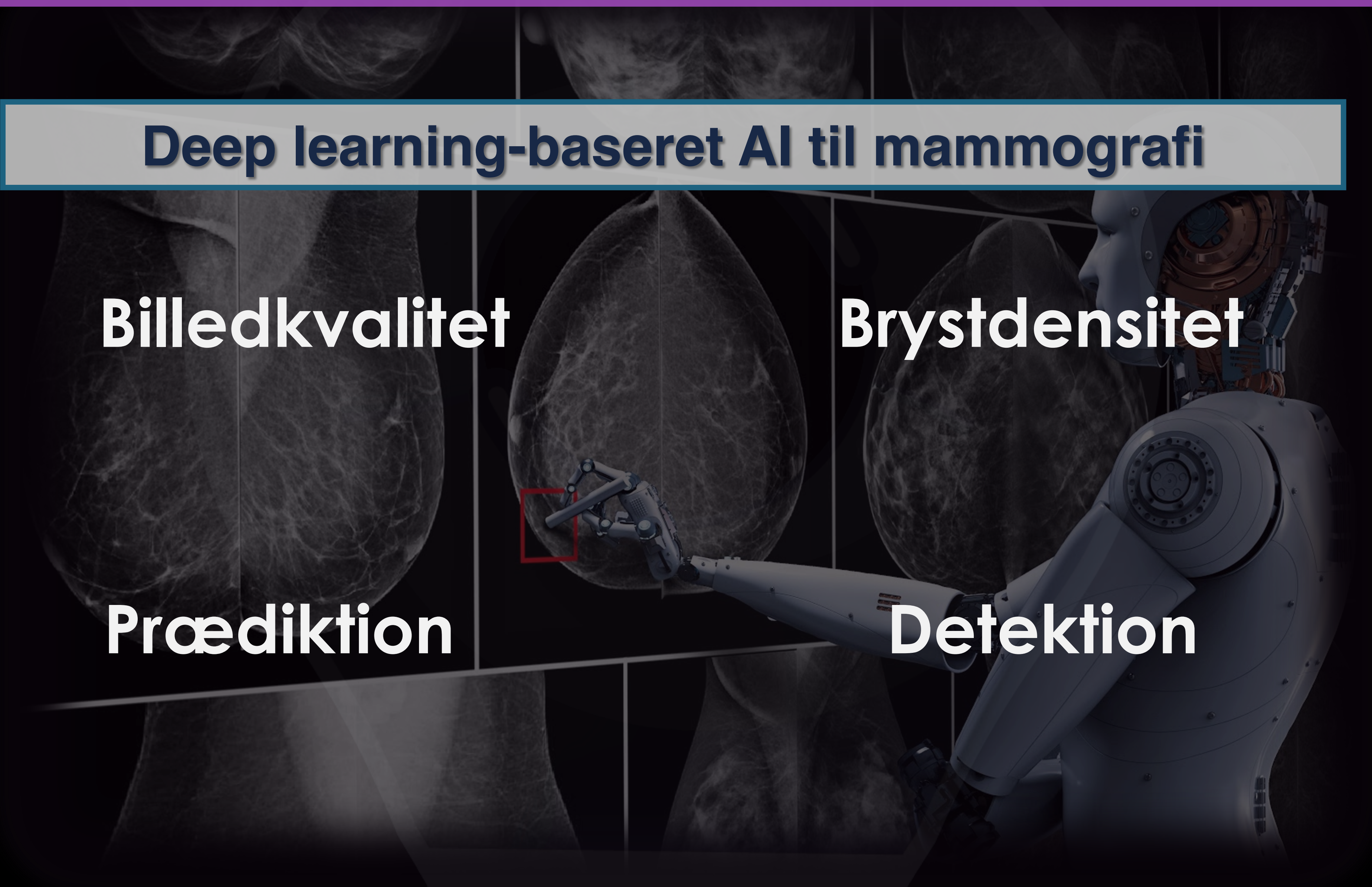
Mammografiscreening

~25%

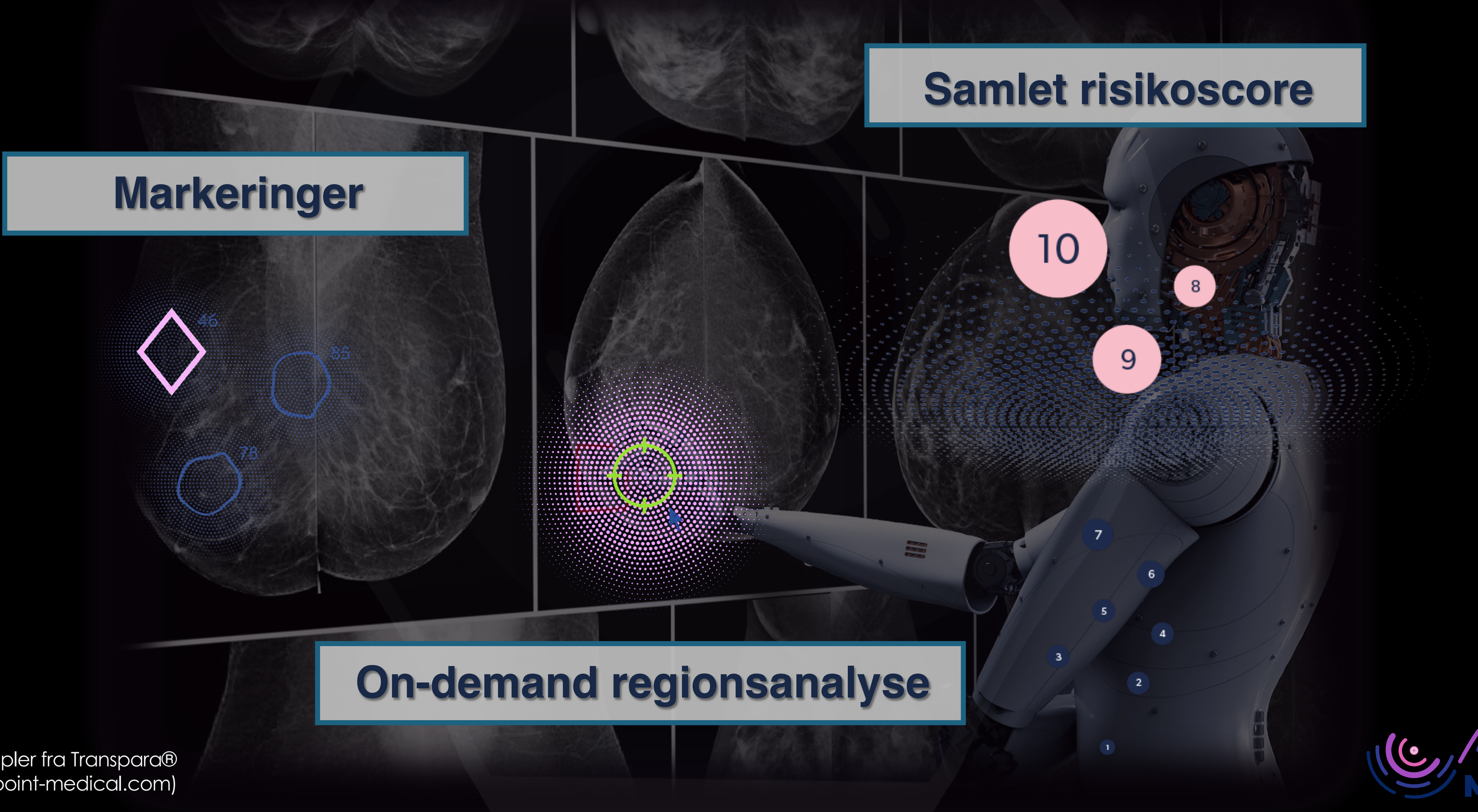


Hvordan? | Hvorfor? | Hvor? | Hvornår?

Deep learning-baseret AI til mammografi

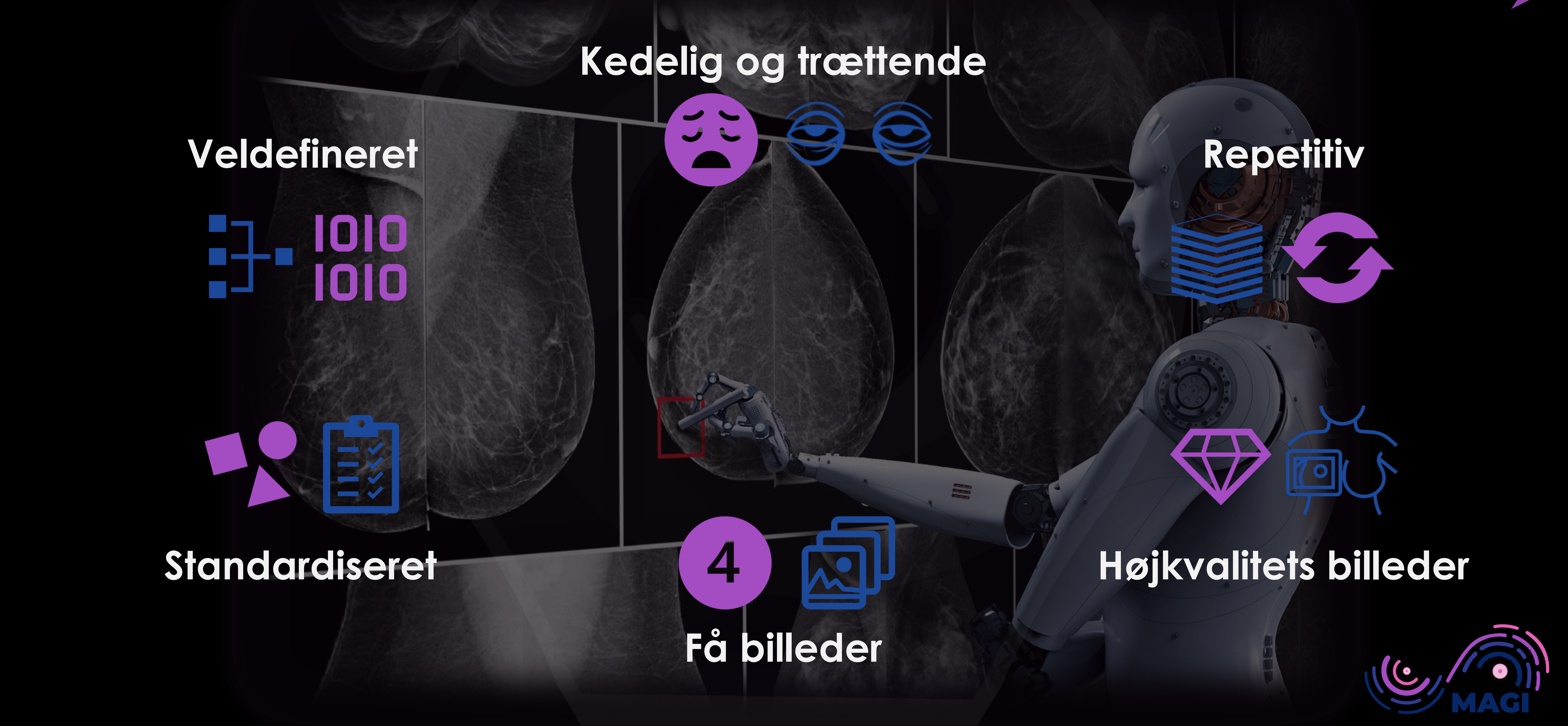


Hvordan? | Hvorfor? | Hvor? | Hvornår?



Funktionseksempler fra Transpara®
(https://screenpoint-medical.com)

Hvordan? | **Hvorfor?** | Hvor? | Hvornår?



Hvordan? | Hvorfor? | **Hvor?** | Hvornår?

BESLUTNINGSSTØTTE - AI som støtteværktøj

+AI +AI

+AI

GENINDKALDELSE
(cancersuspekt)

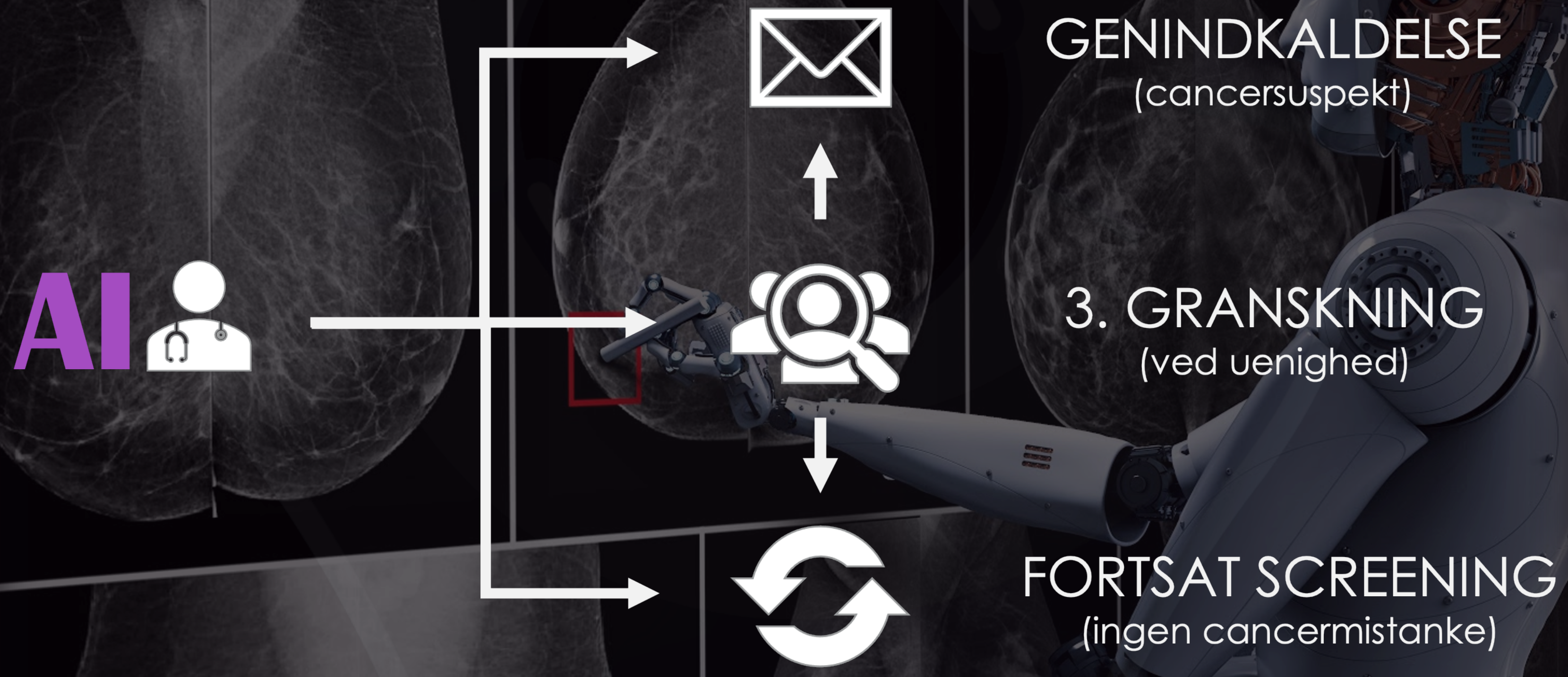
3. GRANSKNING
(ved uenighed)

FORTSAT SCREENING
(ingen cancermistanke)

MAGI

Hvordan? | Hvorfor? | **Hvor?** | Hvornår?

SELVSTÆNDIG GRANSKER – AI som erstatning for én radiolog



Hvordan? | Hvorfor? | **Hvor?** | Hvornår?

TRIAGERING – AI som selvstændig gransker i udvalgte tilfælde

AI

HØJ RISIKO

LAV RISIKO

GENINDKALDELSE
(cancersuspekt)

3. GRANSKNING
(ved uenighed)

FORTSAT SCREENING
(ingen cancermistanke)

MAGI

Hvordan? | Hvorfor? | **Hvor?** | Hvornår?

Slutmål for implementering af AI i brystkræftscreening

Non-inferior specificitet

Færre interval cancere

Ingen skævvridning mod detektion af klinisk ubetydelige cancere

Hvordan? | Hvorfor? | Hvor? | **Hvornår?**

VIDENSKAB

Kunstig intelligens til cancerdiagnostik i brystkræftscreening

Mohammad Talal Elhakim[1,2], Ole Graumann[1,2], Lisbet Brønsro Larsen[1,2], Mads Nielsen[3] & Benjamin Schnack Rasmussen[1,2]

STATUSARTIKEL

1) Radiologisk Afdeling, Odense Universitetshospital
2) Radiologisk Forsknings- og Innovationsenhed, Syddansk Universitet
3) Datalogisk Institut, Københavns Universitet

Ugeskr Læger 2020;182:V06200423

Brystkræft er den hyppigst forekommende cancer hos kvinder i Danmark med den højeste dødelighed efter lungekræft [1]. Organiseret, populationsbaseret mammografiscreening er implementeret med det formål at reducere mortaliteten af brystkræft gennem tidlig detektion, og efter udrulningen af det nationale screeningsprogram i 2007-2010 har antallet af deltagere været støt stigende [2]. Pga. en øget byrde og sparsomme radiologressourcer er området præget af et stort arbejdspres. Desuden er mammografiscreening i litteraturgennemgange af internationale studier forbundet med et vist antal falsk positive [3] samt risiko for overdiagnosticering og dermed potentiel overbehandling [4], om end danske tal efter flere screeningsrunder indikerer et forholdsvist lavere niveau [5].

Kunstig intelligens (AI) har i de senere år tiltrukket sig tiltagende interesse som et potentielt værktøj, der kan øge effektiviteten og kvaliteten inden for billeddiagnostik. Et nyligt publiceret studie fra Google Health i tidsskriftet Nature har bl.a. medført stor medieomtale for konklusionen om, at AI overgår radiologer i at diagnosticere brystkræft [6]. Formålet med denne artikel er at: 1) præsentere læseren for den nyeste viden om AI til cancerdiagnostik i brystkræftscreening, 2) gennemgå metodologiske problemstillinger i den eksisterende litteratur om AI til mammografi og 3) diskutere mulighederne samt udfordringerne i at implementere AI-løsninger i klinisk praksis.

NUVÆRENDE PRAKSIS I BRYSTKRÆFTSCREENING

Alle danske kvinder i aldersgruppen 50-69 år inviteres hvert andet år til screening for brystkræft. Undersøgelbryst i to projektioner. Screeningsbillederne læses ved dobbelt uafhængig granskning af rutinerede mammaradiologer, og i tilfælde af uenighed foretages en tredje granskning mhp. konsensus. Screeningen afsluttes enten ved, at kvinden fortsætter i screeningsprogrammet, eller at hun bliver genindkaldt til nærmere udredning med klinisk mammografi (palpation, supplerende billeddiagnostik og evt. nålebiopsi). Digital brysttomosyntese (DBT), som er en tredimensionel mammografi, foretages i udvalgte tilfælde. Videre udredning og behandling foregår i multidisciplinært organiseret pakkeforløb.

Kvaliteten af screeningsindsatsen monitoreres løbende af Dansk Kvalitetsdatabase for Mammografiscreening [2] ud fra en række organisatoriske krav og kliniske kvalitetsindikatorer iht. europæiske retningslinjer [7] og nationale guidelines [8]. Opgørelserne fra de seneste screeningsrunder viser, at den nationale mammografiscreening har en høj lægefaglig standard, men fortsat ressourcemæssige udfordringer [2].

COMPUTERBASEREDE LØSNINGER TIL MAMMOGRAFI

De første computerassisterede detektions (CAD)-systemer blev i slutningen af 1990'erne lanceret som et støtteværktøj til at bistå radiologer i brystkræftdiagnostik med store forventninger om at reducere oversete cancere samt falsk positive. Med transitionen fra konventionel røntgen på fotografisk film til digital mammografi har CAD inden for de seneste 20 år opnået bred popularitet særligt i USA, hvor der bruges enkeltgranskning i screening [9]. En overvægt af studier peger dog på, at CAD forringer den diagnostiske kvalitet, med bl.a. afledende falsk positive markeringer, uden reel

Ugeskrift For Læger, 2020

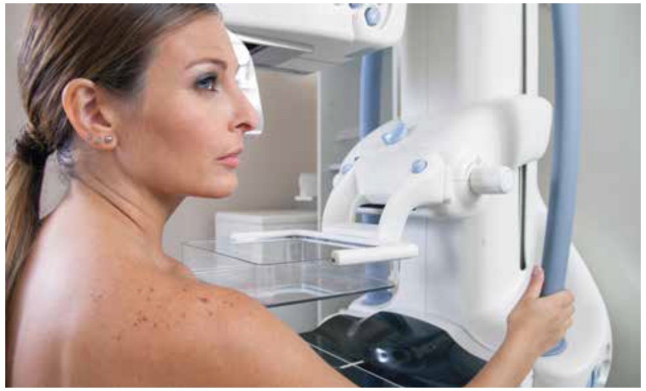
Tema: Billeddiagnostik

Fremtidens screening for brystkræft bruger kunstig intelligens

MAGI er et stort billeddiagnostisk valideringsprojekt, som undersøger, i hvilket omfang kunstig intelligens kan blive en fast del af screeningen for brystkræft. På sigt kan det både øge kvaliteten og aflaste radiologer.

Af Mohammad Talal Elhakim. Læge, ph.d.-studerende[1,2]
Camilla Stryhn. Projektleder[3,4]
Benjamin S. B. Rasmussen. Læge, ph.d., adjunkt[1,2,5]

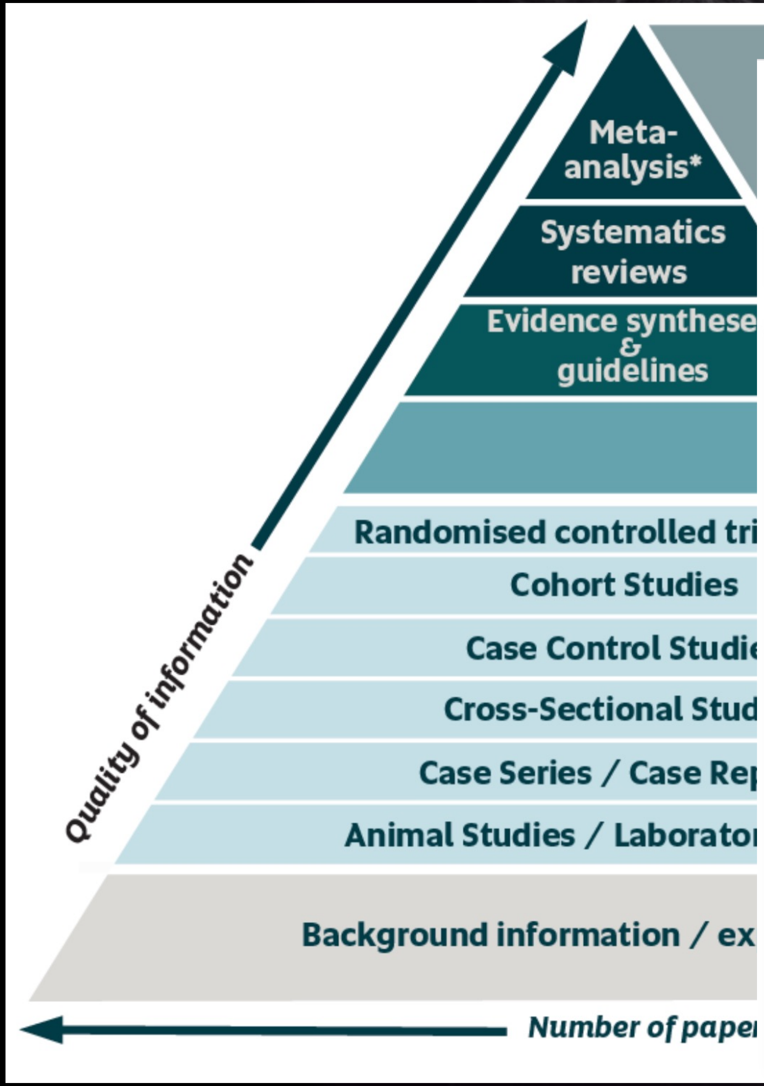
1: Radiologisk Afdeling - Odense Universitetshospital (OUH)
2: Radiologisk Forsknings- og Innovationsenhed - Syddansk Universitet (SDU)
3: Klinisk Udvikling - Innovation og Medicinsk Teknologivurdering, OUH
4: Center for Innovativ Medicinsk Teknologi (CIMT) - OUH/SDU

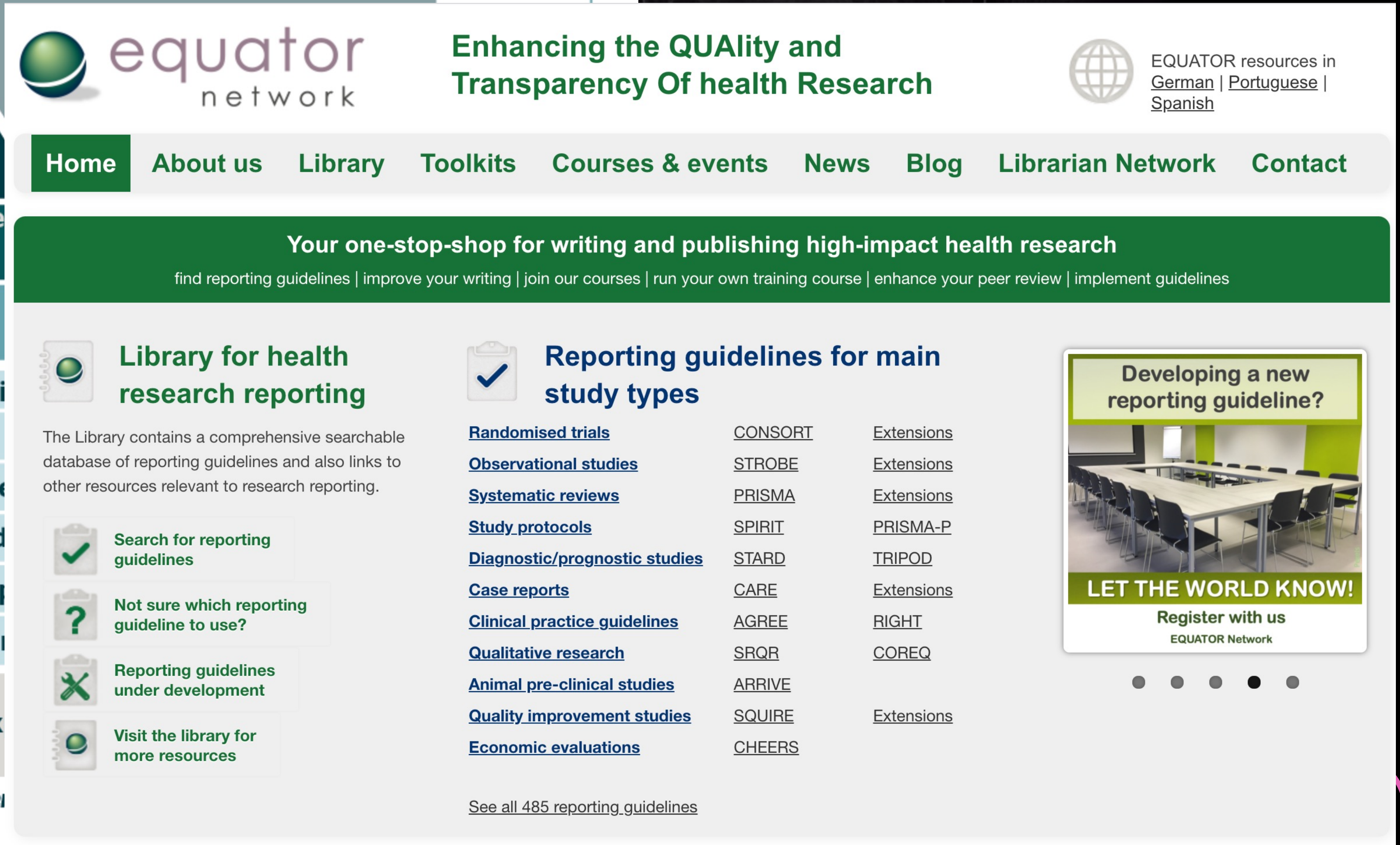


Brystkræft er den hyppigste form for cancer blandt kvinder i Danmark, og omtrent en tredjedel af tilfældene opdages gennem screeningsprogrammet (modelfoto).

Medicoteknik, 2022, Nr. 1

Hvordan? Hvorfor? Hvor? **Hvornår?**





Hvordan? | Hvorfor? | Hvor? | **Hvornår?**

RESEARCH

OPEN ACCESS

Check for updates

Use of artificial intelligence for image analysis in breast cancer screening programmes: systematic review of test accuracy

Karoline Freeman, Julia Geppert, Chris Stinton, Daniel Todkill, Samantha Johnson, Aileen Clarke, Sian Taylor-Phillips

Division of Health Sciences, University of Warwick, Coventry, UK

Correspondence to: S Taylor-Phillips S.Taylor-Phillips@warwick.ac.uk (ORCID 0000-0002-1841-4346)

Additional material is published online only. To view please visit the journal online.

Cite this as: BMJ 2021;374:n1872
http://dx.doi.org/10.1136/bmj.n1872

Accepted: 21 July 2021

ABSTRACT

OBJECTIVE

To examine the accuracy of artificial intelligence (AI) for the detection of breast cancer in mammography screening practice.

DESIGN

Systematic review o

DATA SOURCES

Medline, Embase, W
Database of System
to 17 May 2021.

ELIGIBILITY CRITER

Studies reporting te
or in combination w
women's digital ma
or in test sets. Refer
histology or follow-u
Outcomes included
detected.

STUDY SELECTION A

Two reviewers indep
inclusion and asses
of included studies
Diagnostic Accuracy
single reviewer extr
by a second reviewe
performed.

RESULTS

Twelve studies total
were included. No p
test accuracy of AI i
Studies were of poo

retrospective studies compared AI systems with the clinical decisions of the original radiologist, including 79 910 women, of whom 1878 had screen detected cancer or interval cancer within 12 months of screening. Thirty four (94%) of 36 AI systems

ORIGINAL RESEARCH • EVIDENCE-BASED PRACTICE

Radiology

Machine Learning for Workflow Applications in Screening Mammography: Systematic Review and Meta-Analysis

Sarah E. Hickman, MBBS • Ramona Woitek, MD, PhD • Elizabeth Phuong Vi Le, MA • Yu Ri Im, MB BChir • Carina Mouritsen Luxhøj, MA, MSci • Angelica I. Aviles-Rivero, BSc, MSc, PhD • Gabrielle C. Baxter, PhD • James W. MacKay, MA, MB BChir, PhD • Fiona J. Gilbert, MD

Radiology 2022; 302:88–104 • https://doi.org/10.1148/radiol.2021210391 • Content codes: BR AI

Background: Advances in computer processing and improvements in data availability have led to the development of machine learning (ML) techniques for mammographic imaging.

Purpose: To evaluate the reported performance of stand-alone ML applications for screening mammography workflow.

Materials and Methods: Ovid Embase, Ovid Medline, Cochrane Central Register of Controlled Trials, Scopus, and Web of Science literature databases were searched for relevant studies published from January 2012 to September 2020. The study was registered with the PROSPERO International Prospective Register of Systematic Reviews (protocol no. CRD42019156016). Stand-alone technology was defined as a ML algorithm that can be used independently of a human reader. Studies were quality assessed using the Quality Assessment of Diagnostic Accuracy Studies 2 and the Prediction Model Risk of Bias Assessment Tool, and reporting was evaluated using the Checklist for Artificial Intelligence in Medical Imaging. A primary meta-analysis included the top-performing algorithm and corresponding reader performance from which pooled summary estimates for the area under the receiver operating characteristic curve (AUC) were calculated using a bivariate model.

Results: Fourteen articles were included, which detailed 15 studies for stand-alone detection (n = 8) and triage (n = 7). Triage studies reported that 17%–91% of normal mammograms identified could be read by adapted screening, while "missing" an estimated 0%–7% of cancers. In total, an estimated 185 252 cases from three countries with more than 39 readers were included in the primary meta-analysis. The pooled sensitivity, specificity, and AUC was 75.4% (95% CI: 65.6, 83.2; P = .11), 90.6% (95% CI: 82.9, 95.0; P = .40), and 0.89 (95% CI: 0.84, 0.98), respectively, for algorithms, and 73.0% (95% CI: 60.7, 82.6), 88.6% (95% CI: 72.4, 95.8), and 0.85 (95% CI: 0.78, 0.97), respectively, for readers.

Conclusion: Machine learning (ML) algorithms that demonstrate a stand-alone application in mammographic screening workflows achieve or even exceed human reader detection performance and improve efficiency. However, this evidence is from a small number of retrospective studies. Therefore, further rigorous independent external prospective testing of ML algorithms to assess performance at preassigned thresholds is required to support these claims.

Online supplemental material is available for this article.

ORIGINAL ARTICLE ■ *Data Science*

Check for updates

SA-CME

Independent External Validation of Artificial Intelligence Algorithms for Automated Interpretation of Screening Mammography: A Systematic Review

Anna W. Anderson, MD[a,], M. Luke Marinovich, PhD, MPH[b,*], N...mat Houssami, MBBS, PhD[c,d], Kathryn P. Lowry, MD[a], Joann G. Elmore, MD, MPH[e,f], ...a S. M. Buist, PhD, MPH[g,h], Solveig Hofvind, PhD[i], Christoph I. Lee, MD, MS[a,j]*

...its awarded for this enduring activity are designated "SA-CME" by the American Board of Radiology (ABR) and qualify toward ...ing requirements for Maintenance of Certification (MOC) Part II: Lifelong Learning and Self-assessment. To access the SA-CME ...ty visit https://cortex.acr.org/Presenters/CaseScript/CaseView?Info=ZJSyEc6c6N%2fzqoO3euQnLOeDLx46I0q55AD9Drh5 ...%253d. SA-CME credit for this article expires January 31, 2025. All SA-CME articles can be found on JACR.org by navi...g to SA-CME on th...

...tract

...ose: The aim of th...
...igence (AI) technolo...
...ods: A systematic...
...gh December 10, 20...
...mographic cancer d...
...ting characteristic cu...
...logist and AI interpr...
...lts: After data extra...
...nercially available AI...
5]), or both (15% [...

npj | Digital Medicine

www.nature.com/npjdigitalmed

ARTICLE OPEN

Check for updates

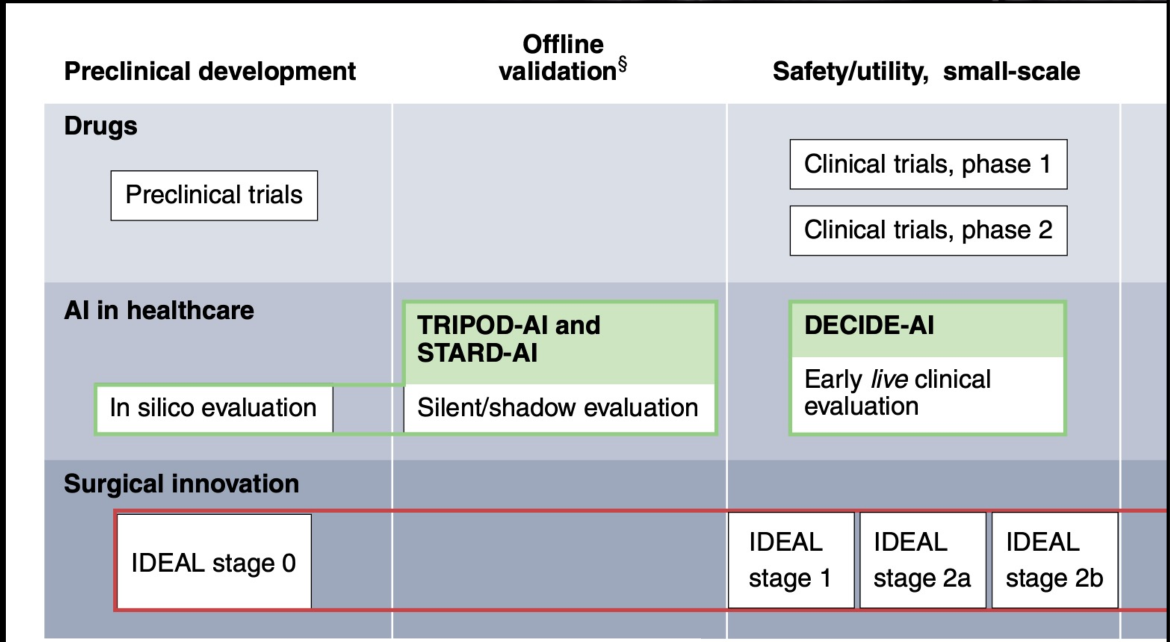
Deep learning in image-based breast and cervical cancer detection: a systematic review and meta-analysis

Peng Xue[1,5], Jiaxu Wang[1,5], Dongxu Qin[1], Huijiao Yan[2], Yimin Qu[1], Samuel Seery[3,4], Yu Jiang[1] and Youlin Qiao[1]

Accurate early detection of breast and cervical cancer is vital for treatment success. Here, we conduct a meta-analysis to assess the diagnostic performance of deep learning (DL) algorithms for early breast and cervical cancer identification. Four subgroups are also investigated: cancer type (breast or cervical), validation type (internal or external), imaging modalities (mammography, ultrasound, cytology, or colposcopy), and DL algorithms versus clinicians. Thirty-five studies are deemed eligible for systematic review, 20 of which are meta-analyzed, with a pooled sensitivity of 88% (95% CI 85–90%), specificity of 84% (79–87%), and AUC of 0.92 (0.90–0.94). Acceptable diagnostic performance with analogous DL algorithms was highlighted across all subgroups. Therefore, DL algorithms could be useful for detecting breast and cervical cancer using medical imaging, having equivalent performance to human clinicians. However, this tentative assertion is based on studies with relatively poor designs and reporting, which likely caused bias and overestimated algorithm performance. Evidence-based, standardized guidelines around study methods and reporting are required to improve the quality of DL research.

npj Digital Medicine (2022)5:19 ; https://doi.org/10.1038/s41746-022-00559-z

Hvordan? | Hvorfor? | Hvor? | **Hvornår?**



the**bmj** | *BMJ* 2022;377:e070904 | doi: 10.1136/bmj-2022-070904 NATURE MEDICINE | VOL 28 | MAY 202

Fig. 1 | Comparison of development pathways for drug therapies, AI in healthcare and guidelines, some of which are study design specific (TRIPOD-AI, STARD-AI, SPIRIT/C (DECIDE-AI and IDEAL). Depending on the context, more than one study design can b

Health Policy

UK National Screening Committee's approach to reviewing evidence on artificial intelligence in breast cancer screening

Sian Taylor-Phillips, Farah Seedat, Goda Kijauskaite, John Marshall, Steve Halligan, Chris Hyde, Rosalind Given-Wilson, Louise Wilkinson, Alastair K Denniston, Ben Glocker, Peter Garrett, Anne Mackie, Robert J Steele

Artificial intelligence (AI) could have the potential to accurately classify mammograms according to the presence or absence of radiological signs of breast cancer, replacing or supplementing human readers (radiologists). The UK National Screening Committee's assessments of the use of AI systems to examine screening mammograms continues to focus on maximising benefits and minimising harms to women screened, when deciding whether to recommend the implementation of AI into the Breast Screening Programme in the UK. Maintaining or improving programme specificity is important to minimise anxiety from false positive results. When considering cancer detection, AI test sensitivity alone is not sufficiently informative, and additional information on the spectrum of disease detected and interval cancers is crucial to better understand the benefits and harms of screening. Although large retrospective studies might provide useful evidence by directly comparing test accuracy and spectrum of disease detected between different AI systems and by population subgroup, most retrospective studies are biased due to differential verification (ie, the use of different reference standards to verify the target condition among study participants). Enriched, multiple-reader, multiple-case, test set laboratory studies are also biased due to the laboratory effect (ie, radiologists' performance in retrospective, laboratory, observer studies is substantially different to their performance in a clinical environment). Therefore, assessment of the effect of incorporating any AI system into the breast screening pathway in prospective studies is required as it will provide key evidence for the effect of the interaction of medical staff with AI, and the impact on women's outcomes.

Lancet Digit Health 2022; 4: e558–65

Warwick Medical School, University of Warwick, Coventry, UK (Prof S Taylor-Phillips PhD); UK National Screening Committee, Office for Health Improvement and Disparities, Department of Health and Social Care, London, UK (F Seedat PhD, G Kijauskaite MSc, J Marshall MA, Prof A Mackie PhD); Centre for Medical Imaging, Division of Medicine, University College London, London, UK (Prof S Halligan FMedSci); Exeter Test Group, College of Medicine and Health, University of Exeter, Exeter, UK

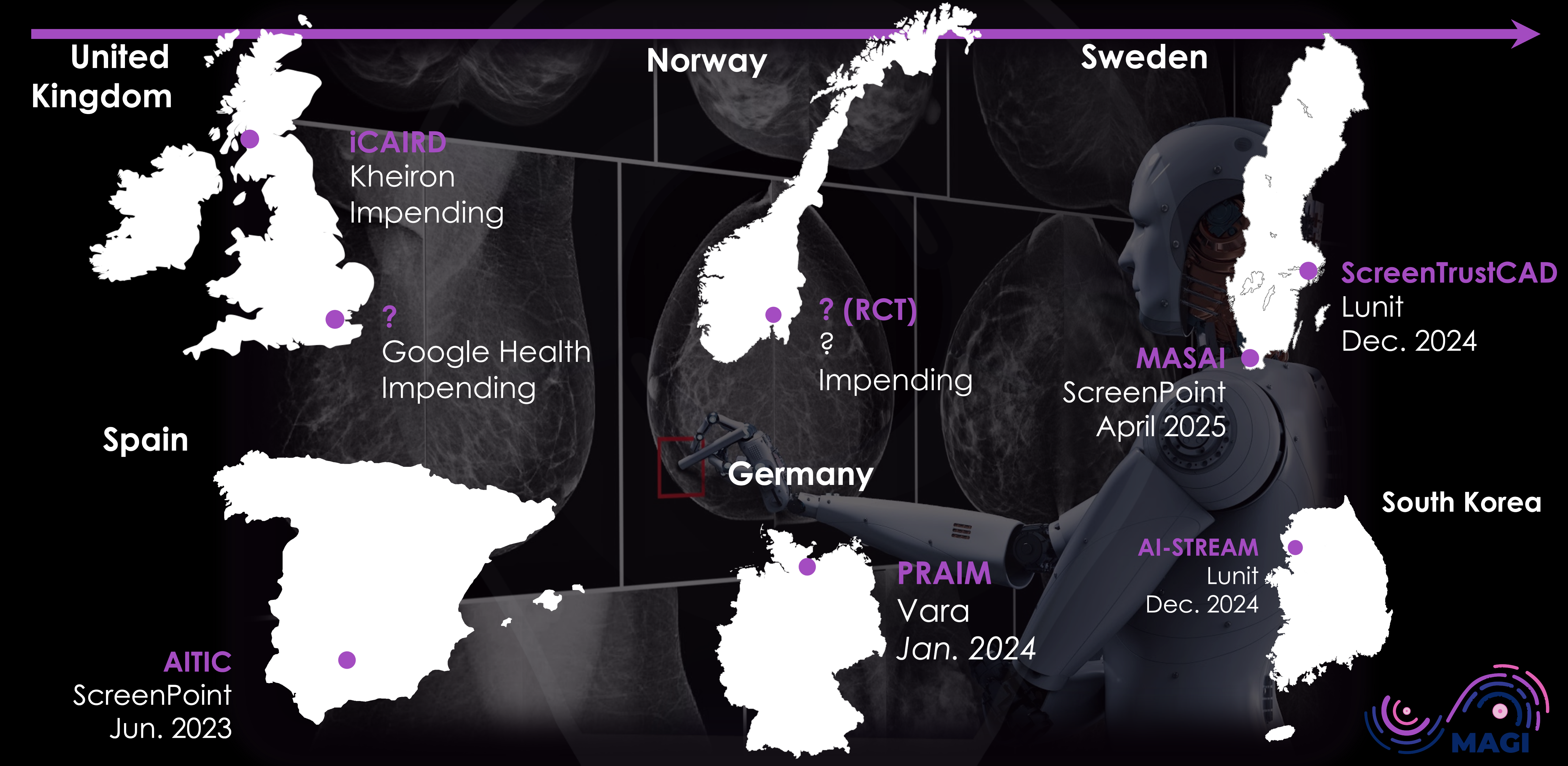
Vasey et al, *BMJ* / *Nat Med* (2022)

Taylor-Philips et al, Lancet Digit Health (2022)

Selvom retrospektive studier spiller en vigtig rolle i tidlig fase evaluering, har de begrænsninger og bias; prospektive studier er derfor nødvendige for at vurdere effekten af implementering af ethvert AI system i screeningsstien

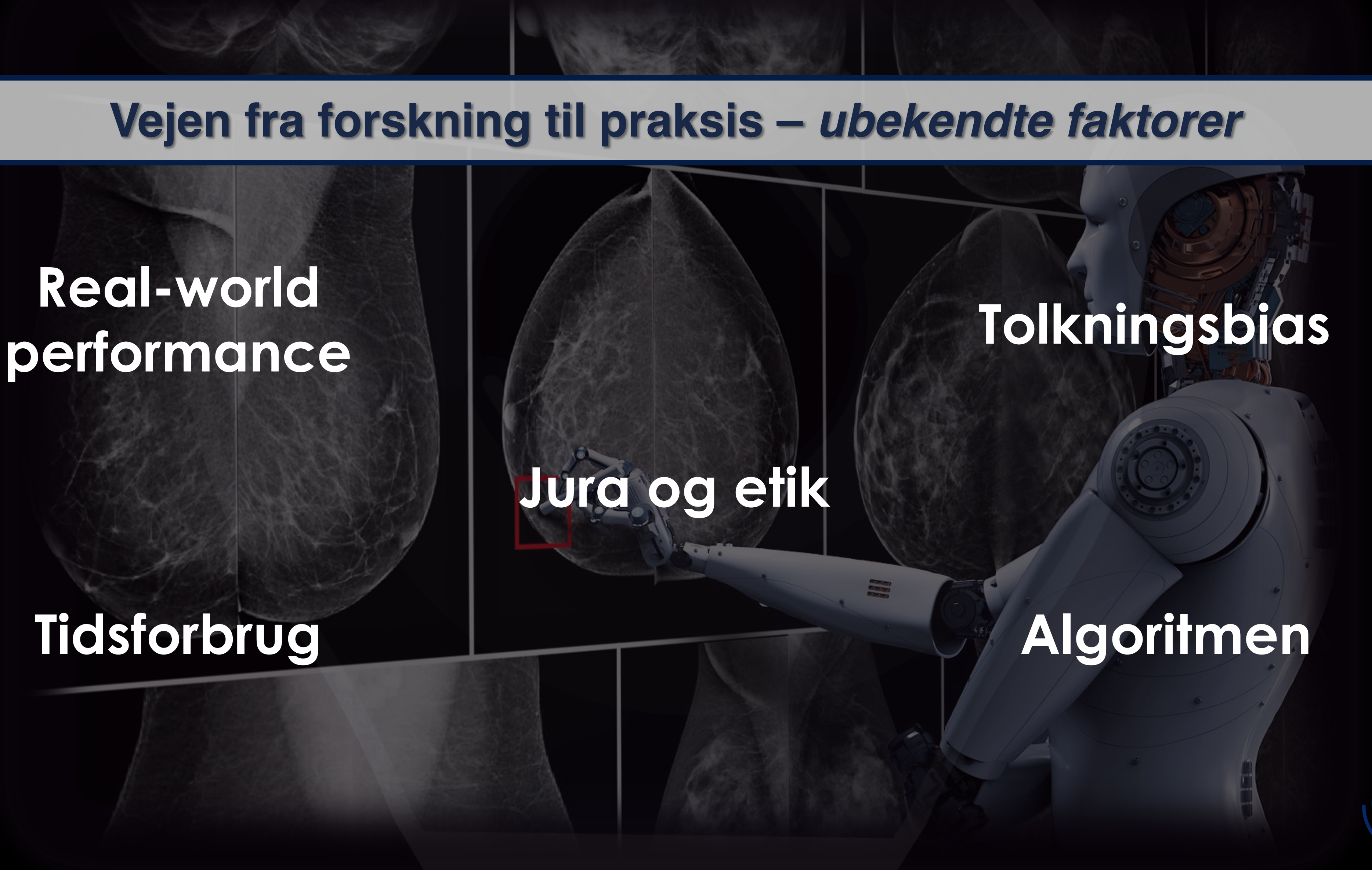
MAGI

Hvordan? | Hvorfor? | Hvor? | **Hvornår?**



Hvordan? | Hvorfor? | Hvor? | **Hvornår?**

Vejen fra forskning til praksis – *ubekendte faktorer*



Real-world performance

Tolkningsbias

Jura og etik

Tidsforbrug

Algoritmen

Hvordan? | Hvorfor? | Hvor? | **Hvornår?**

Fra retrospektiv validering til prospektivt randomiseret forsøg

Multicenter

Hel population

Aug. 2014-2018

≥ 2 års follow-up

Lokal data

RIPA SYD

VNA SYD

Nationale registre

DBCG DANISH BREAST CANCER GROUP

DKMS

MAGI

MAGI

Mammografi & Kunstig Intelligens

Tak for opmærksomheden!

mte@rsyd.dk

Region Syddanmark

CIMT
Center for Innovativ Medicinsk Teknologi

Centre for Clinical Artificial Intelligence

Radiologisk Forsknings- & Innovationsenhed
Syddansk Universitet

OUH
Odense Universitetshospital

KØBENHAVNS UNIVERSITET